

INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE OCCIDENTE



ITESO

Universidad Jesuita
de Guadalajara

DEPARTAMENTO DE MATEMÁTICAS Y FÍSICA

PAP 4J09: Ciencia de Datos Aplicado a Información de Negocios y Desarrollo Social.

Programa de Modelación Matemática para el Desarrollo de Planes y Proyectos de Negocio.

Fecha: 14 de febrero, 2022

Profesor: Pablo Benavides Herrera

Integrantes: Ingeniería Financiera - C. Ricardo Figueroa Acosta
Ingeniería en Sistemas - C. Pedro Javier Herrera Soto
Licenciatura en Derecho - C. Carlos Riggen Castillo
Ingeniería Mecánica - C. Luis Eduardo Vazquez Gonzalez

Tlaquepaque, Jalisco, 14 de mayo del 2022.

Índice Reporte PAP

Presentación Institucional de los Proyectos de Aplicación Profesional

Resumen

1.- Ciclo participativo del Proyecto de Aplicación Profesional

1.1.- Entendimiento del ámbito y del contexto

1.2.- Caracterización de la organización o comunidad

1.3.- Identificación de las problemáticas

1.4. Planeación de las alternativas

1.5. Desarrollo de la propuesta de mejora

1.6. Valoración de productos, resultados e impactos

1.7. Bibliografía y otros recursos

2.- Productos

3. Reflexión crítica y ética de la experiencia

3.1.- Sensibilización ante las realidades

3.2.- Aprendizajes logrados

Presentación Institucional de los Proyectos de Aplicación Profesional

Los Proyectos de Aplicación Profesional (PAP) son experiencias socio-profesionales de los alumnos que desde el currículo de su formación universitaria- enfrentan retos, resuelven problemas o innovan una necesidad sociotécnica del entorno, en vinculación (colaboración) (co-participación) con grupos, instituciones, organizaciones o comunidades, en escenarios reales donde comparten saberes.

El PAP, como espacio curricular de formación vinculada, ha logrado integrar el Servicio Social (acorde con las Orientaciones Fundamentales del ITESO), los requisitos de dar cuenta de los saberes y del saber aplicar los mismos al culminar la formación profesional (Opción Terminal), mediante la realización de proyectos profesionales de cara a las necesidades y retos del entorno (Aplicación Profesional).

El PAP es un proceso acotado en el tiempo en que los estudiantes, los beneficiarios externos y los profesores se asocian colaborativamente y en red, en un proyecto, e incursionan en un mundo social, como actores que enfrentan verdaderos problemas y desafíos traducibles en demandas pertinentes y socialmente relevantes. Frente a éstas transfieren experiencia de sus saberes profesionales y demuestran que saben hacer, innovar, co-crear o transformar en distintos campos sociales.

El PAP trata de sembrar en los estudiantes una disposición permanente de encargarse de la realidad con una actitud comprometida y ética frente a las asimetrías sociales. En otras palabras, se trata del reto de “saber y aprender a transformar”.

El Reporte PAP consta de tres componentes:

El primer componente se refiere al ciclo participativo del PAP, en donde se documentan las diferentes fases del proyecto y las actividades que tuvieron lugar durante el desarrollo de este y la valoración de las incidencias en el entorno.

En caso de requerirse alguna adecuación al nombre de las fases propuestas para este componente, se puede realizar siempre y cuando sea complementario a lo ya establecido.

El segundo componente presenta los productos elaborados de acuerdo con su tipología.

El tercer componente es la reflexión crítica y ética de la experiencia, el reconocimiento de las competencias y los aprendizajes profesionales que el estudiante desarrolló en el transcurso de su labor.

Resumen

El objetivo del equipo dentro de este Proyecto de Aplicación Profesional (PAP), titulado "*Ciencia de datos aplicado a información de negocios y desarrollo social*", es analizar el progreso de la pandemia COVID-19 provocada por el virus del síndrome respiratorio agudo severo "SARS COV 2" a través del desarrollo de modelos analíticos que evalúen las métricas más relevantes para poder determinar la importancia de la aplicación del "*booster*" como herramienta epidemiológica para reducir la mortalidad del virus, así como de modelos predictivos que, utilizando la información de las olas previas, puedan predecir las siguientes curvas de contagios y defunciones en México. Usando los lenguajes de programación "R" se realizó un modelo de regresión para analizar las muertes ocasionadas por el COVID-19. De igual manera, usando el lenguaje "tensor flow" se realizó un modelo para analizar los contagios ocasionados por el COVID-19. Los datos para realizar todos los productos antes dichos, se extrajeron de la página de internet de "Our World in Data". En ambos modelos se puede apreciar la importancia de la vacunación, así como de los refuerzos y cómo estas variables pueden ayudar a ponerle un fin a esta pandemia.

1 Ciclo Participativo del Proyecto de Aplicación Profesional

El PAP es una experiencia de aprendizaje y de contribución social integrada por estudiantes, profesores, actores sociales y responsables de las organizaciones, que de manera colaborativa construyen sus conocimientos para dar respuestas a problemáticas de un contexto específico y en un tiempo delimitado. Por tanto, la experiencia PAP supone un proceso en lógica de proyecto, así como de un estilo de trabajo participativo y recíproco entre los involucrados.

Investigación	
Talleres y asesorías	
Actividades técnicas	

El plan de trabajo, bajo el cual se buscará cumplir con el objetivo será desarrollado de la siguiente forma:

Actividad		Semana 1-4	Semana 5-8	Semana 9-12	Semana 13-16
Sesión de Presentación de PAP					
Selección del Objeto de Estudio					
Investigación el el estado del arte					
Diseño de prototipos					
Difusión en comunidad universitaria de resultados					
Desarrollo de la propuesta					
Simulaciones					
Preparación de productos finales					

1.1- Entendimiento del ámbito y del contexto

En diciembre de 2019, en la ciudad de Wuhan, China, se detectaba el primer caso de un nuevo coronavirus (un virus conocido por ocasionar resfriados leves, llamado así por una capa exterior en forma de corona con proteínas spike, lo que le da su nombre). Con muchas teorías detrás, pero sin un claro origen identificable, el nuevo coronavirus denominado “COVID-19” comenzó a esparcirse rápidamente por la Wuhan, y muy pronto por el mundo.

Si bien en los primeros dos meses del año 2020 la enfermedad no causaba mucha alarma y los datos de China eran prometedores (en Wuhan, con una población de 11 millones de habitantes sólo se habían detectado poco más de 10,000 infecciones y 213 muertes), las acciones del gobierno eran preocupantes: en menos de dos semanas, para poder atender a todos los pacientes, China construyó dos grandes hospitales con más de 2,500 camas de hospital (Miller, 2020).

Poco a poco, la enfermedad abandonó el gigante asiático y llegó a Europa, a dos de los países más visitados del mundo: España e Italia. En pocas semanas y entrando a marzo, Italia se convirtió en el primer país de Occidente en imponer medidas de control, decretando un confinamiento nacional y suspendiendo todo tipo de actividad deportiva, religiosa o laboral para una población de más de 60 millones de personas. El resto de la Unión Europea siguió sus pasos y comenzaron a decretar toques de queda y cuarentenas masivas, suspendiendo por completo las actividades económicas de todos estos países, algo inédito que nadie esperaba.

Los casos, hospitalizaciones y las muertes reportadas aceleraban con una velocidad preocupante. El 11 de marzo de 2020, la Organización Mundial de la Salud declaraba al COVID-19 como pandemia. En menos de un mes, más de la mitad de la población mundial se encontraba bajo algún tipo de cuarentena mientras los hospitales se seguían llenando.

La COVID-19 paralizó al mundo. La mayoría de los gobiernos del mundo se enfocaron en luchar por contener el virus. Los científicos corrían a identificar y encontrar tratamientos que funcionaran contra, así como iniciando una carrera contra el tiempo por desarrollar una vacuna efectiva que pudiera prevenir la enfermedad, las hospitalizaciones y las muertes.

Al día de hoy, se han reportado más de 420 millones de casos y 5.86 millones de muertes según la Universidad John Hopkins. Sin embargo, la realidad es, probablemente, mucho peor. Existe un subregistro en muchos países de las causalidades reales provocadas por la pandemia, como en el caso de México, donde si bien al día de hoy sólo se han detectado 5.37 millones de casos y 315,000 muertes, gracias a la política de la gestión de la pandemia no existe un registro real de la cantidad de casos por la poca disponibilidad de pruebas y en el caso de las muertes se ha analizado el número de defunciones reales vs el número de defunciones esperadas en relación con la media de los 5 años anteriores a la

pandemia y las cifras son alarmantes: se estima que el número real es 3 veces mayor al reportado, según el Instituto de Evaluación y Métricas de Medicina, de la Universidad de Washington (lo que daría un total de casi 1 millón de muertes en los 2 años de la pandemia).

La pandemia ha destruido millones de familias por todo el mundo. Durante el primer año de la pandemia, no había mucha oportunidad ni margen de acción para reducir los casos o las muertes más allá de confinamientos. Sin embargo, en 2021 llegaron las vacunas, desarrolladas por los más altos laboratorios del mundo (Pfizer, AstraZeneca, Moderna, SinoVac, Sinopharm, Sputnik VI, etc.) y ofrecían la primera esperanza de terminar con la pandemia.

Los datos respaldaron inmediatamente la efectividad de las vacunas, las muertes se redujeron sustancialmente en proporción a los casos. Las muertes que se han producido este último año de la pandemia han sido, mayoritariamente, de personas no vacunadas. Sin embargo, mientras más avanzaba el año y nuevas variantes comenzaron a surgir, la inicial eficacia de las vacunas se comenzó a reducir y las muertes en personas vacunadas, si bien seguían siendo bajas en proporción a las de personas no vacunadas, comenzaron a escalar poco a poco.

Bajo esta premisa, las farmacéuticas y los gobiernos del mundo comenzaron a estudiar la posibilidad de aplicar una dosis de refuerzo que fortaleciera la protección que ya otorgaban las vacunas, para intentar reducir lo más posible la mortalidad. Países como Chile, Turquía, Estados Unidos e Israel, desde mediados del año pasado comenzaron a aplicar dosis de refuerzo a personas inmunodeprimidas o en mayor situación de vulnerabilidad. En otoño de 2021, la mayoría de los gobiernos del mundo desarrollado comenzaron con la aplicación masiva de dosis de refuerzo a personas que se habían vacunado al principio de la pandemia, dependiendo de la vacuna específica que se habían colocado.

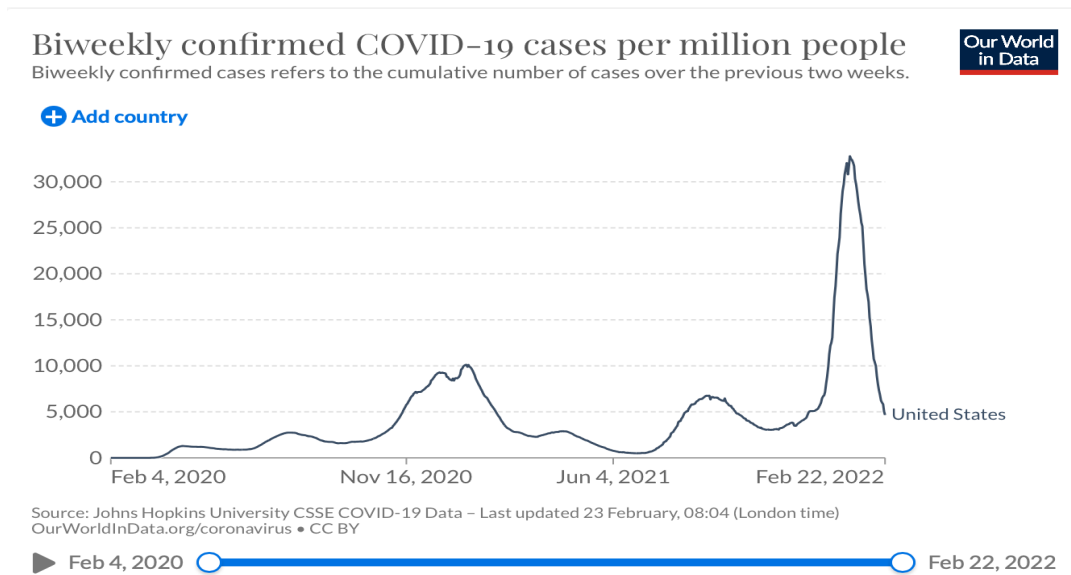
Sin embargo, la recepción para el público en general de la tercera dosis no ha sido tan positiva. A pesar de estar recomendada por las organizaciones de salud más importantes de cada país, existe un rezago importante en la aplicación de los *booster*. En Estados Unidos, el país más afectado por la pandemia, líder en casos y muertes a nivel mundial (78.4 millones de casos y 935 mil muertes), el porcentaje de la población con su dosis de *booster* es inferior al 30% de la población total.

Existe una reticencia generalizada dentro de la población vacunable para aplicarse la dosis de refuerzo, y si bien los esfuerzos para vacunar a la totalidad de la población elegible no han sido suficientes completamente, el rezago es considerablemente más marcado con la aplicación del refuerzo. En este sentido, es fundamental continuar con los esfuerzos de promoción de la vacunación, agregando el matiz adicional sobre el refuerzo y su importancia. La recomendación médica de las instituciones oficiales de salud nacionales e internacionales es aplicar el refuerzo

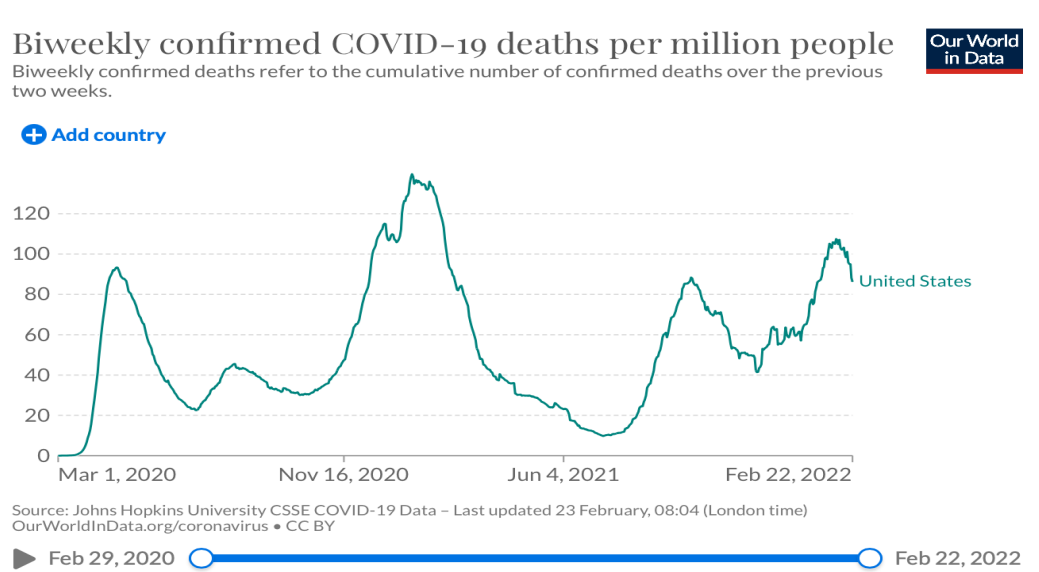
en cuanto la persona sea elegible. Es importante corroborar esto con información accesible para fortalecer la importancia de la aplicación del refuerzo.

1.2.- Caracterización de la organización o comunidad

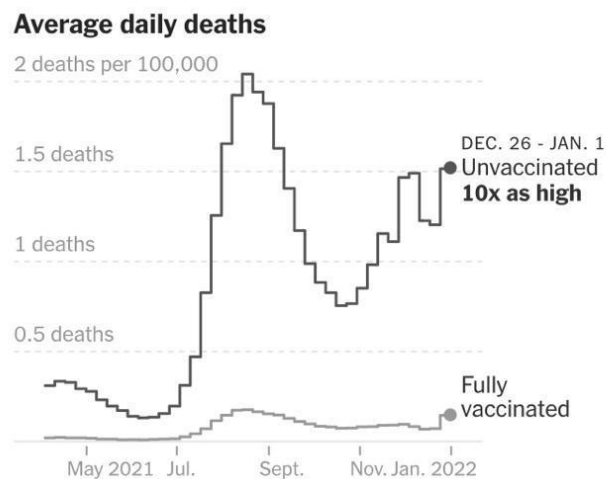
La enfermedad respiratoria provocada por el virus SARS COV 2 ha cobrado millones de vidas alrededor del globo, sin embargo, a diferencia de 2020 donde la pandemia arrasó con todas las personas por igual, en el 2021 y en el transcurso del 2022 las cosas han cambiado. Si bien en países como México y Estados Unidos las muertes totales se han mantenido altas, la tasa de letalidad del virus ha disminuido radicalmente.



De la gráfica anterior se desprenden los casos totales confirmados por millón de habitantes en Estados Unidos. La pandemia ha sido estacional, con picos y olas que han variado en el tiempo como se puede observar. Las dos olas de invierno han sido las que más casos han surgido, especialmente la última ola que provocó la variante Omicrón. Sin embargo, la mortalidad se ha mantenido estable y por debajo de olas que tuvieron menos casos:



Una de las principales razones que justifican la caída de la mortalidad fue que en las últimas olas, aparentemente, la variante del virus era menos letal. Dicha situación en concreto no se ha demostrado completamente, pero es una de las razones que podrían justificar la reducción en la letalidad. Otra razón, más importante y de trascendencia mayor para la eliminación de la pandemia, es la llegada de las vacunas. Las muertes en personas vacunadas en relación con las muertes en personas no vacunadas desde el inicio de la aplicación masiva de vacunas al inicio del año pasado son sustancialmente menores (casi una tasa de 10 a 1) como se desprende de la siguiente gráfica :



La vacuna ha estado disponible para mayores de 18 años, de manera masiva y altamente accesible desde primavera del año 2021. Sin embargo, hasta el día de hoy, existe un 15% de esa población adulta que ha optado por no vacunarse. Ese ha sido el grupo poblacional más afectado por las nuevas olas de la pandemia, ya que hasta el momento las vacunas han demostrado seguir siendo efectivas. En este sentido, la mayoría de las muertes durante 2021 y 2022 han sido en personas que han optado por no vacunarse.

Las estadísticas, informes médicos y autoridades sanitarias han demostrado continuamente que la vacuna es la mejor forma de prevenir la muerte a causa del COVID-19. Si bien aún no es 100% efectiva y existe un riesgo, este es mínimo si se cuenta con un esquema completo. Grupos poblacionales más jóvenes han considerado retrasar la aplicación de su vacuna o de su segunda dosis ya que creen que están en un menor riesgo que las personas mayores, por lo mismo, las campañas se deben reforzar en estos grupos poblacionales, incluyendo en el ITESO con la aplicación del refuerzo que se encuentra próximo.

El escenario del PAP estará centrado en estudiar el desarrollo de la pandemia a través de la estadística, utilizando datos confiables que permitan el desarrollo de modelos que demuestran la importancia de la vacunación activa contra la COVID-19, en particular atención a la aplicación de la dosis de refuerzo. De igual

manera, a través de un análisis de la estacionalidad, se buscará predecir los siguientes ciclos epidémicos.

De igual forma, se realizará un modelo de regresión usando diferentes variables predictoras, con el fin de entender mejor cómo estas afectan las muertes ocasionadas por el COVID-19. De igual manera se usará dicho modelo de regresión para hacer predicciones a futuro de las muertes ocasionadas por el COVID-19.

También se continuará con el modelo de predicción de contagios realizado en Otoño 2021 y se investigará una forma de corregir el problema de la subestimación drástica de contagios debido al cambio de comportamiento al momento de que se genera una nueva variante, entonces mediante modelos basados en redes neuronales se pretende realizar un modelo que haga el ajuste pertinente al aumento en casos debido a nuevas variantes, para esto se pretende investigar algunas variables epidemiológicas así como estacionales con el objetivo de averiguar cuál es la mejor variable para predecir los contagios.

1.3 Identificación de las problemáticas

Existe una reticencia por parte de las personas de aplicarse una dosis de refuerzo en su vacunación activa por la COVID-19. En este sentido, es importante hacer campañas informativas que promuevan la aplicación del refuerzo, usando información y datos científicos para difundir su importancia.

Es importante retomar los análisis de defunciones dentro de los países que hacen un recuento puntual de los vacunados, no vacunados y vacunados con refuerzo, para con esta información, poder generar gráficos accesibles mediante los cuales se pueda evidenciar la importancia del refuerzo.

Conforme a lo anterior, es fundamental difundir esta información para motivar a las personas indecisas a optar por una vacunación activa con refuerzo. Dentro de un estudio realizado en México a finales del año 2020, estimó que la aceptabilidad de la vacunación contra Covid-19 se encontraba en 62.3%, mientras que el rechazo de 28.2% y las personas con duda completaron el 9.5% restante.

En este sentido es fundamental identificar la importancia de la vacunación activa y con refuerzo, ya que si existe evidencia contundente de que esta reduce la mortalidad en las personas enfermas con COVID-19, debe de promoverse con mayor intensidad por parte del gobierno.

De igual forma, las estrategias que se diseñan por parte del Estado para atender para combatir la pandemia, como la implementación de medidas de reducción de movilidad, uso de cubrebocas, reducción de aforos, entre otras, deben de activarse y desactivarse conforme al desarrollo de la pandemia. El impacto económico de estas medidas es grave y no pueden estar presentes todo el tiempo.

Sin embargo, si se cuentan con criterios claros y se tiene identificada la estacionalidad de la pandemia, se pueda actuar de forma oportuna e inteligentemente.

1.4 Planeación de alternativas

Según un reporte emitido por la Organización Mundial de la Salud, el COVID-19 ha cobrado la vida de 14.9 millones de personas, una cifra superior en casi tres veces a la oficial reportada por el gobierno. Existe un subregistro en muchos países de las causalidades reales provocadas por la pandemia, como en el caso de México, donde si bien en abril de 2022 sólo se habían detectado 5.61 millones de casos y 321,000 muertes según las cifras oficiales de Secretaría de Salud, tan sólo en septiembre de 2021, según el análisis realizado por el INEGI existía un exceso de mortalidad de 653,053 personas, antes incluso de que terminara la ola Delta y empezará la de Omicrón.

La pandemia ha causado mucho daño a millones de familias por todo el mundo. Ha sumido al mundo en una crisis económica sin precedentes con estancamiento en el crecimiento, altas cifras de inflación y una destrucción parcial de la cadena de suministros. En la medida en que el mundo y el gobierno tengan la capacidad de entender mejor el funcionamiento epidemiológico del virus, así como de comprender las estrategias más efectivas, poco a poco se podrá regresar a la normalidad.

Bajo esta tesitura se identificaron tres principales problemas, de los cuales se encontró que era necesario utilizar las herramientas tecnológicas más avanzadas para poder desarrollar modelos que generen mejores estrategias para combatir el COVID-19:

En primer lugar, se encontró que existe una reticencia importante en varios países de obtener su vacunación completa con el refuerzo. Esto se puede deber a que no existe suficiente difusión sobre la importancia de obtener la dosis adicional o que las personas que ya cuentan con su esquema completo tienen confianza en que se encuentran perfectamente protegidas contra el virus y no necesitan abonar más a su vacunación.

En segundo lugar, se encontró que los países diseñan sus políticas de combate a la pandemia con base en los casos y muertes que suceden diariamente. Es decir, un país con pocos casos y pocas muertes implementa pocas o nulas medidas de combate y viceversa. En este sentido, es fundamental comprender el trayecto de la pandemia y poder predecir las futuras olas de muertes, para poder diseñar políticas públicas que disminuyan el riesgo, aumenten la capacidad hospitalaria y minimicen el daño a la población.

En tercer lugar, y en relación con la problemática anterior, es fundamental entender el comportamiento de la pandemia en los países, en particular con el aumento de los contagios por temporadas, para así, implementar medidas que reduzcan la propagación de la pandemia y el riesgo a la población general.

A través de estas 3 problemáticas, se comenzó con la planeación de 2 modelos de predicción distintos y un análisis de la vacunación activa a partir de la información brindada por las autoridades sanitarias oficiales de los países, para poder encontrar soluciones efectivas a cada uno.

1.5. Desarrollo de la propuesta de mejora

Para el caso del primer problema, desde el PAP de Ciencia de Datos Aplicado a Información de Negocios y Desarrollo Social comenzamos a estudiar las cifras en los países que transparentan la información de vacunación activa y defunciones según el estatus de vacunación, para analizar la importancia del refuerzo y poder promocionar su importancia. Consideramos importante en consecuencia, compartir esta información para promover activamente la vacunación considerando que según los más recientes reportes del Gobierno Federal, abril será el último mes de la campaña de vacunación.

Se seleccionó a Suiza, Chile y Estados Unidos para realizar el análisis gracias a que cuentan una política efectiva de transparencia, donde es posible conocer respecto de las defunciones quiénes se encontraban vacunados, quiénes no estaban vacunados y quiénes no contaban con el refuerzo, así como qué fármacos y en qué porcentaje fueron aplicados a la población.

Para el segundo problema se realizaron diferentes modelos de regresión, esto con el objetivo de entender la relación que existen entre diferentes variables en la pandemia y las muertes ocasionadas por el COVID-19, así como para realizar pronósticos de posibles muertes futuras a causa del COVID-19. Dado que uno de los objetivos generales del proyecto es investigar el rol de las dosis de refuerzo en la pandemia, al momento de hacer el modelo se optó por únicamente considerar las fechas en las que se aplicaron dosis de refuerzo, es decir el modelo analiza los datos a partir del 13 de agosto del 2021, que fue cuando empezaron a aplicarse las dosis de refuerzo en Estados Unidos. Para este modelo se seleccionó como país de estudio a Estados Unidos, ya que la información es de fácil acceso. La base de datos usada fue extraída de la página web de "Our World in Data".

Primero se realizó un modelo de regresión múltiple. Se tomó como variable de predicción las nuevas muertes promediadas en 7 días y las variables predictoras son: la tendencia de la pandemia, los nuevos casos diarios confirmados promediados en 7 días, los pacientes en terapia intensiva por día, vacunas aplicadas acumuladas, dosis de refuerzo aplicadas acumuladas, pruebas de

COVID-19 realizadas promediadas en 7 días, la tasa positiva de estas pruebas, el grado de exigencia de las regulaciones aplicadas por el país y se crearon dos variables “dummies” correspondientes a las variantes “delta” y “omicron”. Para llegar a estas variable predictoras, se hicieron diferentes pruebas en el modelo de regresión múltiple y usando las medidas de precisión de predicción r^2 ajustada, criterio de información de Akaike (AIC), criterio de información de Akaike corregido (AICc) y el criterio de información de Schwarz-Bayes (BIC) se seleccionaron las variables que arrojaron las mejores medidas de precisión de predicción.

A continuación se presenta el modelo de regresión múltiple:

Coefficientes generados por el modelo de regresión múltiple:

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.745e+03  5.526e+02  -4.968 1.25e-06 ***
trend()      1.258e+01  2.019e+00   6.232 1.92e-09 ***
new_cases_smoothed -2.155e-03  4.840e-04  -4.453 1.27e-05 ***
icu_patients  1.270e-01  9.382e-03  13.540 < 2e-16 ***
total_vaccinations  1.889e-07  1.001e-06   0.189 0.850443
total_boosters -1.997e-06  4.397e-06  -0.454 0.650101
new_tests_smoothed  1.091e-03  1.338e-04   8.153 1.66e-14 ***
positive_rate -4.180e+03  1.145e+03  -3.652 0.000316 ***
stringency_index  6.084e+00  1.945e+00   3.128 0.001963 **
delta         -3.388e+02  1.210e+02  -2.799 0.005515 **
omicron       -3.560e+02  1.142e+02  -3.117 0.002039 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 279.3 on 253 degrees of freedom
Multiple R-squared:  0.7799,    Adjusted R-squared:  0.7712
F-statistic: 89.67 on 10 and 253 DF, p-value: < 2.22e-16

```

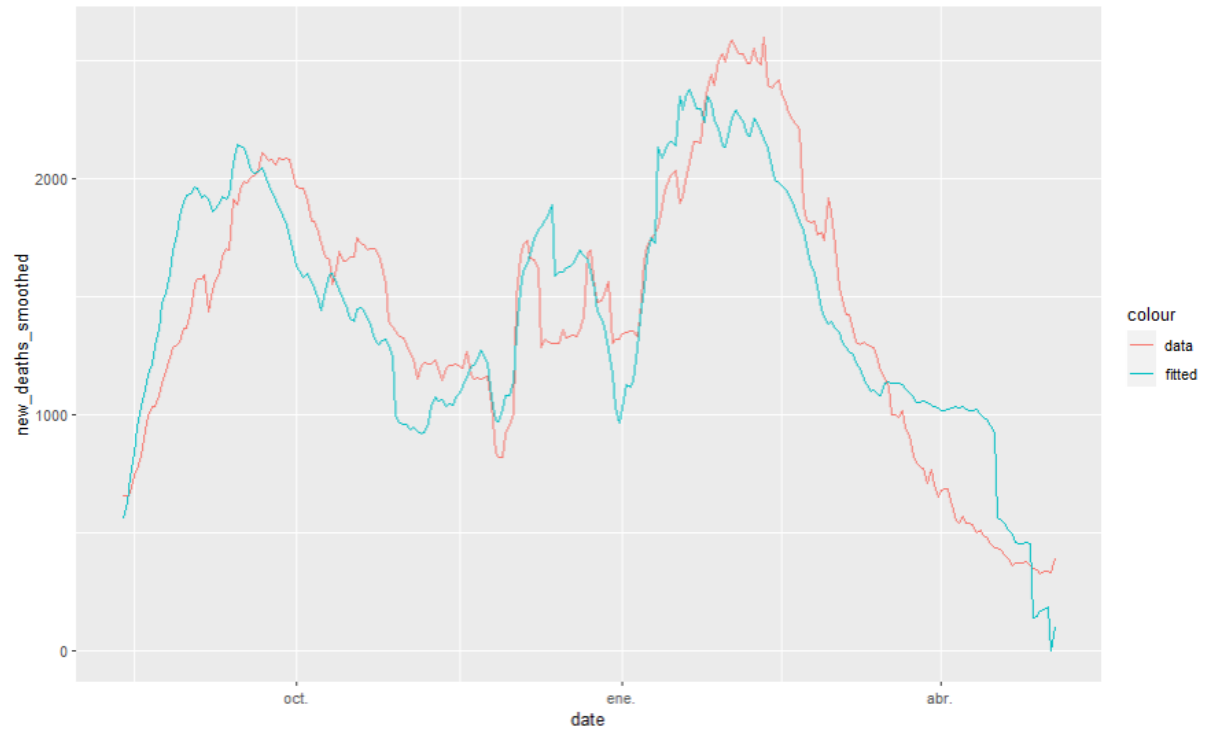
La ecuación generada por el modelo de regresión múltiple es la siguiente:

$$y = -2745 + 12.58x_1 - 0.002155x_2 + 0.1270x_3 + 0.0000001889x_4 - 0.000001997x_5 + 0.001091x_6 - 4180x_7 + 6.084x_8 - 338.8x_9 - 356.0x_{10}$$

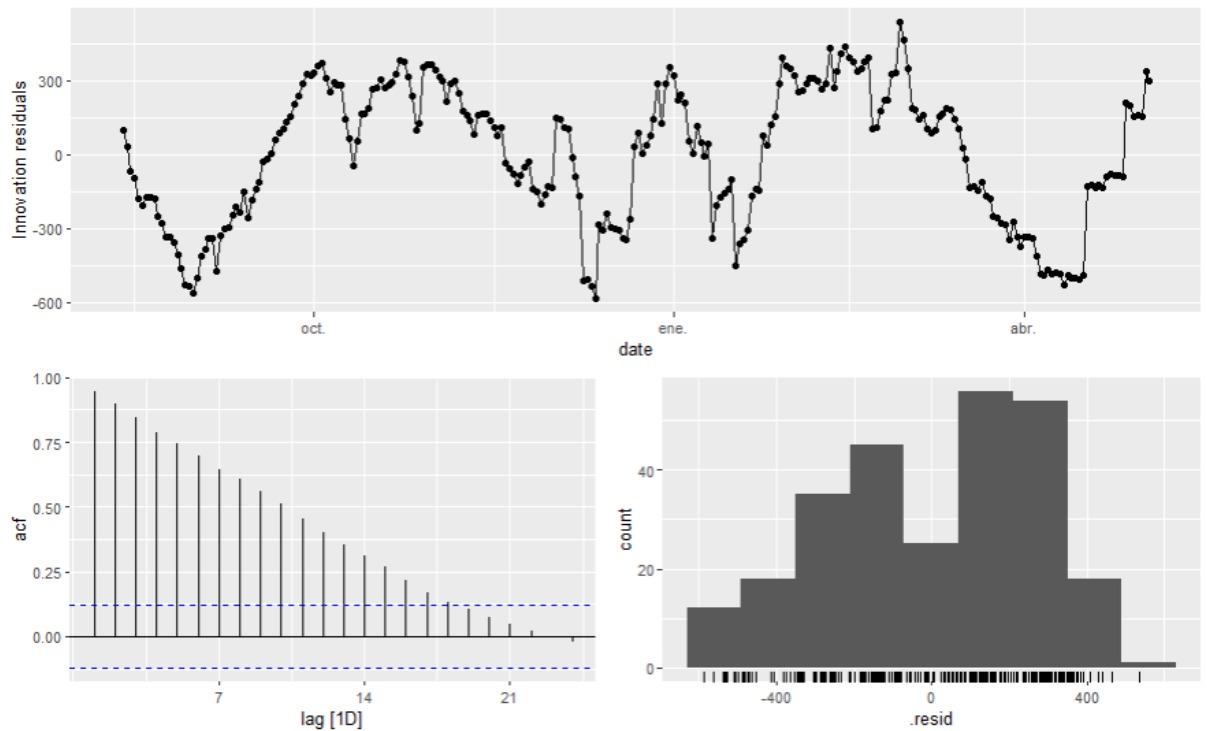
Medidas de precisión de predicción del modelo múltiple:

```
> glance(model_lin)
# A tibble: 1 x 15
  .model r_squared adj_r_squared sigma2 statistic p_value df log_lik AIC AICc BIC CV
  <chr>   <dbl>      <dbl> <dbl> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <dbl>
1 reg_lin 0.907        0.903 30769. 239. 1.58e-120 11 -1687. 2669. 2670. 2711. 34277.
```

Gráfica del ajuste del modelo múltiple a los datos reales:



Gráficas de diagnóstico de residuos del modelo múltiple:



Se puede apreciar que los valores generados por el modelo siguen en general la tendencia de los datos reales, pero no del todo. Esto se puede confirmar en las gráficas de diagnóstico de residuos. Los residuos no parecen tener una media igual a 0, además se presenta correlación y el histograma muestra que los residuos no presentan una distribución normal. Considerando todo esto, se optó por mejorar el modelo de regresión múltiple.

El siguiente paso fue hacer una transformación logarítmica a los datos de la variable predictora (nuevas muertes causadas por el COVID-19) con el fin de mejorar el modelo.

A continuación se presenta el modelo de regresión logarítmica:

Coefficientes generados por el modelo logarítmico:

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.710e+00  4.426e-01  10.642 < 2e-16 ***
trend()      4.956e-03  1.617e-03   3.065 0.002415 **
new_cases_smoothed -1.478e-06  3.877e-07  -3.813 0.000172 ***
icu_patients  7.824e-05  7.515e-06  10.411 < 2e-16 ***
total_vaccinations -1.598e-09  8.018e-10  -1.992 0.047392 *
total_boosters  7.547e-09  3.522e-09   2.143 0.033084 *
new_tests_smoothed 1.082e-06  1.072e-07  10.096 < 2e-16 ***
positive_rate -4.327e+00  9.169e-01  -4.720 3.91e-06 ***
stringency_index  7.263e-03  1.558e-03   4.663 5.06e-06 ***
delta        -1.971e-01  9.693e-02  -2.034 0.043012 *
omicron      -1.995e-01  9.150e-02  -2.181 0.030136 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2237 on 253 degrees of freedom
Multiple R-squared:  0.8152,    Adjusted R-squared:  0.8078
F-statistic: 111.6 on 10 and 253 DF, p-value: < 2.22e-16

```

La ecuación generada por el modelo de regresión logarítmica es la siguiente:

$$y = -4.710 + 0.004956x_1 - 0.000001478x_2 + 0.00007824x_3 - 0.000000001598x_4 + 0.000000007547x_5 + 0.000001082x_6 - 4.327x_7 + 0.007263x_8 - 0.1971x_9 - 0.1995x_{10}$$

Medidas de precisión de predicción del modelo logarítmico:

```

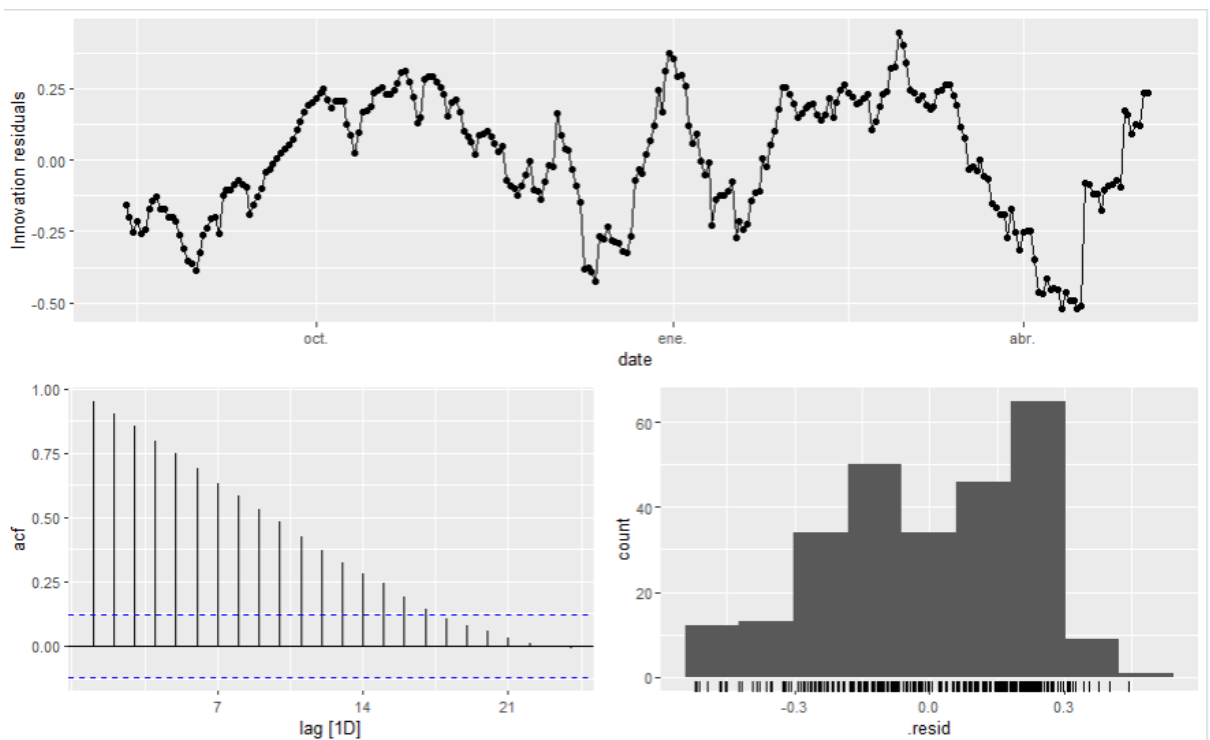
> glance(model_log)
# A tibble: 1 x 15
  .model r_squared adj_r_squared sigma2 statistic p_value df log_lik AIC AICc BIC CV
  <chr>   <dbl>      <dbl>   <dbl>   <dbl>   <dbl> <int> <dbl> <dbl> <dbl> <dbl>
1 reg_log 0.910        0.907 0.0203 250. 9.01e-123 11 142. -989. -987. -946. 0.0224

```

Gráfica del ajuste del modelo logarítmico a los datos reales:



Gráficas de diagnóstico de residuos del modelo logarítmico.



Se puede apreciar como los valores generados por el modelo siguen un poco mejor a los datos reales y como mejoran un poco las gráficas de diagnóstico de

residuos. La media de los residuales parece estar un poco más cercana a cero y el histograma demuestra un comportamiento más parecido a una distribución normal (cargada hacia la izquierda), pero se sigue presentando autocorrelación.

Tomando lo anterior en cuenta, se creó un nuevo modelo usando regresión dinámica con el fin de mejorar el ajuste de los valores generados por el modelo con los datos reales.

A continuación se presenta el modelo de regresión dinámica:

Coefficientes generados por el modelo de regresión dinámica:

```

Coefficients:
      ma1 trend() new_cases_smoothed icu_patients total_vaccinations total_boosters new_tests_smoothed positive_rate
s.e. 0.8308 10.6198      -1e-03      0.1235      0      0e+00      1e-03      -6777.675
      0.0268  3.4778      4e-04      0.0140      0      1e-04      2e-04      1066.310
stringency_index      delta      omicron      intercept
s.e.      5.1708      -130.3798      -220.7545      -2353.5801
      1.7399      109.0855      103.6308      917.2485

sigma^2 estimated as 28542: log likelihood=-1723.25
AIC=3472.5 AICc=3473.95 BIC=3518.99

```

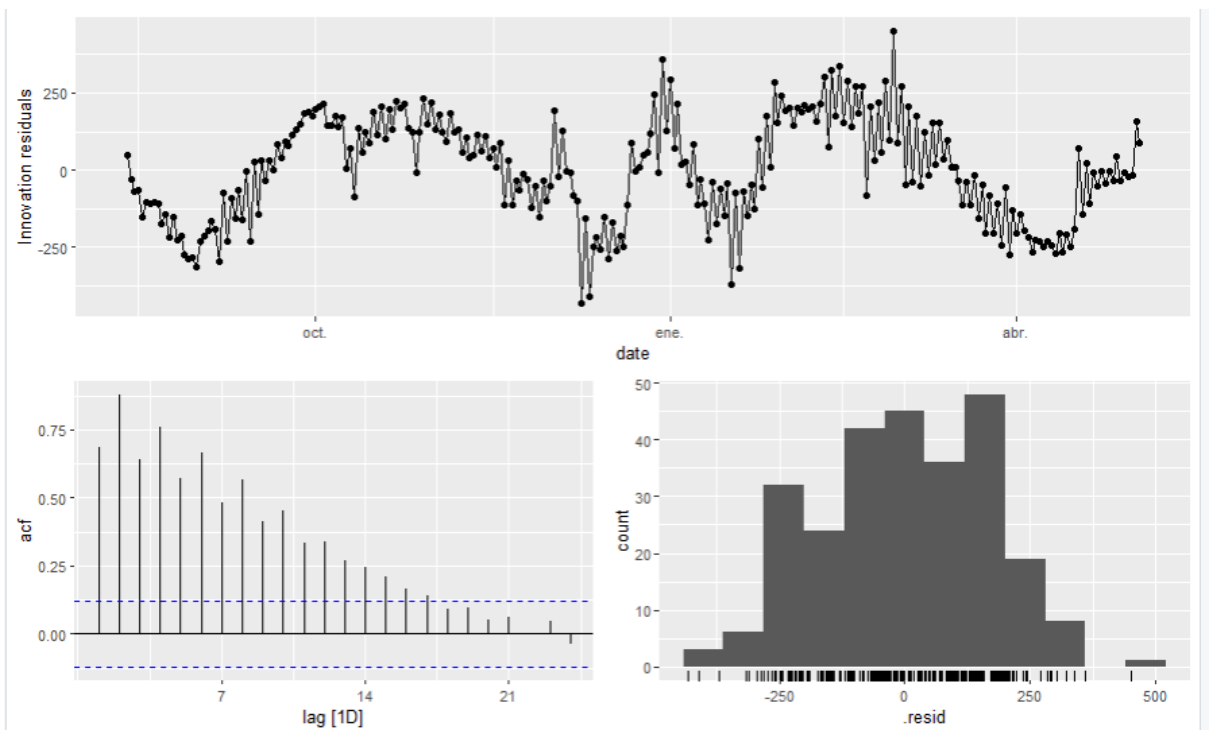
La ecuación generada por el modelo de regresión dinámica es la siguiente:

$$y = -2353.58 + 10.6198x_1 - 0.001x_2 + 0.1235x_3 + 0x_4 + 0x_5 + 0.001x_6 - 6777.68x_7 + 5.1708x_8 - 130.3798x_9 - 220.7545x_{10}$$

Gráfica del ajuste del modelo dinámico a los datos reales:



Gráficas de diagnóstico de residuales del modelo dinámico:

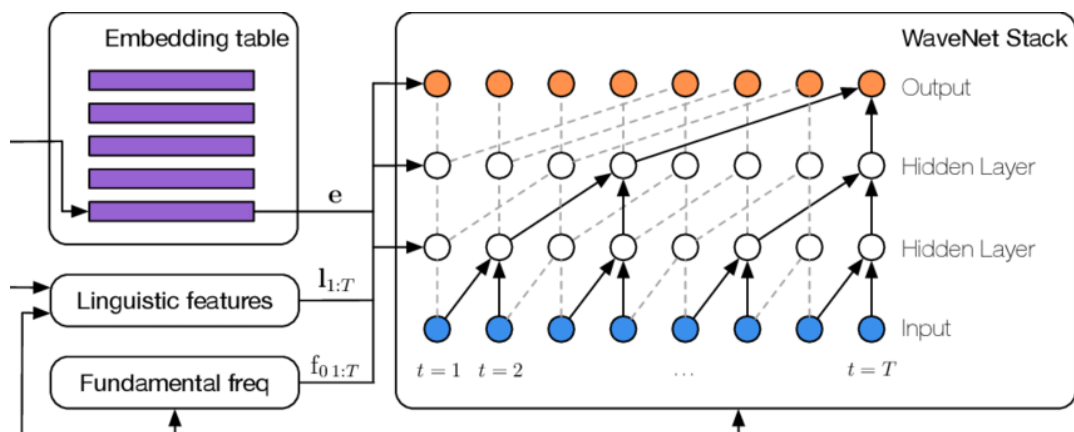


Se puede apreciar como este modelo es el que genera los valores más cercanos a los datos reales y presenta un mejor diagnóstico de residuos: la media

de los residuos es muy cercana a 0, el comportamiento de los residuos es muy parecido a una distribución normal y es el modelo en el que se presenta menos autocorrelación.

Modelo para encontrar los mejores predictores de los casos de COVID

Para el tercer problema que consistía en identificar las variables con mas poder predictivo se realizaron distintos modelos con arquitecturas similares al modelo propuesto el periodo de Otoño 2021, la arquitectura general de los modelos consiste en la aruitectura conocida como “dilated convolutional neural network” llamada tambien “Modelo WaveNet” la cual se define como un técnica que amplía el Kernel (entrada) insertando huecos entre sus elementos consecutivos. En términos más sencillos, es lo mismo que la convolución, pero implica la omisión de píxeles, para cubrir un área mayor de la entrada.



La “dilated convolution” ayuda a ampliar el área cubierta de la imagen de entrada sin necesidad de hacer un pool. El objetivo es cubrir más información de la salida obtenida con cada operación de convolución. Este método ofrece un campo de visión más amplio con el mismo coste computacional. Determinamos el valor del factor de dilatación (l) como viendo cuánta información se obtiene con cada convolución en valores variables de l .

Utilizando este método, podemos obtener más información sin aumentar el número de parámetros del núcleo.

Para probar cuales variables sirven como predictoras de los contagios se utilizaron datasets distintos y que contenian distintas variables y se dieron como entrada al modelo con la arquitectura WaveNet.

Layer (type)	Output Shape	Param #
dense (Dense)	(None, None, 4)	20
dense_1 (Dense)	(None, None, 8)	40
dropout (Dropout)	(None, None, 8)	0
conv1d (Conv1D)	(None, None, 128)	2176
conv1d_1 (Conv1D)	(None, None, 128)	32896
conv1d_2 (Conv1D)	(None, None, 128)	32896
conv1d_3 (Conv1D)	(None, None, 128)	32896
conv1d_4 (Conv1D)	(None, None, 128)	32896
conv1d_5 (Conv1D)	(None, None, 128)	32896
conv1d_6 (Conv1D)	(None, None, 128)	32896
dropout_1 (Dropout)	(None, None, 128)	0
dense_2 (Dense)	(None, None, 8)	1032
dense_3 (Dense)	(None, None, 4)	36
conv1d_7 (Conv1D)	(None, None, 1)	5

=====
Total params: 200,685
Trainable params: 200,685
Non-trainable params: 0
=====

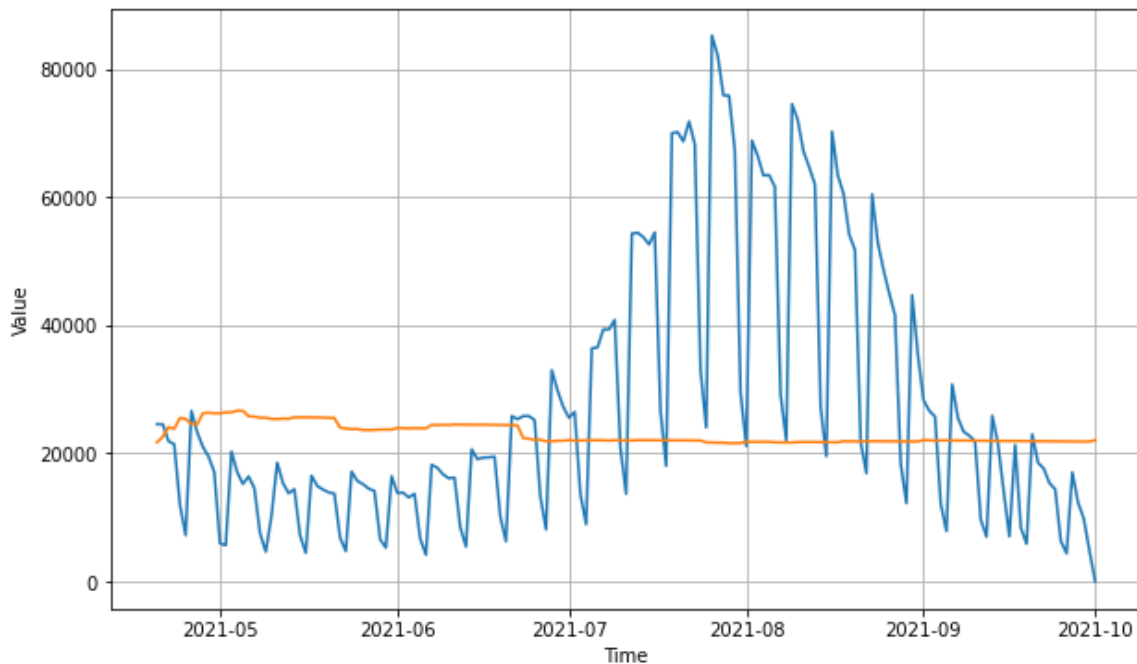
Arquitectura del modelo

En el primer modelo(estacional no regresivo) se decidió usar variables estacionales para intentar corregir el problema de la subestimación en la predicción de casos de COVID , para eso se dividió el dataset separando fecha en Dia, Mes y Año , después esto se le dio como entrada al modelo de red neuronal junto con los casos por dia de forma que cada parte de la fecha predijera un valor numerico que representa el numero de casos.

```
[[[2021 2 5]]] => [[46859]]
[[[2021 2 6]]] => [[24291]]
[[[2021 2 7]]] => [[103714]]
[[[2021 2 8]]] => [[83078]]
[[[2021 2 9]]] => [[77014]]
[[[2021 2 10]]] => [[70365]]
[[[2021 2 11]]] => [[68596]]
[[[2021 2 12]]] => [[35559]]
[[[2021 2 13]]] => [[19086]]
[[[2021 2 14]]] => [[73588]]
[[[2021 2 15]]] => [[69502]]
[[[2021 2 16]]] => [[62842]]
[[[2021 2 17]]] => [[62351]]
```

Ejemplo de la forma de los datos de entrada

Después de entrenar ese modelo nos dimos cuenta que el modelo no alcanzaba a aprender el patrón de predicción por lo cual se modificó la forma de la entrada.



Grafica modelo estacional no regresivo

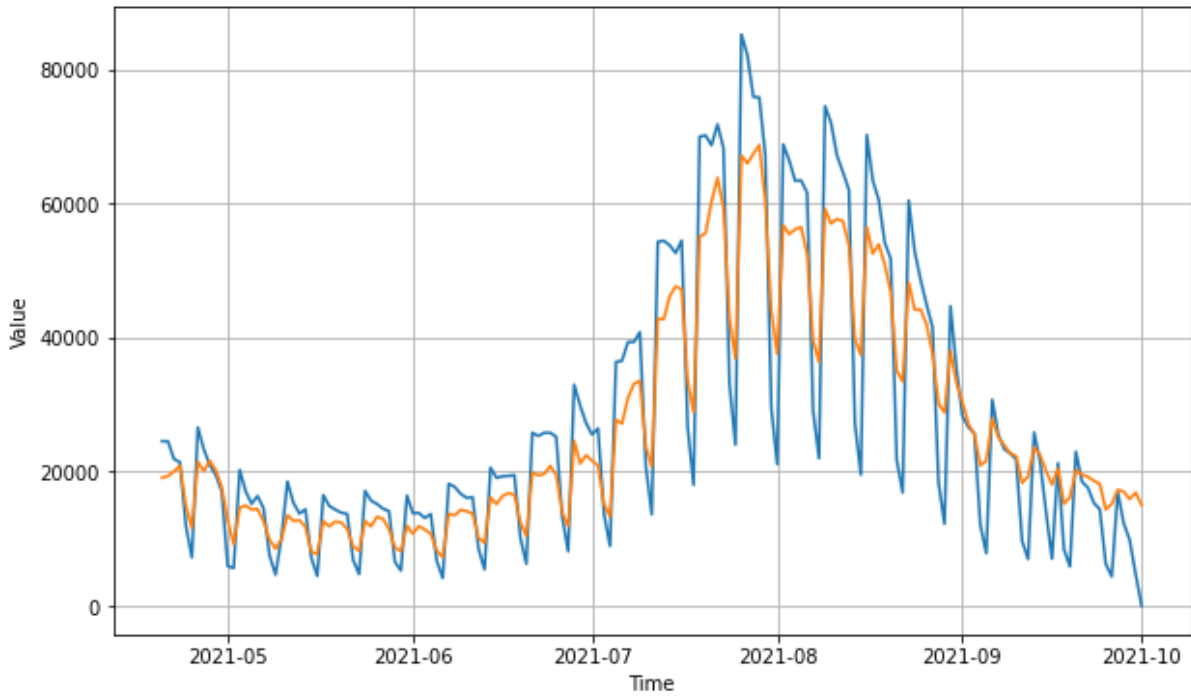
Al ver el resultado tan desfavorable que arrojaba ese modelo se decidió pensar en otra forma de organizar los datos con el objetivo de generar una mejor predicción y para esto se usó el concepto de estacionalidad en conjunto con una autorregresión.

Para esto se le anexó a cada fecha los casos de COVID del día anterior con el objeto de crear un modelo con más capacidad de predicción.

```
[ 2021, 4, 20, 24573],  
[ 2021, 4, 21, 24554],  
[ 2021, 4, 22, 21916],  
[ 2021, 4, 23, 21445],  
[ 2021, 4, 24, 11940],  
[ 2021, 4, 25, 7245],  
[ 2021, 4, 26, 26634],  
[ 2021, 4, 27, 23353],  
[ 2021, 4, 28, 21034],
```

Formato de los datos de entrada al modelo estacional regresivo

Al cambiar la forma de los datos de entrada se tuvo que modificar un poco la arquitectura del modelo ya que la forma de los datos era distinta, después de hacer eso se hizo una predicción para los datos de validación en la cual fue muy exitoso este modelo.



Modelo estacional regresivo

Este modelo fue muy exitoso en la predicción de datos de validación pero a largo plazo perdió precisión ya que solo fue capaz de predecir unos cuantos días a futuro correctamente.

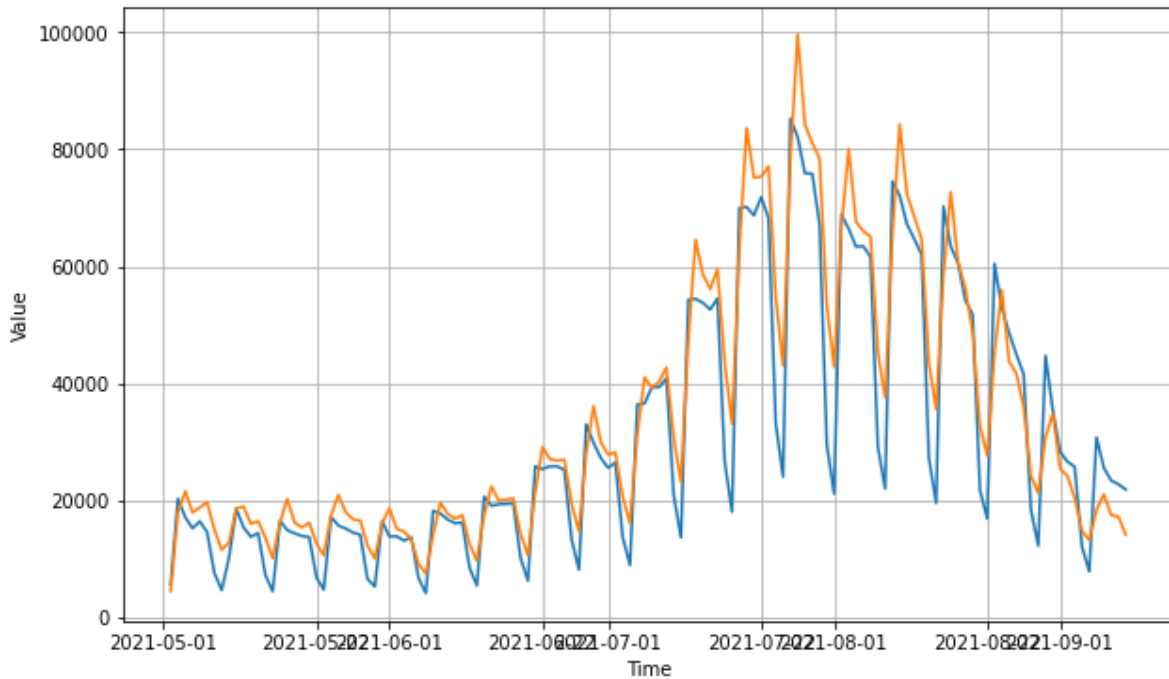
Al descubrir la poca capacidad de predicción de la estacionalidad se optó por utilizar variables relacionadas a la epidemiología como posibles variables de predicción.

Después de realizar una investigación sobre cómo se propagaba la pandemia y de qué dependía desde el punto de vista de la epidemiología, llegamos a lo que se conoce como R o número de reproducción efectiva, el cual se define como el promedio de casos secundarios por cada caso inicial, en otras palabras si una persona se infecta a cuántas personas contagiara en promedio.

Después de realizar esa investigación se empezó con la preparación y limpieza de los datos relacionados al número de reproducción efectiva con respecto al tiempo (fecha) y relacionándolos con los contagios de COVID por día.

	R	ci_95_u	ci_95_l	ci_65_u	ci_65_l	CountFI
FECHA						
2020-03-10	1.5452	1.8795	1.2342	1.6900	1.3883	152
2020-03-11	1.5430	1.8693	1.2394	1.6851	1.3901	376
2020-03-12	1.5408	1.8591	1.2446	1.6802	1.3919	392
2020-03-13	1.5386	1.8489	1.2498	1.6753	1.3937	648
2020-03-14	1.5364	1.8387	1.2550	1.6704	1.3955	514

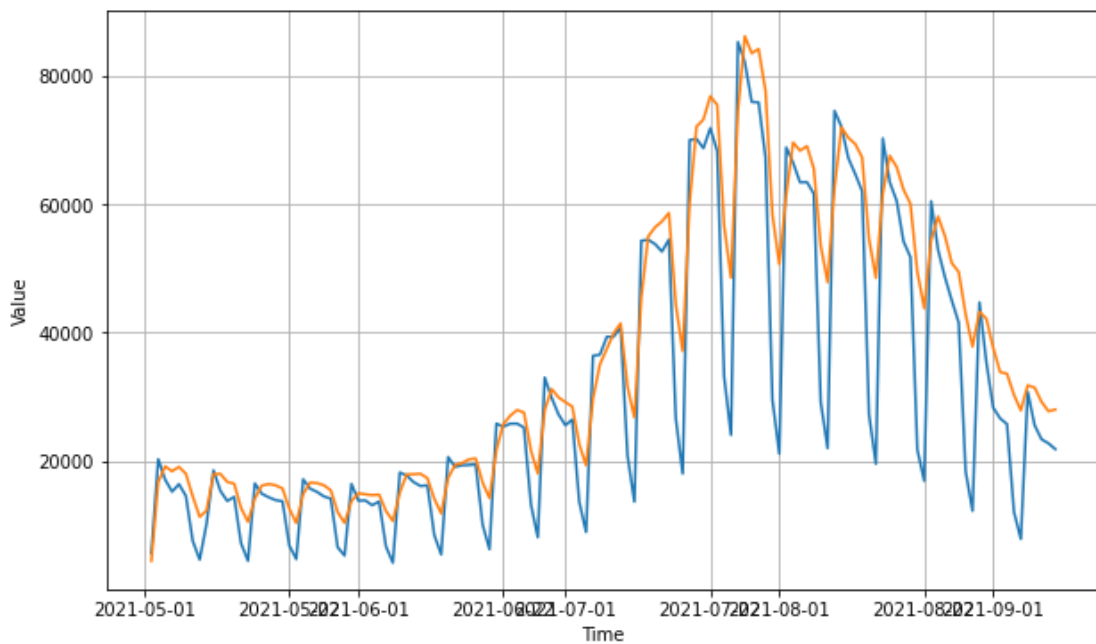
Estructura los datos de reproducción efectiva



Modelo Regresivo con variables de reproducción efectiva

Este modelo fue de los mejores ya que logro predecir correctamente la concavidad y la fecha de los picos de la grafica aunque a largo plazo seguía dando una subestimacion.

Despues se decio solo usar la reproduccion activa(R) por si sola para descartar que el modelo estaba haciendo *overfitting*.



Modelo de regresión usando solo R como variable de apoyo

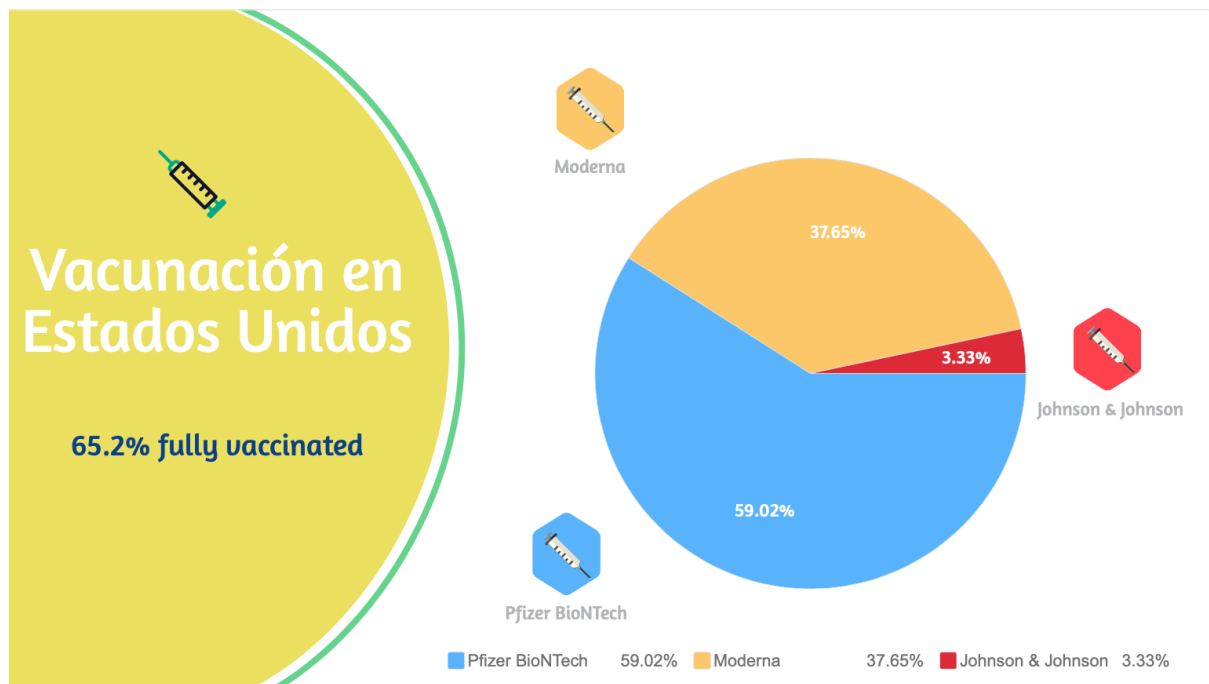
Este modelo fue muy acertado a corto plazo pero no logro predecir adecuadamente la subida drástica de casos debido a omicron.

En general los modelos no tuvieron problema para predecir a corto pero a largo plazo no fueron tan precisos.

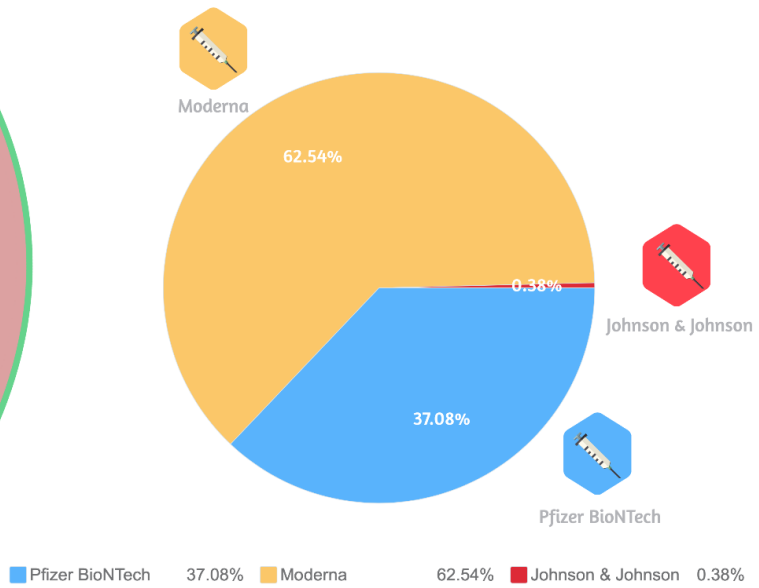
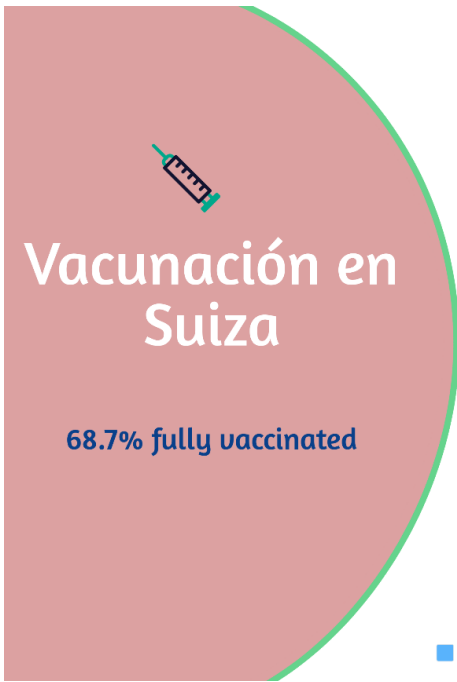
1.6. Valoración de productos, resultados e impactos

Del análisis realizado a partir de la información extraída de las autoridades sanitarias de los países elegidos para el primer problema, fue posible encontrar lo siguiente:

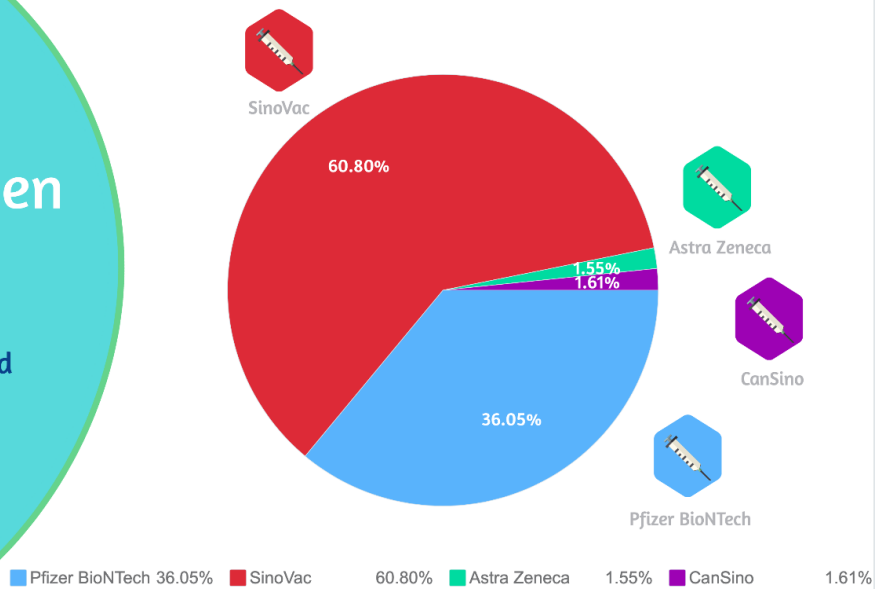
En Estados Unidos, el 65.2% de la población total se encuentra completamente vacunada. Se utilizaron primordialmente vacunas con tecnología de mRNA. El 59.02% lo hizo con el fármaco de Pfizer (fármaco estadounidense-alemán), el 37.65% con Moderna (fármaco estadounidense) y el 3.33% restante con el de Johnson & Johnson (fármaco estadounidense).



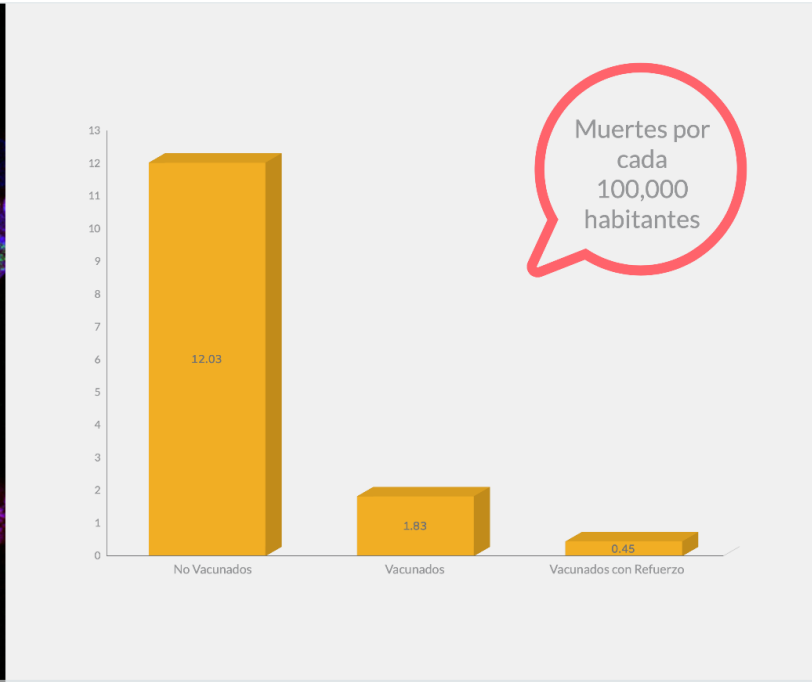
En el caso de Suiza, del 68.7% de la población se encuentra completamente vacunada, el 62.54% lo hizo con el fármaco de Moderna (fármaco estadounidense), el 37.08% lo hizo con el de Pfizer (fármaco estadounidense-alemán) y sólo 0.38% lo hizo con el de Johnson & Johnson (fármaco estadounidense).



En Chile el 90% de la población se encuentra vacunada de las cuáles el 73.92% lo hizo con la vacuna china SinoVac (fármaco chino), el 22.83% con Pfizer (fármaco estadounidense-alemán), el 1.59% con AstraZeneca (fármaco inglés) y el 1.66% con CanSino (fármaco chino).

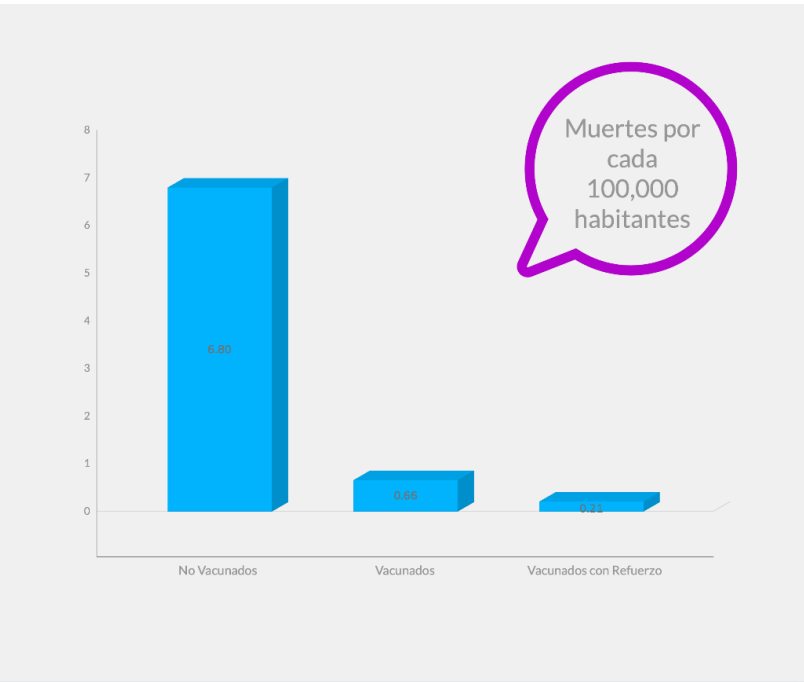


En Estados Unidos, el ratio de muertes semanales en personas no vacunadas es de 12.06 personas por cada 100,000 habitantes. En el caso de personas completamente vacunadas sin refuerzo es de 1.83 personas por cada 100,000 habitantes, mientras que en el caso de personas con refuerzo es de 0.45 personas por cada 100,000 habitantes.



Con estos datos, es posible notar que en Estados Unidos es 6.59 veces más probable fallecer si no te encuentras vacunado en proporción a las personas que se encuentran vacunadas completamente. Pero también es 4.06 veces más probable fallecer en caso de no contar con el refuerzo vs las personas que sí cuentan con este. Respecto de las personas no vacunadas vs las personas con refuerzo la cifra es todavía mayor: las personas no vacunadas fallecen 26.8 veces más en proporción comparado con las personas que cuentan con el refuerzo.

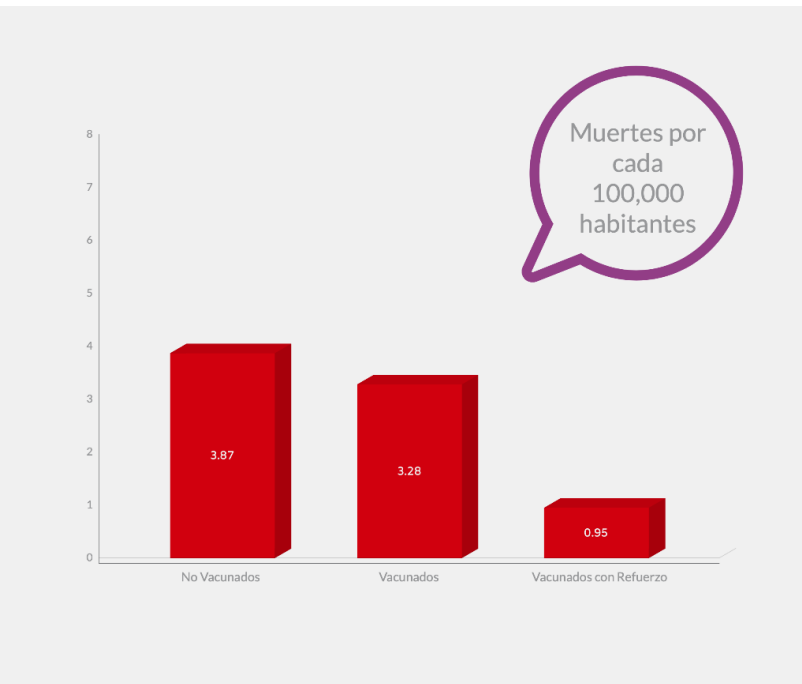
En Suiza, el ratio semanal de muertes de personas no vacunadas es de 6.8 personas por cada 100,000 habitantes. El de personas completamente vacunadas pero sin refuerzo es de 0.66 personas por cada 100,000 habitantes y el de personas completamente vacunadas con refuerzo es de 0.21 por cada 100,000 personas.



En este país, las personas no vacunadas mueren 32.38 veces más en proporción que las personas que cuentan con refuerzo y 10.3 veces más que las personas completamente vacunadas. Adicionalmente, las personas completamente vacunadas fallecen 3.14 veces más que las personas que cuentan con el refuerzo.

Las cifras de Chile son un poco distintas a las de Suiza y Estados Unidos, probablemente por la diferencia de las vacunas aplicadas y el porcentaje de la población completamente vacunada. En el país sudamericano, el ratio semanal de muertes en personas no vacunadas es de 3.87 por cada 100,000 habitantes. De personas completamente vacunadas, son 3.28 personas por cada 100,000 habitantes y de las personas con refuerzo el ratio es de 0.95 por cada 100,000 personas.

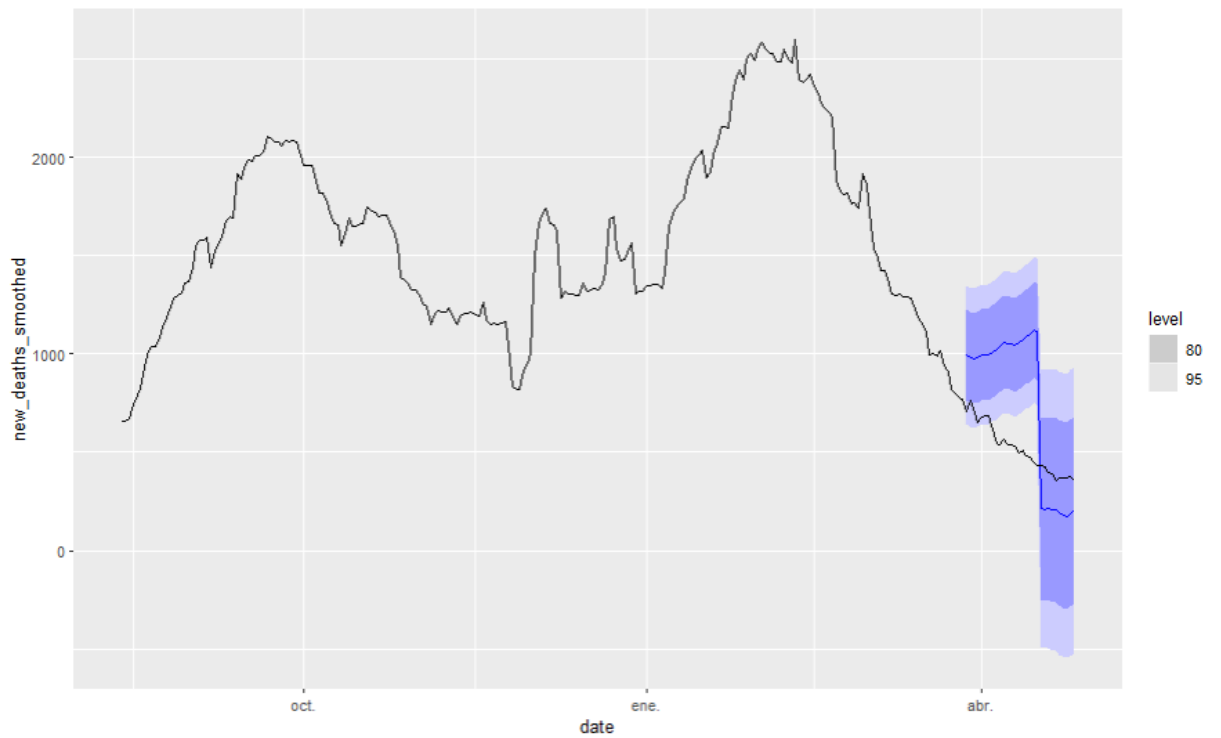
Es importante notar que en el caso de Chile, existió una tendencia a la baja en la efectividad de las vacunas, ya que las cifras del 8 de agosto del año pasado eran muy distintas. En ese momento, sólo 1.2 personas por cada 100,000 personas completamente vacunadas fallecían en comparación con las 8.29 no vacunadas.



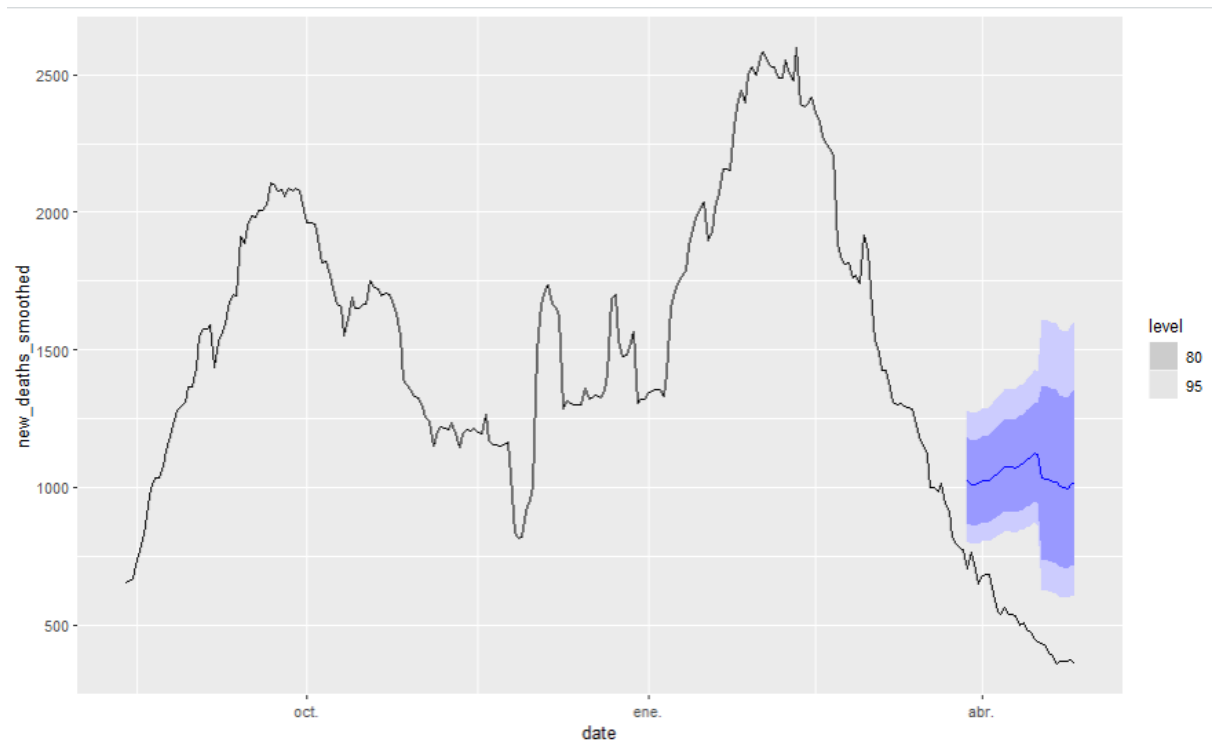
Actualmente, en Chile las personas no vacunadas fallecen 1.18 veces más que las personas completamente vacunadas y 4.07 veces más que las personas que cuentan con el refuerzo. Las personas completamente vacunadas fallecen 3.45 veces más que las personas con refuerzo en proporción. Si bien no es posible concluir con exactitud esto, es posible que la degradación en la efectividad del fármaco de Sinovac que fue el que se usó preponderantemente en este país es mucho mayor a la de las vacunas de mRNA usadas en Estados Unidos y Suiza.

Además del análisis realizado a la vacunación activa en Estados Unidos, Suiza y Chile, se hicieron pronósticos de posibles escenarios de muertes debidas COVID-19 usando los modelos de regresión múltiple, logarítmico y dinámico. Para esto, primero se separaron los datos en datos de entrenamiento y datos de prueba. Los datos de entrenamiento se tomaron a partir de la fecha 13 de agosto del 2021 a la fecha del 26 de marzo del 2022 y los datos de prueba se tomaron de la fecha del 27 de marzo del 2022 en adelante. Esto para visualizar qué tan buenos pronósticos se generaban con los datos de entrenamiento, comparándolos con los datos de prueba. A continuación se muestran las gráficas obtenidas de esta primer prueba de predicción:

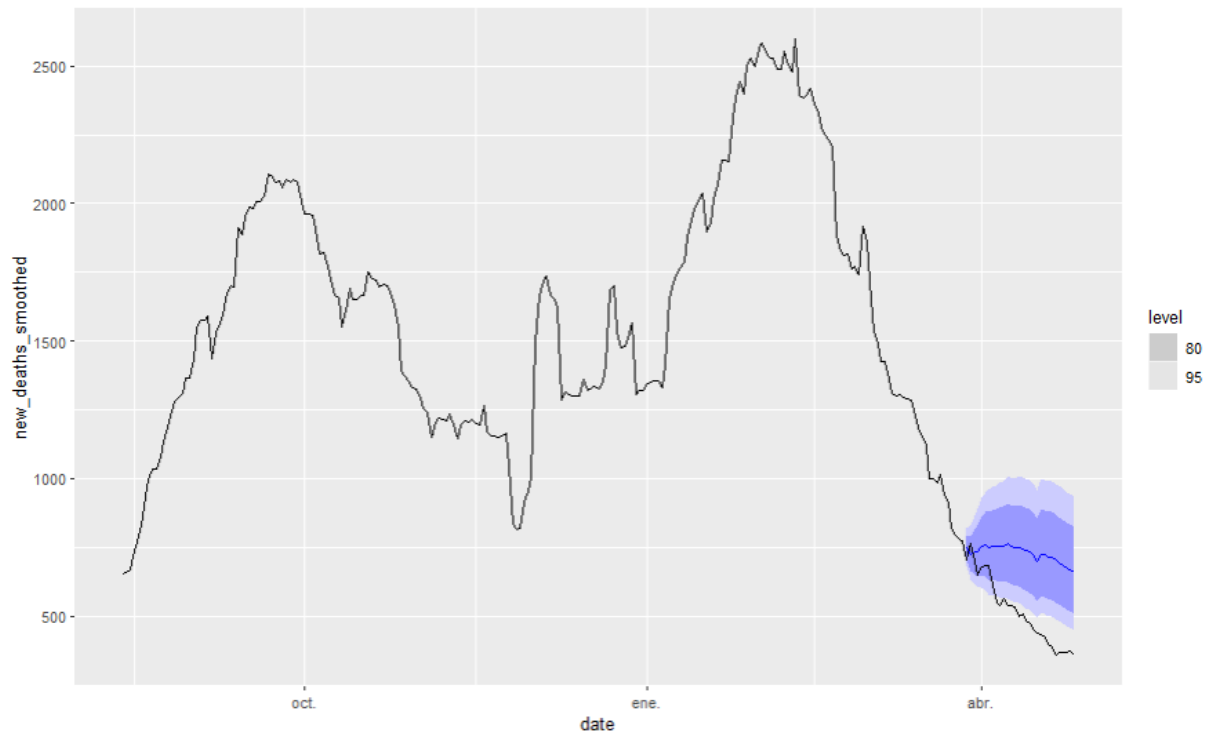
Predicción a partir de los datos de entrenamiento del modelo de regresión múltiple:



Predicción a partir de los datos de entrenamiento del modelo de regresión logarítmico:



Predicción a partir de los datos de entrenamiento del de modelo de regresión dinámica:

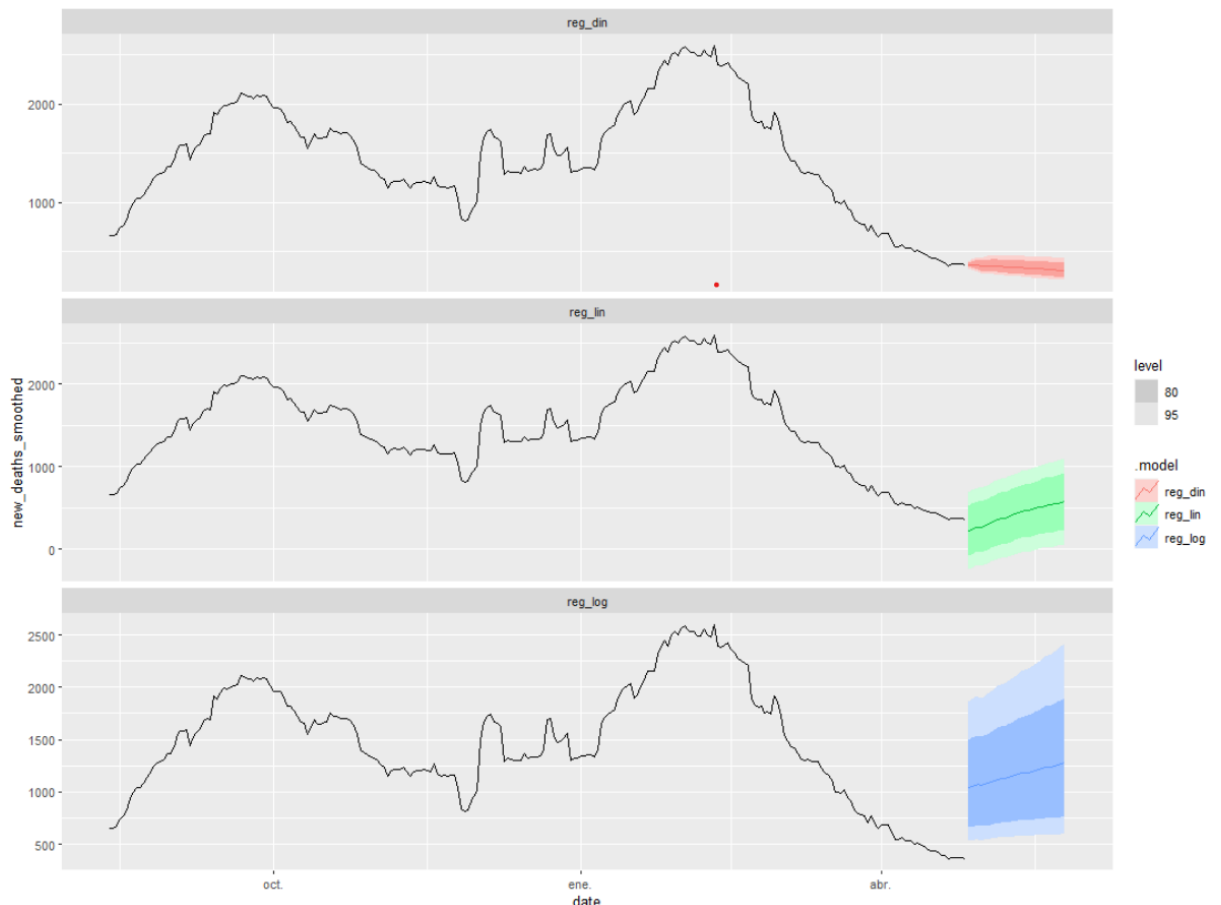


Se puede observar en las anteriores gráficas, que el modelo de regresión dinámica es el que genera una mejor predicción de muertes a causa de COVID-19. El modelo de regresión lineal generó un escalón y predicciones no tan fiables, al igual que el modelo de predicción logarítmica. A pesar de ya tener definidas las variables predictoras, se intentó mejorar estas predicciones jugando con las variables y viendo cuales generaban mejores predicciones, pero no se lograron mejorar dichas predicciones, por lo que se concluye que para mejorar el modelo y las predicciones hacen falta una o más variables predictoras desconocidas.

Usando los parámetros de error de predicción de error absoluto medio (MAE), raíz del error absoluto medio (RMSE), error absoluto de porcentaje (MAPE), y el error absoluto escalado medio (MASE), se concluye que el modelo de regresión dinámica generó las mejores predicciones ya que generó un menor error tanto en los datos de entrenamiento con en los datos de prueba para todos los parámetros. A continuación se presenta una tabla con dichos parámetros generada por los modelos:

.model	RMSE_Training	RMSE_Test	MAE_Training	MAE_Test	MAPE_Training	MAPE_Test	MASE_Training	MASE_Test
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 reg_lin	166.	406.	131.	367.	9.26	72.7	0.575	1.61
2 reg_log	170.	546.	124.	531.	8.21	114.	0.545	2.33
3 reg_din	68.7	237.	47.0	216.	2.99	48.5	0.206	0.946

Una vez realizada la prueba de entrenamiento y concluyendo que para mejorar los modelos hacen falta una o más variables predictoras desconocidas, se optó con proseguir con las predicciones a futuro de las muertes ocasionadas por el COVID-19, considerando que los modelos todavía se pueden mejorar, la predicciones se hicieron a solo un mes en el futuro. Para esto se generaron predicciones de cada una de las variables predictoras usando un modelo combinado ARIMA y suavizado exponencial. A continuación, se muestran las predicciones de cada uno de los modelos:

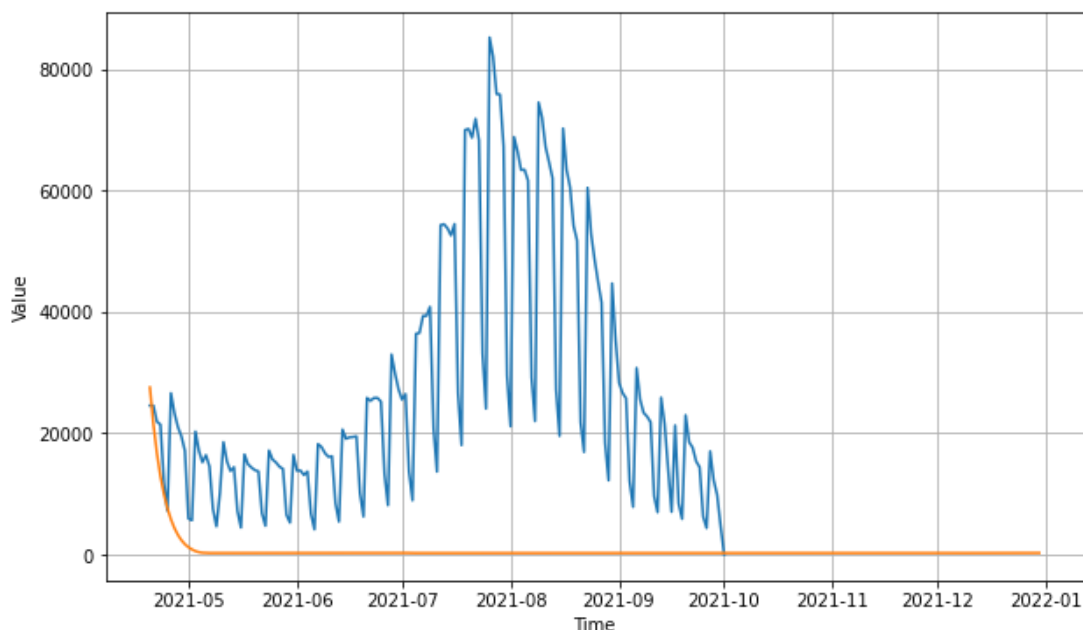


Tomando en cuenta la tendencia actual de la pandemia y las medidas preventivas implementadas por los gobiernos, se podría considerar que el modelo de regresión dinámico es el que genera una mejor predicción, ya que no muestra una alza en las muertes generadas por el COVID-19. Sin embargo hay demasiadas variables en juego y en cualquier momento podría aparecer una nueva variante que genere una alza en las muertes generadas por el COVID-19 y en dicho caso las predicciones generadas por los modelos de regresión múltiple y logarítmico podrían ser más acertadas. De esto se puede concluir que no es momento de bajar la guardia, los gobiernos deberían seguir implementando medidas contra el COVID-19 cómo son las campañas de vacunación de refuerzo y los ciudadanos deberían seguir dichas medidas, esto para evitar un rebrote de contagios y por lo tanto de muertes, como muestran las predicciones de los modelos múltiple y logarítmico.

Identificación de variables con significancia al predecir los contagios.

Como resultado del experimento de realizar distintos modelos de red neuronal para identificar las variables que tienen más significancia a la hora de predecir los aumentos de la curva de contagios.

Como resultados obtuvimos ciertos resultados inesperados como el fracaso de la estacionalidad para predecir la pandemia a futuro, ya que en ambos modelos en los cuales se utilizó tuvo muy poco poder predictivo y solo lograba predecir las primeras semanas con éxito y después predecía un supuesto fin de la pandemia lo cual es erróneo lo cual se muestra en la gráfica posterior.



Predicción a largo plazo usando estacionalidad y autoregresion

Ante este resultado originalmente se tenía planeado hacer lo que se conoce como *model subclassing* que es una técnica de deep learning la cual consiste en juntar 2 o más modelos juntos para hacer un modelo más robusto y con mayor poder de predicción, la propuesta original era crear un *Autoencoder* que hiciera detección de anomalías para detectar nuevas variantes mediante anomalías en los datos y después juntarlo con el modelo *WaveNet* para que trabajaran en conjunto, desafortunadamente nuestros equipos de cómputo no tuvieron los requerimientos de hardware para poder entrenar ese modelo por lo cual se optó por usar variables epidemiológicas que se asemejaran al objetivo anterior.

```

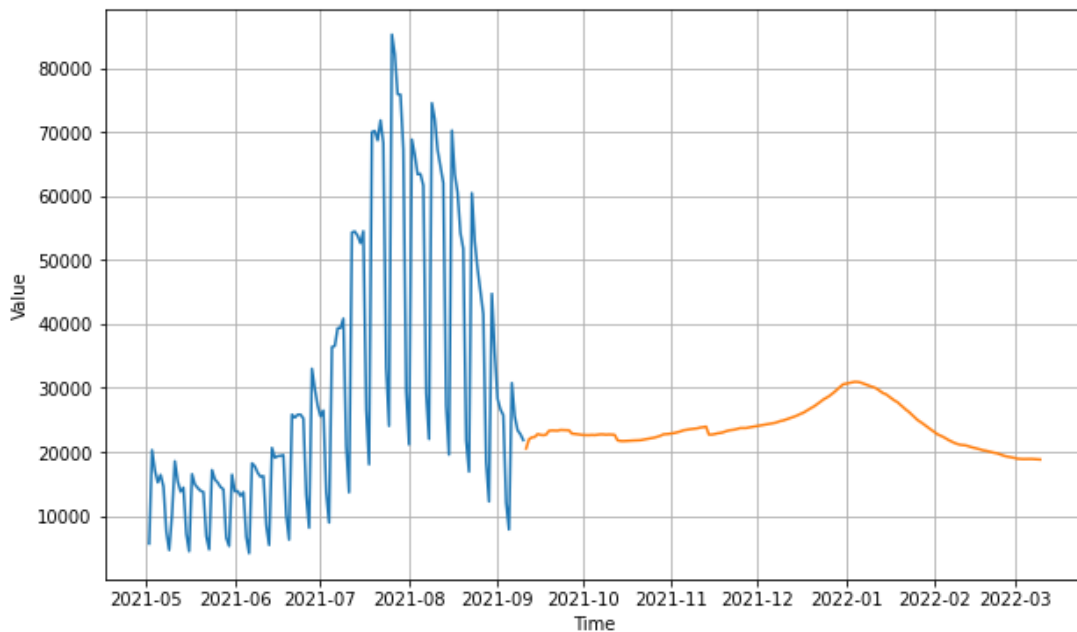
window_size = 1
train_set = TimeseriesGenerator(features_season, x_train, length = window_size, batch_size=1)
valid_set = TimeseriesGenerator(features_season_val, x_valid, length = window_size, batch_size=1)

model = keras.models.Sequential()
model.add(LSTM(128, input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(Dropout(rate=0.2))
model.add(RepeatVector(X_train.shape[1]))
model.add(LSTM(128, return_sequences=True))
model.add(Dropout(rate=0.2))
model.add(TimeDistributed(Dense(X_train.shape[2])))

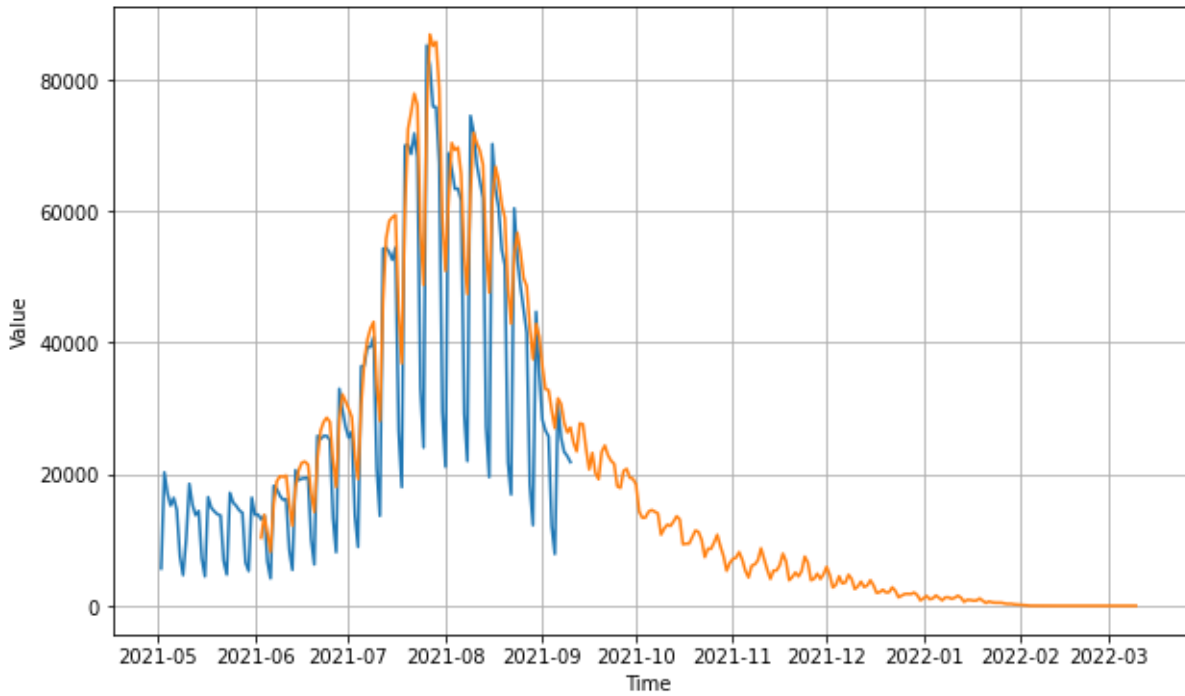
```

Propuesta de arquitectura del modelo de detección de anomalías

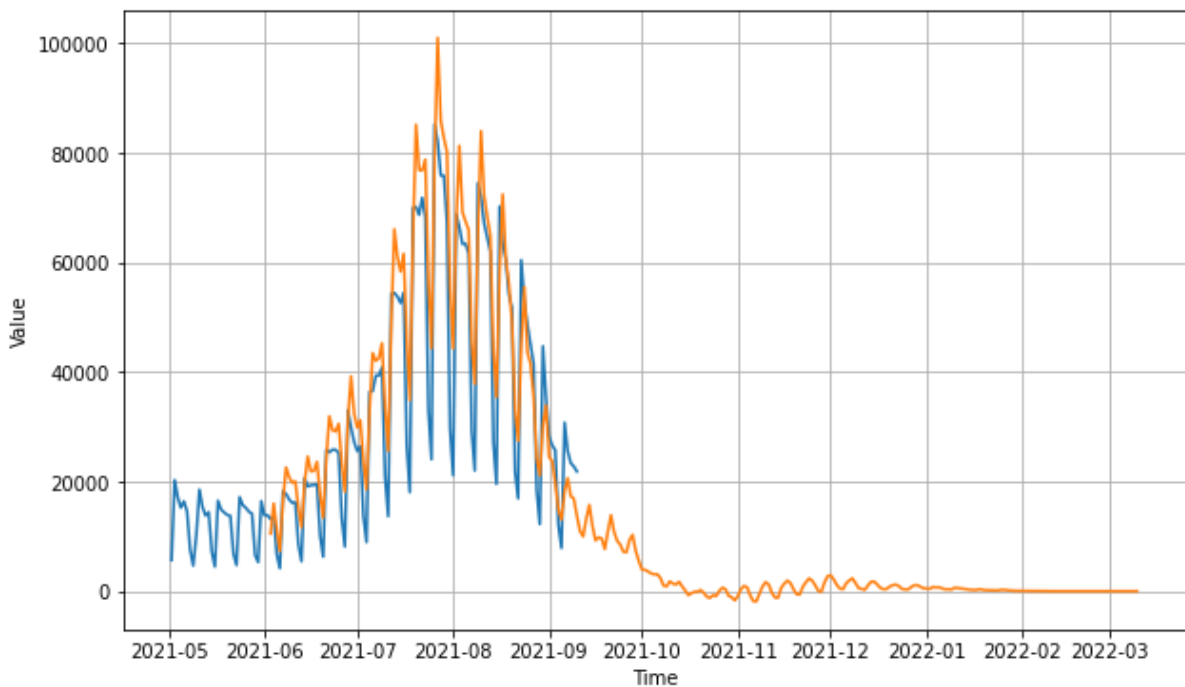
Debido a este problema decidimos redefinir la estrategia y utiliza la variable R o reproducción efectiva que define el promedio de personas a la que un enfermo de COVID puede infectar, esta variable nos buenos resultados tanto a largo plazo como a corto plazo pero no arreglo del todo el problema de la subestimación de contagios ya que solo se acercó un poco más con respecto al modelo de Otoño 2021 y le atino correctamente al periodo en el cual hubo pico en contagios.



Para ver la validez del modelo se hicieron otras predicciones empezando antes del pico que se aprecia en el gráfico anterior con el objetivo de ver que tan fiable era el modelo.



Modelo de regresión usando solo R como variable de apoyo



Modelo Regresivo con variables de reproducción efectiva

Como se puede observar la reproducción efectiva es una buena variable para predecir el comportamiento de la pandemia en el ámbito de los contagios debido a que esta variable cuantifica el aspecto principal que da origen a la transmisión de un virus que es el contacto con otras personas, debido a esto esta variable logra un mejor resultado que la estacionalidad debido que la estacionalidad puede ser generada solo por un retraso en el reporte de los casos y realmente no significar

mucho en cuanto a su relacion con el comportamiento con la pandemia que fue el caso aquí.

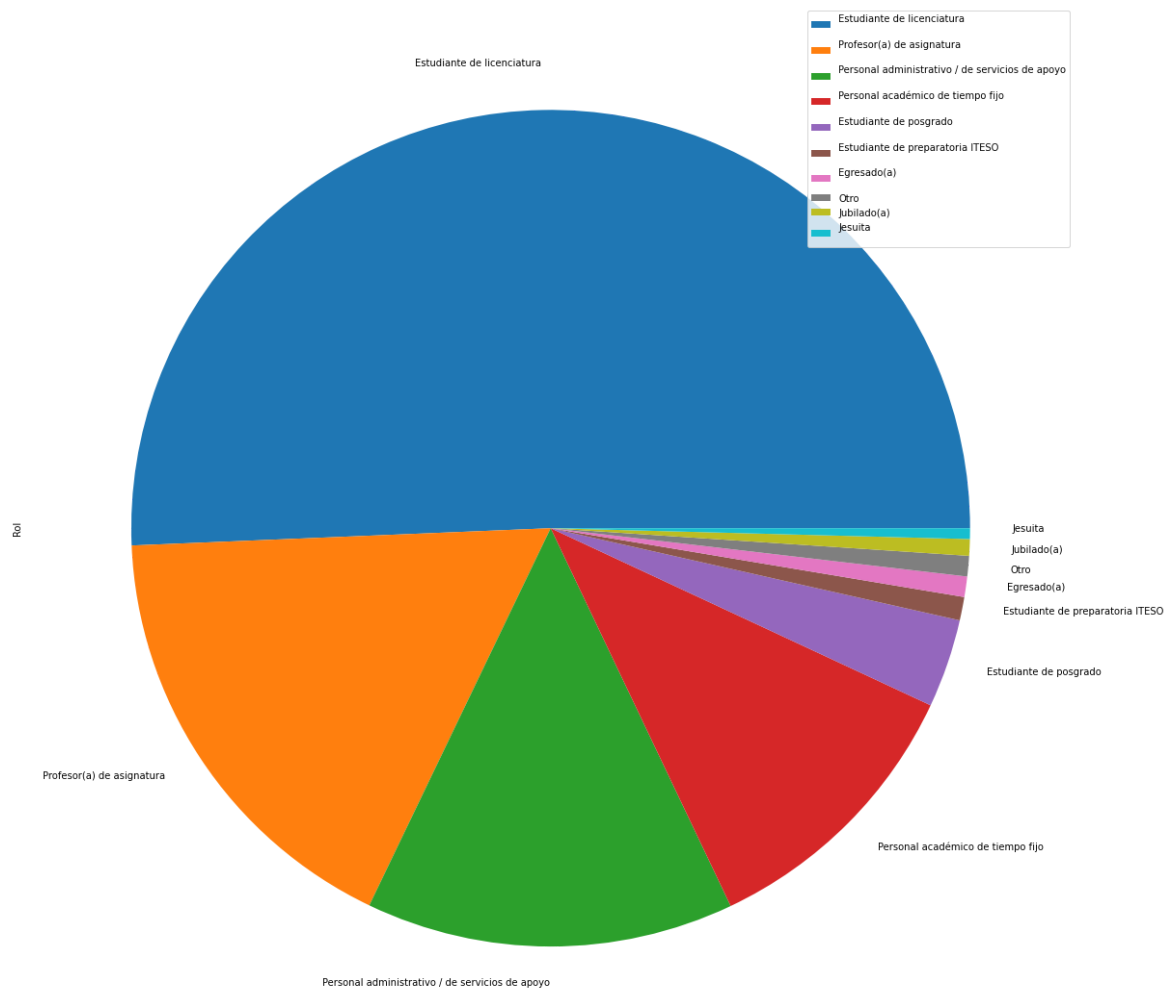
Aunque la reproduccion efectiva probo ser una buena variable predictora, sin embargo no logro predecir adecuadamente la cantidad de casos debido a que pasamos por alto que la pandemia no solo depende del comportamiento y de lo infeccioso que este sea sino que también depende del ser humano y su comportamiento, mas especificamente de cuanto se reúne con otras personas, por lo tanto para la siguiente iteracion de este Proyecto es una propuesta razonable analizar la reproducción efectiva con alguna variable social que defina la cantidad de personas en posibilidad de ser infectadas, un ejemplo podría ser cantidad de pasajeros en el transporte público por unidad de tiempo o cantidad de trafico u ocupación de espacios de ocio como antros, salones de fiesta, centros comerciales, ect. Ya que conforme aumenta la cantidad de personas aglomeradas en un lugar es más probable que alguien se infecte de COVID.

En conclusión predecir los contagios por COVID no es una tarea fácil y que hay demasiadas variables que afectan al comportamiento de esta , esto hace mas viable predecir las muertes ya que como afirma Beth Blauer investigadora de John Hopkings la variante omicron no supuso un aumento en las muertes ya que esta variante es menos letal pero mas infecciosa, lo cual hace mas viable la tarea de predecir muertes debido a que no hay cambios tan drasticos en los datos.

Encuesta de vacunacion en Iteso

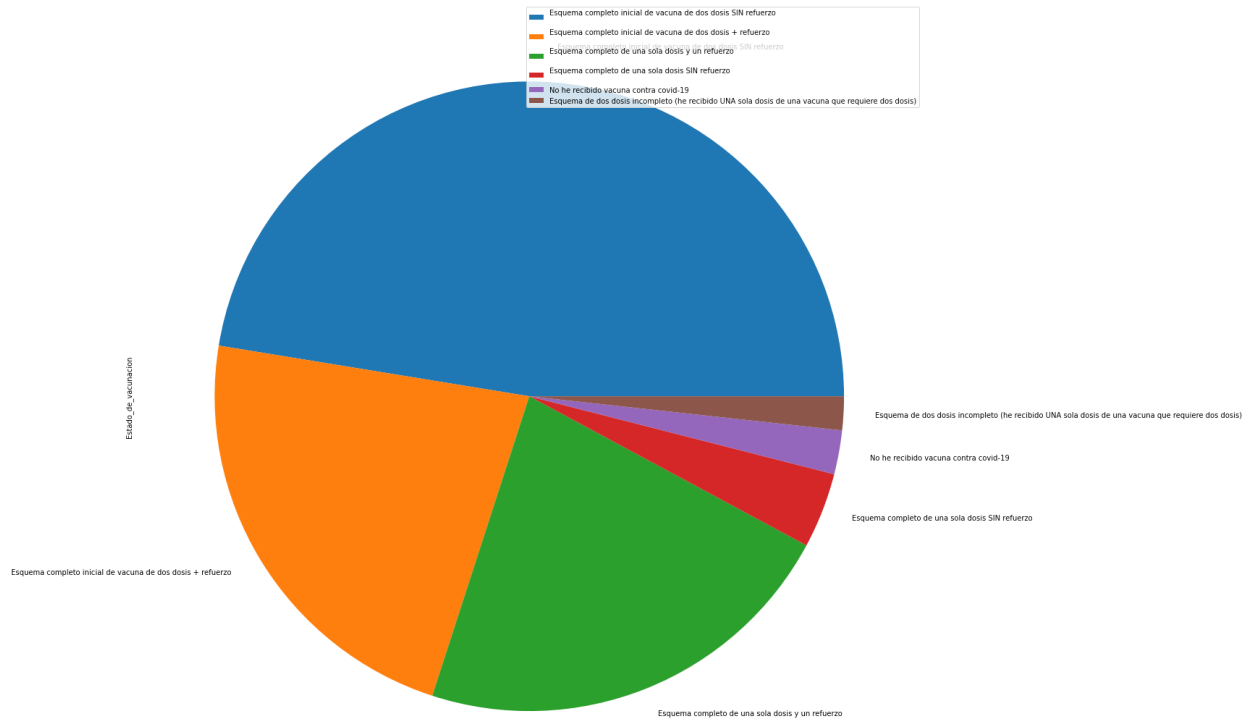
A principios del año de 2022 la administracion y directiva del Iteso lanzo una encuesta con el fin de conocer el perfil del personal y alumnos en cuanto a la vacunacion. Con estos datos se hizo un análisis sobre los datos que se tenian disponibles.

Gráfica de vacunados por rol(profesor, alumno, directivos, etc)



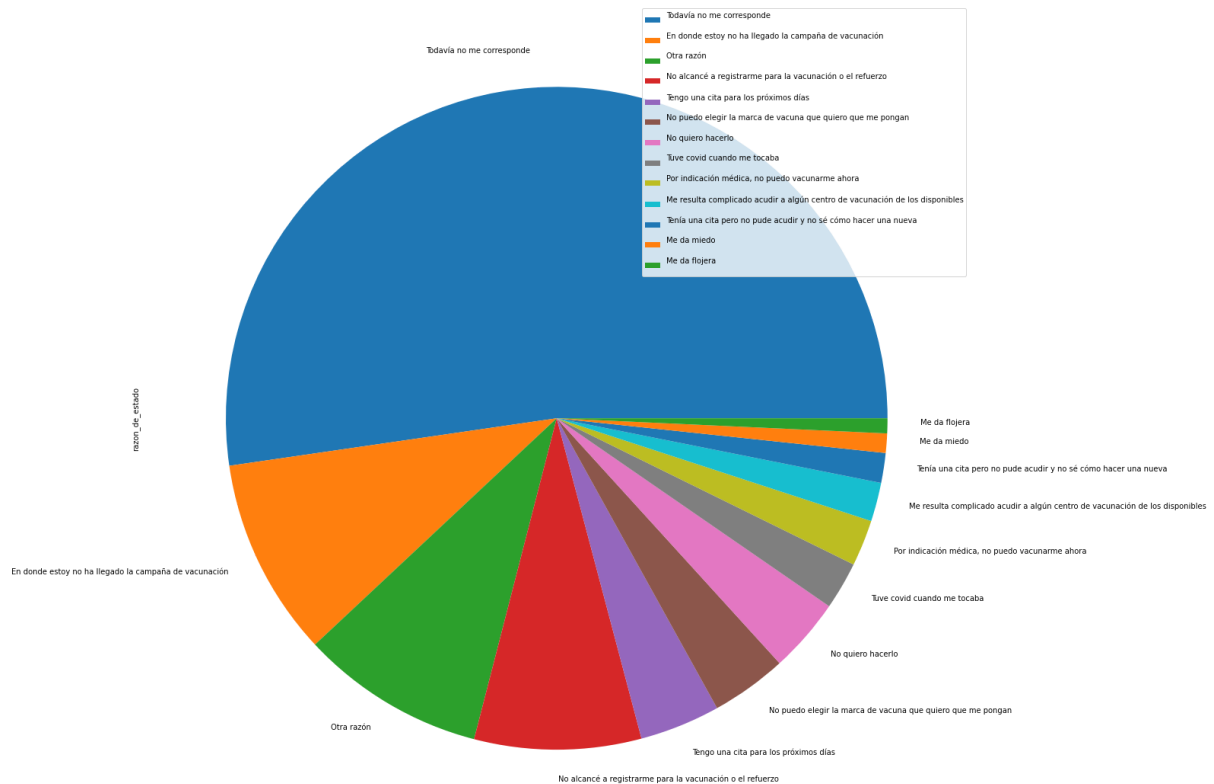
Como se puede apreciar el mayor porcentaje de vacunados son los estudiantes, después están los profesores de asignatura, después está el personal administrativo, después de ellos están los académicos, después están los estudiantes de posgrado y al final están los estudiantes de preparatoria, egresados, jubilados y Jesuitas.

Gráfica de estado de vacunación(si esta vacunado completo o incompleto o si ya cuenta con booster).



En esta gráfica se puede ver que al momento de hacer la encuesta la mayor parte de la comunidad iteso cuenta con el esquema completo de vacunación pero no cuenta con refuerzo, después se encuentran los que tienen 2 dosis y un refuerzo, después están los que recibieron una vacuna unidosis y tienen refuerzo, después están los que recibieron una vacuna unidosis y no tienen refuerzo, después están los que no se han vacunado y por último después se encuentran las personas con solo una dosis de una vacuna que requiere 2 dosis.

Gráfica de las causas/razones por las que no se han vacunado



En esta pregunta la mayor parte respondió que todavía no les toca la fecha de vacunación, después están las personas a las que debido a su ubicación no ha llegado la campaña de vacunación, después están los que no se alcanzaron a registrar, después se encuentran los que tienen cita en los próximos días, después están los que se quejan de que no pueden elegir la marca de vacuna con la que quieren ser vacunados, la siguiente razón es algo alarmante por son personas que dicen que no desean hacerlo, después se encuentran los que tuvieron COVID en día de su cita, luego están las personas que no se pueden vacunar por contraindicación médica, después están los que dicen que los centros de vacunación son inaccesibles, luego están los que perdieron la fecha y no pueden hacer otra cita, al final están los que les da miedo y flojera.

En general pudimos observar que los resultados son favorables y esperanzadores ya que indican que más gente tiene la intención de obtener una vacuna lo cual se traduce en mitigación de riesgos y contagios por COVID y en un estilo de vida más "normal".

1.7. Bibliografía y otros recursos

ADNPolítico. (2021, May 8). Estudio calcula más de 600,000 muertes por COVID en México. <https://politica.expansion.mx/mexico/2021/05/08/estudio-calcula-mas-de-600-000-muertes-por-covid-en-mexico>

Arroyo-Marioli, F., Bullano, F., Kucinskas, S. and Rondón-Moreno, C.

Arroyo-Marioli, F., Bullano, F., Kucinskas, S., & Rondón-Moreno, C. (2021). Tracking R of COVID-19: A new real-time estimation using the Kalman filter. PLOS ONE, 16(1), e0244474. doi: 10.1371/journal.pone.0244474, from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0244474>, <https://github.com/crondonm/TrackingR>

Beth Blauer - Johns Hopkins Coronavirus Resource Center Home - Johns Hopkins Coronavirus Resource Center. (2022). Retrieved 13 May 2022, from <https://coronavirus.jhu.edu/pandemic-data-initiative/data-outlook/comparing-cases-deaths-and-hospitalizations-indicates-omicron-less-deadly>

Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

Miller, H. (2020, February 3). In pictures: China is building two hospitals in less than two weeks to combat coronavirus. CNBC. <https://www.cnbc.com/2020/01/31/pictures-china-builds-two-hospitals-in-days-to-combat-coronavirus.html>

Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A. and Kavukcuoglu, K. Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., & Graves, A. et al. (2016). WaveNet: A Generative Model for Raw Audio. Retrieved 27 November 2021, from <https://arxiv.org/abs/1609.03499>

Salud, S. (2021). Datos Abiertos Bases Históricas. Retrieved 27 November 2021, from <https://www.gob.mx/salud/documentos/datos-abiertos-bases-historicas-direccion-general-de-epidemiologia>

2. Productos

Nombre y código del PAP (como en la carátula):	4J09 Ciencia de datos aplicado a información de negocios y desarrollo social
Nombre del sub proyecto (como en la carátula)	Programa de Modelación Matemática para el Desarrollo de Planes y Proyectos de Negocio
Nombre del producto	Análisis de Vacunación Activa y con Refuerzo en 3 Países Diferentes acompañado de un artículo en CRUCE; Modelo de Regresión Dinámica para Predicción de Decesos Futuros Ocasionados por la COVID-19; Modelo de Red Neuronal para Predicción de Contagios mediante <i>TensorFlow</i> ; Gráficas sobre vacunación Iteso
Descripción (qué es, para quién se realizó y para qué es):	Los modelos y análisis fueron desarrollados para poder comprender mejor el avance de la pandemia en México y en el mundo, como puede seguir afectando en el futuro y promover la vacunación activa con refuerzo.
Autores	Carlos Riggen Castillo, Ricardo Figueroa Acosta, Luis Eduardo Vázquez González y Pedro Javier Herrera Soto

3. Reflexión crítica y ética de la experiencia

El RPAP tiene también como propósito documentar la reflexión sobre los aprendizajes en sus múltiples dimensiones, las implicaciones éticas y los aportes sociales del proyecto para compartir una comprensión crítica y amplia de las problemáticas en las que se intervino.

La pandemia del Virus SARS-CoV-2 ha creado grandes cambios en diferentes ámbitos de la sociedad esto ha traído como consecuencia que los agentes que conforman las diversas esferas sociales tengan diversos planes de acción para poder contener y atacar de manera efectiva problemáticas. Las malas prácticas y las aparentes carencias que reflejaban a nivel mundial denotaba la insuficiente capacidad del sistema de salud y su evidente colapso en diversos lugares del mundo, esta crítica no va dirigida a las esferas gubernamentales, sino que, resalta el crecimiento exponencial de un virus para el cual nadie se encontraba preparado, el problema radica sobre qué tan eficaces fueron las medidas implementadas para poder detener el crecimiento de decesos en la población. Observamos a lo largo de la pandemia diversos escenarios en los cuales permea la desesperación y el descontrol debido a la agresividad con la que este virus se desarrollaba, desde las diversas trincheras se atacó de manera diferente. Los contratiempos generados al momento de proponer opciones de solución para contener esta problemática estuvieron sujetos a muchos factores que salían del espectro de acción de cada sistema debido a las diversas situaciones y los objetivos tan amplios que fueron surgiendo debido al efecto que permeó en cada una de las regiones afectadas.

El estudio realizado es un visionado de la relevancia que ha tenido la campaña de vacunación frente a la Pandemia de COVID-19, además proporciona una manera diferente para abordar situaciones más complejas e ir construyendo, con base en variables madre que en este caso son muertes y contagios, generando escenarios variados que perfilan los acontecimientos posibles en caso de que se muestre una evolución del virus que ocasione una reacción abrupta entre la población, además ofrece la oportunidad de crear estrategias sociales para evitar que, en caso de que el virus golpee con alguna variante más agresiva, contar con un abanico de opciones amplia en función de difundir información fiel y evitando la desinformación y promoviendo la cultura de la salud integral mediante las diferentes ventajas que las herramientas más plausibles y sofisticadas que nos ayudan a ofrecer soluciones contemporáneas que en tiempos anteriores no era posible el hecho de tan solo imaginarlos.

Debemos evidenciar que las problemáticas modernas, necesitan soluciones modernas necesitamos promover las técnicas que son posibles de desarrollar mediante un plan multidisciplinario y la complementación que puede proporcionar

cada perfil profesional, además resaltar que las soluciones no son ajenas a las problemáticas sociales y que se pueden realizar diferentes estudios de manera perpendicular y no paralela, donde pudimos realizar diversas técnicas donde combinamos la herramientas de programación, difusión información y estadística, que hace este estudio un proyecto integral que puede ser utilizado como base de otros proyectos con diferentes objetivos pero tomando como eje la información brindada para contextualizar o mejorar los métodos y resultados obtenidos.

3.1.- Sensibilización ante las realidades

Desafortunadamente, muchas pérdidas humanas vinieron como consecuencia de la pandemia, a causa de que una enfermedad nueva es una variable muy complicada de estudiar, por lo tanto, la desesperación, el colapso del sistema de salud resaltó las deficiencias que muchos gobiernos presentaba de manera histórica, esta pandemia sólo logró que la ciudadanía creara alternativas para poder protegerse de este virus que se desconocía en sus inicios, además de como económicamente varios países sufren un desplome significativo que traería como consecuencia ajustes diversos y permanentes en la vida de la población. El mejoramiento a marchas forzadas del sistema de salud fue evidente y aun así no fue suficiente para no crear un estado de alerta en el cual se veía ya rebasado, es por esto que estudios adyacentes ayudan a mostrar de qué manera se debe mejorar o qué áreas se encuentran vulnerables para evitar tropezar de nuevo con esta situación y hacer una mejora continua para futuras situaciones mundiales de la misma índole.

La pandemia trajo muchos cambios en la sociedad, cambios económicos, de salud y en costumbres. La pandemia trajo diversas transiciones que es imposible evadir, por lo tanto, tenemos que tener presente que el COVID-19 es y seguirá siendo un peligro latente en nuestra sociedad y que no somos ajenos a lo que sucede del otro lado del mundo ya que esto está sujeto a diversos factores que nos hace más cercanos de lo que creemos a las situaciones, más en problemáticas de salud, esto es un aspecto que en el COVID-19 fue más evidente debido a que en el tiempo que nos encontramos el acceso a la información es muchísimo más plausible que hace 20 o 30 años, debido a esto podemos medir de mejor manera con métricas diversas apegadas al eje que se desea atacar, esto es un arma de doble filo, debido a que la información que se ofrece suele ser carente de fidelidad pero las fuentes verificadas ahora están más a la mano de las personas, lo que antes era muy complicado debido a que la difusión de esa información era muy específica y difícil de interpretar para la comunidad ajena al sector científico.

3.2.- Aprendizajes logrados

Utilizar como método de estimación y ajuste motores de programación agrega otro camino por el cual direccionar los objetivos de cualquier problemática, esto

combinándolo con diversos métodos matemáticos y computacionales que nos dan la flexibilidad de manejar los datos disponibles a nuestra manera y apegado a la problemática que queremos mejorar o hacer evidente. Muchas veces el manejo de datos segmentado a una parte del entorno de problema te puede dar el panorama de cómo se comporta el ambiente en general respecto a las variables que se introdujeron, además nos dimos cuenta que la difusión de información de calidad es importantísima para que el estudio tenga un impacto favorable y que no se archive en un nicho específico sino que empape a más de un sector y despierte la necesidad de consumir y realizar proyectos apoyados en esta información y que las personas cuenten con un contexto sólido de lo que se les intenta transmitir sin importar el perfil que tengan o el sector en el que se desenvuelven.