

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 Matemáticas y Física
Maestría en Ciencia de Datos



Pronóstico de costo

TESIS que para obtener el **GRADO** de
MAESTRO EN CIENCIA DE DATOS

Presenta: **JUAN RODRIGO HUERTA MANNING**
Director: **DR. IVÁN ESTEBAN VILLALÓN TURRUBIATES**

Tlaquepaque, Jalisco. Mayo de 2023.

AGRADECIMIENTOS

Quisiera agradecer primero a Dios que me permitió terminar esta meta, así como a mi esposa Monse que me apoyó en este largo trayecto, en todo momento y cuando más lo necesité. También me gustaría dar gracias a mis padres Joaquín y Rosa quien me siguen apoyando en cada meta propuesta y sin duda a Julie Lathrop quien cuando fue mi directora confió en mí y logró el apoyo por parte de HP para poder estudiar este posgrado, misma empresa a la que también estoy eternamente agradecido por brindarme todas las herramientas posibles para seguir desarrollándome profesionalmente.

Adicionalmente, me gustaría agradecer a mi asesor Iván Esteban Villalón Turrubiates quien hizo me apoyó en todo momento. Desde que llevé una materia de programación con él, sabía que debía ser mi asesor, ya que es notable su vocación por apoyar a los profesionistas. Siempre dispuesto, buscando espacios para apoyar. Sin duda, el mejor profe con el que me topé en este posgrado.

Además, me gustaría dar las gracias a mis dos sinodales que me han apoyado en la revisión del documento con sus comentarios y su valiosa retroalimentación. El maestro Victor Martinez colaborador en Intel y el maestro Eduardo Correa García colaborador en HP Inc. Sin sus aportaciones, no hubiera podido terminar el documento de la mejor manera.

Por último, me gustaría dar gracias al ITESO que siempre a través de sus profesores, pude resolver mis dudas y siempre bajo un ambiente cordial fomentando el aprendizaje colaborativo.

DEDICATORIA

Dedico esta tesis a mi esposa Monse, quien es mi compañera de vida. No ha sido fácil combinar el desarrollo profesional y personal, pero con la persona correcta, todo es posible. Ahora es tiempo de dedicar más tiempo al desarrollo personal. Gracias Monse por tu paciencia y verás que ahora sí los mejores años están por venir.

También me gustaría dedicar este texto a cualquier mexicano que tenga la meta de seguirse preparando, no importa si en principio no es el campo que dominas, siempre la búsqueda del conocimiento traerá los frutos. Solamente hay que apostar por el mediano y largo plazo. El cortoplacismo es muy efímero y no nos lleva a conseguir nada de valor intrínseco.

RESUMEN

El problema principal que se logra resolver es el establecer un pronóstico mejor fundamentado referente a los costos no controlables de la organización en la cual laboro. Cabe mencionar que el registro de dichos costos también es relativamente nuevo y esto hizo retador el llevar a cabo el estudio. Sin embargo, el enfoque del trabajo fue el costo no controlable del Overhead, dada la relevancia e impacto en el estado de resultados de éste, ya que este costo de la organización tiene que ver con el recurso humano disponible para principalmente crecer el equipo de ventas y soportar el diverso análisis financiero.

Primero se dará un poco de contexto acerca de lo que es la organización para la cuál será de utilidad este estudio. En esta sección también se menciona lo que representa un costo no controlable y cómo es que se puede clasificar, así como la definición de estos y la importancia de mantener una estructura de costos adecuada a la situación actual de la economía mundial y del negocio que se busca impulsar.

Posteriormente se desarrollará el método de pronóstico más adecuado para poder elegir la mejor solución. Por lo tanto, aquí también se van a describir las herramientas implementadas que son las librerías de Pandas, Matplotlib, NumPy y StatsModels, así como el método ARIMA que significa un promedio móvil integrado por un autoregresivo y se especifica mediante parámetros de orden (p,d,q) .

Finalmente se realizará la implementación del modelo utilizando solamente el Overhead, siendo así consistente con el enfoque del trabajo. Los datos utilizados en este estudio resultaron ser estacionarios y eso facilitó su análisis.

TABLA DE CONTENIDO

MAESTRÍA EN CIENCIA DE DATOS	1
1. INTRODUCCIÓN	7
1.1. ANTECEDENTES.....	9
1.2. JUSTIFICACIÓN.....	9
1.3. PROBLEMA.....	9
1.4. NOVEDAD CIENTÍFICA, TECNOLÓGICA O APORTACIÓN.....	9
2. ESTADO DEL ARTE O DE LA TÉCNICA	11
3. MARCO TEÓRICO/CONCEPTUAL	13
4. DESARROLLO METODOLÓGICO	15
5. RESULTADOS Y DISCUSIÓN	22
6. CONCLUSIONES	26
BIBLIOGRAFÍA	29

LISTA DE ACRÓNIMOS Y ABREVIATURAS

ARIMA		Autoregressive Integrated Moving Average
OH		Overhead (Costos Generales de Gente)
HW		Hardware: Equipos de Cómputo
E2E		End to End (Proceso Completo)
Costo No controlable		Costo que no depende del control de la gerencia pero que debe ser supervisado para evitar impactos grandes en el estado de resultados.

1. INTRODUCCIÓN

En esta sección se da contexto de la relevancia del objeto de estudio y la razón detrás de la selección de problema a resolver.

La organización de servicios de cómputo de la empresa HP ha visto que existe un negocio potencial de 3 billones de dólares que aún no se aprovecha por completo. No importando si el HW deja de venderse como en los últimos años, la penetración de los servicios es un hecho que no se puede dejar de lado. La empresa Apple ya cuenta con el 40% de sus ingresos totales derivados de los negocios de servicios. Por lo tanto, es momento de entrar en este negocio y para ello se necesita independencia y enfoque. Esto solo puede ser logrado con enfoque e independencia. Por esa razón, desde noviembre del año pasado se ha conformado una nueva organización buscando aumentar la tasa de penetración de servicios. En otras palabras, no importando si se vende más o menos HW, sí buscar que por cada HW vendido, se vende a su vez un servicio (*HP - Home, s. f.*).

El servicio se refiere a una garantía extendida entre otras cosas. La principal característica de es esta garantía es que busca protección de los productos de los clientes. Sin embargo, el verdadero reto no es conceptual, sino financiero.

El separar todos los estados financieros de cómputo y servicios fue retador y agotador. Al final se ha logrado pero el reporte sigue siendo muy manual y los procesos aún no logran establecerse por completo con un sentido E2E. Es decir, no se ha logrado implementar procesos sólidos, ya que la información aún no se encuentra trabajada de la mejor manera

posible y en su mayoría no ha sido posible tener una clasificación idónea y se ha ido ajustando bajo la marcha.

Los procesos y datos relevantes para mi estudio son los costos financieros. Mismos que pueden ser considerados como controlables y no controlables. En específico me centraré en los costos no controlables porque son los que la organización necesita entender más a fondo.

1.1. Antecedentes

Los pronósticos de los costos no controlables se basan en promedios de un trimestre sin considerar ningún método específico y esto ha dado resultados muy poco precisos. Solo hay información de máximo 2 años porque la organización es reciente y solo se he podido separar de ese modo. Cómputo no siempre ha vendido servicios de la forma en la que la nueva organización lo requiere y por ello solo se tienen registros limitados de los costos.

1.2. Justificación

La visión de la compañía ante la caída de ventas del HW, es buscar fuentes de ingreso alterna y los servicios representan un negocio fresco de alrededor de 3 billones de dólares potenciales. Por lo tanto, asegurarse de vender un servicio cada que se venda un dispositivo, será de gran relevancia para el futuro de la compañía.

1.3. Problema

El problema o reto es establecer un mejor método para pronosticar los costos no controlables de la organización de servicios, ya que actualmente es complicado tener una idea clara de los márgenes de utilidad porque no se puede estimar un costo de manera razonable. El no tener control sobre los elementos que impactan el estado de resultados genera incertidumbre en los inversionistas y dueños del negocio. Lo que se pretende resolver es la falta de un método estándar para poder pronosticar los costos no controlables, específicamente los de costos de Overhead. Teniendo esta metodología y agregando la mayor cantidad de datos posibles con el transcurso del tiempo, se podrá lograr la estabilidad necesaria para pronosticar y evitar sorpresas financieras.

1.4. Novedad científica, tecnológica o aportación

La aportación más importante de este trabajo es implementar un método con más sustento que solamente promediar dos trimestres consecutivos para determinar el costo no

controlable de los siguientes trimestres. No existe actualmente ningún equipo dedicado a esto y lograrlo conllevará a un mejor control de costos y por con siguiente una mejor planeación de los márgenes financieros. El método por utilizar es la aplicación de ARIMA. Este método busca que los datos sean estacionarios para que, al tener una media y varianza constante, la línea del pronóstico sea semejante a los datos reales y se logre una estimación estándar y significativa.

2. ESTADO DEL ARTE O DE LA TÉCNICA

En esta sección se da contexto tanto de lo que significan los costos no controlables para las finanzas y el impacto que tiene su buen manejo para el éxito de las organizaciones.

En general no hay mucha información detallada más allá de lo conceptual de lo que es un costo no controlable. Por lo tanto, tampoco hay mucha bibliografía de cómo poder hacer un buen pronóstico de este. Sin embargo, se ha determinado mediante el estudio de las series de tiempo y de la guía de mi asesor, que el ARIMA podría ser una buena solución para poder pronosticar y sin duda será mejor que el método actual que solo radica en un promedio de los últimos 2 trimestres consecutivos, mismo que pierde de vista sustento estadístico.

A continuación, algunas definiciones encontradas:

Los costos no controlables son aquellos que no se puede tener una autoridad para su control, como por ejemplo el valor del arrendamiento a pagar depende del dueño del inmueble.

También este tipo de costos podrían incluir la depreciación y tienen un control limitado por parte de los gerentes (OBALearn, 2016).

Otro ejemplo de costo no controlable podría ser el seguro que contrata una empresa. El gerente no debería ser penalizado si hay muchos accidentes o si el costo de ese seguro aumenta de un periodo a otro (What are Uncontrollable Costs? - Definition | Meaning | Example, 2023).

El pronóstico de los costos no controlables puede llegar a ser algo relativamente nuevo, es por ello por lo que, al poder tratarlos como series de tiempo, buscando aplicaciones de diversos métodos, encontramos el ARIMA que data de 1970. Sin embargo, en 2009 este modelo destacó por su gran versatilidad y simplicidad de implementación. Este método puede aplicarse tanto a estudios de variabilidad climática como para pronosticar el PIB de un país, siempre y cuando se cuente con la información suficiente (Amaris et al., 2017).

Cabe mencionar que este método debido a su robustez estadística es útil para realizar predicciones en el corto plazo cuando existe un componente fuerte estacional. Por ejemplo, realizar predicciones bursátiles es ideal. Sin embargo, también tiene al menos dos grandes limitaciones. Cuando se pretende realizar predicciones de mediano y largo plazo se pierde la precisión y la segunda, dado su carácter univariante, se debe analizar variable por variable para obtener mejores resultados (*idUS - Depósito de Investigación de la Universidad de Sevilla, s. f.*).

3. MARCO TEÓRICO/CONCEPTUAL

En esta sección se explican las características principales, así como el sustento teórico de los elementos fundamentales para la utilización del método ARIMA.

El método ARIMA se construye a través de una media móvil integrado ajustado a series de tiempo con componentes de tendencia estacionales que se van suavizando exponencialmente. Se puede determinar manualmente los mejores parámetros para obtener un mejor resultado, pero esto requiere mucho tiempo y conocimientos (IBM Documentation, s. f.).

Para la aplicación de este método, se debe contemplar 5 etapas principales que son:

Estacionariedad – Una serie de tiempo es estacionaria cuando su media y varianza son constantes a través del tiempo. En otras palabras, la serie es estable y por lo tanto su comportamiento puede ser predecible.

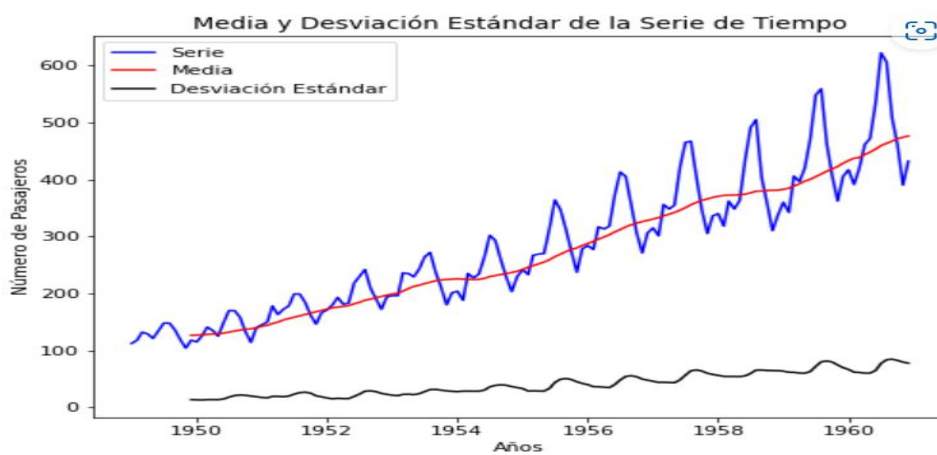


Figura 1: Gráfico de serie de tiempo estacionaria

Identificación - En esta etapa se comparan las funciones de autocorrelación calculadas contra las teóricas (UNAM, s. f.).

Estimación - Se hace el cálculo de los parámetros por medio de la estimación de máxima verosimilitud utilizando el criterio de mínimos cuadrados. Es decir, que la estimación de los parámetros del modelo ARMA se realiza minimizando la suma de los cuadrados de los residuales (UNAM, s. f.).

Evaluación - Aquí se va a determinar cuál de los modelos es estadísticamente correcto y más apto para aplicarlo al problema del costo no controlable (UNAM, s. f.).

Pronóstico - Esta fase es la más importante porque nos permite pronosticar valores futuros mediante las predicciones puntuales y asumiendo diversos intervalos de confianza (UNAM, s. f.).

Por otro lado, es importante mencionar que las siguientes librerías fueron fundamentales para el análisis y pronóstico con ARIMA:

```
#Importación de Librerías  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns
```

Pandas: Es una librería que ayuda a poder hacer dataframes y manejar colecciones de datos para poder hacer cálculos o estimaciones con ellas (*pandas documentation — pandas 2.0.1 documentation*, s. f.).

NumPy: Es una librería que permite utilizar la estadística y hacer arreglos de datos para su mejor uso (*NumPy*, s. f.).

Matplotlib: Es una librería que crea visualizaciones estáticas, interactivas y animadas en Python (*Matplotlib — Visualization with Python*, s. f.).

Seaborn: Es una poderosa librería que utiliza interfaz de alto nivel para generar gráficas estadísticas (*seaborn: statistical data visualization — seaborn 0.12.2 documentation*, s. f.).

4. DESARROLLO METODOLÓGICO

En esta sección se explica el método y código implementados para resolver la problemática del trabajo.

A continuación, se describe el método que se siguió a través de Python:

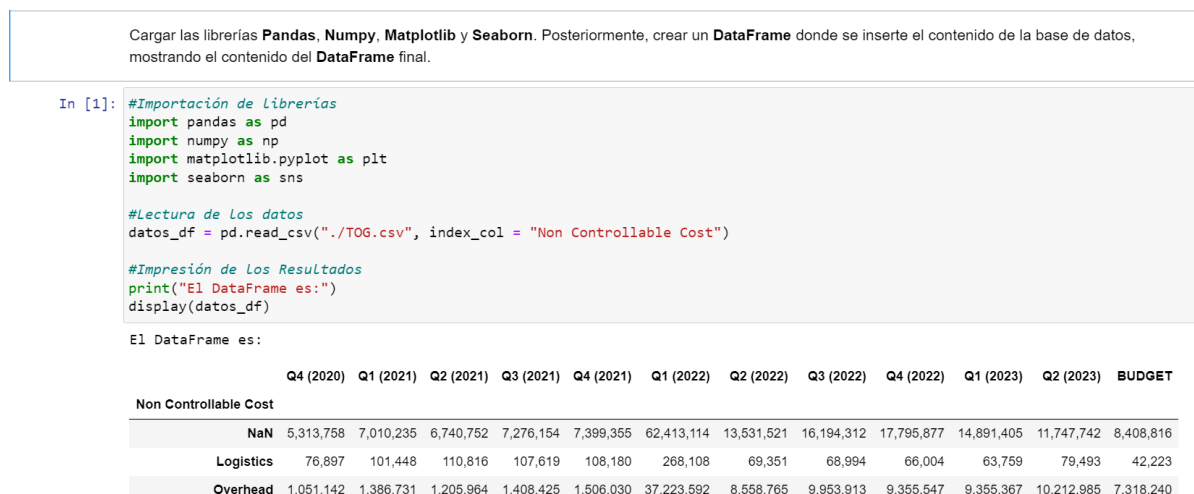


Figura 2: Base de datos y carga de librerías

Se cargan las librerías a utilizar y se genera el dataframe de los costos utilizando el índice que indican los diferentes tipos de costos no controlables. Es muy importante que la base de datos se haga un data frame, ya que de esta manera es posible hacer operaciones y aplicar diferentes herramientas estadísticas sobre la colección de datos. También es importante que se le de este tratamiento a los datos para poder guardarlo en archivos de Excel más adelante.

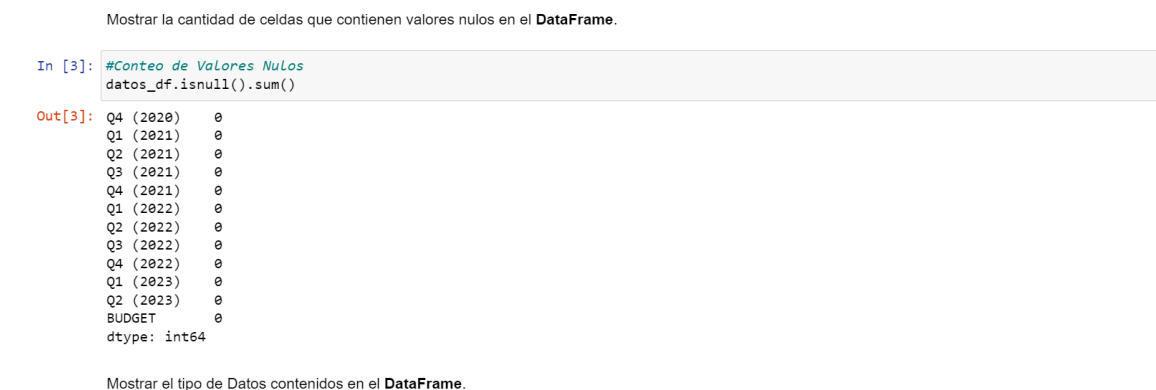


Figura 3: Tabla de datos nulos

Posteriormente, se observa que no hay valores nulos en la base de datos. Esto facilitará el tratamiento de los datos, ya que se podrán realizar operaciones y métodos de manera directa.

```
In [4]: #Tipo de Datos
datos_df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 37 entries, nan to Federal
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Q4 (2020)    37 non-null     object
1   Q1 (2021)    37 non-null     object
2   Q2 (2021)    37 non-null     object
3   Q3 (2021)    37 non-null     object
4   Q4 (2021)    37 non-null     object
5   Q1 (2022)    37 non-null     object
6   Q2 (2022)    37 non-null     object
7   Q3 (2022)    37 non-null     object
8   Q4 (2022)    37 non-null     object
9   Q1 (2023)    37 non-null     object
10  Q2 (2023)    37 non-null     object
11  BUDGET       37 non-null     object
dtypes: object(12)
memory usage: 3.8+ KB
```

Como los datos insertados en el **DataFrame** son del tipo *objeto*, se debe hacer un cambio a valores numéricos. Para ello, primeramente se eliminan las `,` y los paréntesis.

Figura 4: Tabla descriptiva de los datos

```
#Eliminación de comas y paréntesis a todo el DataFrame
datos_df = datos_df.apply(lambda x: x.str.replace(',', ''), regex = True)
datos_df = datos_df.apply(lambda x: x.str.replace('(', ''), regex = True)
datos_df = datos_df.apply(lambda x: x.str.replace(')', ''), regex = True)
datos_df = datos_df.apply(lambda x: x.str.replace('-', '0', regex = True))
datos_df.head(15)
```

```
Out[5]:
```

	Q4 (2020)	Q1 (2021)	Q2 (2021)	Q3 (2021)	Q4 (2021)	Q1 (2022)	Q2 (2022)	Q3 (2022)	Q4 (2022)	Q1 (2023)	Q2 (2023)	BUDGET
Non Controllable Cost												
NaN	5313758	7010235	6740752	7276154	7399355	62413114	13531521	16194312	17795877	14891405	11747742	8408816
Logistics	76897	101448	110816	107619	108180	268108	69351	68994	66004	63759	79493	42223
Overhead	1051142	1386731	1205964	1408425	1506030	37223592	8558765	9953913	9355547	9355367	10212985	7318240
Operations	189212	249620	244830	305949	304944	9684252	2238102	2539191	2574862	2332097	2511487	1831594
Supply Chain	17	22	1	12	94	29	3	7	15	4	0	0
Proc	3463	4569	4397	4402	4681	50124	11763	12198	13147	13016	11774	7891
MSCP	16	21	11	0	23	463	24	24	153	262	0	0
SCT	88235	116404	108514	140374	150852	637452	173429	182634	151534	129855	147984	132448
Print	2149	2835	3824	3248	3393	9384	2329	2357	2356	2342	9164	6562
SCG	10837	14296	15265	20373	16729	133076	13537	13250	48396	57892	26622	8197

Figura 5: Tabla de base de datos convertida

El dataframe es transformado a un formato sin comas para posteriormente ser transformado a formato numérico.

```
In [6]: #Conversión de datos de una columna del DataFrame
#datos_df['Q1 (2022)'] = datos_df['Q1 (2022)'].astype('float')

#Conversión de datos de todo el DataFrame
datos_df = datos_df.apply(lambda x: x.astype(float))

#Mostrar la información del tipo de datos en el DataFrame
datos_df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 37 entries, nan to Federal
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Q4 (2020)    37 non-null    float64
1   Q1 (2021)    37 non-null    float64
2   Q2 (2021)    37 non-null    float64
3   Q3 (2021)    37 non-null    float64
4   Q4 (2021)    37 non-null    float64
5   Q1 (2022)    37 non-null    float64
6   Q2 (2022)    37 non-null    float64
7   Q3 (2022)    37 non-null    float64
8   Q4 (2022)    37 non-null    float64
9   Q1 (2023)    37 non-null    float64
10  Q2 (2023)    37 non-null    float64
11  BUDGET       37 non-null    float64
dtypes: float64(12)
```

Figura 6: Tabla de datos formato numérico

Se convierten los datos a formato número para poder hacer operaciones y aplicar métodos estadísticos y así lograr generar el pronóstico buscado.

```
In [7]: #Valores Estadísticos de Las Columnas
datos_df.describe().applymap('{:.2f}'.format)
```

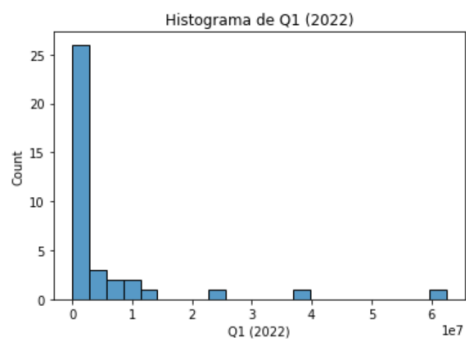
Out[7]:

	Q4 (2020)	Q1 (2021)	Q2 (2021)	Q3 (2021)	Q4 (2021)	Q1 (2022)	Q2 (2022)	Q3 (2022)	Q4 (2022)	Q1 (2023)	Q2 (2023)	BUDGET
count	37.00	37.00	37.00	37.00	37.00	37.00	37.00	37.00	37.00	37.00	37.00	37.00
mean	428804.57	565705.24	543552.46	589316.70	597140.59	5057430.11	1095336.46	1311187.57	1458969.16	1205846.78	950593.19	680755.70
std	1134853.31	1497168.06	1459016.20	1565626.02	1582505.10	12314304.86	2684909.13	3199933.24	3498186.33	2950373.66	2545579.49	1821645.86
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	17.00	22.00	707.00	12.00	94.00	2863.00	1117.00	254.00	258.00	2342.00	100.00	41.00
50%	10837.00	14296.00	16193.00	12081.00	7933.00	73976.00	13537.00	13250.00	28764.00	25237.00	23868.00	10121.00
75%	189212.00	249620.00	110816.00	140374.00	160615.00	4739012.00	935705.00	1269899.00	1090001.00	1273529.00	649378.00	447569.00
max	5313758.00	7010235.00	6740752.00	7276154.00	7399355.00	62413114.00	13531521.00	16194312.00	17795877.00	14891405.00	11747742.00	8408816.00

Figura 7: Descriptivos de la base de datos

Se calcularon unos descriptivos para tener una visión general de la información. También se generó el siguiente histograma para ver la dispersión de los datos y ver los posibles sesgos.

```
In [8]: #Gráfica de Histograma de Datos
sns.histplot(data=datos_df['Q1 (2022)'])
plt.title('Histograma de Q1 (2022)')
plt.plot();
```



Se puede realizar un análisis de Correlación en el **DataFrame** y guardar el gráfico en un archivo **PNG**.

Figura 8: Histograma



Figura 9: Gráfico de correlación

Posteriormente, se sacó una matriz de correlación que nos muestra que prácticamente todos los cuartos tienen una correlación positiva y alta con respecto a los valores de los demás. Esto quiere decir que los trimestres pueden estar bajo un efecto de estacionalidad.

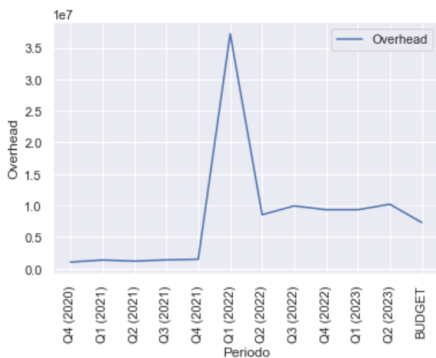
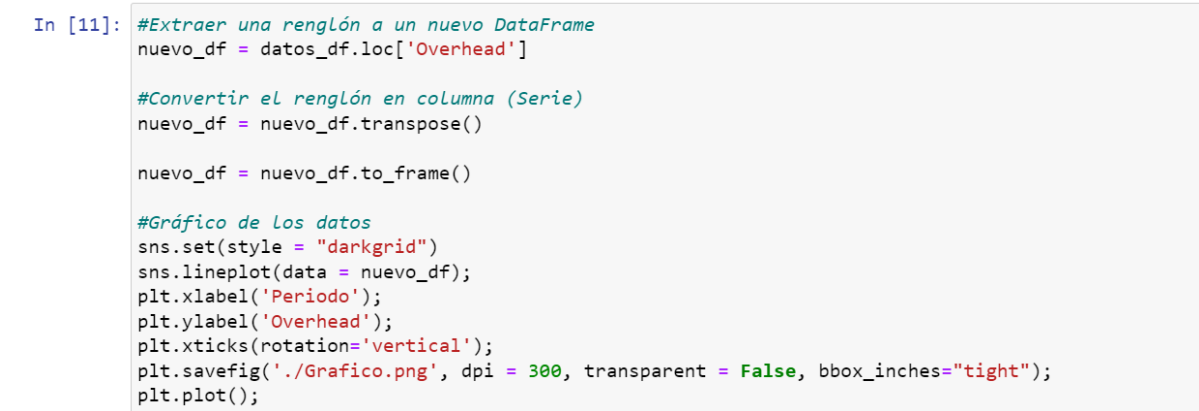


Figura 10: Gráfico de costo Overhead

Se hace la selección del costo objetivo Overhead porque es el costo más importante para la organización, debido a la cantidad de personas que se deben contratar para lograr los planes ambiciosos de expansión anteriormente descritos.

```
In [12]: #Función para La prueba Dicker-fuller de Estacionariedad
import pandas as pd
from statsmodels.tsa.stattools import adfuller

def prueba_estacionariedad(serie_tiempo):
    #Determinación de Estadísticos
    rolmean = serie_tiempo.rolling(12).mean()
    rolstd = serie_tiempo.rolling(12).std()
    #Prueba Dickey-Fuller
    print('Resultados de la Prueba Dickey-Fuller de Estacionariedad:\n')
    dftest = adfuller(serie_tiempo, autolag='AIC')
    dfoutput = pd.Series(dftest[0:4], index=['Prueba Estadística', 'Valor-p', 'Número de Retardos', 'Número de Observaciones'])
    for key,value in dfoutput[4].items():
        dfoutput['Valor Crítico (%)'%key] = value
    print(dfoutput)
```

Una vez definida la función `prueba_estacionariedad`, se aplica a la `Serie`.

```
In [13]: #Prueba de Estacionariedad para La Serie
prueba_estacionariedad(nuevo_df);
```

Resultados de la Prueba Dickey-Fuller de Estacionariedad:

```
In [13]: #Prueba de Estacionariedad para La Serie
prueba_estacionariedad(nuevo_df);
```

Resultados de la Prueba Dickey-Fuller de Estacionariedad:

Prueba Estadística	-4.001964
Valor-p	0.001400
Número de Retardos	4.000000
Número de Observaciones	7.000000
Valor Crítico (1%)	-4.938690
Valor Crítico (5%)	-3.477583
Valor Crítico (10%)	-2.843868

dtype: float64

Figura 11: Tabla de prueba de estacionariedad

Aquí propiamente ya comienza la metodología que se utilizará para calcular el pronóstico. El primer paso es aplicar la prueba de estacionariedad para, corroborar que la información puede ser tratada como una serie de tiempo cuya característica principal es que su varianza y media son constantes.

Se observa que el valor de prueba estadística es mayor que el valor del 5% de Valor Crítico, lo cual indica que esta Serie es Estacionaria con una confiabilidad de 95%. Es decir, se comprueba que se puede predecir valores futuros utilizando series de tiempo.

Modelo de Promedios Móviles Integrados Auto Regresivos (ARIMA)

Se implementa el modelo **Modelo de Promedios Móviles Integrados Auto Regresivos (ARIMA)** para realizar las predicciones de los valores.

```
In [64]: #Importación de librerías
from statsmodels.tsa.arima.model import ARIMA

# Ignore harmless warnings
import warnings
warnings.filterwarnings("ignore")

# (2, 1, 2)
# (4, 1, 4)
# (10, 1, 10)
# (10, 5, 10)
# (0, 1, 10) - Sin Auto Regresivo (AR)
# (10, 1, 0) - Sin Promedio Móvil (MA)

#Modelo de Promedios Móviles Integrados Auto Regresivos (ARIMA)
modelo_ARIMA = ARIMA(nuevo_df['Overhead'], order = (10, 1, 10))
resulta_ARIMA = modelo_ARIMA.fit()
RSS_ARIMA = sum((resulta_ARIMA.fittedvalues - nuevo_df['Overhead'])**2)
resulta_ARIMA.summary()
```

Figura 12: Tabla método ARIMA

Por último, se aplica el método ARIMA con diversos escenarios como lo son (2,1,2), (4,1,4) y (10,1,10) entre los más importantes. Considerando que p especifica el número de términos autorregresivos, del número de diferenciaciones aplicadas y q el número de términos de media móvil.

5. RESULTADOS Y DISCUSIÓN

En esta sección se muestran los resultados obtenidos aplicando conceptos de Series de Tiempo y programación.

Aquí podemos observar los resultados del ARIMA 10,1,10. Este ARIMA fue el seleccionado porque cumple con la premisa importante de replicar el comportamiento como se verá en el gráfico más adelante.

El 10 es el parámetro autorregresivo, el 1 es el número de diferenciación o rezago y el otro 10 es un promedio móvil.

ar.L6	-0.1438	397.557	-0.000	1.000	-779.341	779.054
ar.L7	0.6564	168.792	0.004	0.997	-330.170	331.483
ar.L8	0.0081	192.800	4.2e-05	1.000	-377.872	377.888
ar.L9	-0.0919	88.721	-0.001	0.999	-173.982	173.798
ar.L10	-0.0685	187.578	-0.000	1.000	-367.716	367.578
ma.L1	-0.3140	37.445	-0.008	0.993	-73.705	73.076
ma.L2	-0.2122	55.232	-0.004	0.997	-108.465	108.040
ma.L3	-0.0661	207.603	-0.000	1.000	-406.961	406.829
ma.L4	0.0351	501.502	7e-05	1.000	-982.892	982.962
ma.L5	0.0006	112.455	5.12e-06	1.000	-220.408	220.409
ma.L6	-0.0032	1295.094	-2.45e-06	1.000	-2538.340	2538.334
ma.L7	-0.0008	366.375	-2.19e-06	1.000	-718.082	718.081

Adicionalmente, se generó un gráfico individual de la siguiente manera:

```
In [28]: #Gráfico de La Serie de Tiempo
plt.plot(nuevo_df['Overhead']);
plt.plot(resulta_ARIMA.fittedvalues, color='red');
plt.title("Promedios Móviles Integrados Auto Regresivos (ARIMA)");
plt.xticks(rotation='vertical');
```

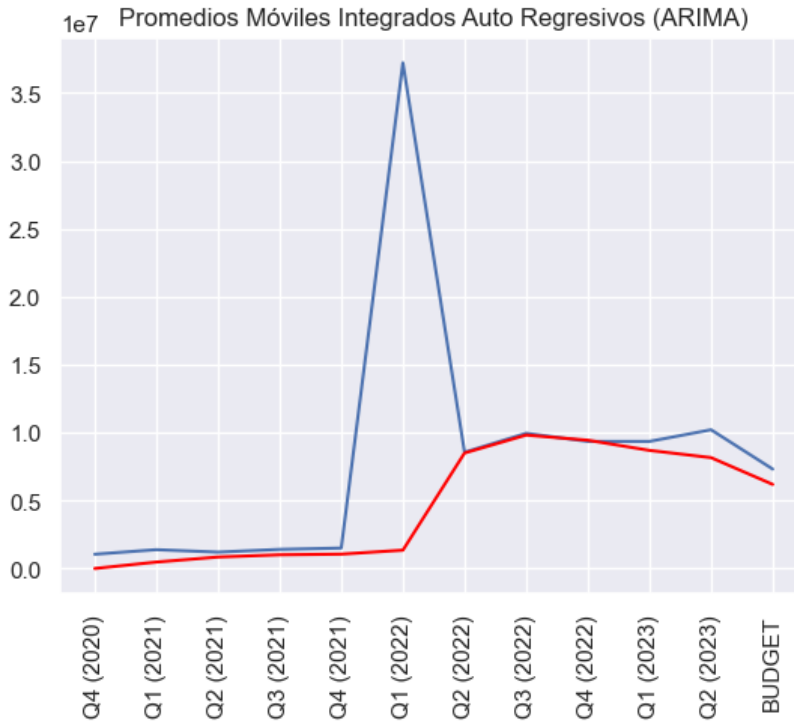


Figura 13: Gráfico ARIMA

Aquí finalmente se muestra el gráfico de los resultados del pronóstico y el código empleado.

```
In [29]: # Ignore harmless warnings
import warnings
warnings.filterwarnings("ignore")

# Forecast for the next 3 years
forecast = resulta_ARIMA.predict(start = 0,
                                end = (len(nuevo_df['Overhead']) + 3),
                                typ = 'levels')

# Plot the forecast values
nuevo_df.plot(figsize = (12, 5), legend = True);
forecast.plot(legend = True);
plt.xticks(rotation='vertical');
```

El gráfico de los promedios móviles muestra en qué periodos hay datos atípicos en el trimestre 1 del 2022 mediante una línea azul y el pronóstico lo muestra a través de una línea roja que trata de emular el comportamiento basado en los parámetros (10,1,10).

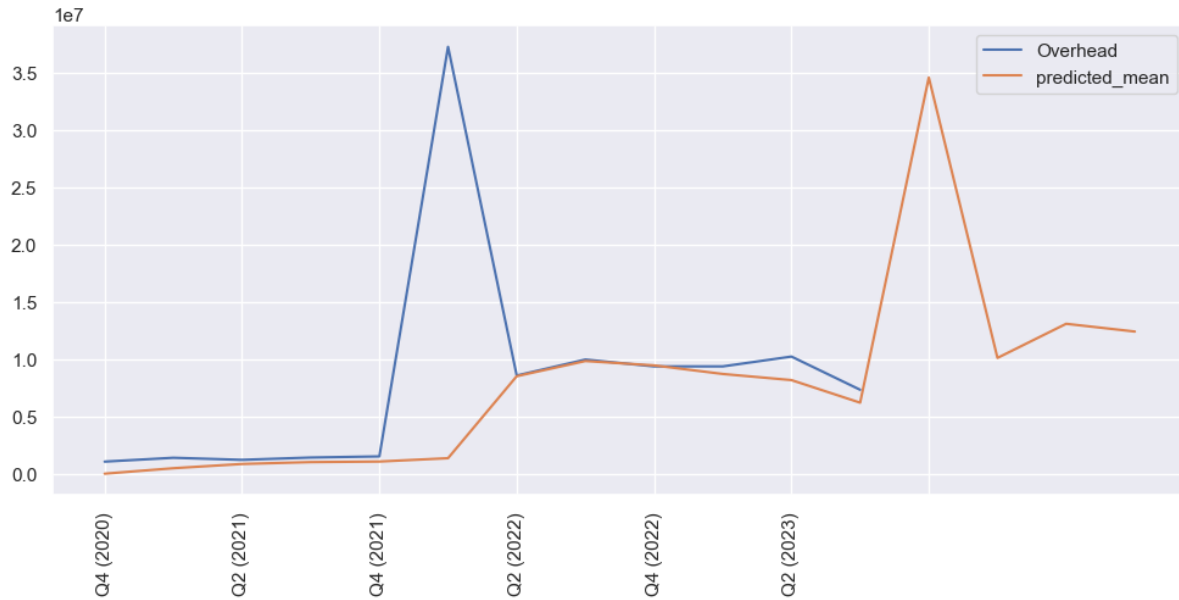


Figura 14: Gráfico de pronóstico

Se observa que la línea del pronóstico sigue la tendencia de los datos reales respetando el dato más atípico de la serie de tiempo. En este caso, se deja este dato atípico porque es relevante para el costo no controlable Overhead y es representativo de lo que se espera de la organización en trimestres similares.

Se pudo utilizar el método ARIMA porque los datos de la base de datos de los costos no controlables son estacionarios. Es decir, tienen varianza y media constantes.

También, utilizando el método forecast se analiza directamente en Excel los resultados del pronóstico del costo:

```
In [30]: display(nuevo_df)
display(forecast.to_frame())

#Guardado del Archivo en formato XLSx
nuevo_df.to_excel('./DataFrame.xlsx')

#Guardado del Archivo en formato XLSx
forecast.to_frame().to_excel('./DataFrame_Estimado.xlsx')
```

	A	B
1		predicted_mean
2	0	-
3	1	475,252.75
4	2	845,133.41
5	3	1,009,569.85
6	4	1,058,346.28
7	5	1,349,872.05
8	6	8,502,064.08
9	7	9,823,396.23
10	8	9,443,224.69
11	9	8,695,552.48
12	10	8,161,088.29
13	11	6,190,951.04
14	12	34,550,626.92
15	13	10,093,857.56
16	14	13,071,058.72
17	15	12,394,766.18

Figura 15: Tabla de resultados del pronóstico

El único dato atípico del periodo 12 pero de manera general los costos no controlables son razonables y se ha logrado el objetivo de tener un modelo con más fundamento que el actual, calculando hasta 1 año por anticipado. Este dato atípico será evaluado en el futuro y se irá ajustando conforme se tengan más datos. El objetivo final es el establecer un método para pronosticar máximo 2 trimestres hacia adelante para tener suficientes datos dentro del modelo y no diluir la estacionalidad de este.

6. CONCLUSIONES

Como ya se mencionó al inicio del trabajo, actualmente no existe una metodología estándar para realizar un pronóstico del costo no controlable. Esto ha traído como consecuencia gran dificultad para asegurar el margen financiero en el estado de resultados y poder alcanzar las metas establecidas de crecimiento.

La gran apuesta de HP es invertir lo más que se pueda en el área de servicios, ya que, ante una economía bastante golpeada, el aumentar volumen en periféricos y en HW luce muy complicado. Las ventas han caído alrededor de 20% año sobre año y no se ve algún repunte para este año. Por lo tanto, la relevancia de poder establecer un método estándar de pronóstico en los costos no controlables para la organización de servicios es sumamente relevante para las metas y planes de contingencia de la compañía.

A través de la aplicación de métodos para tratar series de tiempo como lo es el ARIMA, se pudo establecer primeramente que los costos no controlables de al menos un año hacia atrás son estacionarios y presentan una estacionalidad importante. Lo cual quiere decir, que se puede llegar a un pronóstico.

Por lo tanto, se puede concluir que se logró el objetivo de conseguir un modelo más formal y con sustento estadístico para elaborar un pronóstico del costo controlable Overhead a través de código en Python.

El pronóstico se logró con un ARIMA (10,1,10) emulando el comportamiento de los datos atípicos de Overhead.

También es importante mencionar que se ha pronosticado hasta 1 año hacia adelante, pero se pretende solamente predecir 2 trimestres consecutivos hacia adelante, con el fin de tener mayor precisión.

A pesar de lograr el objetivo, hay que reconocer que el modelo creado tiene una limitación de manejo de información, ya que aún no se cuenta con muchos trimestres de historia. Sin embargo, esto irá autocorrigiéndose a medida que se vayan integrando más datos.

Por otro lado, el modelo creado es suficientemente flexible para tratar por separado el comportamiento de cada una de las líneas del costo no controlable. Es decir, que se buscará en el futuro expandir el campo de acción del modelo a más líneas del costo y no sólo la más relevante como lo es Overhead.

Esto va a generar posibilidad de tener analíticos de gran relevancia para la toma de decisiones incluso por la línea de producto, ya que se analizará en específico cada línea de costo asociado a una determinada unidad de negocio.

El aplicar este método a las otras líneas del costo no controlable, no garantiza que todas las ARIMAS deban ser (10,1,10). Por lo tanto, se recomienda que se analice por separado cada línea y se genere un ARIMA por cada línea de costo no controlable y al final se seleccione la que tenga un gráfico que se mueva con relación a los datos actuales.

Las recomendaciones para el trabajo a realizar posteriormente son:

- Utilizar *streamlit* para implementar un tablero dinámico, mostrando mejor forma los resultados y lograr tener diferentes vistas y segmentadores para el usuario y a su vez.

- Agregar más datos con el paso del tiempo para ir mejorando la predictibilidad del modelo.
- Aplicar el método del costo no controlable *overhead* a los demás costos, pero personalizando el ARIMA.

La implementación del modelo está en revisión con la gerencia y se espera hacer testeos más profundos por cada costo en el siguiente trimestre.

Por último, se puede concluir que las finanzas y la ciencia de datos tienen un camino muy próspero por recorrer juntos, ya que se pueden crear modelos muy interesantes que ayuden a la toma de decisiones.

BIBLIOGRAFÍA

HP - Home. (s. f.). <https://investor.hp.com/home/default.aspx>

OBAlearn. (2016, 5 febrero). ¿Qué son los costos no controlables? -

OBAlearn. <https://www.obalearn.com/es/q-and-a/que-son-los-costos-no-controlables/>

What are Uncontrollable Costs? - Definition | Meaning | Example. (2023, 26 marzo). My

Accounting Course. https://www-myaccountingcourse-com.translate.googleusercontent.com/translation/what-are-uncontrollable-costs?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=rq#:~:text=Depreciation%20is%20another%20example%20of%20an%20uncontrollable%20cost.

dictionary/uncontrollable-

costs?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=rq#:~:text=Depreciation%20is%20another%20example%20of%20an%20uncontrollable%20cost.

IBM Documentation. (s. f.). [https://www.ibm.com/docs/es/spss-](https://www.ibm.com/docs/es/spss-modeler/saas?topic=series-arima)

modeler/saas?topic=series-arima

Modelación ARIMA (2021).

<http://www.ptolomeo.unam.mx:8080/jspui/bitstream/132.248.52.100/363/7/A7.pdf>

pandas documentation — pandas 2.0.1 documentation.

(s. f.). <https://pandas.pydata.org/docs/#module-pandas>

NumPy. (s. f.). <https://numpy.org/>

Matplotlib — Visualization with Python. (s. f.). <https://matplotlib.org/>

seaborn: statistical data visualization — seaborn 0.12.2 documentation.

(s. f.). <https://seaborn.pydata.org/>

Amaris, G., Avila, H., & Guerrero, T. (2017). Applying ARIMA model for annual volume time series of the Magdalena River. DOAJ (DOAJ: Directory of Open Access

Journals). <https://doi.org/10.14483/udistrital.jour.tecnura.2017.2.a07>

idUS - Depósito de Investigación de la Universidad de Sevilla. (s. f.). idUS - Depósito de Investigación Universidad de Sevilla. <https://idus.us.es/>