

# **Instituto Tecnológico y de Estudios Superiores de Occidente**

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Electrónica, Sistemas e Informática  
**Maestría en Sistemas Computacionales**



## **Análisis de datos y tendencias sociales de las donaciones de sangre en Guadalajara**

---

**TRABAJO RECEPCIONAL** que para obtener el **GRADO** de  
**MAESTRO EN SISTEMAS COMPUTACIONALES**

Presenta: **JORGE APOLINAR ZÚÑIGA CISNEROS**

Asesor: **DR. IVÁN ESTEBAN VILLALÓN TURRUBIATES**

Tlaquepaque, Jalisco. Octubre de 2019.



## AGRADECIMIENTOS

Quiero agradecer a mis padres, Guadalupe y Agustín, por todo el cariño, el apoyo y la motivación que me brindaron en estos dos años de maestría que, aunque el camino no fue sencillo, siempre tuvieron las mejores palabras para mantenerme fuerte y lleno de determinación. No ha sido fácil vivir en ciudades diferentes y todos los días me hicieron sentir su amor y su apoyo como si estuvieran a mi lado, sin duda, fueron una pieza importante para poder alcanzar esta meta en mi desarrollo profesional y personal.

Durante estos dos años tuve el privilegio de trabajar el proyecto con mi asesor Iván Villalón que, con paciencia, buenos consejos y una excelente guía me ayudó a lograr el objetivo de terminar el Trabajo de Obtención de Grado y realizar las actividades necesarias para terminar la maestría.

También quiero agradecer a Oracle y al ITESO por el apoyo económico que me brindaron para poder lograr este sueño, siempre encontré la ayuda del personal de ambas instituciones cuando lo necesité y me sentí completamente apoyado para no interrumpir mis estudios y terminarlos de la mejor manera posible.

Por último, pero no menos importante, quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca otorgada durante los dos años de maestría, con el número de apoyo 487251.

# DEDICATORIA

## **A Guadalupe y Agustín:**

Toda la vida, con su ejemplo y su actuar diario, me han enseñado que las metas que nos planteamos las vamos construyendo día a día con dedicación, persistencia, resiliencia y esfuerzo.

Gracias, los amo con todo mi corazón.

# RESUMEN

La sangre es un tejido que principalmente se compone de plasma, pero también de plaquetas, eritrocitos y leucocitos que ayudan en múltiples funciones vitales. Debido a su compleja composición, cuando una persona necesita sangre para un trasplante, dificultades en el parto, alguna emergencia, entre otras circunstancias, esta sólo se puede obtener por medio de una donación de sangre.

Como se podrá observar a lo largo de este Trabajo de Obtención de Grado (TOG), en el mundo las donaciones de sangre son muy irregulares, existen países que son capaces de recolectar el suministro necesario para atender todas las necesidades de su población y existen otros países, como México, que deben priorizar en qué situaciones utilizar la sangre que se ha recolectado.

A través de las secciones de este trabajo se busca emplear el conocimiento adquirido a lo largo de la maestría para poder, con ayuda de herramientas y métodos computacionales, analizar un problema social como lo es: el bajo porcentaje de donaciones de sangre en México.

Se empezará por analizar cómo diferentes estrategias de *Machine Learning* han sido aplicadas en distintas investigaciones para encontrar tendencias, analizar datos de los centros de donación y mejorar procedimientos actuales. Tomando en consideración los resultados obtenidos en los diferentes trabajos, se determinarán estrategias para analizar la información obtenida sobre las donaciones de sangre en Guadalajara que forma parte de una de las 5 entidades federativas que mayor número de donaciones tienen anualmente y, sin embargo, no es suficiente para tener el suministro necesario de sangre y sus componentes.

Con el resultado de tales estrategias, se espera concientizar sobre la situación actual de México sobre las donaciones de sangre en el país y considerar el valor agregado que se puede obtener al utilizar distintos métodos tecnológicos para entender y buscar soluciones a problemas sociales.

# TABLA DE CONTENIDO

<b>AGRADECIMIENTOS</b> .....	<b>3</b>
<b>DEDICATORIA</b> .....	<b>4</b>
<b>RESUMEN</b> .....	<b>5</b>
<b>TABLA DE CONTENIDO</b> .....	<b>6</b>
<b>LISTA DE FIGURAS</b> .....	<b>9</b>
<b>LISTA DE TABLAS</b> .....	<b>11</b>
<b>LISTA DE ACRÓNIMOS Y ABREVIATURAS</b> .....	<b>12</b>
<b>1. INTRODUCCIÓN</b> .....	<b>13</b>
1.1. ANTECEDENTES .....	14
1.2. JUSTIFICACIÓN .....	15
1.3. PROBLEMA .....	15
1.4. HIPÓTESIS .....	16
1.5. OBJETIVOS .....	16
1.5.1. Objetivo General: .....	16
1.5.2. Objetivos Específicos: .....	16
1.6. NOVEDAD CIENTÍFICA, TECNOLÓGICA O APORTACIÓN .....	16
<b>2. ESTADO DEL ARTE O DE LA TÉCNICA</b> .....	<b>17</b>
2.1. ESTUDIOS SOBRE LAS PRINCIPALES RAZONES QUE MOTIVAN O DESMOTIVAN A LAS PERSONAS A DONAR .....	18
2.2. ANÁLISIS DE LOS MOTIVOS POR LOS CUALES UN POSIBLE DONADOR ES RECHAZADO PARA DONAR SANGRE .....	21
2.3. UTILIZAR ANÁLISIS DE DATOS PARA AYUDAR A TOMAR DECISIONES EN TEMAS DE SALUD .....	24
2.4. PREDICCIÓN DE DONACIONES UTILIZANDO MÉTODOS DE <i>MACHINE LEARNING</i> .....	26
2.4.1. Competiciones .....	26
2.4.2. Investigaciones .....	27
<b>3. MARCO TEÓRICO/CONCEPTUAL</b> .....	<b>29</b>
3.1. SANGRE .....	30
3.1.1. ¿Qué es la sangre? .....	30
3.1.2. Grupos sanguíneos .....	31
3.2. DONACIÓN DE SANGRE .....	31
3.2.1. Requisitos .....	32
3.2.2. Tipos de donadores .....	32
3.2.3. ¿Para qué se utiliza la sangre donada? .....	34
3.2.4. Beneficios de donar .....	34
3.2.5. Donaciones de sangre en México .....	35
3.3. GENERACIONES SOCIALES .....	37
3.3.1. Generación silenciosa .....	38
3.3.2. Generación baby boomers .....	38
3.3.3. Generación X .....	39

3.3.4.	Generación millennials.....	39
3.3.5.	Generación Z.....	39
3.4.	MACHINE LEARNING (APRENDIZAJE AUTÓMATICO).....	39
3.4.1.	Aprendizaje supervisado.....	39
3.4.2.	Aprendizaje no supervisado.....	40
3.5.	MODELOS DE REGRESIÓN.....	40
3.6.	PYTHON.....	41
3.7.	NUMPY.....	41
3.8.	PYTHON DATA ANALYSIS LIBRARY (PANDAS).....	42
3.9.	MATPLOTLIB.....	42
3.10.	SCIKIT-LEARN.....	43
<b>4.</b>	<b>DESARROLLO METODOLÓGICO.....</b>	<b>44</b>
4.1.	ENCUESTA – DONACIÓN DE SANGRE.....	45
4.1.1.	Estructura de la encuesta.....	45
4.2.	OBTENCIÓN DE DATOS.....	45
4.3.	ANÁLISIS DE DATOS.....	46
4.4.	TRANSFORMACIÓN DE DATOS.....	47
4.5.	LIMPIEZA DE DATOS.....	47
4.6.	PREPARACIÓN DE DATOS.....	50
4.7.	MODELO DE REGRESIÓN LINEAL CON SCIKIT-LEARN.....	54
4.7.1.	Regresión lineal sin normalización.....	54
4.7.2.	Regresión lineal con normalización.....	54
4.8.	MODELO DE REGRESIÓN POLINOMIAL CON SCIKIT-LEARN.....	55
<b>5.</b>	<b>RESULTADOS Y DISCUSIÓN.....</b>	<b>56</b>
5.1.	RESULTADOS DE LAS ENCUESTAS.....	57
5.1.1.	¿En qué año naciste?.....	57
5.1.2.	¿Género?.....	57
5.1.3.	¿Cuál es tu grupo sanguíneo?.....	58
5.1.4.	¿Has donado sangre alguna vez?.....	58
5.1.5.	La primera vez que donaste, fue:.....	58
5.1.6.	¿Cuántas veces has donado?.....	59
5.1.7.	¿Te generó un sentimiento positivo donar sangre?.....	59
5.1.8.	Si no has donado, ¿a qué se debe?.....	60
5.1.9.	En México menos del 4% de la población dona sangre de manera altruista, ¿a qué crees que se deba? 60	
5.2.	RESULTADOS MODELO DE REGRESIÓN LINEAL.....	61
5.2.1.	Regresión lineal sin normalización.....	62
5.2.2.	Regresión lineal con normalización.....	63
5.2.3.	Predicción.....	63
5.3.	RESULTADOS MODELO DE REGRESIÓN POLINOMIAL.....	64
5.4.	MODELOS DE REGRESIÓN LINEAL POR TIPO DE SANGRE.....	66
5.4.1.	Grupo sanguíneo A+.....	67
5.4.2.	Grupo sanguíneo A-.....	67
5.4.3.	Grupo sanguíneo B+.....	68
5.4.4.	Grupo sanguíneo B-.....	68
5.4.5.	Grupo sanguíneo AB+.....	69

5.4.6.	Grupo sanguíneo O+ .....	69
5.4.7.	Grupo sanguíneo O- .....	70
5.5.	DISCUSIÓN .....	70
<b>6.</b>	<b>CONCLUSIONES .....</b>	<b>72</b>
6.1.	CONCLUSIONES .....	73
6.2.	TRABAJO FUTURO .....	74
	<b>BIBLIOGRAFÍA.....</b>	<b>75</b>

# LISTA DE FIGURAS

Figura 2-1 Grupo de edades de los donadores noruegos comparado con el porcentaje de población [6]	19
Figura 2-2 Tabla de resultados de los motivos de descalificación obtenidos con las pruebas físicas [8]	21
Figura 2-3 Técnica de <i>Cupping</i> [11]	22
Figura 2-4 Distribución por género de los voluntarios de sangre en Gurajat, India [10]	23
Figura 2-5 Motivos de descalificación de los voluntarios en Gurajat, India [10]	23
Figura 2-6 Información geográfica de Hong Kong y los centros de donación [13]	24
Figura 2-7 Edad y género de las donaciones en Hong Kong [13]	25
Figura 2-8 Análisis de correlación [13]	25
Figura 2-9 Área bajo la curva de los distintos modelos [16]	27
Figura 2-10 Imagen 3D con la distribución de los pacientes que han donado o no [17]	27
Figura 2-11 Método del codo para K-Means [17]	28
Figura 3-1 Distribución de los componentes de la sangre [18]	30
Figura 3-2 Tabla de compatibilidad entre los grupos sanguíneos [20]	31
Figura 3-3 Donantes voluntarios y por reposición en México [22]	34
Figura 3-4 Población en Jalisco [25]	36
Figura 3-5 Gráfica de donaciones en Jalisco con línea de tendencia	37
Figura 3-6 Generaciones sociales [26]	38
Figura 3-7 Ejemplos de gráficas con matplotlib [31]	42
Figura 3-8 Algoritmos incluidos en <i>scikit-learn</i> [32, 23]	43
Figura 4-1 Código para cargar el <i>dataset</i> de las encuestas y generar un <i>DataFrame</i>	47
Figura 4-2 Función para limpiar irregularidades con las fechas de donación	49
Figura 4-3 <i>DataFrame</i> después de la limpieza de datos	49
Figura 4-4 <i>DataFrame</i> para donadores y no donadores	50
Figura 4-5 Función para obtener el año de una fecha	50
Figura 4-6 Año de la primera y última donación y total de donaciones	51
Figura 4-7 Función <i>generate_donations_by_year</i>	52
Figura 4-8 Arreglo de las donaciones por año	53
Figura 4-9 Creación del modelo y función <i>fit</i>	54
Figura 4-10 Datos del modelo de regresión	54
Figura 4-11 Creación del modelo de regresión con normalización	54
Figura 4-12 Datos del modelo de regresión con normalización	55
Figura 4-13 Función <i>create_polynomial_reg_model</i>	55
Figura 5-1 Participación en la encuesta por generación social	57
Figura 5-2 Donadores versus No donadores	58
Figura 5-3 Total de personas y donaciones realizadas	59
Figura 5-4 Sentimiento generado después de realizar la donación	59
Figura 5-5 Razones por las cuales la gente no ha donado	60
Figura 5-6 ¿A qué se debe el bajo porcentaje de donaciones altruistas en México?	61
Figura 5-7 Donaciones vs Año y línea de tendencia obtenida de la regresión lineal	62
Figura 5-8 Resultados modelo de regresión lineal	62
Figura 5-9 Resultados regresión lineal con normalización	63

Figura 5-10 Donaciones de sangre actuales y predicción a 10 años .....	64
Figura 5-11 Modelos de regresión polinomial .....	65
Figura 5-12 Predicción de los siguientes 10 años con un modelo polinomial .....	66
Figura 5-13 Regresión Lineal (A+) .....	67
Figura 5-14 Regresión Lineal (A-) .....	67
Figura 5-15 Regresión Lineal (B+) .....	68
Figura 5-16 Regresión Lineal (B-) .....	68
Figura 5-17 Regresión Lineal (AB+) .....	69
Figura 5-18 Regresión Lineal (O+) .....	69
Figura 5-19 Regresión Lineal (O-) .....	70

## LISTA DE TABLAS

Tabla 3-1 Top 5 de entidades federativas con mayor número de donaciones.....	35
Tabla 3-2 Entidades federativas con el menor número de donaciones .....	36
Tabla 3-3 Donaciones de sangre en Jalisco .....	36
Tabla 3-4 Porcentaje de las donaciones de sangre en Jalisco respecto a la población total.....	37
Tabla 4-1 Preguntas y distribución en el análisis .....	46
Tabla 5-1 Resultados ¿En qué año naciste? .....	57
Tabla 5-2 Total de participantes y donadores por género .....	57
Tabla 5-3 Distribución de los grupos sanguíneos y los donadores por grupo.....	58
Tabla 5-4 Principal razón para haber realizado la primera donación .....	59
Tabla 5-5 Datos utilizados para el modelo .....	61

## LISTA DE ACRÓNIMOS Y ABREVIATURAS

OMS	Organización Mundial de la Salud
TOG	Trabajo de Obtención de Grado
VIH	Virus de la Inmunodeficiencia Humana
UCI	Universidad de California, Irving
CSV	Comma-Separated Values

---

# 1. INTRODUCCIÓN

---

**Resumen:** *En este capítulo se presenta la problemática de las donaciones de sangre a nivel mundial, recomendaciones de la Organización Mundial de la Salud sobre el tema y se delimita el objeto de estudio a la ciudad de Guadalajara, México. Se plantea la situación que tiene el país ante este problema y se proveen algunas soluciones que se esperan alcanzar con el desarrollo de este trabajo.*

## 1.1. Antecedentes

Las donaciones de sangre y sus productos son vitales en un amplio número de servicios médicos, no contar con el abastecimiento necesario de este elemento puede costar la vida de miles de pacientes. De acuerdo con la Organización Mundial de la Salud (OMS), la tasa de donación de sangre por cada 1000 personas es de 32.6% en los países de ingresos altos, 15.1% en los de ingresos medio altos, 8.1% en los de ingresos medio bajos y 4.4% en los de ingresos bajos [1]. Debido a eso los países desarrollados utilizan su abastecimiento de sangre para tratar eventos como cuidados en cirugías cardiovasculares o trasplantes, traumas masivos, terapia para neoplasias hematológicas y para servicios de adultos mayores, mientras que los países no desarrollados utilizan la sangre para tratar complicaciones relacionadas con embarazos o anemia en los niños [2].

El hecho de que exista una diferencia tan grande entre los países respecto a las tasas de donación ha llevado a que distintos investigadores analicen los principales factores sociales que motivan o desmotivan a las personas a donar sangre. Con los resultados de esas investigaciones se busca mejorar los métodos y procesos que los países tienen actualmente para poder incrementar el número de donadores voluntarios no remunerados, debido a que se ha comprobado que la sangre de este tipo de donadores es la que es menos propensa a transmitir enfermedades como Hepatitis B o C, VIH u otras enfermedades de transmisión sexual.

Además de conocer las motivaciones que llevan a las personas a donar, es importante entender los factores que pueden ocasionar que un donador se convierta en alguien no apto para donar sangre y en este ámbito pueden influir distintos factores culturales como ciertos tipos de medicina alternativa, factores personales como datos sobre su vida sexual, tratamientos médicos o dentales y factores médicos respecto al historial que cada donador tiene sobre las enfermedades que ha tenido. Las investigaciones que se han realizado buscando identificar los factores por lo cuales alguien no es apto para donar han mostrado que en los países no desarrollados la gente está muy desinformada sobre los requisitos para ser donador, sobre los bancos de sangre a los que pueden acudir a donar y también sobre campañas de donación. En estos países una gran parte de las donaciones se realizan por reemplazo, es decir, que primero se provee de sangre al paciente y sus familiares consiguen cierto número de donadores para suplir la sangre que fue otorgada pero no es lo más recomendado para mantener un abastecimiento de sangre segura ya que existe mayor riesgo de que la sangre pueda venir infectada en comparación con la sangre de los donadores voluntarios no remunerados.

Sumado a lo anterior, también se han realizado investigaciones que partiendo de resultados similares a los anteriores buscan predecir las donaciones para años futuros o la mejor ubicación para situar un nuevo centro de donación de sangre. Este tipo de objetivos se logran como resultado de un análisis exhaustivo a los datos y las relaciones que existen entre ellos, enseguida se somete la información a métodos de *Machine Learning* como regresiones lineales, *clustering*, redes neuronales, entre otros, con el objetivo de poder encontrar estrategias que ayuden a generar mejores campañas de donación, mejorar los servicios en los bancos de sangre para que los donadores actuales recomienden y motiven a más personas a donar y buscar reducir los miedos y mitos que tiene la población debido a la falta de información sobre el tema.

## 1.2. Justificación

Este Trabajo de Obtención de Grado (TOG) pretende enriquecer el contenido que existe en México respecto a las donaciones de sangre. Busca comprender las emociones y motivos sociales o culturales que favorecen a que las personas realicen una donación voluntaria no remunerada y también los motivos por los cuales los mexicanos, en la ciudad de Guadalajara, deciden no realizar donaciones.

Debido a los índices tan bajos que tiene México respecto a las donaciones voluntarias no remuneradas, se decidió realizar esta investigación para poder encontrar estrategias que ayuden a aumentar la cantidad de este tipo de donaciones en los Centros de Transfusión de Sangre Mexicanos y que motiven a los actuales donadores para que sigan donando constantemente y se logre mantener un suministro adecuado para satisfacer las necesidades de la población.

También como parte del TOG, se busca utilizar los conocimientos obtenidos durante la maestría para poder solucionar problemas sociales, por lo cual, con los datos adquiridos de las encuestas realizadas a una muestra de los habitantes de la ciudad de Guadalajara se aplicarán algunos algoritmos de *Machine Learning* para poder predecir el comportamiento de distintos grupos de personas, poder determinar quienes son más propensos a donar en el futuro y tener un estimado de la cantidad de donaciones que podemos esperar para los años por venir y así enfocar la atención en mantener ese grupo de donadores recurrentes.

## 1.3. Problema

Es complicado determinar un valor exacto para el índice de donaciones voluntarias no remuneradas que se realizan en México anualmente, en 2017 una noticia de El Universal mencionó que en México sólo el 2.7% de las donaciones se realizan de manera altruista y que el principal motivo de donación es el conocido como reposición o reemplazo que surge sólo cuando algún familiar necesita sangre [3]. También, un reportaje de Noticieros Televisa comentó que según el Centro Nacional de la Transfusión Sanguínea de 2 millones de donaciones sólo un 3.8% son realizadas de manera altruista [4], lo que es alarmante ya que la OMS recomienda que el abastecimiento de sangre sea totalmente generado por donadores voluntarios no remunerados debido a que se tienen datos de que su sangre es menos propensa a transmitir enfermedades.

Contar con un número tan bajo de donaciones altruistas repercute en costos para los bancos de sangre ya que cada unidad de sangre obtenida tiene que pasar por un proceso para verificar que es 100% saludable y que no transmitirá ninguna enfermedad al receptor, cuando la sangre es encontrada con cierto nivel de riesgo ya se generaron pérdidas de tiempo al personal, recursos gastados en la evaluación de la unidad de sangre y, lo más importante, seguirá faltando esa sangre para poder satisfacer alguna emergencia. En países como México que no cuentan con un suministro adecuado de sangre, la poca cantidad que se tiene se prioriza para administrarla a complicaciones relacionadas con el embarazo, anemia en niños o traumatología.

En países que no cuentan con el suministro necesario, normalmente son los países que no son considerados de ingresos altos, se han realizado investigaciones para intentar comprender los factores sociales o culturales por los cuales las personas no donan o dejan de donar. También se han realizado diferentes estudios aplicando distintas metodologías para intentar predecir si las personas donarán o no en un momento específico dependiendo de su historial de donaciones.

Por lo cual, es preciso comprender los factores culturales y sociales de la población mexicana que ocasionan que se tenga un porcentaje tan bajo de donaciones altruistas y que la mayor parte de ellas sea debido a reposición o reemplazo, ya que no se encuentran estudios respecto al tema en el país, también se analizarán los datos para poder encontrar tendencias y características de los grupos sociales más propensos a donar sangre.

## 1.4. Hipótesis

La falta de calidad en el proceso de donación de sangre y la poca información que se le transmite a la sociedad genera mitos y desconfianza en la población, por lo cual, las personas se niegan a acudir a un banco de sangre para donar, como consecuencia el índice de donaciones voluntarias no remuneradas en México se mantiene entre los niveles más bajos a nivel mundial.

## 1.5. Objetivos

### *1.5.1. Objetivo General:*

Aplicar los conocimientos adquiridos durante la maestría en busca de generar una solución o mejora a un problema social. A lo largo de este trabajo se aplicarán distintas metodologías de *Machine Learning* a un conjunto de datos para obtener tendencias y predicciones sobre el comportamiento actual y futuro de las donaciones de sangre en México.

### *1.5.2. Objetivos Específicos:*

1. Determinar, de acuerdo con la muestra obtenida, qué tipo de donaciones son más frecuentes en la población mexicana.
2. Analizar los factores sociales y culturales por los cuales las personas realizan donaciones.
3. Analizar los factores sociales y culturales por los cuales las personas no están dispuestas a donar.
4. Clasificar a los donadores en diferentes grupos, basándose en características en común, para determinar si cierto grupo es más propenso a donar en el futuro.

## 1.6. Novedad científica, tecnológica o aportación

Los distintos trabajos de investigación que se revisarán en la sección Estado del Arte donde se aplicaron métodos de *Machine Learning* utilizaron un *dataset* del 2008 para poder realizar su análisis y además la información fue generada en el país de Taiwán por lo cual las costumbres, la manera de pensar, la cultura, la importancia que el gobierno y los centros de salud le dan a las donaciones de sangre y entre muchas otras cosas divergen de cómo es la cultura de la sociedad mexicana y la atención e importancia que se le da a las donaciones de sangre en México por parte del sistema de salud y el gobierno.

Para realizar el desarrollo de este TOG se realizaron 515 encuestas a habitantes de la ciudad de Guadalajara durante el año de 2019, lo que otorga al trabajo tendencias actuales de la población y ayuda a predecir un posible panorama para los años por venir. Además, contar con información actual y tomada de la sociedad mexicana, otorga la oportunidad de conocer y analizar los causas que motivan a los mexicanos tanto a donar sangre, como a no hacerlo, con el objetivo de ayudar a generar estrategias para mejorar los índices de donaciones de sangre nacionales, ya que con los resultados se descubrirán los mitos, las motivaciones, las complicaciones, las limitantes, entre otras causas que le suceden a la sociedad mexicana cuando deben decidir si donar sangre o no hacerlo.

---

## 2. ESTADO DEL ARTE O DE LA TÉCNICA

---

**Resumen:** *En este capítulo se presenta un resumen de los trabajos relacionados con el tema de donaciones de sangre en distintos países, se describe el enfoque de cada investigación y los resultados encontrados.*

## 2.1. Estudios sobre las principales razones que motivan o desmotivan a las personas a donar

Es importante conocer las razones que motivan a la gente tanto a donar como a dejar de hacerlo, de acuerdo a un estudio de la OMS, de las 117.4 millones de donaciones extraídas en el mundo en el 2015, el 42% se obtuvieron de los países de altos ingresos, según el Banco Mundial para que un país se considere de alto ingreso debe tener un Ingreso Nacional Bruto per cápita de \$12,376 o más [5], estos países constituyen el 16% de la población mundial [1], lo que ocasiona que mientras unos países tienen el abasto necesario de sangre y sus productos hay otros que no cuentan con el abastecimiento necesario para cubrir las necesidades de la población.

Noruega es un país considerado de alto ingreso por el Banco Mundial [5] y desde 1982 se han mantenido como auto suficientes en productos de sangre y plasma utilizando sólo donadores voluntarios no remunerados lo que ayuda a reducir el riesgo de transmitir alguna enfermedad al receptor debido a que se ha demostrado que la sangre que es menos propensa a transmitir alguna infección proviene de las personas que donan sangre de manera voluntaria y no remunerada y de las personas que son donadores frecuentes en comparación con las personas que donan por primera vez [6].

Por tal motivo, el sistema de bancos de sangre que tiene Noruega fue considerado para realizar una investigación sobre los principales factores que intervienen en el desarrollo de un compromiso como donador voluntario no remunerado. El estudio realizado por Misje et al. [6] consistía en un cuestionario que incluía información sociodemográfica (género, edad, estado social, educación, ingresos anuales, etc.), historial de donaciones, método por el cual fue reclutado para donar, una evaluación sobre el servicio del banco de sangre y 19 oraciones sobre los motivos para donar. Estas oraciones aluden a un factor de motivación, para este análisis se consideraron cuatro factores:

- Valores: se refiere a que la donación fue por motivos altruistas o empatía.
- Sociales: influencia de amigos, familia o un grupo social que motivo al individuo a donar.
- Autoestima: se refiere a que las razones para donar son sentirse mejor consigo mismo.
- Entendimiento: experiencias positivas asociadas con voluntariados.

Los resultados del estudio fueron muy significativos, se encontró que:

- El 55% de los donadores fueron reclutados por contactos interpersonales como: vecinos, familiares, amigos o pareja.
- Un 20% de los donadores se enteró vía publicidad en periódicos, televisión, radio, revistas, etcétera.
- Respecto a las preguntas para calificar el servicio del banco de sangre, el 94.5% de los donadores estaban satisfechos con los servicios y el profesionalismo de los trabajadores. El 5% restante mencionó que en ocasiones tienen problemas para encontrar estacionamiento. A pesar de que la donación es no remunerada, en ocasiones el banco de sangre les otorga a los donadores un regalo como agradecimiento, puede ser una gorra, un paraguas, el dinero del transporte público, etcétera.
- 89% de todos los donadores señalan como principal motivación el factor orientado a los valores, es decir, motivos altruistas o empatía.

- Los donadores más recurrentes rechazan que la donación se haga por motivos de autoestima, este factor es más común en gente que dona por primera vez o lo ha hecho sólo en pocas ocasiones. Los donadores más recurrentes atribuyen el acto de donar a factores como el entendimiento y obligación moral.
- Cerca del 50% de los donadores noruegos recurrentes han donado más de 20 veces lo que resalta que el compromiso de donar sangre es alto entre los habitantes del país.
- Existe una mayor cantidad de hombres donantes que de mujeres (53% contra 47%) y la mayoría de los donadores se encuentran entre los 26 años y 55 años, debajo de 26 años y arriba de 56 años se presentaron muy poco donadores como se aprecia en la Figura 2-1.

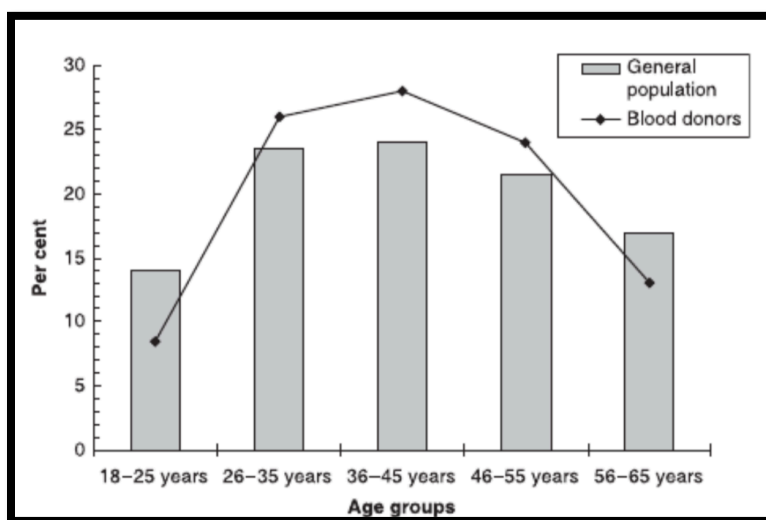


Figura 2-1 Grupo de edades de los donadores noruegos comparado con el porcentaje de población [6]

En otra investigación realizada en Brasil, en la ciudad de Belo Horizonte, se buscó encontrar los motivos por los cuales la gente accede a donar sangre y también los motivos por los cuales dejan de hacerlo. A diferencia de Noruega, Brasil no es considerado un país de alto ingreso, Brasil está considerado como un país de ingreso medio-alto que de acuerdo con el Banco Mundial un país en este nivel es aquel que tiene un Ingreso Nacional Bruto per cápita entre \$3,996 y \$12,375 [5], por lo cual es interesante conocer las diferencias que existen desde el momento de reclutar donadores hasta el servicio que se les ofrece a los voluntarios en los bancos de sangre.

En Brasil sólo el 1.78% de la población son donadores de sangre, tal dato coloca al país un 1.22% debajo del índice recomendado que comienza en 3% y va hasta 5% para poder tener un abastecimiento suficiente de sangre. El porcentaje tan bajo probablemente se deba a que existe bastante desinformación en Brasil respecto a la donación de sangre ya que en los resultados del estudio se obtuvo que:

- La gente no dona por miedo, miedo del dolor, miedo a contagiarse de alguna enfermedad, entre otros motivos que pueden generar miedo en el donador.
- Poca frecuencia de las campañas de donación.
- Las campañas no mencionan los requerimientos completos para donar lo que genera una pérdida de tiempo para el donador y el personal del banco de sangre ya que cuando llega el donador se le especifican los requerimientos completos y al no ser apto en uno tiene que irse.

- En ocasiones no se le menciona al voluntario las razones por la cuales no es apto y al no ser apto en una, se marcha del lugar con un sentimiento de tristeza, por lo cual, no vuelve en el futuro a realizar donaciones ya que le ha generado una sensación negativa.

Sin embargo, las motivaciones que mueven a la sociedad de Belo Horizonte a donar son muy similares a las de las personas en Noruega, saber que están realizando una acción altruista que favorecerá a alguien en una situación de emergencia es de los principales factores por los cuales la gente se acerca a donar y también el hecho de ayudar a personas cercanas, amigos o familia, cuando estos necesitan donadores de repuesto por la sangre que ya han recibido como parte del tratamiento [7].

Si bien no hay que generalizar, ambos estudios confirman los datos de la OMS de por qué el 42% de las donaciones mundiales se realiza en los países de altos ingresos ya que se invierte más en campañas de donación y se identificó que el mejor medio de reclutar nuevos donadores y mantener a los actuales es a través de ofrecerles una buena experiencia cuando visitan el banco de sangre, además de ser transparentes con la información que se les brinda ya sea compartiendo los requerimientos necesarios para que puedan ser considerados aptos para donar y, en caso de ser rechazado, explicar los motivos del rechazo para evitar causar una experiencia negativa en el donador.

## 2.2. Análisis de los motivos por los cuales un posible donador es rechazado para donar sangre

En muchas ocasiones se dice que el proceso de donar sangre es algo que puede consumir demasiado tiempo del donador. Sin embargo, el largo proceso busca asegurar que la sangre o los componentes de la sangre cumplan con las condiciones necesarias para favorecer la salud del receptor y evitar nuevas infecciones en él y en el donador.

El proceso de elección de un donador consta, al menos, de dos pasos: un conjunto de pruebas físicas y un cuestionario sobre el historial médico. En el conjunto de pruebas físicas se revisa el peso, pulso, presión sanguínea, temperatura, el nivel de hemoglobina [8] y otras distintas causas que pueden descalificar a alguien para donar sangre como se observa en la Figura 2-2. Por otro lado, el cuestionario explora las enfermedades que el donador ha tenido, buscando evitar infecciones como hepatitis B o C, VIH, herpes, sífilis, entre otras. En algunos países, por ejemplo, en Arabia Saudita el cuestionario también ayuda a identificar si han visitado ciertos países como África o Sudán que son conocidos como áreas endémicas de malaria y VIH y así intentar eliminar el riesgo de posibles infecciones [9].

Deferral category	First-time (rate)	Repeat (rate)	Missing (rate)	Total (rate)
Low Hb	9,582 (673)	3,177 (223)	468 (33)	13,227 (929)
High blood pressure	2,263 (159)	871 (61)	32 (2)	3,166 (222)
Medical diagnosis/medication <sup>a</sup>	1,684 (118)	492 (35)	0 (0)	2,176 (153)
Fever/flu/cough	1,440 (101)	418 (29)	40 (3)	1,898 (133)
High-risk donor <sup>b</sup>	388 (27)	1,309 (92)	24 (2)	1,721 (121)
Infectious exposure <sup>c</sup>	655 (46)	147 (10)	0 (0)	802 (56)
Other <sup>d</sup>	3,262 (229)	1,032 (72)	320 (22)	4,614 (324)
Total number of deferred visits	19,274 (1,353)	7,446 (523)	884 (62)	27,604 (1,938)

**Notes:** <sup>a</sup>Medical diagnosis category includes deferrals due to cancer, glucose-6-phosphate dehydrogenase deficiency, surgical procedures, and/or conditions related to cardiac, gastroenterology, neurology, pulmonary, endocrinology, and urology diseases. Medication includes antibiotics. <sup>b</sup>Donors positive for infectious disease markers or at risk for exposure. <sup>c</sup>Donor at risk of infectious disease due to exposure to dentist, endoscopy, history of malaria, tuberculosis, typhoid, UK visit, tattoo, and needle prick. <sup>d</sup>Includes other deferral reasons such as blood transfusion, positive antibody screening, pregnancy/abortion, apprehensive/cannot wait, age/weight out of the acceptable range, and/or no identification document, vaccination.

**Abbreviation:** Hb, hemoglobin.

Figura 2-2 Tabla de resultados de los motivos de descalificación obtenidos con las pruebas físicas [8]

En la actualidad, nos encontramos lejos de tener un suministro adecuado de sangre que nos permita cumplir con la demanda [10], por ello entender las razones por las cuales las personas son descalificadas para donar nos podría ayudar a planear de manera más eficiente estrategias de reclutamiento y mejorar la evaluación de los criterios para la selección de donadores. En muchas ocasiones las causas por las cuales una persona no está calificada para donar puede ser una condición corregible y con la guía y educación adecuada la persona podría convertirse en un donador apto para el futuro y así mantener un suministro de sangre continuo. Es importante tener tacto y hacer conocer a los donadores la razón por la cual no son aptos en ese momento ya que en el estudio de Shaer et al. [8] se demuestra que el donador pudiera generar sentimientos negativos al ser rechazado y no volver de nuevo si no conoce la razón por la cual es descalificado para donar sangre.

Además de los motivos ya mencionados antes por los cuales alguien pudiera ser descalificado para donar, dependiendo de las regiones geográficas, distintas costumbres pudieran influir en el rechazo para la donación. Por ejemplo, en el estudio llevado a cabo en Jeddah, una ciudad en Arabia Saudita, se tomó de

muestra a 500 personas que ya habían sido calificadas como no aptas para donar, de las 500 personas, 428 eran hombres y 72 eran mujeres en un rango de edad de 16 a 76 años. El estudio divide las causas de descalificación en tres categorías; causas personales con un 10 por ciento, examen médico con un 47.20 por ciento y el historial médico con un 42.60 por ciento. Dentro de estas categorías las principales causas fueron:

- Causas personales:
  - Falta de descanso (dormir) – 29 personas.
  - Mayor de edad – 8 personas.
  - Venas delgadas – 6 personas.
- Examen médico:
  - Baja presión sanguínea – 68 personas.
  - Bajo nivel de hemoglobina – 54 personas.
  - Alto nivel de hemoglobina – 35 personas.
- Historial médico:
  - Cupping – 58 personas.
  - Tomando antibióticos – 32 personas.
  - Procedimiento dental – 18 personas.

En esos resultados se puede observar que la práctica de *cupping*, técnica con gran popularidad en la medicina tradicional de China y algunos otros países de Medio Este que consiste en generar succiones en la piel utilizando vidrio, cerámica o vasos de plástico [11], como se puede observar en la Figura 2-3, refleja las creencias religiosas de las personas en Arabia Saudita respecto a que la técnica de *cupping* puede ayudar a curar distintas enfermedades.

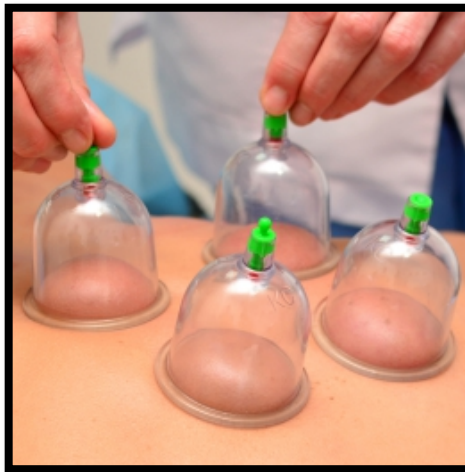


Figura 2-3 Técnica de *Cupping* [11]

Como resultados de estas investigaciones también podemos observar como las causas pueden ir variando dependiendo de las tradiciones de la sociedad o distintas condiciones que hacen a alguien no apto para donar. Por ejemplo, en el estudio realizado en Gurajat, India, se realizó un seguimiento de un mes a una campaña de donación en la cual se obtuvieron 161 posibles donadores, la distribución de hombres y

mujeres se puede observar en la Figura 2-4, del total de personas 132 eran hombres y 29 eran mujeres, entre las edades de 21 a 30 años.

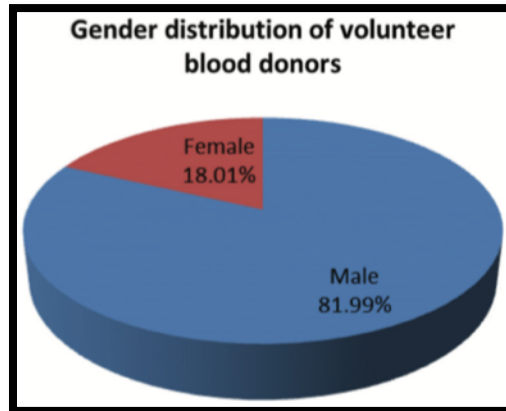


Figura 2-4 Distribución por género de los voluntarios de sangre en Gurajat, India [10]

De estos donadores sólo 28 fueron rechazados, los principales motivos fueron: rastros de ictericia, término utilizado para describir un tono amarillo en la piel y en las partes blancas de los ojos causado por exceso de bilirrubina [12], presión sanguínea alta, enfermedades con la tiroides, menorragia y anemia. También, el estudio menciona que el bajo número de mujeres en la donación puede deberse a altos niveles de anemia en las mujeres de la India [10]. En la Figura 2-5 se pueden observar las demás causas de descalificación del estudio realizado en Gurajat.

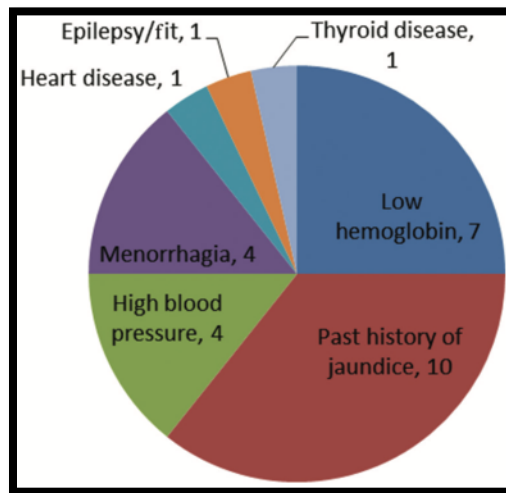


Figura 2-5 Motivos de descalificación de los voluntarios en Gurajat, India [10]

## 2.3. Utilizar análisis de datos para ayudar a tomar decisiones en temas de salud

El análisis de datos nos ayuda a identificar tendencias en los datos y, al mismo tiempo, las tendencias ayudan a interpretar ciertas maneras de actuar de la población y con base en esa información se pueden predecir ciertos eventos.

En Hong Kong se buscaba colocar un nuevo centro de donación de sangre y para ello se decidió realizar un análisis de los datos históricos de donaciones en los distintos centros de donación existente para poder determinar la mejor ubicación del nuevo centro de donación. El estudio identificó la relación entre la edad de los donadores, género, tipo de sangre y frecuencia de donación, la relación entre la frecuencia de las donaciones y la distancia al centro de donación, las características de los diferentes centros de donación y la tendencia de la donación de sangre [13].

El conjunto de datos constaba de 775,690 registros que tenía información como el ID del donador, género, fecha de nacimiento, día de donación, dirección, dirección del centro de donación y tipo de sangre. La dirección se agrupó en los 18 distritos definidos por el gobierno de Hong Kong y esos 18 distritos se agruparon en 5 distritos considerando la distancia de los distritos a los centros de donación actuales. En la Figura 2-6 se observan los 18 distritos y la agrupación que se realizó.

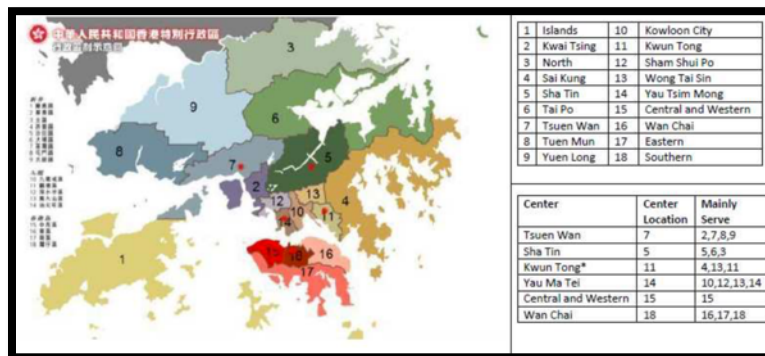


Figura 2-6 Información geográfica de Hong Kong y los centros de donación [13]

En la primera parte del estudio se analizó la información para buscar relaciones entre los datos, se obtuvieron los siguientes resultados:

- El promedio del total de donaciones por el total del número de donadores es 2.23, lo que significa que en los cuatro años de información la gente se acercó a donar 2.23 veces. Si se quitan los resultados obtenidos en las unidades móviles para donar el promedio del total de donaciones aumenta un 19.44%, da un total de 2.79, eso nos refleja que la gente prefiere acudir a los centros de donación a donar.
- Del total de donaciones, la cantidad de hombres que donan es del 54% por un 46% de mujeres.
- En la Figura 2-7 se observa que las mujeres jóvenes donan más sangre que los hombres, pero al aumentar la edad los hombres comienzan a ser más participativos

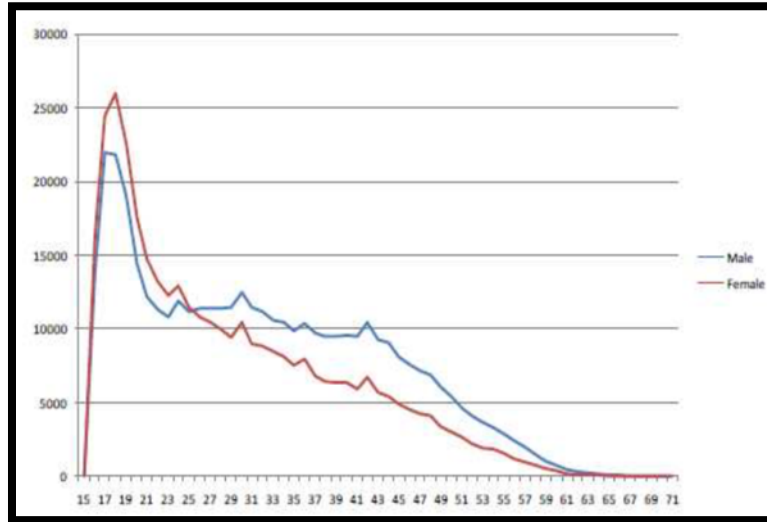


Figura 2-7 Edad y género de las donaciones en Hong Kong [13]

En la segunda parte del estudio el objetivo fue encontrar la correlación entre el número de donaciones y la distancia entre la dirección residencial del donador con la del centro de donaciones [13], para ello sólo se tomaron en consideración las donaciones realizadas en los centros y por cada distrito se calculó la distancia entre la dirección residencial del donador y la dirección del centro y se contó el número de donaciones por cada ubicación, después se generó una regresión lineal con la distancia en el eje x y el total de donaciones en el eje y. La Figura 2-8 muestra la correlación entre la Distribución Geográfica y la frecuencia de las donaciones. Los puntos en el círculo azul indican una posible opción de una buena ubicación para un centro de donación ya que demuestran que de acuerdo con la correlación existente generarían un aumento de las donaciones de aproximadamente 14,279 nuevas donaciones.

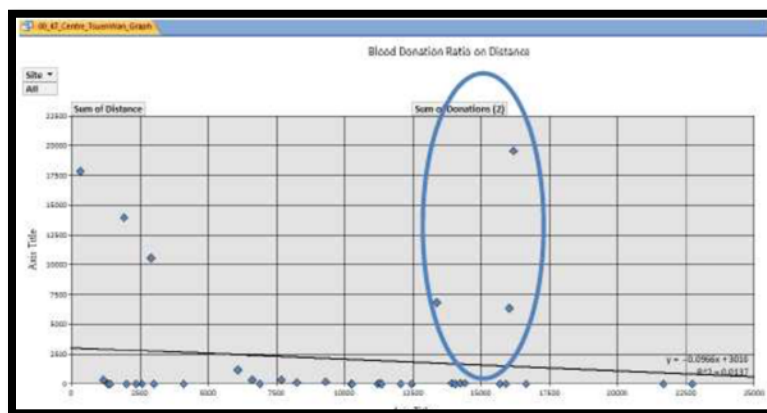


Figura 2-8 Análisis de correlación [13]

A través de este estudio, se muestra un ejemplo de cómo el análisis de datos y el uso de un método de análisis espacial, como la regresión lineal, pueden proveer información importante para tomar decisiones en temas del cuidado de la salud y además mucha información sobre las tendencias en la población y los factores que influyen en su comportamiento para donar sangre.

## 2.4. Predicción de donaciones utilizando métodos de *Machine Learning*

Debido a los bajos porcentajes de donación de sangre en varios países y lo crítico que puede ser no contar con el abastecimiento necesario de este elemento vital, más y más investigadores e ingenieros han aplicado distintos métodos y algoritmos de *Machine Learning* con el objetivo de intentar predecir en los años por venir si la población es más propensa a realizar donaciones y a convertirse en donadores recurrentes o si la situación puede mantenerse igual o incluso empeorar y en estos caso se busca encontrar patrones o tendencias que pudieran favorecer al comportamiento negativo de los datos para generar acciones y contrarrestar ese comportamiento.

Aunque existe gran interés en los resultados, no existen suficientes datos para poder aplicar distintas metodologías, por lo cual los trabajos que analizaremos se han realizado utilizando el *dataset* Blood Transfusion Service Center Data Set [14] el cual fue donado en el año 2008 por el Prof. I-Cheng Yeh a la Universidad de California, Irving (UCI). Este *dataset* se generó de la base de datos del Centro de Servicio de Transfusiones de Sangre en Hsin-Chu, Taiwán. El centro utilizaba su servicio móvil para asistir a una universidad de la ciudad y recolectar sangre cada tres meses, de la base de datos de donadores se tomaron 748 registros para crear el *dataset*. La información incluye los siguientes campos:

- R: meses desde la última donación.
- F: número total de donaciones.
- M: sangre donada en total en centímetros cúbicos.
- T: meses desde la primera donación.
- DM: variable binaria, si la persona donó sangre en marzo de 2007 es un 1, de lo contrario 0.

### 2.4.1. Competiciones

Debido a la importancia del tema el sitio web DrivenData, que se encarga de fusionar el conocimiento de científicos de datos con problemas sociales como educación, salud, medio ambiente, entre otros, creó una competición utilizando este *dataset* y tuvo más de 5,500 participantes que subían sus mejores propuestas de algoritmos para intentar obtener el que pudiera dar mejores predicciones acerca de si una persona volverá a donar o no, en total hubieron más de 16,500 propuestas [15]. Para comprender mejor la competición se puede analizar la participación de Philip Kalinda, *Data Scientist*, la metodología que siguió consistió en dividir los datos en dos conjuntos uno de entrenamiento y el otro de pruebas. A los datos de entrenamiento le aplicó algunos modelos como *Logistic Regression*, *Support Vector Machines*, *ADA Boost*, *XG Boost* y *Random Forest*. Una vez que obtuvo los resultados de cada uno de ellos calculó el área bajo la curva para obtener el método que le generó los mejores resultados y ese fue *Random Forest* como se puede observar en la Figura 2-9.

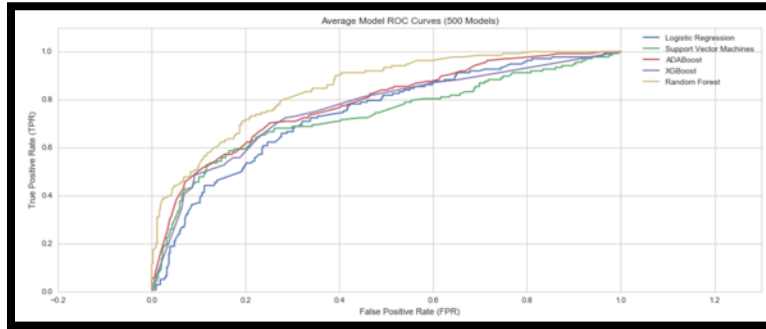


Figura 2-9 Área bajo la curva de los distintos modelos [16]

### 2.4.2. Investigaciones

Bahel et al. [17] realizó un análisis sobre los distintos enfoques que se le han realizado al *dataset* del Prof. I-Cheng Yeng y con los resultados obtenidos determinó una estrategia diferente que consistió en aplicar el algoritmo *K-Means* para generar clústers y también mejorar la elaboración de los clústers manipulando la información con una versión mejorada de *Naive Bayes*.

Durante su investigación encontró que otros investigadores utilizaron algoritmos como Redes Neuronales Artificiales y Probabilísticas, *Clustering*, Clasificación, Árboles de Decisión, *Cross Validation*, *Random Forest*, *Logistic Regression*, entre otros. Con la ayuda de este análisis encontró que existía mucha correlación entre algunas variables de los datos por lo cual decidió sólo trabajar con F (número total de donaciones), T (meses desde la primera donación) y R (meses desde la última donación) para predecir DM (donará en marzo de 2007).

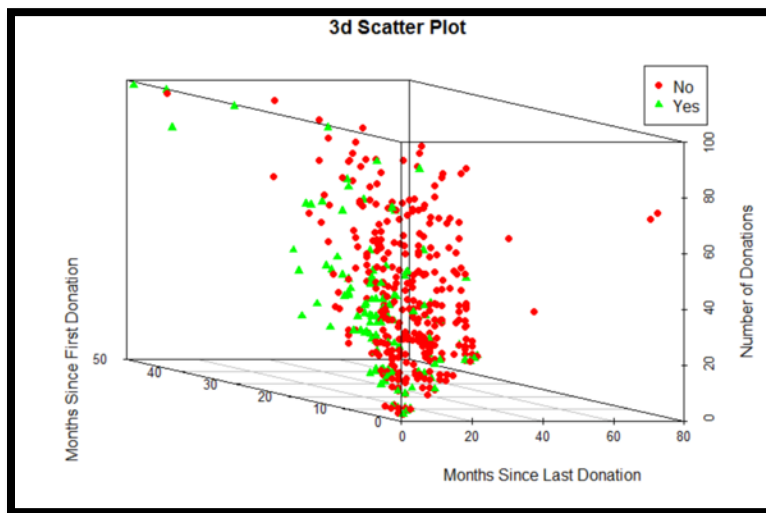


Figura 2-10 Imagen 3D con la distribución de los pacientes que han donado o no [17]

En la Figura 2-10 podemos observar que los datos se encuentran muy relacionados en el espacio y no existen grandes diferencias entre su posición por lo cual el autor decidió utilizar *Cross Validation* para obtener distintos conjuntos de la información de entrenamiento y de pruebas y así realizar algunas iteraciones al algoritmo de *K-Means* y poder garantizar que los resultados no estuvieran siendo manipulados por cierta parte del *dataset* y que si estuviera considerando la mayor parte de la información

en los resultados. Por último, determinó con distintos valores de K y siguiendo la metodología del codo (*elbow method*) que el número adecuado para ejecutar el algoritmo era  $K=4$ , como podemos observar en la Figura 2-11.

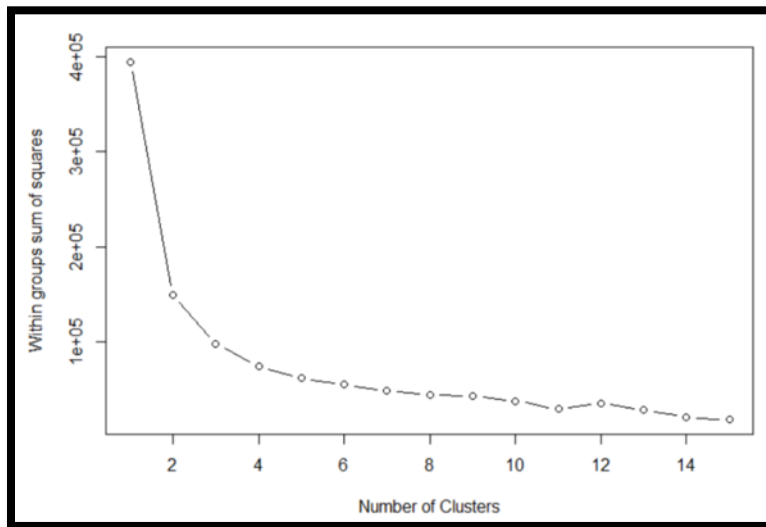


Figura 2-11 Método del codo para K-Means [17]

---

## 3. MARCO TEÓRICO/CONCEPTUAL

---

**Resumen:** *En este capítulo se presentan las bases teóricas y conceptuales sobre la sangre para comprender su composición y la importancia que tiene, además se explica detalladamente el proceso de la donación de sangre y se analizan estadísticas sobre México. También, el capítulo ayudará a entender las distintas generaciones sociales y las tecnologías utilizadas para llevar a cabo este TOG.*

## 3.1. Sangre

### 3.1.1. ¿Qué es la sangre?

La sangre es un tejido que se encuentra en todo el cuerpo, proveyéndolo de nutrientes, oxígeno, eliminando los desechos que se van generando por las células, transportando hormonas de una parte del cuerpo a otra y regulando los niveles de acidez (pH) y temperatura del cuerpo. En su mayoría es líquida, con gran cantidad de células y proteínas suspendidas en ella que la hacen más ligera que el agua pura [18].

En promedio una persona tiene 5 litros de sangre en el cuerpo, que dependiendo del sistema circulatorio se va adaptando a las diferentes necesidades, por ejemplo, cuando se realiza ejercicio el corazón late más fuerte y rápido por lo cual se necesita proveer de más sangre y oxígeno a los músculos. También, cuando se tiene una infección la sangre comienza a generar más glóbulos blancos en el lugar de la infección para poder atacarla.

La sangre está compuesta por un 60% de plasma y un 40% de glóbulos rojos, glóbulos blancos y plaquetas, en la Figura 3-1 podemos observar la distribución de cada uno de los componentes.

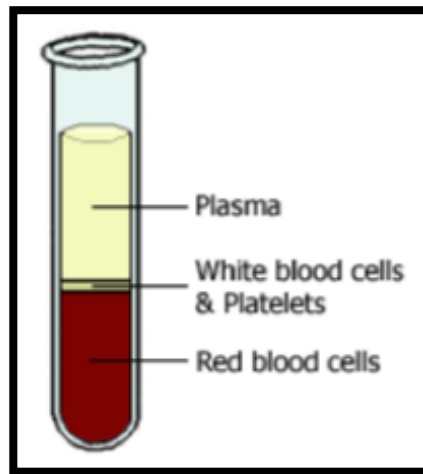


Figura 3-1 Distribución de los componentes de la sangre [18]

El plasma en su mayoría es agua, pero también está conformado por proteínas, dióxido de carbono, glucosa, vitaminas y grasas. Por otra parte, los glóbulos rojos, los glóbulos blancos y las plaquetas se generan en la médula ósea. Los glóbulos rojos, también conocidos como eritrocitos, son los que se encargan de transportar oxígeno y son los más comunes de encontrar en la sangre. Los glóbulos blancos, también conocidos como leucocitos, son parte del sistema inmune y se encargan de formar defensas en contra de enfermedades o infecciones, los leucocitos circulan por toda la sangre hasta que reciben una señal de que una parte del cuerpo está dañada. Por último, las plaquetas son fragmentos de formas irregulares que circulan en la sangre hasta que tienen que actuar para formar un coágulo de sangre y detener el sangrado [18].

### 3.1.2. Grupos sanguíneos

El grupo sanguíneo es determinado por los genes heredados por nuestros padres, es importante considerar estos grupos sanguíneos al momento de las transfusiones de sangre debido a que algunos antígenos podrían disparar el sistema inmunológico del paciente para atacar la sangre de la transfusión ya que podría ser no compatible con la del paciente. En la Figura 3-2 se puede observar una tabla de compatibilidad de los diferentes grupos sanguíneos.

Existen cuatro principales grupos sanguíneos determinados por la presencia o ausencia de dos antígenos – A y B – en la superficie de los glóbulos rojos. En adición a los antígenos A y B, hay una proteína llamada factor Rh, el cual puede estar presente (+) o ausente (-), creando los 8 tipos de sangre más comunes (A+, A-, B+, B-, O+, O-, AB+, AB-) [19].

- **Grupo A:** Sólo tiene antígenos A en los glóbulos rojos y anticuerpos B en el plasma.
- **Grupo B:** Sólo tiene antígenos B en los glóbulos rojos y anticuerpos A en el plasma.
- **Grupo AB:** Sólo tiene antígenos A y B en los glóbulos rojos y no tiene anticuerpos de ningún tipo en el plasma.
- **Grupo O:** Tiene anticuerpos de tipo A y B en el plasma y no tiene antígenos en los glóbulos rojos.

Grupo sanguíneo	puede donar a				puede recibir de			
A+	A+	AB+			O+	O-	A+	A-
A-	A+	AB+	A-	AB-	O-	A-		
B+	B+	AB+			O+	O-	B+	B-
B-	B+	B-	AB+	AB-	O-	B-		
AB+ (Receptor universal)	AB+				Todos			
AB-	AB+	AB-			AB-	A-	B-	O-
O+	A+	B+	AB+	O+	O+	O-		
O- (Donante universal)	Todos				O-			

Figura 3-2 Taba de compatibilidad entre los grupos sanguíneos [20]

## 3.2. Donación de sangre

La donación de sangre es un proceso médico en el que se le extraen al donante entre 405 a 450 mililitros de su sangre, a través de una punción en la vena del pliegue del brazo, para ello se utiliza un equipo estéril, nuevo y desechable.

El proceso de donación que recomienda la OMS consta de los siguientes pasos:

1. Registro del donador.
2. Breve introducción antes de la donación en la que se busca dar a conocer al donador los requisitos de la donación y en qué consiste el proceso.
3. Completar un cuestionario médico.
4. Entrevista para repasar las respuestas del cuestionario, aclarar dudas y realizar algunas preguntas extra.

5. Evaluación médica al donador para conocer su estado de salud y ver si está en las condiciones recomendadas para realizar la donación.
6. Solicitar consentimiento del donador para poder analizar la sangre obtenida y una última oportunidad para que confirme que toda la información otorgada en los pasos previos es verídica y su sangre es segura.

### *3.2.1. Requisitos*

De acuerdo a la OMS [21], la mayoría de la gente es candidata a donar sangre, sólo necesitan estar saludables y cumplir con una serie de criterios básicos como:

- Tener entre 18 y 65 años. En algunos países la legislación permite a jóvenes donar a partir de los 16 años con el consentimiento de sus padres.
- Pesar al menos 50 kg. Algunos países aceptan donadores desde 45 kg para donar 350 ml.
- Tener los niveles de hemoglobina mayor a 12.5 g/dl en el caso de las mujeres y mayor a 13.5 g/dl en el caso de los hombres.
- Tener el pulso, temperatura corporal y la presión sanguínea en rangos aceptables.

No es posible donar sangre si:

- Se tiene un resfriado, gripe, dolor de garganta, herpes labial, infección en el estómago o cualquier otra infección.
- El donador se ha realizado un tatuaje y aún no han pasado más de 6 meses desde el procedimiento.
- Se realizó un procedimiento dental menor se deben esperar 24 horas y por un trabajo más complejo se debe esperar mínimo un mes.
- El donador ha viajado a un país donde las infecciones por mosquito como malaria, dengue o zika son endémicas.
- En los 12 meses anteriores se tuvieron relaciones sexuales de riesgo o con múltiples parejas.
- Alguna vez se obtuvo positivo en una prueba de VIH.
- Alguna vez se tuvo Hepatitis B o C, Sífilis, Chagas o Herpes.
- El donador ha estado bajo el efecto de drogas intravenosas o inhaladas.
- En caso de las mujeres, no deben encontrarse embarazadas, en lactancia o en menstruación.

### *3.2.2. Tipos de donadores*

Los donadores se pueden clasificar en tres categorías dependiendo de las motivaciones por las cuales donen:

- **Donador voluntario:** Es aquel cuya principal motivación es la de poder ayudar a otras personas a través de la donación de cualquier componente sanguíneo: sangre, plasma o plaquetas. La donación se realiza de manera voluntaria sin recibir ningún pago o recompensa, en algunos bancos de sangre se ofrecen refrigerios al terminar la donación o símbolos de agradecimiento como una sombrilla, un bolígrafo con el logo del banco de sangre, entre otras cosas, pero esto no es considerado como un pago o recompensa. Regularmente este tipo de donadores son los que recomienda la OMS debido a que no se encuentran bajo presión para donar sangre por lo cual tienden a no mentir sobre situaciones que los pudieran llevar a no ser aceptados y la mayor parte

de las veces reúnen los criterios necesarios para la donación. También, diferentes estudios demuestran que la sangre de estos donadores es la menos propensa a transmitir infecciones ya que este tipo de donadores cuidan sus hábitos para seguirse manteniendo como donadores a lo largo del tiempo y además su sangre se ha evaluado varias veces en cada donación realizada. Es importante mantener un registro de este tipo de donadores porque ellos pueden ayudar a garantizar el suministro de sangre de un país debido a que se encuentran donando constantemente.

- **Donador de reposición o reemplazo:** Es aquella donación que surge ante la solicitud de un familiar, amigo o conocido que necesitar ser transfundido o sometido a alguna intervención quirúrgica. En estos casos el paciente es el que se encarga de buscar a los donadores de reposición y dependiendo del tratamiento el hospital solicita cierto número de donadores. La desventaja de este tipo de donadores es que la necesidad de donar los puede llevar a ocultar información que pudiera hacer que los rechazaran y esto puede causar que su sangre tenga alguna infección y se desechen en las pruebas posteriores a la donación. En algunos casos cuando el paciente no junta el total de donadores solicitados puede recurrir a pagar para conseguir donadores, pero esta circunstancia podría generar los mismos problemas por la necesidad de realizar la donación para obtener la remuneración.
- **Donador remunerado:** Es aquella persona que dona sangre a cambio de un pago que en la mayoría de las ocasiones es monetario. Las desventajas de este tipo de donadores son similares a las de los donadores de reposición ya que la necesidad de querer el beneficio podría llevarlos a ocultar información que es necesaria para garantizar la salud del receptor de la sangre. Muchos de este tipo de donadores se encuentran en sectores sociales bajos de la población, por lo cual, debido a la falta de educación sobre salud, buenos hábitos y acceso restringido a los servicios de salud son más vulnerables a estar expuestos a condiciones de riesgo que pongan en peligro su salud y sus condiciones para ser considerados como donadores aptos.

En México, se necesita buscar estrategias para fomentar la donación altruista y conservar a los donadores voluntarios no remunerados en un ciclo de donación constante para ir teniendo un suministro adecuado que ayude a satisfacer las necesidades de la población, en la Figura 3-3 se puede observar un artículo de Merca 2.0 [22] en el cual se muestra la cantidad de donadores voluntarios y por reposición en México en los años de 2012 a 2017.

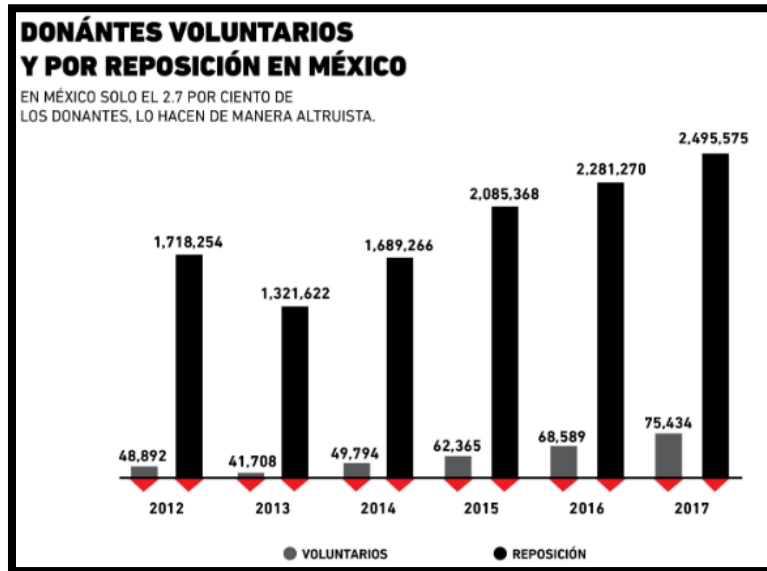


Figura 3-3 Donantes voluntarios y por reposición en México [22]

### 3.2.3. ¿Para qué se utiliza la sangre donada?

Las donaciones de sangre contribuyen a salvar vidas en casos de accidentes, cirugías de emergencia o transfusiones, sin embargo, de acuerdo con información que provee la OMS [23] también son de gran utilidad en los siguientes casos:

- Niños con anemia grave, a menudo causada por el paludismo o la malnutrición.
- Mujeres con complicaciones obstétricas como embarazos ectópicos, hemorragias antes, durante o después del parto.
- Personas con traumatismos graves provocados por las catástrofes naturales y las causadas por el hombre.
- Intervenciones quirúrgicas y médicas de alta complejidad.
- Enfermos de cáncer.
- Trasplantes de órganos.
- Se utiliza para realizar transfusiones periódicas en personas afectadas por enfermedades como la talasemia o la drepanocitosis.
- Elaboración de diversos productos, por ejemplo, factores de coagulación para hemofílicos.

Es importante mencionar que en la mayoría de las ocasiones y dependiendo de la infraestructura de cada banco de sangre, clínica u hospital, la sangre se separa por componentes – glóbulos rojos, plaquetas y plasma – para poder darle a cada paciente el componente que necesita y también obtener el mayor provecho de cada unidad de sangre obtenida.

### 3.2.4. Beneficios de donar

El proceso de donación de sangre no sólo le genera beneficios al receptor, también ayuda a mantener la buena salud del donador. Los beneficios para el donador son:

- Se reduce el riesgo de contraer hemocromatosis debido a que la donación de sangre ayuda a nivelar los niveles de hierro. La hemocromatosis es causada debido a niveles de hierro altos en el cuerpo.
- Reduce el riesgo de enfermedades cardíacas y hepáticas.
- Estimula la producción de nuevas células sanguíneas en el cuerpo.
- Depura los triglicéridos
- Sin embargo, el beneficio más relevante es en factor de la salud emocional. Este tipo de eventos altruistas en los que se dona algo de uno mismo, en este caso la propia sangre, tiene beneficios positivos en la felicidad ya que se genera una sensación de solidaridad y bienestar cuando se termina de realizar la donación.

### 3.2.5. Donaciones de sangre en México

Es interesante analizar las tendencias de las donaciones de sangre en México en los últimos años para poder comprender los resultados de las labores y campañas que se realizan en el país y el impacto que provocan en la sociedad. A continuación, se analizará un conjunto de datos recopilados por el Centro Nacional de la Transfusión Sanguínea, esta institución forma parte de la Secretaría de Salud de México, donde se puede observar la cantidad de donadores que ha existido cada año durante los años de 2005 y hasta 2017 en los bancos de sangre de cada una de las 32 entidades federativas que integran la república mexicana [24]. Los bancos de sangre que se consideran son:

- Cruz Roja.
- Centros Estatales de Transfusión Sanguínea.
- Secretaría de la Defensa Nacional.
- Desarrollo Integral para la Familia.
- Instituto Mexicano del Seguro Social.
- Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado.
- Privado.
- Servicios Médicos de la Entidad.
- Secretaría de Salud.
- Universitarios.

Las 5 entidades federativas con mayor número de donaciones son:

Entidad federativa	Número de donaciones (2005 - 2017)
<b>Ciudad de México</b>	5,034,129
<b>Jalisco</b>	1,695,478
<b>Estado de México</b>	1,380,725
<b>Nuevo León</b>	1,318,405
<b>Veracruz</b>	1,062,745

Tabla 3-1 Top 5 de entidades federativas con mayor número de donaciones

Las 5 entidades federativas con el menor número de donaciones son:

Entidad federativa	Número de donaciones (2005 - 2017)
Campeche	146,138
Colima	162,647
Baja California Sur	179,610
Tlaxcala	182,021
Nayarit	213,003

Tabla 3-2 Entidades federativas con el menor número de donaciones

Debido a que el trabajo se realizó considerando a la población de la ciudad de Guadalajara, se analizará más a fondo la tendencia de donaciones durante estos años en el estado de Jalisco.

Año	Donaciones de sangre en Jalisco
2005	113,504
2006	114,121
2007	112,120
2008	116,876
2009	120,722
2010	120,376
2011	128,778
2012	135,892
2013	141,024
2014	84,614
2015	69,615
2016	268,487
2017	169,349

Tabla 3-3 Donaciones de sangre en Jalisco

Como se puede observar de la información de la Tabla 3-3, se tiene un promedio de 130,421 donaciones anuales en el estado de Jalisco, estas donaciones consideran las donaciones voluntarias y las de reemplazo, aunque pudiera parecer una cantidad significativa de donaciones anuales si se considera el total de la población durante esos años, el porcentaje de donaciones resulta muy bajo.

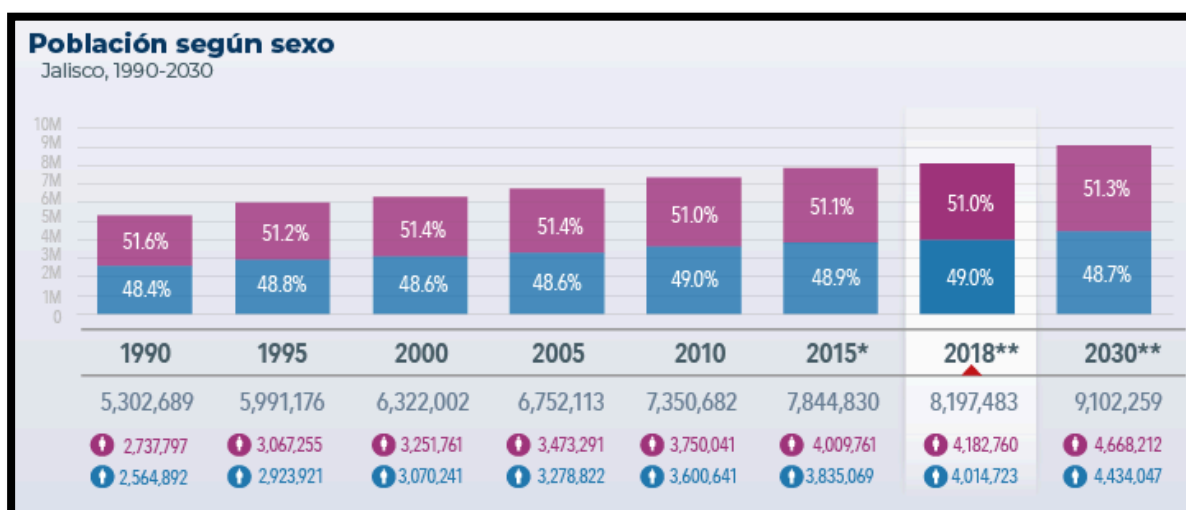


Figura 3-4 Población en Jalisco [25]

Considerando la información del Instituto de Información Estadística y Geográfica que se muestra en la Figura 3-4, se obtiene que el porcentaje de donaciones en los años 2005, 2010 y 2015 fue:

Año	Población	Donaciones	Porcentaje
2005	6,752,113	113,504	1.68 %
2010	7,350,682	120,376	1.64 %
2015	7,844,830	69,615	0.89 %

Tabla 3-4 Porcentaje de las donaciones de sangre en Jalisco respecto a la población total

Con los porcentajes anteriores se observa que Jalisco, siendo una de las 5 entidades federativas con mayor número de donaciones de sangre en México, está por debajo del porcentaje que recomienda la OMS para que un país pueda tener un suministro adecuado de sangre para las enfermedades y emergencias. Además, es importante mencionar que, aunque el número de donaciones se encuentra estable con el paso de los años es necesario aumentarlo porque la población está en constante crecimiento y parece que las donaciones de sangre no están siguiendo ese crecimiento. En la Figura 3-5 se puede observar el total de donaciones sanguíneas en los últimos años y la ecuación lineal que ejemplifica la tendencia de las donaciones.

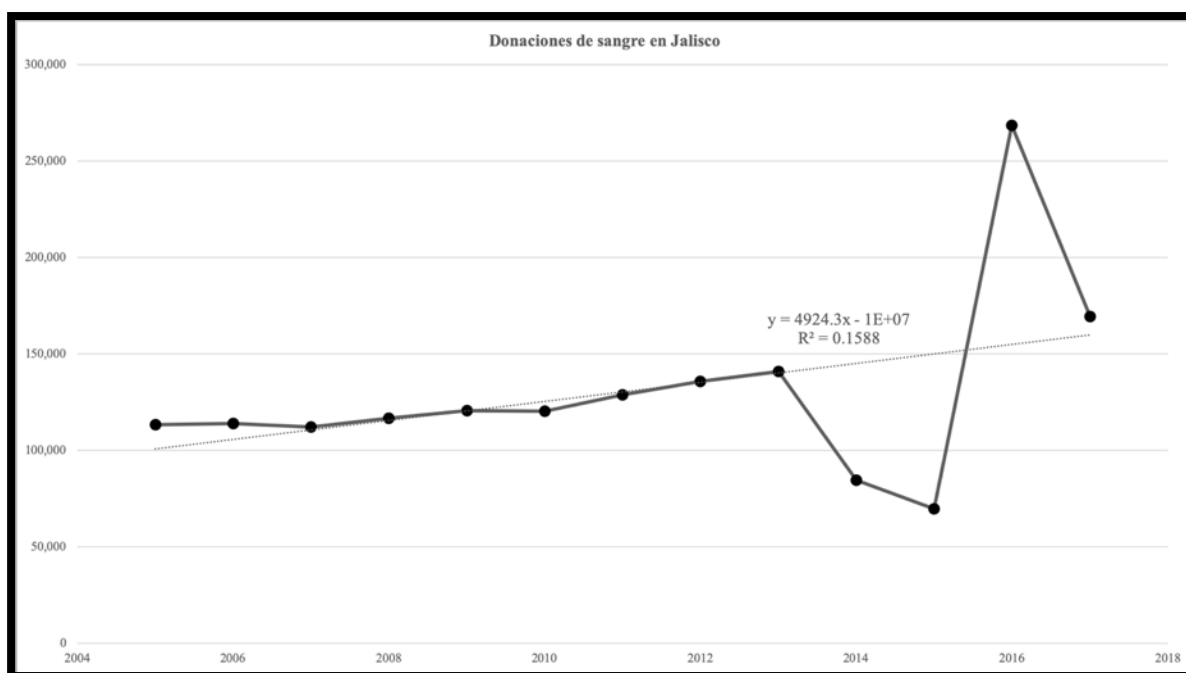


Figura 3-5 Gráfica de donaciones en Jalisco con línea de tendencia

### 3.3. Generaciones sociales

Denominar ciertas generaciones sociales ayuda a observar similitudes en actitudes y puntos de vista de grupos de personas que nacieron en un tiempo similar y como ciertos eventos mundiales, tecnológicos, económicos y sociales influyen con su ciclo de vida y la percepción que las personas van formando del mundo. Por ejemplo: la opinión que tiene un *millennial* y un *baby boomer* sobre el matrimonio entre personas del mismo sexo probablemente sea diferente.

Analizar este tipo de tendencias ayuda a conocer la sensibilidad que van desarrollando las personas ante los problemas sociales y cómo colaboran, como generación, en buscar revertir los problemas sociales y naturales a los que se enfrentan como sociedad.

De acuerdo con la organización estadounidense *Pew Research Center*, que se encarga de realizar investigaciones en temas sociales, de opinión pública y tendencias demográficas, las distintas generaciones sociales que han existido en los últimos años se pueden clasificar como se observa en la Figura 3-6.

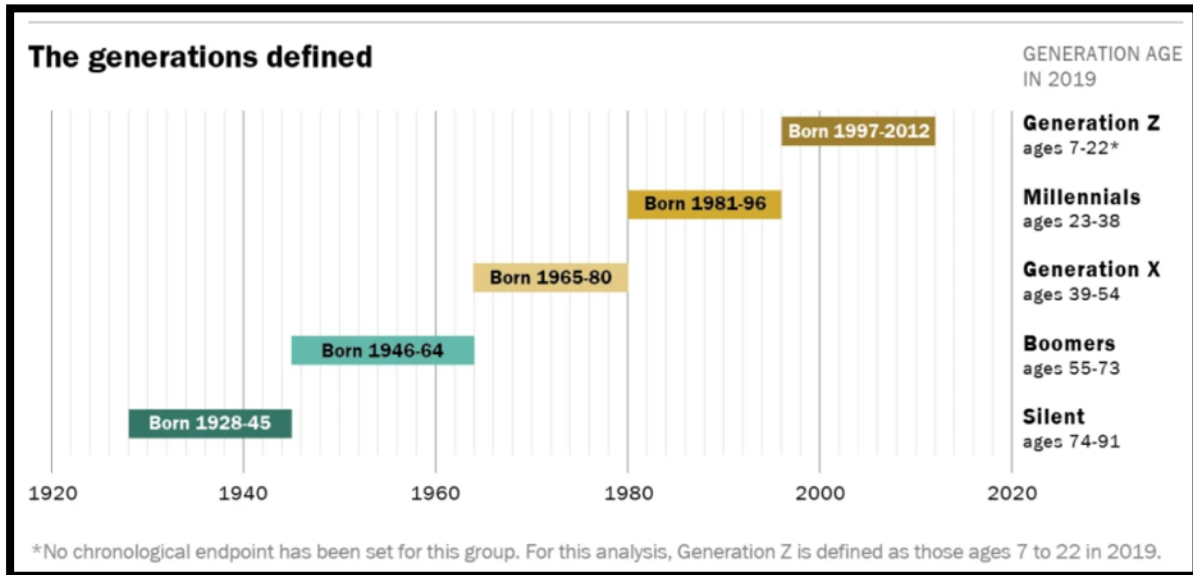


Figura 3-6 Generaciones sociales [26]

### 3.3.1. Generación silenciosa

Los integrantes de esta generación nacieron entre 1928 y 1945. Esta generación se caracteriza por haber sido marcada por acontecimientos con gran impacto mundial como la Gran Depresión de Norteamérica que afectó las economías mundiales, el Nazismo y la Segunda Guerra Mundial. Es una generación que debido a los problemas económicos de la época se hizo cuidadosa con sus gastos y buscan limitar sus compras a lo necesario. En México es una generación en la que el 28% ahorra dinero cada mes y están seguros sobre su futuro financiero. Esta generación tiene como principal fuente de información la televisión, periódicos, revistas, etcétera.

### 3.3.2. Generación baby boomers

Los integrantes de esta generación nacieron entre 1946 y 1964. A esta generación la acompañan sucesos importantes como la llegada del hombre a la luna, el movimiento por los derechos civiles, movimientos feministas, la liberación sexual, entre otros. Esta generación recibió este nombre debido a que ha registrado la mayor tasa de nacimientos, es una generación que ha buscado la estabilidad en sus trabajos, aproximadamente un 74% se encuentra satisfecho con su ocupación actual y también se encuentran interesados en ahorrar para su futuro. Dedicar gran parte de su tiempo al cuidado de sus padres e hijos y

a pesar de que han visto varios adelantos tecnológicos generalmente no están tan actualizados y les cuesta trabajo interactuar con ellos.

### 3.3.3. *Generación X*

Los integrantes de esta generación nacieron entre 1965 y 1980. Esta generación vio los sucesos de la Guerra Fría y los inicios de las computadoras personales. También se les conoce como *latchkey*, término en inglés que caracteriza a los niños que regresan de la escuela a una casa vacía o se tenían que quedar solos la mayor parte del tiempo debido a que sus padres se encontraban trabajando. Si bien esta generación aún sigue utilizando medios de información como revistas, periódicos, radio y televisión es una generación que también dedica demasiado tiempo a redes sociales como Facebook.

### 3.3.4. *Generación millennials*

Los integrantes de esta generación nacieron entre 1981 y 1996. Esta generación vivió importantes acontecimientos como la Gran Recesión, el ataque a las Torres Gemelas en Estados Unidos (9/11) y la explosión tecnológica del Internet y las redes sociales. Estos acontecimientos hacen que esta generación dedique parte de sus ingresos a los dispositivos móviles y computadoras y es algo con lo que se sienten cómodos. Respecto a lo económico, es una generación que se encuentra pagando las deudas de su educación y a pesar de que consiguen buen crecimiento profesional deciden pasar eventos como comprar casa, casarse o tener una familia a años por venir y no en su presente, en vez de eso dedican su tiempo a viajar y salir con sus amigos.

### 3.3.5. *Generación Z*

Los integrantes de esta generación son todos aquellos que nacieron a partir de 1997. Lo que caracteriza a esta generación es que desde muy pequeños han tenido acceso a Internet y a dispositivos móviles por lo cual quieren utilizarlos para resolver la mayoría de sus conflictos en la vida. Sus principales objetivos son hacer dinero, tener una carrera satisfactoria, estar en forma y saludables y dedicar tiempo a su familia. Esta generación ha visto los problemas financieros de sus padres (Generación X) y como los *Millennials* enfrentan sus deudas por lo cual intentan ser más precavidos con sus gastos y tener un mejor manejo con sus finanzas, se dice que un 47% ahorra dinero sin saber para qué lo podría necesitar.

## 3.4. Machine learning (Aprendizaje automático)

*Machine Learning* es una rama de la Inteligencia Artificial. Mediante el uso de la informática, se diseñan sistemas que pueden aprender de los datos de forma capacitada. Los sistemas pueden aprender y mejorar con la experiencia y, con el tiempo, refinar un modelo que puede ser utilizado para predecir los resultados de las preguntas en función del aprendizaje previo [27].

Existen distintos tipos de algoritmos que se pueden emplear en *Machine Learning*, principalmente se clasifican en dos categorías: algoritmos de aprendizaje supervisado y no supervisado.

### 3.4.1. *Aprendizaje supervisado*

Los algoritmos de esta categoría se caracterizan por recibir un entrenamiento con información etiquetada, es decir que a cada entrada o registro del *dataset* le corresponde un valor de salida, lo que va a ayudar al algoritmo a aprender de los datos de entrenamiento y poder clasificar futuras entradas con lo aprendido de

la información de entrenamiento. Los dos principales usos para los algoritmos de estas categorías son: clasificación o regresión.

Es importante evaluar la cantidad y tipo de información con la que se entrena el algoritmo ya que puede quedar entrenado para sólo identificar cierto tipo de información y todo categorizarlo de la misma manera o pudiera tener poco entrenamiento que la clasificación que realice, en la mayoría de las ocasiones, sería errónea.

### 3.4.2. *Aprendizaje no supervisado*

Este tipo de algoritmos se utilizan normalmente para hacer *clustering*. Trabajan sobre un conjunto de datos de los cuales no se conoce nada al respecto, es decir, no se tiene una etiqueta o algo para identificar la información o los datos. El algoritmo no supervisado busca encontrar patrones ocultos de los datos para poder crear clústeres a partir de sus descubrimientos.

Este tipo de algoritmos son realmente útiles en análisis de exploración porque pueden identificar estructura o relaciones en los datos y también dan grandes resultados cuando se busca reducir dimensiones de un conjunto de datos.

## 3.5. Modelos de regresión

El análisis de regresión es un modelo estadístico que posee un conjunto de técnicas que ayudan a investigar la relación que existe entre un predictor (que también es conocido como variable independiente, variable explicativa o variable de control y comúnmente se denotan con la letra  $x$ ) y una o mas variables responsivas (que también son conocidas como variables dependientes, variables explicadas y comúnmente se denotan con la letra  $y$ ).

De acuerdo con Yan et al. [28], hay tres tipos de regresiones. La primera es la regresión lineal simple que se utiliza para modelar la relación entre dos variables. Una de ellas es la variable dependiente  $y$  y la otra es la variable independiente  $x$ . Por ejemplo, la regresión lineal simple podría ayudar a modelar la relación entre el peso de las personas con su estatura, o un modelo sobre la cantidad de horas de estudio y los resultados de un examen, etcétera. El modelo de regresión lineal simple a menudo se escribe con la siguiente fórmula:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

donde  $y$  es la variable dependiente,  $\beta_0$  es la intercepción con  $y$ ,  $\beta_1$  es el gradiente o la pendiente de la línea de regresión, es uno de los valores más importantes en un análisis de regresión debido a que un valor muy cercano a 0 indica que la relación es poca o nula y un valor grande indica que existe una amplia relación entre las variables ya sea positiva o negativa depende del valor,  $x$  es la variable independiente, y  $\varepsilon$  es el error aleatorio. Usualmente se asume que el error  $\varepsilon$  es distribuido normalmente con  $E(\varepsilon) = 0$  y una varianza constante  $Var(\varepsilon) = \sigma^2$  en la regresión lineal simple [28].

El segundo tipo de regresión es la regresión lineal múltiple que consiste en un modelo de regresión lineal que tiene una variable dependiente y más de una variable independiente. Este tipo de regresiones pueden ayudar a modelar escenarios como, por ejemplo, encontrar la cantidad de alumnos que desertan de la escuela si se tienen como datos la cantidad de asistencias, las calificaciones, si tienen o no una beca, entre otros datos. La regresión lineal múltiple asume que la variable dependiente es una función lineal de los

parámetros del modelo y existe más de una variable independiente en el modelo. La forma general para este tipo de regresiones es de la siguiente manera:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$$

donde  $y$  es la variable dependiente,  $\beta_0, \beta_1, \beta_2, \dots, \beta_p$  son los coeficientes de regresión, y  $x_1, x_2, \dots, x_n$  son variables independientes en el modelo. En una configuración clásica de la regresión lineal múltiple usualmente se asume que el error  $\varepsilon$  sigue una distribución normal con  $E(\varepsilon) = 0$  y una varianza constante  $Var(\varepsilon) = \sigma^2$ .

La regresión lineal múltiple implica más problemas que la regresión lineal simple, como la colinealidad, la inflación de la varianza, la visualización gráfica del diagnóstico de regresión y la detección de la regresión atípica [28].

El tercer tipo de regresión son las regresiones no lineales, las cuales se caracterizan porque la relación que existe entre la variable dependiente e independiente no es lineal cuando se toman en cuenta los parámetros de la regresión. Algunos tipos de modelos que entran en esta categoría son:

- Regresión polinómica.
- Regresión logarítmica.
- Regresión asintótica.
- Regresión exponencial.

### 3.6. Python

Es un lenguaje de programación de propósito general, de alto nivel e interpretado que fue creado por Guido van Rossum y se lanzó en 1991. Python puede ser utilizado con distintos paradigmas de programación como el procedural, orientado a objetos y funcional. Es un lenguaje que debido a su sintaxis y a la amplia cantidad de librerías que se han desarrollado se caracteriza por ser fácil de leer y de dar mantenimiento. Es sencillo de aprender y tiene compatibilidad con distintos sistemas operativos y plataformas.

Debido a las ventajas antes mencionadas, Python puede ser utilizado para:

- Crear aplicaciones web de lado del servidor.
- Crear flujos o *jobs* de sistemas existentes.
- Realizar cálculos matemáticos complejos.
- Junto con ciertas librerías se puede utilizar para manejar grandes cantidades de información y realizar operaciones y cálculos sobre esos conjuntos de datos.

### 3.7. NumPy

Es un paquete fundamental que se utiliza para cómputo científico con Python. Algunas de sus características principales son:

- Posee una poderosa manera para manejar arreglos de N-dimensiones.
- Tiene un conjunto de sofisticadas funciones que se pueden aplicar a todo el conjunto de datos.
- Se integra con otros lenguajes como C y C++.

- Útil para cálculos matemáticos como álgebra lineal y transformada de Fourier, además de que posee un algoritmo muy adecuado para calcular números aleatorios.

Además de sus usos científicos, NumPy puede ser utilizado como un contenedor multi-dimensional eficiente de información genérica, lo que permite que se pueda integrar sin problemas y rápidamente con una amplia variedad de base de datos [29].

### 3.8. Python Data Analysis Library (pandas)

Es una librería de código abierto y alto rendimiento que facilita el uso de estructuras de datos y tiene distintas herramientas para llevar a cabo la preparación y análisis de los datos, lo que permite que se dedique menos tiempo al manejo de la información previo a ejecutar un algoritmo sobre ella. Pandas se caracteriza debido a que posee un objeto llamado **DataFrame** que es muy eficiente para manipular grandes cantidades de información y se encuentra indexado para facilitar ciertas operaciones y además cuenta con distintas funciones para escribir y leer información entre las estructuras de datos almacenadas en memoria y diferentes formatos: CSV, archivos de texto, archivos de Microsoft Excel, bases de datos, entre otros.

Python con pandas es utilizado en una amplia variedad de dominios académicos y comerciales, incluyendo Finanzas, Neurociencia, Economía, Estadística, Publicidad, *Web Analytics*, y más [30].

### 3.9. Matplotlib

Es una librería que permite crear gráficas en 2D con Python, es ampliamente utilizada debido a la calidad de las gráficas y a que es sencillo de utilizar. Con esta librería se pueden realizar histogramas, gráficas de barras, gráficas de errores, gráficas de puntos, gráficas circulares, tablas, entre otros.

En la figura 3-7 se pueden observar algunas gráficas que se pueden generar utilizando esta librería.

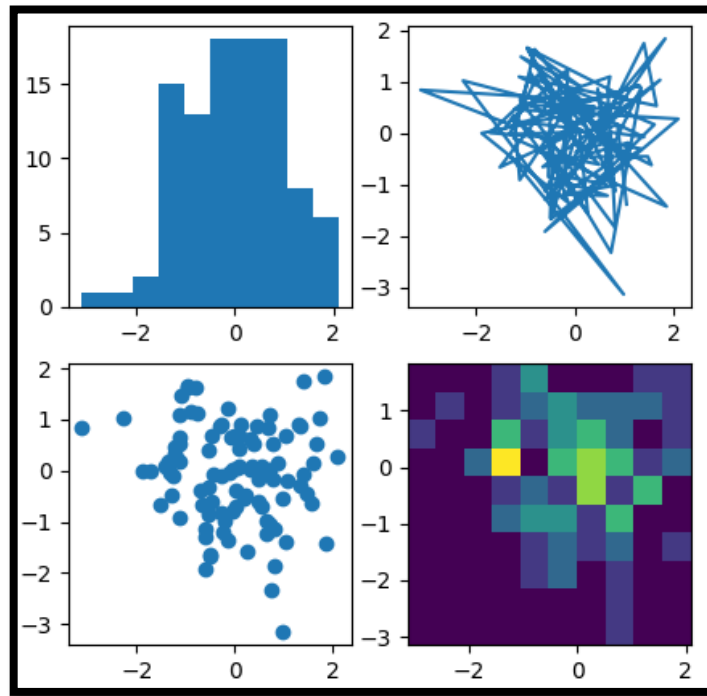


Figura 3-7 Ejemplos de gráficas con matplotlib [31]

### 3.10. Scikit-learn

Es una librería muy popular para *Machine Learning* debido a que posee una amplia cantidad de herramientas eficientes y simples de utilizar para hacer minería y análisis de datos. *Scikit-learn* fue desarrollada por David Cournapeau en el 2007 y a partir de ahí se hizo un proyecto libre donde más programadores comenzaron a realizar aportaciones, en la actualidad la librería tiene una amplia gama de algoritmos supervisados y no supervisados.

Una de las grandes ventajas de la librería es que se ha desarrollado utilizando algunas de las librerías más comunes de Python para manejar datos y realizar operaciones matemáticas, por lo cual, se puede integrar fácilmente con librerías como: NumPy, Matplotlib, Pandas, SciPy, entre otras.

En la Figura 3-8 se observan distintos ejemplos de los algoritmos incluidos en esta librería pero algunos de los algoritmos más utilizados son:

- Regresión.
- Árboles de decisión.
- Redes Neuronales.
- *Clustering*.
- *Support Vector Machine*.
- *Naive Bayes*.

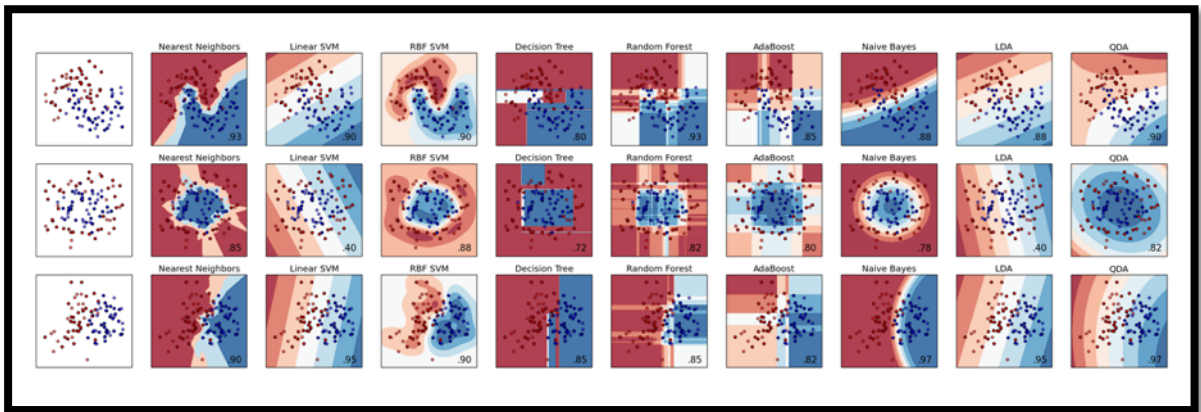


Figura 3-8 Algoritmos incluidos en *scikit-learn* [32, 23]

---

## 4. DESARROLLO METODOLÓGICO

---

**Resumen:** *En este capítulo se presenta en detalle el desarrollo metodológico que incluye una introducción a la encuesta, cómo se encuentra conformada, la estrategia de obtención de los datos de las encuestas a través de un muestreo estadístico, el análisis de los datos obtenidos y los valores que se tomarán en consideración para, posteriormente, realizar el análisis estadístico y predictivo.*

## 4.1. Encuesta – Donación de sangre

Esta encuesta tiene el objetivo de obtener información que permita conocer, a grandes rasgos, las características de un donador recurrente en México y en particular en la ciudad de Guadalajara. Las preguntas contenidas en esta encuesta nos ayudarán a conocer las generaciones sociales que tienen más participación activa en las donaciones de sangre, determinar si se tienen más donantes hombres o mujeres, ayudará a conocer las principales razones que motivan a las personas a donar, el número de veces que han donado y cómo se sienten después de realizar una donación y respecto a las personas que no son donadores ayudará a encontrar algunas de las principales razones por las cuales no se motivan a donar o las causas por las cuales fueron considerados no aptos para realizar una donación.

### 4.1.1. Estructura de la encuesta

La encuesta consta de cuatro secciones con doce preguntas en total, en las que se busca conocer la información general de la persona, información sobre las donaciones que ha realizado y en caso de que no haya donado entender las causas, las secciones son las siguientes:

- **Información general:** Esta sección busca obtener información básica sobre la persona como el año en que nació para poder categorizarlo dentro de una de las generaciones sociales que se han analizado en este trabajo, su género debido a que se ha encontrado cierta tendencia que muestra mayor participación de los hombres y el grupo sanguíneo para conocer cuál es el más común dentro de la población debido a que suele variar dependiendo de la clasificación racial de cada persona.
- **Donador:** En esta sección se busca conocer qué tan frecuente la persona realiza donaciones a través de una aproximación entre su primera donación, la última donación y el total de donaciones que realizó en ese periodo de tiempo. También, busca identificar el motivo que llevó a la persona a donar la primera vez, ya sea una donación altruista o una donación de reemplazo y, por último, se busca conocer que sensación generó la donación de sangre en el donante.
- **No donador:** Esta sección busca entender los motivos por los cuales las personas no han realizado alguna donación o si al momento de asistir al banco de sangre fueron rechazados por algún motivo y qué motivo fue.
- **México y las donaciones:** Por último, esta sección consta de dos preguntas donde se puede ver la opinión social sobre apoyar a una persona que no conocemos ante una situación de emergencia y cuál es la razón por la que México tiene un porcentaje muy reducido de donaciones altruistas desde la perspectiva social.

## 4.2. Obtención de datos

Debido a que no es posible aplicar la encuesta a todos los habitantes de la ciudad de Guadalajara, para la obtención de datos se optó por utilizar los fundamentos del muestreo estadístico el cual se basa en el principio de equiprobabilidad o aleatoriedad, que indica que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño  $n$  tienen la misma probabilidad de ser elegidas [33]. Además, permite realizar un estudio con un costo bajo, gran rapidez de obtención de los datos y obtener resultados de buena calidad de una muestra con individuos aleatorios que pueden representar en gran porcentaje las tendencias, hábitos y actitudes de la población total.

Tomando en consideración la aleatoriedad del muestreo estadístico, se distribuyó la encuesta a través de distintas redes sociales como *Facebook* y *Twitter* con el objetivo de generar un espacio muestral que incluya personas de las generaciones: X, *Millennials* y Z. También, para lograr añadir personas de las generaciones: *baby boomers* y silenciosa, se decidió compartir la encuesta en una lista de distribución de la empresa y se distribuyó a través de grupos familiares, grupos de vecinos, entre otros grupos de temas diversos, a través de aplicaciones como *WhatsApp* y *Telegram*.

### 4.3. Análisis de datos

En esta parte del desarrollo metodológico se busca identificar las preguntas que tienen relación para poder categorizarlas y analizarlas.

La encuesta está conformada de 12 preguntas que permiten conocer características de los donadores, de las personas que no han donado aún y una perspectiva de ambos acerca de cómo ven a México en el tema. La información se dividirá en 3 secciones para el análisis: donadores, no donadores y una sección de análisis de emociones y la perspectiva que tienen de México, para ello cada pregunta se identificará por una clave como se muestra en la Tabla 4-1 y además se observa en qué parte del análisis se utilizará.

	Clave	Pregunta	Donadores	No donadores	Análisis emocional
1	timestamp	Hora en que contestó la encuesta			
2	age	¿En qué año naciste?	X	X	X
3	gender	¿Género?	X	X	X
4	bloodType	¿Cuál es tu grupo sanguíneo?	X	X	X
5	donator	¿Has donado sangre alguna vez?	X		
6	firstDonation	¿Recuerdas cuándo fue la primera vez que donaste?	X		
7	beneficiary	La primera vez que donaste, fue:	X		
8	donationTimes	¿Cuántas veces has donado?	X		
9	lastDonation	¿Cuándo fue la última vez que donaste?	X		
10	feelingAfterDonation	¿Te generó un sentimiento positivo donar sangre?	X		X
11	reasonNotDonate	Si no has donado, ¿a qué se debe?		X	X
12	wouldDonate	Si pudieras donar sangre a un desconocido, ¿lo harías?	X	X	
13	mexicoOpinion	En México menos del 4% de la población dona sangre de manera altruista, ¿a qué crees que se deba?			X

Tabla 4-1 Preguntas y distribución en el análisis

## 4.4. Transformación de datos

Ya que se han identificado las preguntas relacionadas, se analizará la información obtenida con ayuda de *Jupyter Notebook*, aplicación web que permite escribir y compartir documentos que contienen código de programación y, al mismo tiempo, permite ejecutar el código dentro del mismo documento para poder observar los resultados y agregar anotaciones.

El siguiente paso consiste en exportar las respuestas de la encuesta de *Google Forms* a un archivo de tipo *Comma-Separated Values* (CSV) para poder importar la información dentro del ambiente que genera *Jupyter Notebook* y comenzar el análisis. Para poder manipular la información de las respuestas se realizará el *mapping* con las claves de la Tabla 4-1.

En la Figura 4-1 se puede observar el código para realizar la carga de la información y una representación de cómo se cargó la información a un objeto tipo **DataFrame** de la librería *pandas*, de esa manera se podrá manipular y aplicar diferentes operaciones en él.

```
bda_initial_headers = ["timestamp", "age", "gender", "bloodType",
                      "donator", "firstDonation", "beneficiary",
                      "donationTimes", "lastDonation", "feelingAfterDonation",
                      "reasonNotDonate", "wouldDonate", "mexicoOpinion"]

blood_initial_dataset = pd.read_csv('BloodDonationAnalysis.csv', sep=',',
                                   names=bda_initial_headers,
                                   header=0)

print("Data set loaded. Total responses: ", len(blood_initial_dataset))
blood_initial_dataset
```

Data set loaded. Total responses: 515

	timestamp	age	gender	bloodType	donator	firstDonation	beneficiary	donationTimes	lastDonation	feelingAfterDonation	reasonNotDonate	wouldDonate
0	2019/02/07 1:48:02 pm GMT-5	Entre 1981 y 1996	Mujer	B+	Sí	2018-05-30	Algún familiar	1.00	2018-05-30	Sí, me genera satisfacción	NaN	S
1	2019/02/07 8:32:31 pm GMT-5	Entre 1981 y 1996	Mujer	B+	No	NaN	NaN	0.00	NaN	NaN	No he buscado donde hacerlo	S
...	...	...	...	...	...	...	...	...	...	...	...	...
513	2019/04/10 11:48:51 am GMT-5	Entre 1981 y 1996	Mujer	A+	No	NaN	NaN	nan	NaN	NaN	No he tenido la necesidad	Nc
514	2019/04/11 5:11:02 pm GMT-5	Entre 1981 y 1996	Mujer	No sé	No	NaN	NaN	nan	NaN	NaN	No he tenido la necesidad	S

515 rows x 13 columns

Figura 4-1 Código para cargar el *dataset* de las encuestas y generar un **DataFrame**

## 4.5. Limpieza de datos

Con el objetivo de mantener cierto nivel de estandarización en el *dataset* la mayoría de las preguntas se crearon como opción múltiple, sin embargo, para los objetivos de la investigación algunas se mantuvieron como preguntas abiertas por lo cual se obtuvieron distintas respuestas e incluso cuando las personas no recordaban el dato correcto dejaron la pregunta vacía, para que ello no afectara con el trabajo antes de realizar el análisis se normalizó la información para poder manipularla más adelante. Los casos que se arreglaron en esta etapa fueron:

- Cuando una persona no ha donado sangre, las preguntas que corresponden a los donadores se cargaron al *DataFrame* con valores de **NaN** y **0**, esta situación se solucionó otorgando un valor de 0 en los campos con valor **NaN**. De esta manera durante el desarrollo del análisis se puede identificar que un 0 en ciertos campos como *firstDonation*, *beneficiary*, *donationTimes*, *lastDonation* y *feelingAfterDonation* significa que la fila corresponde a una persona que no ha donado.
- Cuando una persona ha donado sangre el valor del campo *reasonNotDonate* fue cargada con valor **NaN**, el cual también se sustituyó por **0** para identificar que la fila corresponde a una persona que ha donado.
- Cuando una persona ha donado sólo una vez, en ocasiones, sólo colocaba el año en el campo de *firstDonation* o *lastDonation*, debido a eso se tuvo que revisar si *donationTimes* era igual a uno y buscar en cuál de los dos campos venía el año de la donación y en el campo vacío colocar el mismo año, ya que ambos campos se utilizan más adelante.
- Cuando una persona sólo ha donado una vez, en ocasiones, no recordaba el año en que realizó la donación. Para solucionar este problema se tomó la media de su rango de edad a la cual se sumaron 18 años, edad que se solicita en México para comenzar a donar, y se generó un valor aleatorio entre el año obtenido y el año de la última donación de la encuesta (2019), con el objetivo de que estos casos simularan aleatoriedad en las predicciones y no descartar ninguna donación. El resultado se agregó como valor a *firstDonation* y *lastDonation*.
- Los valores de *firstDonation* menores a 1946, se cambiarían al límite inferior que era 1946 el cual considera que si alguien nació en 1928 y pasaron los 18 que debe tener para donar, estaría comenzando a partir de 1946 a donar.
- Los valores de *lastDonation* mayores al año presente, 2019, se modificarían al año presente. En caso de que para cierta encuesta el valor de *firstDonation* tuviera un valor entre 1946 – 2019, se utilizaría ese año en vez del año presente.
- Cuando el valor de *firstDonation* tiene un valor de 0 pero la encuesta indica que la persona ha donado en múltiples ocasiones y proveyó un valor para *lastDonation*, en ese caso se calcula el valor para *firstDonation* como el resultado del valor en *lastDonation* menos *donationTimes*. Esto ayuda a realizar un aproximado de que al menos 1 vez al año la persona ha realizado una donación.
- Cuando el valor de *lastDonation* tiene un valor de 0 pero la encuesta indica que la persona ha donado en múltiples ocasiones y proveyó un valor para *firstDonation*, en ese caso se calcula el valor para *lastDonation* como el resultado del valor en *firstDonation* más *donationTimes*. Esto ayuda a realizar un aproximado de que al menos 1 vez al año la persona ha realizado una donación.

En la Figura 4-2 se observa la función que se utilizó para poder llevar a cabo las reglas antes mencionadas y el resultado de este proceso se puede observar en la Figura 4-3 la cual contiene la información del *DataFrame* después de la limpieza de los datos.

```

min_year = get_first_donation_year(first_donation_years)
print("The first donation was on: ", min_year)
max_year = max(last_donation_years)
print("The last donation was on: ", max_year)

The first donation was on: 1980
The last donation was on: 2019

YEARS_GAP = max_year - min_year

# When someone has donated more than once over the years we need to know in which range of years he/she has donated
# blood. With this range we can, randomly, assign a year for those intermediates donations.

def get_years_difference(first_donation_years, last_donation_years, YEARS_GAP, donation_times):
    years_difference = []
    for i in range(0, len(first_donation_years)):
        if (first_donation_years[i] > last_donation_years[i]):
            (first_donation_years[i], last_donation_years[i]) = (last_donation_years[i], first_donation_years[i])
        if (abs(last_donation_years[i] - first_donation_years[i]) > YEARS_GAP):
            if (donation_times[i] == 1):
                max_value = max(last_donation_years[i], first_donation_years[i])
                last_donation_years[i] = max_value
                first_donation_years[i] = max_value
            if (first_donation_years[i] == 0 and donation_times[i] > 1):
                first_donation_years[i] = last_donation_years[i] - donation_times[i]
            if (first_donation_years[i] > 0 and last_donation_years[i] == 0 and donation_times[i] > 0):
                last_donation_years[i] = first_donation_years[i] + donation_times[i]

        years_difference.append(last_donation_years[i] - first_donation_years[i])
    return years_difference

```

Figura 4-2 Función para limpiar irregularidades con las fechas de donación

	timestamp	age	gender	bloodType	donator	firstDonation	beneficiary	donationTimes	lastDonation	feelingAfterDonation	reasonNotDonate	wouldDonate
0	2019/02/07 1:48:02 pm GMT-5	Entre 1981 y 1996	Mujer	B+	Sí	2018-05-30	Algún familiar	1.00	2018-05-30	Sí, me genera satisfacción	0	Sí
1	2019/02/07 8:32:31 pm GMT-5	Entre 1981 y 1996	Mujer	B+	No	0	0	0.00	0	0	No he buscado donde hacerlo	Sí
2	2019/02/08 12:20:59 pm GMT-5	Entre 1965 y 1980	Hombre	O+	Sí	2019-01-01	Algún conocido	1.00	2019-01-01	Sí, me gustó ayudar	0	Sí
3	2019/02/08 1:31:47 pm GMT-5	Entre 1981 y 1996	Hombre	A+	No	0	0	0.00	0	0	No he tenido la necesidad	No
...	...	...	...	...	...	...	...	...	...	...	...	...
511	2019/04/10 10:20:40 am GMT-5	Entre 1981 y 1996	Hombre	No sé	Sí	2014-08-01	Algún familiar	2.00	2018-01-30	Sí, me gustó ayudar	0	Sí
512	2019/04/10 11:10:26 am GMT-5	Entre 1981 y 1996	Mujer	O+	No	0	0	0.00	0	0	Miedo	Sí
513	2019/04/10 11:48:51 am GMT-5	Entre 1981 y 1996	Mujer	A+	No	0	0	0.00	0	0	No he tenido la necesidad	No
514	2019/04/11 5:11:02 pm GMT-5	Entre 1981 y 1996	Mujer	No sé	No	0	0	0.00	0	0	No he tenido la necesidad	Sí

Figura 4-3 DataFrame después de la limpieza de datos

## 4.6. Preparación de datos

En esta parte del proceso, se manipularán los datos para que tengan el formato que esperan las funciones de *scikit-learn* para ello primero se dividirá el **DataFrame** en dos, uno contendrá la información de los donadores y otro la de los no donadores, para realizar esta separación se utilizaran los resultados de la columna *donator* como se observa en la Figura 4-4.

```
# Separate the dataset to get the donators information
# and substitute the NaN values with zero
donators_headers = ["timestamp", "age", "gender", "bloodType",
                   "firstDonation", "beneficiary", "donationTimes",
                   "lastDonation", "feelingAfterDonation", "wouldDonate"]

not_donators_headers = ["timestamp", "age", "gender", "bloodType",
                       "reasonNotDonate", "wouldDonate"]

blood_initial_dataset.fillna(0, inplace=True)

donators = blood_initial_dataset.loc[blood_initial_dataset['donator'] == 'Si']
not_donators = blood_initial_dataset.loc[blood_initial_dataset['donator'] == 'No']
```

Figura 4-4 DataFrame para donadores y no donadores

Hasta ahora el archivo tiene las fechas en formato *yyyy-mm-dd*, muchos de los encuestados comentaron que era difícil recordar el mes y el día exactos de su donación, por lo cual, se determinó sólo considerar el año para el análisis, debido a eso en el siguiente paso se busca obtener el año de la primera donación (*firstDonation*) y la última donación (*lastDonation*), utilizando la función de la Figura 4-5, y el total de donaciones realizadas.

```
def get_donation_year(arr):
    yearsArray = []
    for i in range(0, len(arr)):
        if (arr[i] != 0):
            yearsArray.append(int(arr[i].split("-")[0]))
        else:
            yearsArray.append(0)
    return yearsArray
```

Figura 4-5 Función para obtener el año de una fecha

Conociendo el año en que la persona donó por primera y última vez, además del total de veces que ha donado se generará una predicción aleatoria de en qué año se realizaron sus demás donaciones, en caso de tener más de una, para poder obtener una tendencia del crecimiento o disminución de donaciones a lo largo de los años. El procedimiento para obtener lo anterior será:

1. Extraer el año de *firstDonation* y *lastDonation* además del total de donaciones realizadas como se puede observar en la Figura 4-6.
2. Obtener el año de la primera donación de la cual se tiene registro en las respuestas de las encuestas.
3. Obtener el año de la última donación de la cual se tiene registro en las repuestas de las encuestas.
4. Generar el arreglo *years\_difference*, el cual utiliza los valores de *firstDonation* y *lastDonation* para calcular la diferencia entre ambos años.
5. El siguiente paso consiste en utilizar lo obtenido, previamente, para generar un arreglo que contenga las donaciones de cada año desde el año de la primera donación hasta el año de la última

donación. Este arreglo, utilizará los valores de *years\_difference* para que cuando la persona haya donado más de una vez, se genere una predicción aleatoria del año en que se realizaron el resto de sus donaciones, pero tal predicción debe encontrarse dentro del margen de la primera y la última donación de tal persona. Por ejemplo, si un encuestado ha realizado 5 donaciones; su primera donación fue en 2007 y la última en 2017, aún quedarían 3 donaciones realizadas dentro de esos años, entonces con la función *generate\_donations\_by\_year* se predecirá un año aleatorio para cada donación faltante dentro del rango de la primera y la última donación. En la Figura 4-7 se observa la función *generate\_donations\_by\_year*.

	<b>firstDonation</b>	<b>lastDonation</b>	<b>donationTimes</b>
<b>0</b>	2018.00	2018	1.00
<b>1</b>	2019.00	2019	1.00
<b>2</b>	2010.00	0	1.00
<b>3</b>	1984.00	0	1.00
...	...	...	...
<b>207</b>	2018.00	2018	1.00
<b>208</b>	2010.00	2015	2.00
<b>209</b>	2018.00	2018	4.00
<b>210</b>	2014.00	2018	2.00

211 rows x 3 columns

Figura 4-6 Año de la primera y última donación y total de donaciones

```

def generate_donations_by_year(first_donation_years, last_donation_years, min_year, max_year,
                               donation_times, YEARS_GAP, years_difference):
    donations_by_year = {}
    for i in range(min_year, max_year + 1):
        donations_by_year[i] = 0

    for i in range(0, len(donation_times)):
        if (first_donation_years[i] == 0):
            first_donation_years[i] = min_year + random.randint(0, YEARS_GAP)
            donations_by_year[first_donation_years[i]] += 1

        if (donation_times[i] > 1):
            if (last_donation_years[i] == 0):
                donations_by_year[max_year - random.randint(0, YEARS_GAP)] += 1
            else:
                donations_by_year[last_donation_years[i]] += 1

        if (donation_times[i] > 2):
            for j in range(0, int(donation_times[i]) - 2):
                random_year = random.randint(0, years_difference[i])
                donations_by_year[first_donation_years[i] + random_year] += 1
    print(sum(donations_by_year.values()))
    return donations_by_year

```

Figura 4-7 Función *generate\_donations\_by\_year*

Al completar esos pasos, el resultado tiene la estructura que se observa en la Figura 4-8 y ya se encuentra listo para poder analizarlo con la librería *scikit-learn*.

```
1 {
2   "1980": 2,
3   "1981": 0,
4   "1982": 0,
5   "1983": 0,
6   "1984": 1,
7   "1985": 1,
8   "1986": 1,
9   "1987": 0,
10  "1988": 1,
11  "1989": 2,
12  "1990": 4,
13  "1991": 2,
14  "1992": 0,
15  "1993": 3,
16  "1994": 1,
17  "1995": 6,
18  "1996": 1,
19  "1997": 1,
20  "1998": 3,
21  "1999": 3,
22  "2000": 17,
23  "2001": 6,
24  "2002": 12,
25  "2003": 12,
26  "2004": 10,
27  "2005": 20,
28  "2006": 14,
29  "2007": 14,
30  "2008": 28,
31  "2009": 22,
32  "2010": 44,
33  "2011": 27,
34  "2012": 38,
35  "2013": 29,
36  "2014": 43,
37  "2015": 45,
38  "2016": 52,
39  "2017": 72,
40  "2018": 77,
41  "2019": 45
42 }
```

Figura 4-8 Arreglo de las donaciones por año

Por último, para mantener la consistencia de los resultados, respecto a los datos que se generan aleatoriamente, se optó por almacenar el resultado que se observa en la Figura 4-8 en un formato CSV y cada que se ejecutan las funciones de *scikit-learn* se leen los datos del archivo CSV, a menos que en la función se especifique que se desea generar un nuevo modelo de datos, en ese caso se reemplazan los datos que se tenían previamente almacenados con los valores nuevos y se utilizan los valores nuevos en el desarrollo de las funciones.

## 4.7. Modelo de regresión lineal con scikit-learn

Después de realizar los pasos anteriores tenemos la información lista para generar el modelo de regresión lineal con las distintas funciones que provee la librería *scikit-learn*. Se analizarán los resultados de la regresión lineal con normalización de los datos y sin normalización con el objetivo de comparar en cual se tienen mejores resultados.

### 4.7.1. Regresión lineal sin normalización

Para el modelo se considerará al año como la variable independiente, es decir, el valor de  $x$  y la variable dependiente,  $y$ , será el total de donaciones realizadas en cada año. Con estos valores se hará el *fit* del modelo como se observa en la Figura 4-9.

```
# Creating the Linear Regression Model
LR_model = LinearRegression()
LR_model.fit(x, y)
```

Figura 4-9 Creación del modelo y función *fit*

El siguiente paso consiste en evaluar que tan bueno es el resultado del modelo, para ello se obtiene el valor del coeficiente de determinación, también conocido como  $r^2$ , y se obtendrán los valores de  $\beta_0$  que es la intercepción con  $y$ , y el valor de la pendiente de la línea de regresión  $\beta_1$ . Los resultados de estos valores se pueden observar en la Figura 4-10.

```
# Obtaining coefficient of determination (R^2)
r_sq = LR_model.score(x, y)
print('coefficient of determination:', r_sq)

# y = b0 + b1x
print ('Intercept (b0): ', LR_model.intercept_)
print ('Slope (b1): ', LR_model.coef_)
```

Figura 4-10 Datos del modelo de regresión

### 4.7.2. Regresión lineal con normalización

La normalización es un proceso que se utiliza para que los valores que se encuentran fuera de la escala que muestran la mayoría de los datos se ajusten a la escala sin afectar la diferencia que existe entre los valores actuales.

Para generar una regresión lineal con normalización de los datos sólo se tiene que especificar la opción al momento de generar el modelo, como se observa en la Figura 4-11.

```
LR_Norm_model = LinearRegression(normalize=True)
LR_Norm_model.fit(x, y)
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=True)
```

Figura 4-11 Creación del modelo de regresión con normalización

Por último, para poder comparar cual método da mejores resultados se calcula el valor del coeficiente de determinación y los valores de,  $\beta_0$  que es la intercepción con  $y$ , y el valor de la pendiente de la línea de regresión  $\beta_1$ . Los resultados de estos valores se pueden observar en la Figura 4-12.

```
# Obtaining coefficient of determination (R^2)

r_sq_n = LR_Norm_model.score(x, y)
print('coefficient of determination:', r_sq_n)

# y = b0 + b1X
print ('Intercept (b0): ', LR_Norm_model.intercept_)
print ('Slope (b1): ', LR_Norm_model.coef_)
```

Figura 4-12 Datos del modelo de regresión con normalización

#### 4.8. *Modelo de regresión polinomial con scikit-learn*

Utilizando otra función de la librería de *scikit-learn* se calculará un modelo de regresión polinomial para comparar cuál modelo da la mejor aproximación a la tendencia que llevan las donaciones de sangre a lo largo del tiempo.

Para realizar el análisis se seguirá considerando los años como la variable independiente,  $x$ , y el total de donaciones como la variable dependiente  $y$  y se utilizará la función *fit\_transformation* de la librería para poder convertir los datos en el formato esperado para el modelo polinomial.

Con el objetivo de poder calcular polinomios de distintos grados se creó la función, *create\_polynomial\_reg\_model*, de la Figura 4-13, a la cual sólo se le indica el grado del polinomio que se desea calcular haciendo uso del parámetro *degree* y se generara el modelo polinomial con ese grado.

```
def create_polynomial_reg_model (degree = 2):
    polynomial_features = PF(degree = degree)
    X_polynomial = polynomial_features.fit_transform(x)
    polynomial_model = LinearRegression()
    polynomial_model.fit(X_polynomial, y)

    y_train_predicted = polynomial_model.predict(X_polynomial)

    y_test_predicted = polynomial_model.predict(polynomial_features
                                                .fit_transform(np.array(range(2020, 2030))
                                                                .reshape(-1, 1)))

    rmse = math.sqrt(mean_squared_error(y, y_train_predicted))
    r2 = r2_score(y, y_train_predicted)

    print("RMSE is {}".format(rmse))
    print("R2 score is {}".format(r2))
    print(y_test_predicted)

    return (x, y_train_predicted)
```

Figura 4-13 Función *create\_polynomial\_reg\_model*

En la siguiente sección se compararán los resultados obtenidos de ambos modelos para poder identificar cuál es el que provee una línea de tendencia que se ajusta mejor a los resultados de las encuestas. Con el resultado, también, se podrá generar una predicción de cómo, de acuerdo con la información obtenida, se espera que sean las donaciones de sangre en los próximos años en México.

---

## 5. RESULTADOS Y DISCUSIÓN

---

**Resumen:** *En este capítulo se presentan los resultados del desarrollo de este trabajo; desde los resultados obtenidos en las encuestas hasta los resultados de los distintos análisis realizados, además de una discusión sobre las donaciones de sangre en México tomando en consideración los resultados obtenidos del trabajo de investigación.*

## 5.1. Resultados de las encuestas

La encuesta se distribuyó en distintos medios, anteriormente mencionados, y con ayuda de amigos y familiares se lograron obtener 515 encuestas respondidas en la ciudad de Guadalajara. A continuación, se detallan los resultados más representativos de la encuesta.

### 5.1.1. ¿En qué año naciste?

Los rangos de edad se especificaron considerando la investigación realizada por la organización estadounidense *Pew Research Center* [26]. En la Figura 5-1 se puede observar la distribución de los encuestados respecto a los años en los que nacieron, cada rango hace referencia a una generación social.

Generación social	Total de respuestas
Silenciosa	1
Baby boomers	21
X	99
Millennials	367
Z	27
<b>TOTAL</b>	<b>515</b>

Tabla 5-1 Resultados ¿En qué año naciste?

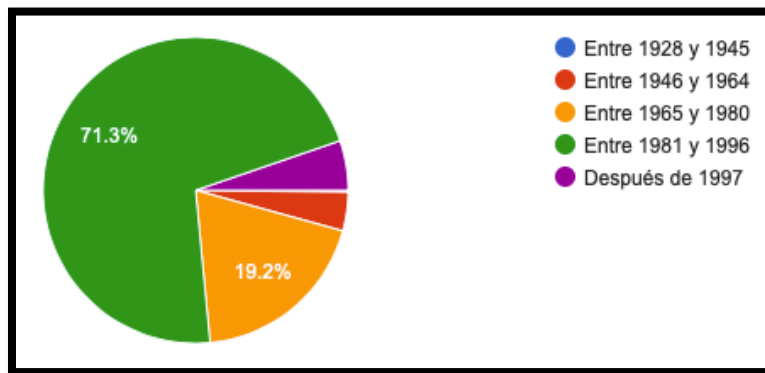


Figura 5-1 Participación en la encuesta por generación social

### 5.1.2. ¿Género?

En las investigaciones que se analizaron en el Estado del Arte se observó cierta tendencia a que los hombres realizan mayor cantidad de donaciones que las mujeres. En Jalisco la población de mujeres es mayor a la población de hombres y los resultados obtenidos muestran que, al menos, en la ciudad de Guadalajara existe una mayor participación de los hombres en las donaciones.

Género	Donadores	Total	Porcentaje de donación
Hombre	142	295	48.13 %
Mujer	69	220	31.36 %
<b>TOTAL</b>	<b>211</b>	<b>515</b>	<b>40.97 %</b>

Tabla 5-2 Total de participantes y donadores por género

### 5.1.3. ¿Cuál es tu grupo sanguíneo?

Es importante conocer el tipo de sangre que somos para en caso de una emergencia poder recibir la sangre adecuada y también es importante saber que los tipos de sangre no se encuentran repartidos equitativamente en el país, de los resultados obtenidos se puede observar cuál es el más predominante en nuestra sociedad.

Grupo sanguíneo	Donadores	Total	Porcentaje de donación
A+	48	132	36.4 %
A-	7	8	87.5 %
B+	20	44	45.5 %
B-	2	5	40.0 %
O+	104	246	42.3 %
O-	11	23	47.8 %
AB+	7	12	58.3 %
AB-	0	2	0 %
No sé	12	43	27.9 %
<b>TOTAL</b>	<b>211</b>	<b>515</b>	<b>40.97 %</b>

Tabla 5-3 Distribución de los grupos sanguíneos y los donadores por grupo

### 5.1.4. ¿Has donado sangre alguna vez?

De los encuestados sólo 211 habían donado sangre en una ocasión o más, eso equivale al 41 % de las personas que realizaron la encuesta. En la Figura 5-2 se observa la distribución de las personas que sí han donado y las personas que no han donado.

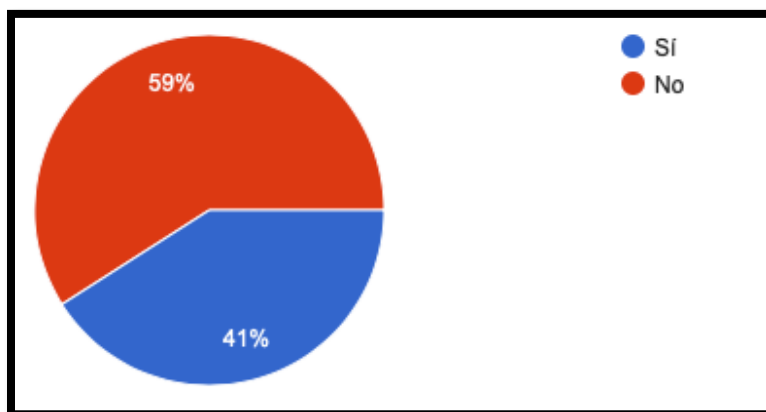


Figura 5-2 Donadores versus No donadores

### 5.1.5. La primera vez que donaste, fue:

Como se ha observado a lo largo del trabajo, en México el porcentaje de donaciones altruistas es muy bajo, pero en los resultados obtenidos de la encuesta se obtuvo que de los 211 donadores un 28.4% tuvo su primer acercamiento con las donaciones de una manera altruista.

Motivo de donación	Total	Porcentaje
Algún familiar	96	45.5 %
Algún conocido	55	26.1 %
Altruista	60	28.4 %

Tabla 5-4 Principal razón para haber realizado la primera donación

### 5.1.6. ¿Cuántas veces has donado?

Es importante conocer la cantidad de donaciones que las personas han realizado porque eso ayuda a obtener una tendencia de qué tan frecuente una persona se puede llegar a convertir en un donador recurrente. En la Figura 5-3 se observa en el eje x la cantidad de donaciones realizadas y en el eje y se encuentra el total de personas que han realizado ese número de donaciones.

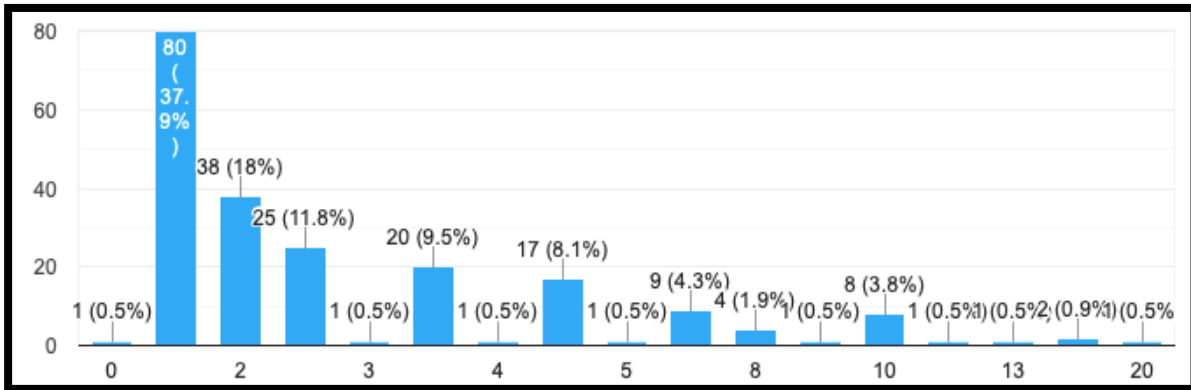


Figura 5-3 Total de personas y donaciones realizadas

### 5.1.7. ¿Te generó un sentimiento positivo donar sangre?

En general, se obtuvieron respuestas positivas sobre la experiencia de donar sangre lo cual es importante para mantener la disposición y que el donador se sienta motivado a donar en otra ocasión. Sólo 18 personas de los 211 donadores tuvieron una experiencia negativa debido a que la donación les generó malestar físico, dolor o simplemente fue algo que realizaron por cumplir el requisito.

La mayoría consideró la donación como algo satisfactorio en lo emocional y lo personal, este resultado es similar a lo que se obtuvo en las investigaciones analizadas anteriormente, que mencionan que en las primeras ocasiones que la gente se acerca a donar obtienen una gran satisfacción personal por saber que están ayudando a alguien que lo necesita donando una porción de su sangre y también se observó que entre más donaciones se van realizando el sentimiento se va convirtiendo en algo más normal. En la Figura 5-4 se observa el porcentaje de las personas que encontró satisfacción personal en ayudar y algunas otras de las respuestas que se obtuvieron en la encuesta.

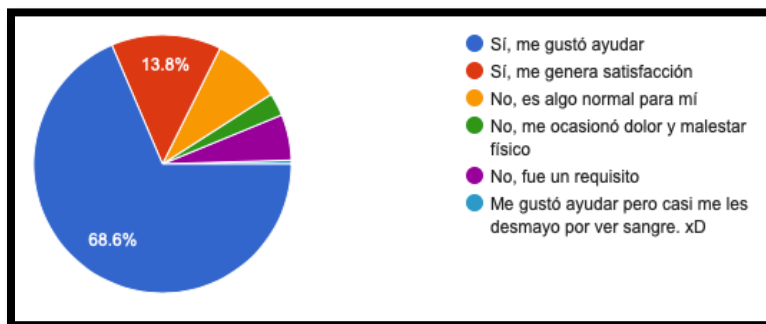


Figura 5-4 Sentimiento generado después de realizar la donación

### 5.1.8. Si no has donado, ¿a qué se debe?

Entender las distintas causas que hacen que una persona no haya donado aún ayuda de distintas maneras a mejorar el proceso, por ejemplo, se pueden mejorar los medios de información sobre el tema para que la gente conozca mejor los requisitos y cuide su salud para ser un candidato válido cuando quiera realizar una donación y que si no se encuentra apto no acuda al banco de sangre para cuidar su tiempo y el tiempo del personal del banco de sangre.

Los resultados nos permiten ver que muchas personas no han podido donar por problemas relacionados con la salud como presión baja, problemas con la presión, pesan menos de lo indicado, hipertensión y la mayoría de estas personas se enteró de esta situación hasta el momento en que acudieron al banco de sangre. También el miedo y no haber tenido la necesidad de donar hasta ahora, son razones por las cuales la gente en Guadalajara no realiza donaciones.

En la figura 5-5 podemos observar la distribución de los resultados.

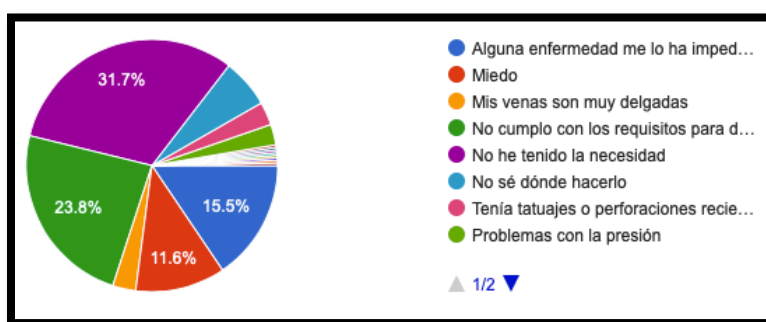


Figura 5-5 Razones por las cuales la gente no ha donado

### 5.1.9. En México menos del 4% de la población dona sangre de manera altruista, ¿a qué crees que se deba?

Aunque los resultados en esta investigación nos dieron un porcentaje alto de gente que ha donado de manera altruista, otras investigaciones que se analizaron en este trabajo nos mencionaron que el porcentaje de donaciones altruistas en México era muy bajo, menor al 4%, lo cual es preocupante porque es un porcentaje menor al porcentaje que sugiere la OMS. Debido a eso, también se buscó conocer la opinión de los habitantes de Guadalajara sobre este tema y cuál consideraban que era la razón de tener un porcentaje tan bajo de donaciones de sangre en el país.

Los resultados de la encuesta nos dicen que en la sociedad falta solidaridad, empatía y cultura sobre las donaciones, el proceso de donación es muy tardado, muchas veces la gente pasa 4 o 5 horas realizando la donación y, en ocasiones, a la mitad del proceso o más es cuando se les informa que no son aptos lo cual genera descontento. También, se encontró que la falta de información sobre el proceso, los requerimientos y los lugares donde la gente puede donar es una razón por la cual la gente no realiza donaciones.

En la Figura 5-6 se observan las principales respuestas de las personas ante esta pregunta.

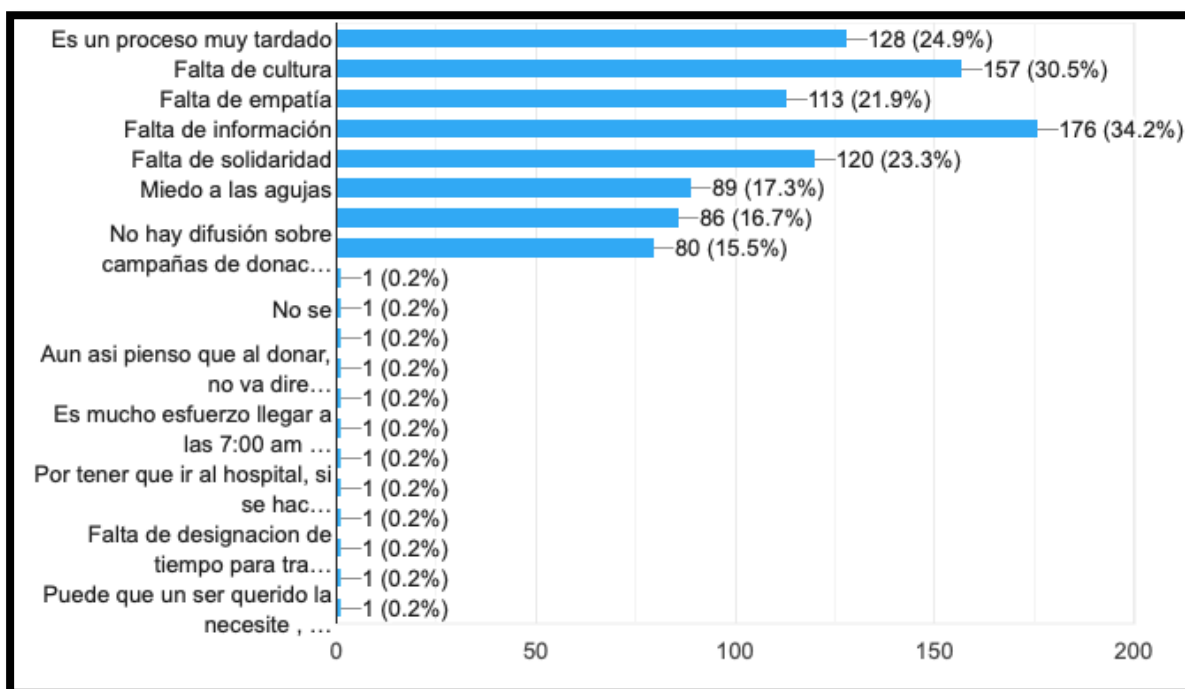


Figura 5-6 ¿A qué se debe el bajo porcentaje de donaciones altruistas en México?

## 5.2. Resultados modelo de regresión lineal

Para la regresión lineal se realizó el proceso considerando un modelo de datos sin normalizar y uno con normalización con el objetivo de observar cuál pudiera generar el resultado que mejor se ajustara a la tendencia de nuestros datos. De cada modelo se obtendrán los siguientes datos:

- Coeficiente de determinación ( $r^2$ ).
- Intercepción ( $\beta_0$ ) y pendiente ( $\beta_1$ ).
- *Mean squared error*.
- Gráfica con los datos y la línea de tendencia.
- Predicción de los siguientes 10 años.

Los datos que se obtuvieron de la predicción y se utilizaron como modelo de datos para este análisis se pueden observar en la Tabla 5-5.

Año	Donaciones	Año	Donaciones	Año	Donaciones	Año	Donaciones
1980	3	1990	3	2000	15	2010	29
1981	0	1991	2	2001	6	2011	32
1982	0	1992	1	2002	11	2012	31
1983	1	1993	1	2003	13	2013	43
1984	1	1994	3	2004	16	2014	34
1985	1	1995	5	2005	14	2015	50
1986	0	1996	1	2006	20	2016	58
1987	0	1997	2	2007	19	2017	71
1988	0	1998	6	2008	20	2018	70
1989	5	1999	6	2009	25	2019	41

Tabla 5-5 Datos utilizados para el modelo

Como se observa en los datos de la Tabla 5-5, en los primeros años se tuvieron pocas donaciones y eso tiene sentido debido a que sólo se tienen 21 registros de personas que nacieron entre 1946 y 1964 (*Generación baby boomer*). Esta generación es la que a partir de 1980 ya tenía la edad solicitada para comenzar a donar, las personas de la Generación X seguían siendo menores de edad.

Debido a eso, y a que en los últimos años hay gran participación por parte de los *Millennials*, fue un poco complicado adecuar una tendencia lineal para nuestro modelo de datos.

### 5.2.1. Regresión lineal sin normalización

Los resultados que se obtuvieron muestran una línea de tendencia que deja afuera los resultados de los primeros años y, difícilmente, se adecua al crecimiento de los últimos años como se observa en la Figura 5-7.

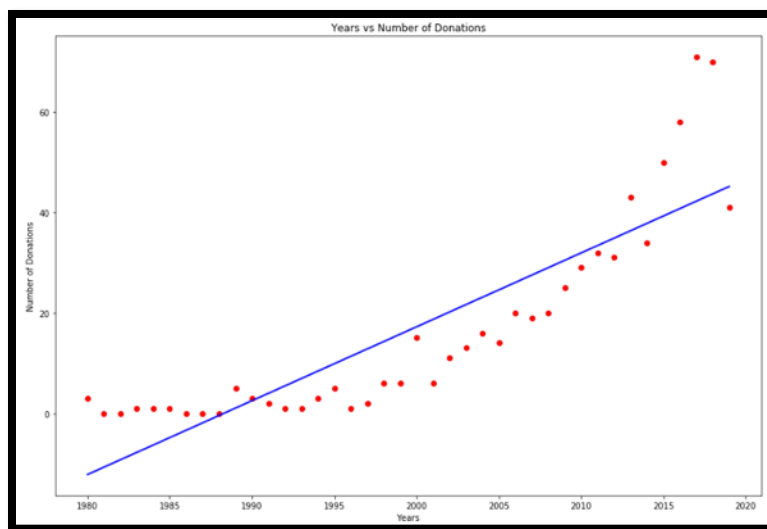


Figura 5-7 Donaciones vs Año y línea de tendencia obtenida de la regresión lineal

Para entender mejor el resultado, analicemos los valores obtenidos que se presentan en la Figura 5-8:

```
# Obtaining coefficient of determination (R^2)
r_sq = LR_model.score(x, y)
print('coefficient of determination:', r_sq)

# y = b0 + b1x
print ('Intercept (b0): ', LR_model.intercept_)
print ('Slope (b1): ', LR_model.coef_)

# MSE
print("Mean Squared Error: ", mse)

coefficient of determination: 0.7491090257972236
Intercept (b0): -2926.3163227016885
Slope (b1): [1.4717636]
Mean Squared Error: 9.831995502107945
```

Figura 5-8 Resultados modelo de regresión lineal

El coeficiente de determinación es un valor que oscila entre 0 y 1, cuando es más cercano a 1 refleja un modelo fiable. En nuestro análisis se obtuvo un valor de 0.749 lo que es bueno debido a la diferencia mencionada con los datos de los primeros años y el aumento que se tiene en los últimos años.

La fórmula para nuestra línea de tendencia es:

$$y = 1.4717636x - 2926.316322$$

Y respecto al error se obtuvo un valor de 9.8, lo que nos indica que en las predicciones podemos tener un resultado de más o menos 9 donaciones.

### 5.2.2. Regresión lineal con normalización

Como se observa en la Figura 4-11, en esta ocasión el modelo se realizó aplicando la opción de normalización a los datos. Sin embargo, los resultados no fueron los que se esperaban ya que se obtuvieron los mismos resultados que en el desarrollo sin normalización. El valor del coeficiente de determinación, la línea de tendencia y el error fueron los mismos, en la Figura 5-9 se pueden observar los resultados de la Regresión lineal con normalización.

```
Linear Regression with Normalization

LR_Norm_model = LinearRegression(normalize=True)
LR_Norm_model.fit(x, y)

LinearRegression(copy_X=True, fit_intercept=True, n_jobs=1, normalize=True)

# Obtaining coefficient of determination (R^2)
r_sq_n = LR_Norm_model.score(x, y)
print('coefficient of determination:', r_sq_n)

# y = b0 + b1X
print ('Intercept (b0): ', LR_Norm_model.intercept_)
print ('Slope (b1): ', LR_Norm_model.coef_)

# MSE
print("Mean Squared Error: ", mse)

coefficient of determination: 0.7491090257972236
Intercept (b0): -2926.3163227016885
Slope (b1): [1.4717636]
Mean Squared Error: 9.831995502107945
```

Figura 5-9 Resultados regresión lineal con normalización

### 5.2.3. Predicción

El siguiente paso consiste en utilizar los resultados obtenidos para predecir el comportamiento de las donaciones de sangre en Guadalajara en los siguientes años.

Debido a que ambos métodos dieron los mismos resultados, se pudo realizar la predicción utilizando los valores de cualquier modelo. En la Figura 5-10 se puede observar los datos obtenidos con las encuestas (puntos azules), la fórmula para la línea de tendencia que describe el comportamiento de los datos a lo largo del tiempo y la predicción de donaciones de sangre en Guadalajara que, de acuerdo con la línea de tendencia, se espera para los siguientes 10 años y se representa con los puntos de color anaranjado.

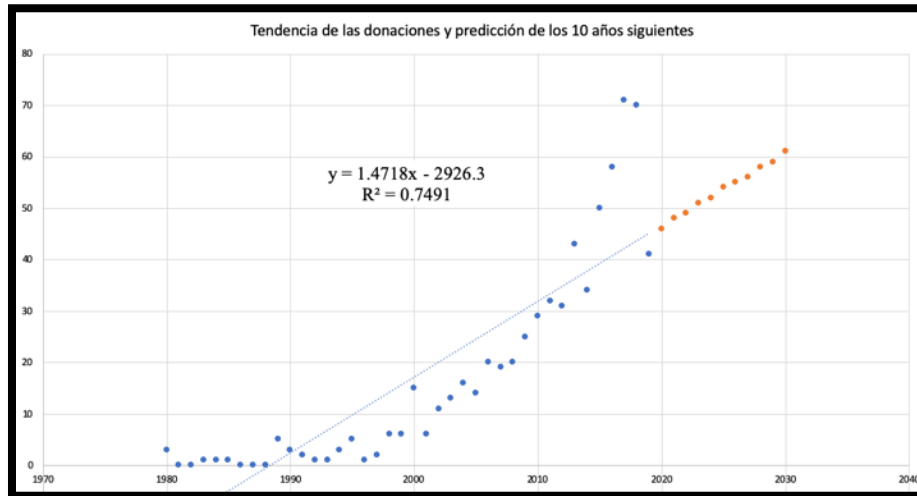


Figura 5-10 Donaciones de sangre actuales y predicción a 10 años

Como se puede observar en la Figura 5-10, aproximadamente a partir de 2015, la cantidad de donaciones de sangre en la ciudad de Guadalajara aumentó considerablemente respecto a lo que venía sucediendo en años anteriores, como consecuencia de eso, la línea de tendencia que tenemos y nuestras predicciones quedan apartadas de lo que parece un futuro prometedor con un gran aumento de donaciones anuales.

De acuerdo con la predicción que nos otorga este modelo, sería en 2028 cuando alcanzaríamos la cantidad de 58 donaciones anuales que se presentaron en 2016, por lo cual, entre más avancen los años y la tendencia de donaciones en la sociedad de Guadalajara mantenga el mismo ritmo, el modelo irá quedando cada vez más alejado de la realidad y el error será más y más grande.

Realizar una regresión lineal de este estilo nos ayuda a observar la tendencia que van teniendo nuestros datos y nos ayuda a determinar si es el modelo adecuado o se debe considerar buscar un modelo que se apegue mejor a los datos que se están analizando, en este caso, se decidió utilizar otro modelo que permita obtener una línea de tendencia más apegada al comportamiento de los datos. En la siguiente sección se evalúan los mismos datos con un modelo polinomial.

### 5.3. Resultados modelo de regresión polinomial

El modelo de regresión polinomial generó mejores resultados que el modelo de regresión lineal debido a que las ecuaciones de este tipo tienden a tener un crecimiento como el que se observa en los datos de donaciones de sangre obtenidos en las encuestas.

Sin embargo, hay que tener cuidado en escoger un modelo de un orden considerable para no agregarle demasiada complejidad matemática a la ecuación. Para el análisis se comparó un modelo cuadrático y uno cúbico.

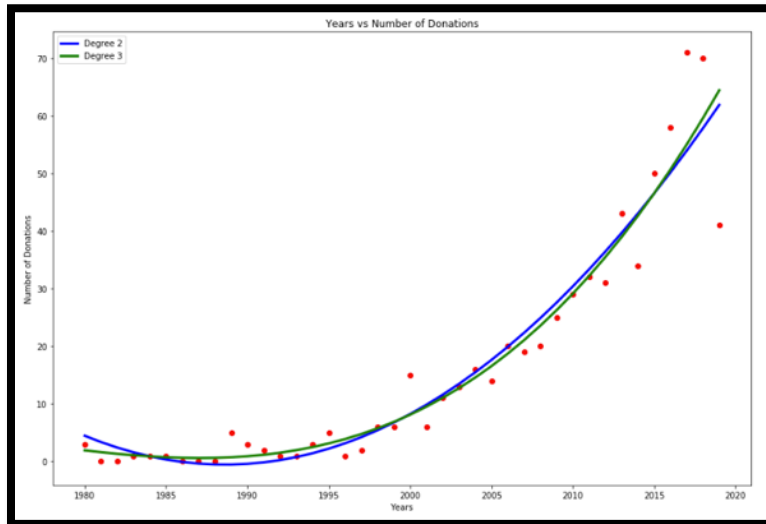


Figura 5-11 Modelos de regresión polinomial

En la Figura 5-11 se pueden observar los resultados de ambos modelos, en color azul se encuentra el modelo cuadrático y en color verde el modelo cúbico. Como se observa en la imagen, ambos modelos se ajustan muy bien a la tendencia que van generando los datos a lo largo del tiempo, debido a que los dos modelos otorgan resultados muy buenos se decidió utilizar el modelo cuadrático para realizar las predicciones ya que al ser cuadrático su complejidad matemática es más sencilla que un modelo cúbico.

El modelo cuadrático dio como resultado un error de 5.6544 y un coeficiente de determinación ( $r^2$ ) con valor de 0.917, que como mencionamos con anterioridad entre más cercano a 1 se encuentre nos ayuda a verificar qué tan fiable es el resultado que estamos obteniendo. Por el otro lado, el modelo cúbico dio un error menor de 5.5434 y un coeficiente de determinación ( $r^2$ ) con valor de 0.92024.

Utilizando los resultados obtenidos del método cuadrático se obtiene una predicción como la que se muestra en la Figura 5-12, al mismo tiempo podemos observar como los datos de los últimos años se ven mejor representados con este modelo y las predicciones mantienen la tendencia de crecimiento que se observa a partir del año 2016. En la figura se observan de color azul los datos obtenidos por la encuesta y de color anaranjado las predicciones del crecimiento de donaciones que se espera en la ciudad de Guadalajara para los próximos 10 años.

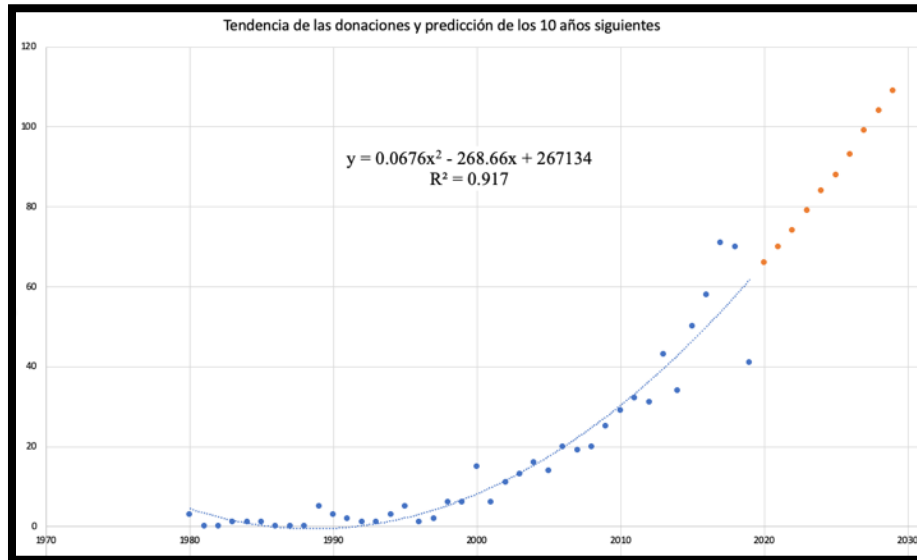


Figura 5-12 Predicción de los siguientes 10 años con un modelo polinomial

#### 5.4. Modelos de regresión lineal por tipo de sangre

Debido a la importancia que tiene conocer qué tipo de sangre prevalece en la sociedad y la participación que tienen las personas de cierto grupo sanguíneo en el total de donaciones de sangre anuales. En esta sección se enriquecerá la información de la Tabla 5-3 donde se dio a conocer el total de donadores por grupo sanguíneo, para ello se obtuvo de las respuestas de cada grupo sanguíneo:

- Total de donadores de sangre.
- Total de donaciones que los donadores han realizado.
- Año de la primera donación.
- Año de la última donación.
- Coeficiente de determinación ( $r^2$ ).
- Intercepción ( $\beta_0$ ) y pendiente ( $\beta_1$ ).
- Error.
- Predicción de donaciones en los siguientes años.

### 5.4.1. Grupo sanguíneo A+

En la Figura 5-13 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo A+.

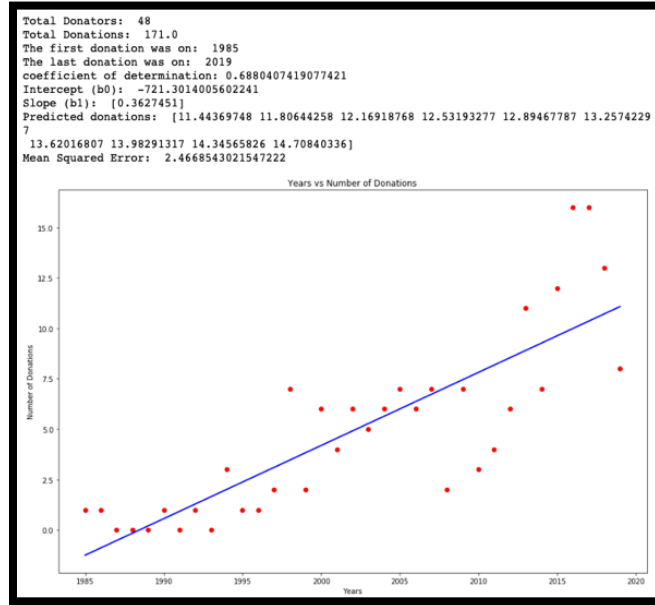


Figura 5-13 Regresión Lineal (A+)

### 5.4.2. Grupo sanguíneo A-

En la Figura 5-14 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo A-.

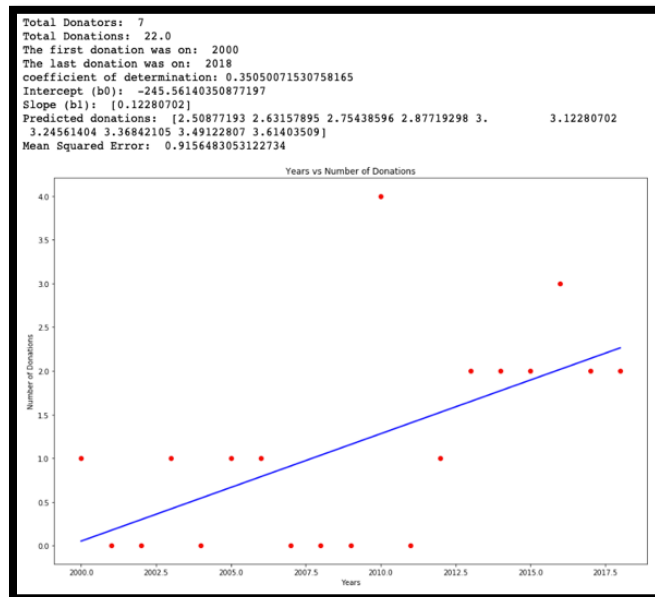


Figura 5-14 Regresión Lineal (A-)

### 5.4.3. Grupo sanguíneo B+

En la Figura 5-15 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo B+.

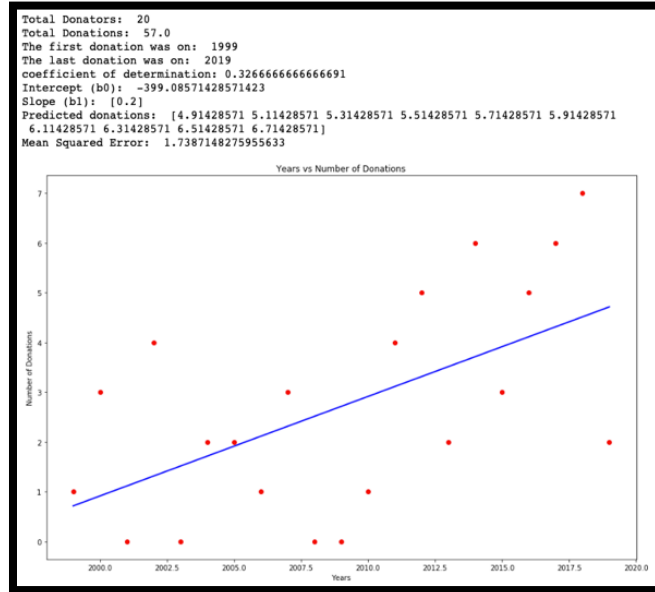


Figura 5-15 Regresión Lineal (B+)

### 5.4.4. Grupo sanguíneo B-

En la Figura 5-16 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo B-.

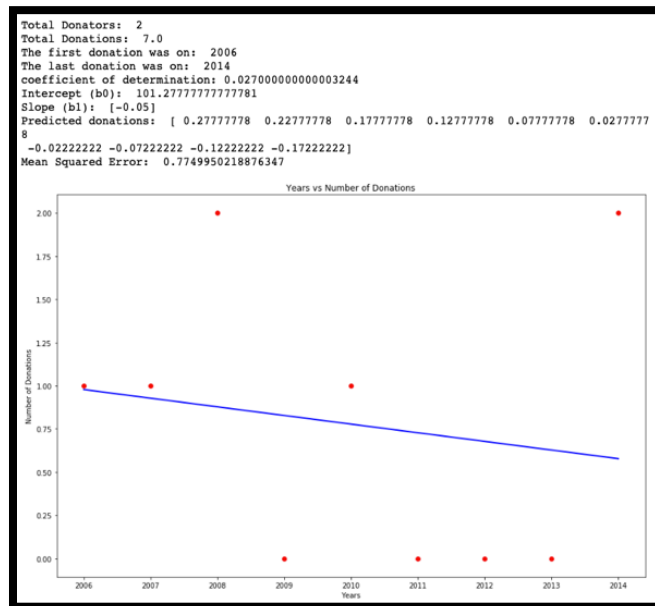


Figura 5-16 Regresión Lineal (B-)

### 5.4.5. Grupo sanguíneo AB+

En la Figura 5-17 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo AB+.

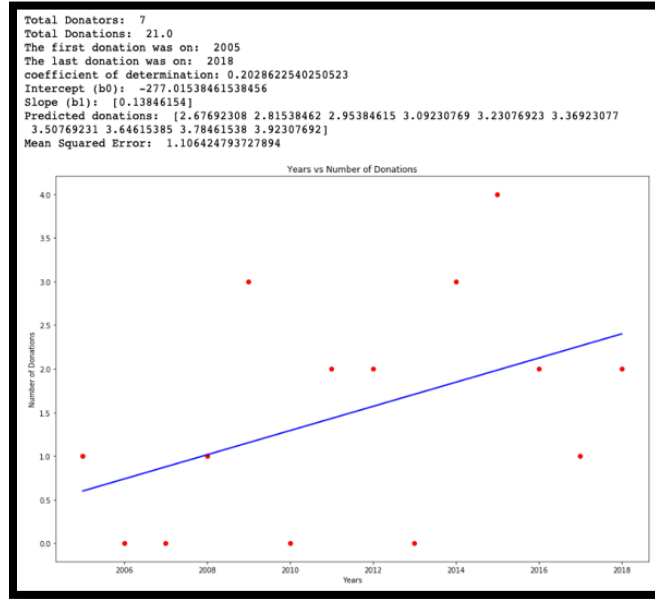


Figura 5-17 Regresión Lineal (AB+)

### 5.4.6. Grupo sanguíneo O+

En la Figura 5-18 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo O+.

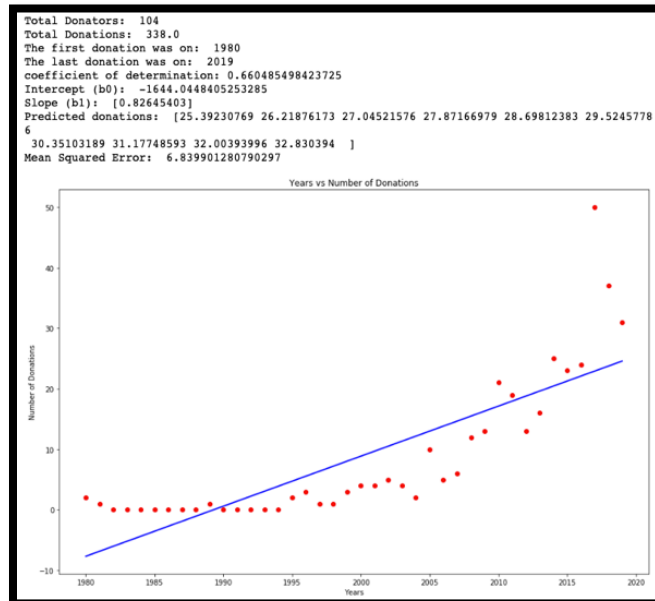


Figura 5-18 Regresión Lineal (O+)

### 5.4.7. Grupo sanguíneo O-

En la Figura 5-19 se observan los resultados de la regresión lineal que se aplicó a los donadores del grupo sanguíneo O-.

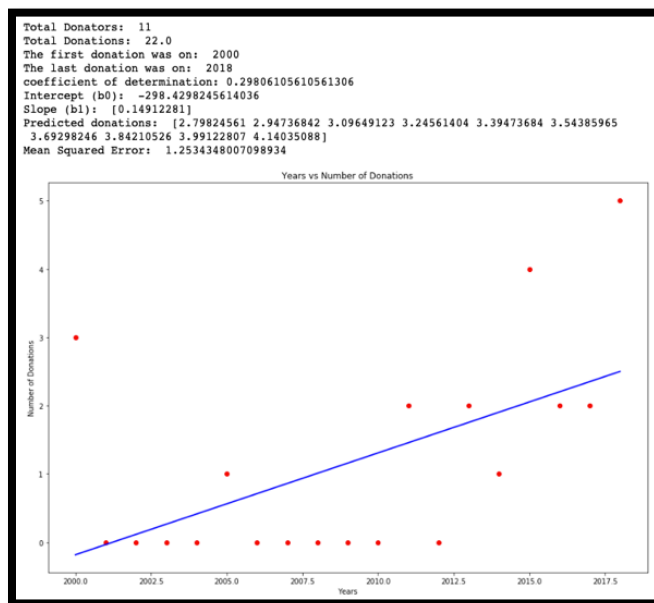


Figura 5-19 Regresión Lineal (O-)

Con los datos que se recopilaron se puede observar que, en general, se tiene el doble de donaciones realizadas que el total donadores en la mayoría de los grupos sanguíneos, eso es importante porque nos indica que hay una parte de donadores que son recurrentes.

Por otra parte, se tiene que poner atención en grupos sanguíneos como AB-, B-, A-, AB+ y O- ya que son grupos sanguíneos que tienen poca participación para donar y, además, son de los grupos más reducidos en la sociedad, por lo cual, en un momento de emergencia las personas pertenecientes a alguno de esos grupos podrían tener mayores problemas para conseguir una persona que pueda donarles el tipo de sangre que necesitan.

También, se observa que la mayoría de las líneas de tendencia apuntan a cierto crecimiento, lo que es bueno para aumentar las donaciones de sangre realizadas anualmente.

## 5.5. Discusión

Aumentar la cantidad de donaciones en México va a ser una labor que tome algunos años y esfuerzo por parte de distintas organizaciones de salud del país. A través de campañas de donación y proveyendo a la sociedad de información transparente y completa se debe intentar concientizar a las personas sobre la importancia que tiene donar su sangre para ayudar a otras personas en momentos de emergencia y prevención, y así también, poder disminuir los mitos y miedos que existen y que evitan que nuevas personas se acerquen a donar y se conviertan en donadores recurrentes.

Como se pudo observar en el análisis de los resultados, algunas personas no se animan a donar porque el servicio es muy tardado y complicado y probablemente esos comentarios provengan de personas que ya

han realizado donaciones, por lo cual, se debe buscar mejorar el servicio para que las personas salgan del banco de sangre con comentarios positivos y más gente se anime a donar así como sucede en Noruega, trabajo de investigación que se analizó en el Estado del Arte, que la principal razón por la que la gente se acerca a donar es por la recomendación de otras personas sobre el servicio eficiente de los bancos de sangre del país y el gran aporte emocional que les genera ayudar a otras personas en necesidad. Esa misma sensación se pudo observar también en los resultados de los encuestados, sólo queda mejorar el servicio para poder ir adquiriendo más donadores.

Sería de gran valor que en cada banco de sangre se otorgara a los donadores una encuesta sobre su experiencia en la donación con el objetivo de ir mejorando el servicio que se provee, de esa manera se podrían encontrar las principales áreas de oportunidad y también se podrían incrementar o mejorar aquellas cosas que la gente ya disfruta sobre las donaciones.

También es de gran relevancia enseñar a la población que hay ciertos grupos sanguíneos que prevalecen en la sociedad y que no todos los grupos sanguíneos pueden recibir cualquier tipo de sangre, de esa manera las personas que son de algún grupo sanguíneo complicado de encontrar podrían acudir a donar sangre sabiendo que su sangre es difícil de encontrar y que en alguna situación de emergencia podrían marcar la diferencia.

Por último, se observó en los resultados que las generaciones sociales más recientes son más propensas a donar sangre, sería importante profundizar en conocer qué factores están motivando a estas generaciones a realizar más donaciones que generaciones pasadas y buscar estrategias que mantengan el interés de los más jóvenes y busquen atraer a personas de generaciones anteriores.

---

## 6. CONCLUSIONES

---

**Resumen:** *En este capítulo se presentan las conclusiones y trabajo futuro que pudieran ser de utilidad a ampliar los resultados de la investigación y generar modelos y sugerencias que ayuden a incrementar las donaciones de sangre en México.*

## 6.1. Conclusiones

El análisis de datos es una herramienta que nos puede dar mucha información que en la mayoría de las ocasiones pasamos por alto porque se encuentra mezclada con datos poco relevantes y no se le da la importancia adecuada pero como observamos en el desarrollo de este TOG desde el momento en que se comienzan a analizar los datos para ver cuáles nos pueden mostrar ciertas tendencias o qué datos podemos usar para aplicarles algunos algoritmos, desde ese momento se comienzan a notar hábitos, tendencias, clasificaciones que vienen en los datos y se comienza a generar una idea de hasta donde se puede llegar con cierto *dataset*. Si aunado al análisis se aplican algoritmos o métodos para analizar los datos como se realizó en la investigación el resultado puede ser aún más representativo y producir resultados de valor para atacar un problema.

En este TOG, se buscó aplicar el conocimiento adquirido en la maestría a un problema social como lo es las donaciones de sangre en el país. Constantemente, en las redes sociales se encuentran publicaciones de personas buscando donadores para sus seres queridos o sus conocidos y en pocas ocasiones se ha realizado un análisis en México para conocer las razones de por qué la gente no acude a estos llamados de ayuda si por lo general México es un país que en situaciones de emergencia suele ser muy unido y solidario como nuestro pasado reciente nos lo ha demostrado ante desastres naturales que han sucedido en los últimos años.

El presente trabajo ha ayudado a identificar algunas de las razones del bajo porcentaje de donaciones de sangre que se tiene en México y ha buscado comparar el sistema actual de donaciones con el sistema de otros países que año con año tienen el suministro necesario para satisfacer las necesidades de su población, como Noruega, para poder identificar las cosas que se están haciendo bien y generan resultados positivos e intentar aplicarlas a nuestro sistema y mejorar las condiciones actuales.

Sin embargo, no todo es malo en nuestro actual sistema de donaciones, esta investigación nos permitió observar que la gente obtiene satisfacción personal al donar sangre y que parece existir una tendencia de aumento de donaciones en las generaciones más recientes y en los próximos años. Incluso, conocer las cosas que le disgustan a las personas del proceso actual de donación de sangre permite generar nuevas estrategias y mejoras al procedimiento para adquirir nuevos donadores y mantener a los que se tienen actualmente.

Por último, este trabajo busca generar más conciencia en el valor que tienen los datos e intentar aprovechar mejor la información que se tiene archivada en la actualidad para obtener resultados de valor que aporten ideas para mejorar el sistema de salud y otros problemas sociales, así como la investigación que se realizó en Hong Kong donde se utilizó la información que el hospital había recopilado por años para colocar un nuevo centro de donaciones basándose en los datos de los donadores con el objetivo de buscar la mejor ubicación y aumentar la cantidad de donadores recurrentes.

## 6.2. Trabajo Futuro

Para poder mejorar esta investigación sería recomendable buscar ampliar las preguntas de la encuesta con el objetivo de identificar los perfiles de las personas que donan, las que donan más frecuentemente y las que no donan. Se buscaría agregar preguntas relacionadas con lo socioeconómico, el estilo de vida y algunos hábitos para poder generar perfiles y realizar una clasificación sobre los datos.

También, con el objetivo de ser más acertados respecto a los resultados, se buscaría generar una alianza con algún hospital que pudiera proveer información sobre sus donadores y las personas que ha rechazado para donación y así se ampliaría el volumen de datos. Con una cantidad de datos mayor se podría representar mejor las tendencias que existen en la sociedad y se podrían aplicar modelos de predicción más robustos y que se vayan adaptando conforme se van recibiendo datos nuevos.

# BIBLIOGRAFÍA

- [1] Organización Mundial de la Salud, «Disponibilidad y seguridad de la sangre a nivel mundial,» 14 Junio 2019. [En línea]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/blood-safety-and-availability>. [Último acceso: 25 Julio 2019].
- [2] World Health Organization, «Global Status Report on Blood Safety and Availability,» World Health Organization, 2016.
- [3] El Universal, «En México sólo 2.7% dona sangre de manera altruista,» 12 Junio 2017. [En línea]. Available: <https://www.eluniversal.com.mx/articulo/ciencia-y-salud/salud/2017/06/12/en-mexico-solo-27-dona-sangre-de-manera-altruista>. [Último acceso: 11 Agosto 2019].
- [4] Noticieros Televisa, «En México sólo el 3.8% de las donaciones de sangre son altruistas,» 14 Junio 2017. [En línea]. Available: <https://noticieros.televisa.com/ultimas-noticias/en-mexico-solo-3-8-donaciones-sangre-son-altruistas/>. [Último acceso: 11 Agosto 2019].
- [5] The World Bank Group, «World Bank Country and Lending Groups,» [En línea]. Available: <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>. [Último acceso: 25 Julio 2019].
- [6] A. H. Misje, V. Bosnes, O. Gasdal y H. Heier, «Motivation, recruitment and retention of voluntary non-remunerated blood donors: a survey-based questionnaire study,» *Vox Sanguinis*, vol. 89, n° 4, pp. 236-244, 2005.
- [7] J. R. Pereira, C. V. e. Sousa, E. Bragança de Matos, L. B. Oliveira, N. X. Buenos y A. Machado, «To donate or not donate, that is the question: an analysis of the critical factors of blood donation,» *Ciência & Saúde Coletiva*, vol. 21, n° 8, pp. 2475-2484, 2016.
- [8] L. Al Shaer, R. Sharma y M. AbdulRahman, «Analysis of blood donor pre-donation deferral in Dubai: characteristics and reasons,» *Journal of Blood Medicine*, vol. 8, pp. 55-60, 2017.
- [9] A. K. AlNouri, L. A. Maghrabi, S. S. Hamdi, S. M. Abd El-Ghany y K. A. AlNouri, «Analysis of the most common causes of blood donor deferral in northern Jeddah: a single-center study,» *Journal of Blood Medicine*, vol. 10, pp. 47-51, 2019.
- [10] A. Shah, P. Joshi, G. Aghera y K. Shah, «A study on analysis of blood donation deferral during blood donation camp at tertiary-care teaching hospital in south Gujarat region,» *International Journal of Medical Science and Public Health*, vol. 5, n° 5, pp. 894-897, 2016.
- [11] NCCIH, «Cupping,» [En línea]. Available: <https://nccih.nih.gov/health/cupping>. [Último acceso: 16 July 2019].

- [12] MedlinePlus, «Ictericia,» 4 Junio 2019. [En línea]. Available: <https://medlineplus.gov/spanish/jaundice.html>. [Último acceso: 22 Julio 2019].
- [13] M. Chau, C.-K. Lee, C. W. Chan y E. Cheng, «Data Analysis for Healthcare: A Case Study in Blood Donation Center Analysis,» *Research Gate*, 2010.
- [14] I.-C. Yeng, «Blood Transfusion Service Center Data Set,» University of California, Irving, 3 Octubre 2008. [En línea]. Available: <https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center>.
- [15] Driven Data, «Predicting Blood Donations,» [En línea]. Available: <https://www.drivendata.org/competitions/2/warm-up-predict-blood-donations/page/5/>. [Último acceso: 04 Agosto 2019].
- [16] P. Kalinda, «Predicting Blood Donations,» 6 Septiembre 2016. [En línea]. Available: <http://www.philipkalinda.com/ds5.html>. [Último acceso: 4 Agosto 2019].
- [17] D. Bahel, P. Ghosh, A. Sarkar y M. A. Lanham, 2017. [En línea]. Available: [http://www.matthewalanham.com/Students/2017\\_MWDSI\\_Final\\_Bahel.pdf](http://www.matthewalanham.com/Students/2017_MWDSI_Final_Bahel.pdf). [Último acceso: 17 Julio 2019].
- [18] L. Dean, Blood Groups and Red Cell Antigens, Estados Unidos: National Center for Biotechnology Information, 2005.
- [19] American Red Cross, «Facts About Blood and Blood Types,» 2019. [En línea]. Available: <https://www.redcrossblood.org/donate-blood/blood-types.html>. [Último acceso: 20 Agosto 2019].
- [20] Fundación Banco de Sangre y Tejidos de las Islas Baleares, «Dona Sangre,» 2019. [En línea]. Available: [http://www.donasang.org/que-es-la-sang/es\\_quadre-de-compatibilitat.html](http://www.donasang.org/que-es-la-sang/es_quadre-de-compatibilitat.html). [Último acceso: 21 Agosto 2019].
- [21] Organización Mundial de la Salud, «Who can give blood,» 2018. [En línea]. Available: <https://www.who.int/campaigns/world-blood-donor-day/2018/who-can-give-blood>. [Último acceso: 21 Agosto 2019].
- [22] Merca 2.0, «Donantes voluntarios y por reposición en México,» 16 Junio 2018. [En línea]. Available: <https://www.merca20.com/donantes-voluntarios-y-por-reposicion-en-mexico/>. [Último acceso: 21 Agosto 2019].
- [23] Organización Mundial de la Salud, «¿Por qué es importante donar sangre?,» Abril 2016. [En línea]. Available: <https://www.who.int/features/qa/61/es/>. [Último acceso: 22 Agosto 2019].
- [24] Secretaría de Salud, «Banco de Sangre,» [En línea]. Available: [http://www.cnts.salud.gob.mx/interior/trans\\_bancodesangre.html](http://www.cnts.salud.gob.mx/interior/trans_bancodesangre.html). [Último acceso: 30 Septiembre 2019].

- [25] Instituto de Información Estadística y Geográfica, «Población en Jalisco,» 10 Julio 2018. [En línea]. Available: <https://iieg.gob.mx/strategos/portfolio/poblacion-en-jalisco-2018/>. [Último acceso: 2019 Septiembre 30].
- [26] Pew Research Center, «Defining generations: Where Millennials end and Generation Z begins,» 17 Enero 2019. [En línea]. Available: <https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/>. [Último acceso: 27 Agosto 2019].
- [27] J. Bell, Machine Learning: Hands-On for Developers and Technical Professionals, Indianapolis: Wiley, 2015.
- [28] X. Yan, X. Su y World Scientific (FIRM), Linear Regression Analysis: Theory And Computing, Singapore: World Scientific, 2009.
- [29] NumPy developers, «NumPy,» 2019. [En línea]. Available: <https://numpy.org>. [Último acceso: 25 Septiembre 2019].
- [30] NUMFOCUS, «Python Data Analysis Library,» 22 Agosto 2019. [En línea]. Available: <https://pandas.pydata.org>. [Último acceso: 25 Septiembre 2019].
- [31] Matplotlib, «Sample plots in Matplotlib,» 26 Agosto 2019. [En línea]. Available: [https://matplotlib.org/tutorials/introductory/sample\\_plots.html](https://matplotlib.org/tutorials/introductory/sample_plots.html). [Último acceso: 27 Septiembre 2019].
- [32] Scikit-learn developers, «Classifier comparison,» 2019. [En línea]. Available: [https://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html). [Último acceso: 1 Octubre 2019].
- [33] G. Castro, «Muestreo estadístico,» Monografías.com, [En línea]. Available: <https://www.monografias.com/trabajos39/muestreo-estadistico/muestreo-estadistico.shtml>. [Último acceso: 06 Octubre 2019].

## APÉNDICE A. Resultados completos de la encuesta

A continuación, los resultados generados por *Google Forms* de la encuesta con todas las preguntas realizadas.

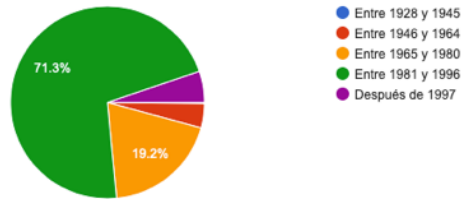
515 respuestas

SUMMARY INDIVIDUAL

Accepting responses

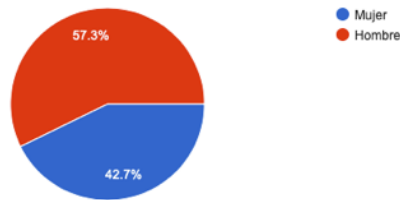
### ¿En qué año naciste?

515 respuestas



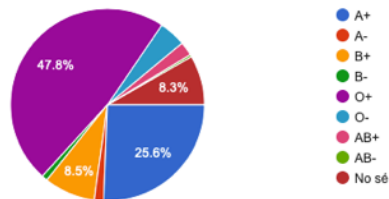
### Eres:

515 respuestas



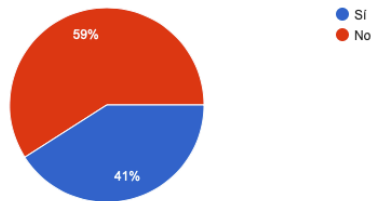
### ¿Cuál es tu grupo sanguíneo?

515 respuestas



### ¿Has donado sangre alguna vez?

515 responses

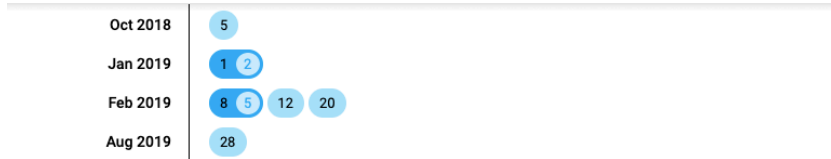


### Sí he donado

### ¿Recuerdas cuándo fue la primera vez que donaste? (mes y año) El día no es importante

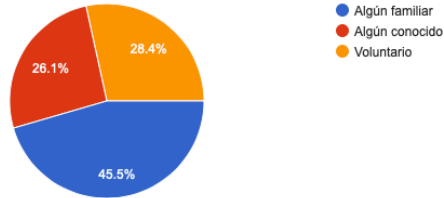
192 responses, 61 not shown





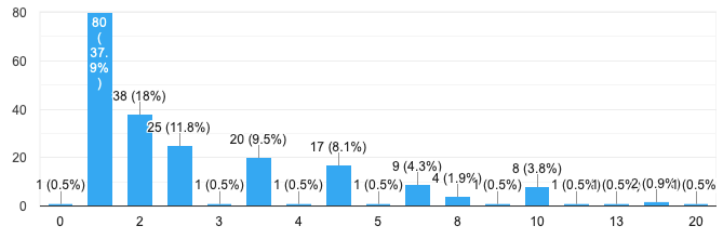
### La primera vez que donaste, fue:

211 responses



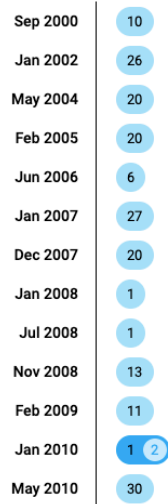
### ¿Cuántas veces has donado?

211 responses



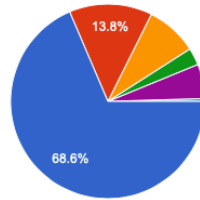
### ¿Cuándo fue la última vez que donaste? (mes y año) El día no es importante

191 responses, 52 not shown



### ¿Te generó un sentimiento positivo donar sangre?

210 responses

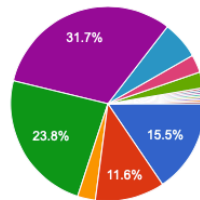


- Sí, me gustó ayudar
- Sí, me genera satisfacción
- No, es algo normal para mí
- No, me ocasionó dolor y malestar físico
- No, fue un requisito
- Me gustó ayudar pero casi me les desmayo por ver sangre. xD

## No he donado

### Si no has donado, ¿a qué se debe?

303 responses



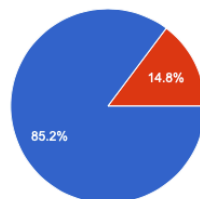
- Alguna enfermedad me lo ha imped...
- Miedo
- Mis venas son muy delgadas
- No cumplo con los requisitos para d...
- No he tenido la necesidad
- No sé dónde hacerlo
- Tenía tatuajes o perforaciones recie...
- Problemas con la presión

▲ 1/2 ▼

## Estadísticas México

### Si pudieras donar sangre a un desconocido, ¿lo harías?

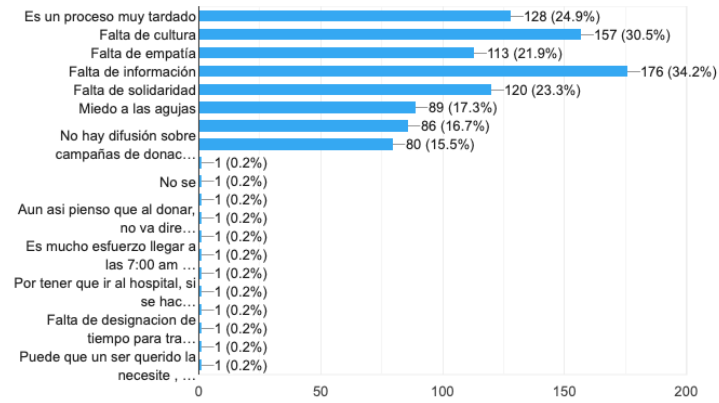
515 responses



- Si
- No

**En México menos del 4% de la población dona sangre de manera altruista, ¿a qué crees que se deba?**

515 respuestas



## APÉNDICE B. Código

### Blood Donation Analysis

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import random
import math

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split

In [ ]: # Set pandas output display to have one digit for decimal places and limit it to
# printing 15 rows.
pd.options.display.float_format = '{:.2f}'.format
pd.options.display.max_rows = 8
plt.rcParams['figure.figsize'] = [15, 10]

In [ ]: # Columns from the dataset
bda_initial_headers = ["timestamp", "age", "gender", "bloodType",
                       "donator", "firstDonation", "beneficiary",
                       "donationTimes", "lastDonation", "feelingAfterDonation",
                       "reasonNotDonate", "wouldDonate", "mexicoOpinion"]

blood_initial_dataset = pd.read_csv('BloodDonationAnalysis.csv', sep=',',
                                   names=bda_initial_headers,
                                   header=0)

print("Data set loaded. Total responses: ", len(blood_initial_dataset))
blood_initial_dataset

In [ ]: # Separate the dataset to get the donators information
# and substitute the NaN values with zero
donators_headers = ["timestamp", "age", "gender", "bloodType",
                    "firstDonation", "beneficiary", "donationTimes",
                    "lastDonation", "feelingAfterDonation", "wouldDonate"]

not_donators_headers = ["timestamp", "age", "gender", "bloodType",
                        "reasonNotDonate", "wouldDonate"]

blood_initial_dataset.fillna(0, inplace=True)

donators = blood_initial_dataset.loc[blood_initial_dataset['donator'] == 'Si']
not_donators = blood_initial_dataset.loc[blood_initial_dataset['donator'] == 'No']

In [ ]: # Donators information
print("Total Donators: ", len(donators))
donators[donators_headers]

In [ ]: # Donations by gender
men = donators.loc[donators['gender'] == 'Hombre']
women = donators.loc[donators['gender'] == 'Mujer']
print("Total Men: ", len(men))
print("Total Women: ", len(women))

In [ ]: blood_initial_dataset

In [ ]: def get_donation_year(arr):
yearsArray = []
for i in range(0, len(arr)):
    if (arr[i] != 0):
        yearsArray.append(int(arr[i].split("-")[0]))
    else:
        yearsArray.append(0)
return yearsArray

In [ ]: first_donation_years = get_donation_year(np.array(donators["firstDonation"]))
donation_times = np.array(donators["donationTimes"])
last_donation_years = get_donation_year(np.array(donators["lastDonation"]))
print("First Donation Years: ", len(first_donation_years))
print("Last Donation Years: ", len(last_donation_years))
print("Total Donators: ", len(donation_times))
print("Total Donations: ", sum(donation_times))

In [ ]: df = pd.DataFrame()
df['firstDonation'] = first_donation_years
df['lastDonation'] = last_donation_years
df['donationTimes'] = donation_times
df
```

```

In [ ]: def get_first_donation_year(arr):
        min_result = 10000
        for i in range(0, len(arr)):
            if(arr[i] < min_result and arr[i] > 0):
                min_result = arr[i]
        return min_result

In [ ]: min_year = get_first_donation_year(first_donation_years)
        print("The first donation was on: ", min_year)
        max_year = max(last_donation_years)
        print("The last donation was on: ", max_year)

In [ ]: YEARS_GAP = max_year - min_year

In [ ]: # When someone has donated more than once over the years we need to know in which range of years he/she has donated
        # blood. With this range we can, randomly, assign a year for those intermediates donations.

        def get_years_difference(first_donation_years, last_donation_years, YEARS_GAP, donation_times):
            years_difference = []
            for i in range(0, len(first_donation_years)):
                if (first_donation_years[i] > last_donation_years[i]):
                    (first_donation_years[i], last_donation_years[i]) = (last_donation_years[i], first_donation_years[i])
                if (abs(last_donation_years[i] - first_donation_years[i]) > YEARS_GAP):
                    if (donation_times[i] == 1):
                        max_value = max(last_donation_years[i], first_donation_years[i])
                        last_donation_years[i] = max_value
                        first_donation_years[i] = max_value
                    if (first_donation_years[i] == 0 and donation_times[i] > 1):
                        first_donation_years[i] = last_donation_years[i] - donation_times[i]
                    if (first_donation_years[i] > 0 and last_donation_years[i] == 0 and donation_times[i] > 0):
                        last_donation_years[i] = first_donation_years[i] + donation_times[i]

                years_difference.append(last_donation_years[i] - first_donation_years[i])
            return years_difference

In [ ]: years_difference = get_years_difference(first_donation_years, last_donation_years, YEARS_GAP, donation_times)
        print(years_difference)

In [ ]: def generate_donations_by_year(first_donation_years, last_donation_years, min_year, max_year,
        donation_times, YEARS_GAP, years_difference):

        donations_by_year = {}
        for i in range(min_year, max_year + 1):
            donations_by_year[i] = 0

        for i in range(0, len(donation_times)):
            if (first_donation_years[i] == 0):
                first_donation_years[i] = min_year + random.randint(0, YEARS_GAP)
                donations_by_year[first_donation_years[i]] += 1

            if (donation_times[i] > 1):
                if (last_donation_years[i] == 0):
                    donations_by_year[max_year - random.randint(0, YEARS_GAP)] += 1
                else:
                    donations_by_year[last_donation_years[i]] += 1

            if (donation_times[i] > 2):
                for j in range(0, int(donation_times[i]) - 2):
                    random_year = random.randint(0, years_difference[i])
                    donations_by_year[first_donation_years[i] + random_year] += 1
        print(sum(donations_by_year.values()))
        return donations_by_year

In [ ]: print(generate_donations_by_year(first_donation_years, last_donation_years, min_year, max_year, donation_times, YEARS_
        GAP, years_difference))

In [ ]: def get_model_data(first_donation_years, last_donation_years,
        min_year, max_year, donation_times, YEARS_GAP, years_difference, filename="", new_data = False):

        if (new_data):
            new_dict = generate_donations_by_year(first_donation_years,
                last_donation_years,
                min_year,
                max_year,
                donation_times,
                YEARS_GAP,
                years_difference)

            np.save('dictionaries/'+filename+'.npy', new_dict)
            return new_dict
        else:
            try:
                read_dict = np.load('dictionaries/'+filename+'.npy', allow_pickle = True).item()
            except FileNotFoundError:

```

```

except FileNotFoundError:
    print(filename, 'does not exist!')
    read_dict = generate_donations_by_year(first_donation_years,
                                          last_donation_years,
                                          min_year,
                                          max_year,
                                          donation_times,
                                          YEARS_GAP,
                                          years_difference)
    np.save('dictionaries/'+filename+'.npy', read_dict)
return read_dict

```

```

In [ ]: #get_model_data(first_donation_years,
#                   last_donation_years,
#                   min_year,
#                   max_year,
#                   donation_times,
#                   filename = "completeDonators",
#                   new_data = False)

```

```

In [ ]: #get_model_data(first_donation_years,
#                   last_donation_years,
#                   min_year,
#                   max_year,
#                   donation_times,
#                   filename = "completeDonators",
#                   new_data = True)

```

## Linear Regression Model

```

In [ ]: # Get X and y values.
# X will represent the year and Y the total of donations by year
all_donators = get_model_data(first_donation_years,
                              last_donation_years,
                              min_year,
                              max_year,
                              donation_times,
                              YEARS_GAP,
                              years_difference,
                              filename = "completeDonators",
                              new_data = False)
x = np.array(range(min_year, max_year+1)).reshape(-1, 1)
y = np.array(list(all_donators.values()))

```

```

In [ ]: # Creating the Linear Regression Model

LR_model = LinearRegression()

LR_model.fit(x, y)

```

```

In [ ]: # Obtaining coefficient of determination (R^2)
r_sq = LR_model.score(x, y)
print('coefficient of determination:', r_sq)

# y = b0 + b1X
print ('Intercept (b0): ', LR_model.intercept_)
print ('Slope (b1): ', LR_model.coef_)

```

```

In [ ]: # Prediction
y_pred_next_10_years = LR_model.predict(np.array(range(2020, 2030)).reshape(-1, 1))
print('Predicted donations: ', y_pred_next_10_years, sep=' ')

```

```

In [ ]: import math

# Predict donations from 1980 (min year) to 2019(max year) to evaluate our LR model
y_pred_model = LR_model.predict(np.array(range(1980, 2020)).reshape(-1, 1))

mse = math.sqrt(mean_squared_error(y, y_pred_model))
print("Mean Squared Error: ", mse)

```

```

In [ ]: plt.scatter(x, y, color='red')
plt.plot(x, LR_model.predict(x), color='blue', linewidth=2)
plt.title('Years vs Number of Donations')
plt.xlabel('Years')
plt.ylabel('Number of Donations')

plt.show()

```

## Linear Regression with Normalization

```
In [ ]: LR_Norm_model = LinearRegression(normalize=True)
LR_Norm_model.fit(x, y)

In [ ]: # Obtaining coefficient of determination (R^2)
r_sq_n = LR_Norm_model.score(x, y)
print('coefficient of determination:', r_sq_n)

# y = b0 + b1X
print ('Intercept (b0): ', LR_Norm_model.intercept_)
print ('Slope (b1): ', LR_Norm_model.coef_)

In [ ]: # Prediction
y_pred_next_10_years = LR_model.predict(np.array(range(2020, 2030)).reshape(-1, 1))
print('Predicted donations: ', y_pred_next_10_years, sep=' ')

In [ ]: import math

# Predict donations from 1980 (min year) to 2019(max year) to evaluate our LR model
y_pred_model = LR_Norm_model.predict(np.array(range(1980, 2020)).reshape(-1, 1))

mse = math.sqrt(mean_squared_error(y, y_pred_model))
print("Mean Squared Error: ", mse)

In [ ]: plt.scatter(x, y, color='red')
plt.plot(x, LR_Norm_model.predict(x), color='blue', linewidth=2)
plt.title('Years vs Number of Donations')
plt.xlabel('Years')
plt.ylabel('Number of Donations')

plt.show()

In [ ]: print('coefficient of determination without normalization:', r_sq)
print('coefficient of determination with normalization:', r_sq_n)
```

## Polynomial Regression Model

```
In [ ]: from sklearn.preprocessing import PolynomialFeatures as PF

In [ ]: def create_polynomial_reg_model (degree = 2):
    polynomial_features = PF(degree = degree)
    X_polynomial = polynomial_features.fit_transform(x)
    polynomial_model = LinearRegression()
    polynomial_model.fit(X_polynomial, y)

    y_train_predicted = polynomial_model.predict(X_polynomial)

    y_test_predicted = polynomial_model.predict(polynomial_features
                                                .fit_transform(np.array(range(2020, 2030))
                                                                .reshape(-1, 1)))

    rmse = math.sqrt(mean_squared_error(y, y_train_predicted))
    r2 = r2_score(y, y_train_predicted)

    print("RMSE is {}".format(rmse))
    print("R2 score is {}".format(r2))
    print(y_test_predicted)

    return (x, y_train_predicted)

In [ ]: plt.scatter(x, y, color='red')
x_p, y_p = create_polynomial_reg_model(2)
plt.plot(x_p, y_p, color='blue', linewidth=3, label="Degree 2")
x_p, y_p = create_polynomial_reg_model(3)
plt.plot(x_p, y_p, color='green', linewidth=3, label="Degree 3")
plt.title('Years vs Number of Donations')
plt.xlabel('Years')
plt.ylabel('Number of Donations')
plt.legend()
plt.show()
```

## Analysis by blood type

```
In [ ]: donators.head()

In [ ]: # Get the total by each blood type
a_positive = donators.loc[donators["bloodType"] == "A+"]
a_negative = donators.loc[donators["bloodType"] == "A-"]
b_positive = donators.loc[donators["bloodType"] == "B+"]
b_negative = donators.loc[donators["bloodType"] == "B-"]
o_positive = donators.loc[donators["bloodType"] == "O+"]
o_negative = donators.loc[donators["bloodType"] == "O-"]
ab_positive = donators.loc[donators["bloodType"] == "AB+"]
ab_negative = donators.loc[donators["bloodType"] == "AB-"]
unknown_donator = donators.loc[donators["bloodType"] == "No sé"]

print ("A+: ", len(a_positive))
print ("A-: ", len(a_negative))
print ("B+: ", len(b_positive))
print ("B-: ", len(b_negative))
print ("O+: ", len(o_positive))
print ("O-: ", len(o_negative))
print ("AB+: ", len(ab_positive))
print ("AB-: ", len(ab_negative))
print ("Unkown: ", len(unknown_donator))

In [ ]: # filenames by blood type
apositiveFile = "ApositiveDonators"
anegativeFile = "AnegativeDonators"
bpositiveFile = "BpositiveDonators"
bnegativeFile = "BnegativeDonators"
opositiveFile = "OpositiveDonators"
onegativeFile = "OnegativeDonators"
abpositiveFile = "ABpositiveDonators"

In [ ]: def generate_prediction(donatorsArray, filename):
    first_donation_years = get_donation_year(np.array(donatorsArray["firstDonation"]))
    donation_times_arr = np.array(donatorsArray["donationTimes"])
    last_donation_years = get_donation_year(np.array(donatorsArray["lastDonation"]))
    print("First Donation Years(len): ", len(first_donation_years))
    print("Last Donation Years(len): ", len(last_donation_years))
    print("Total Donators: ", len(donation_times_arr))
    print("Total Donations: ", sum(donation_times_arr))

    min_year = get_first_donation_year(first_donation_years)
    max_year = max(last_donation_years)

    YEARS_GAP = max_year - min_year
    years_difference = get_years_difference(first_donation_years, last_donation_years, YEARS_GAP, donation_times_arr)

    min_year = get_first_donation_year(first_donation_years)
    print("The first donation was on: ", min_year)
    max_year = max(last_donation_years)
    print("The last donation was on: ", max_year)

    YEARS_GAP = max_year - min_year

    donators = get_model_data(first_donation_years,
                              last_donation_years,
                              min_year,
                              max_year,
                              donation_times_arr,
                              YEARS_GAP,
                              years_difference,
                              filename = filename,
                              new_data = False)

    x = np.array(range(min_year, max_year+1)).reshape(-1, 1)
    y = np.array(list(donators.values()))

    LR_model = LinearRegression()

    LR_model.fit(x, y)
    r_sq = LR_model.score(x, y)
    print('coefficient of determination:', r_sq)

    # y = b0 + b1X
    print ('Intercept (b0): ', LR_model.intercept_)
    print ('Slope (b1): ', LR_model.coef_)
    # Prediction
    y_pred_next_10_years = LR_model.predict(np.array(range(2020, 2030)).reshape(-1, 1))
    print('Predicted donations: ', y_pred_next_10_years, sep=' ')

    # Predict donations from 1980 (min year) to 2019(max year) to evaluate our LR model
    y_pred_model = LR_model.predict(np.array(range(min_year, max_year+1)).reshape(-1, 1))

    mse = math.sqrt(mean_squared_error(y, y_pred_model))
    print("Mean Squared Error: ", mse)

    plt.scatter(x, y, color='red')
    plt.plot(x, LR_model.predict(x), color='blue', linewidth=2)
    plt.title('Years vs Number of Donations')
    plt.xlabel('Years')
    plt.ylabel('Number of Donations')

    plt.show()
```

### **A+**

```
In [ ]: a_positive
```

```
In [ ]: generate_prediction(a_positive, apositiveFile)
```

### **A-**

```
In [ ]: a_negative
```

```
In [ ]: generate_prediction(a_negative, anegativeFile)
```

### **B+**

```
In [ ]: b_positive
```

```
In [ ]: generate_prediction(b_positive, bpositiveFile)
```

### **B-**

```
In [ ]: b_negative
```

```
In [ ]: generate_prediction(b_negative, bnegativeFile)
```

### **AB+**

```
In [ ]: ab_positive
```

```
In [ ]: generate_prediction(ab_positive, abpositiveFile)
```

### **O+**

```
In [ ]: o_positive.head(14)
```

```
In [ ]: generate_prediction(o_positive, opositiveFile)
```

### **O-**

```
In [ ]: o_negative
```

```
In [ ]: generate_prediction(o_negative, onegativeFile)
```