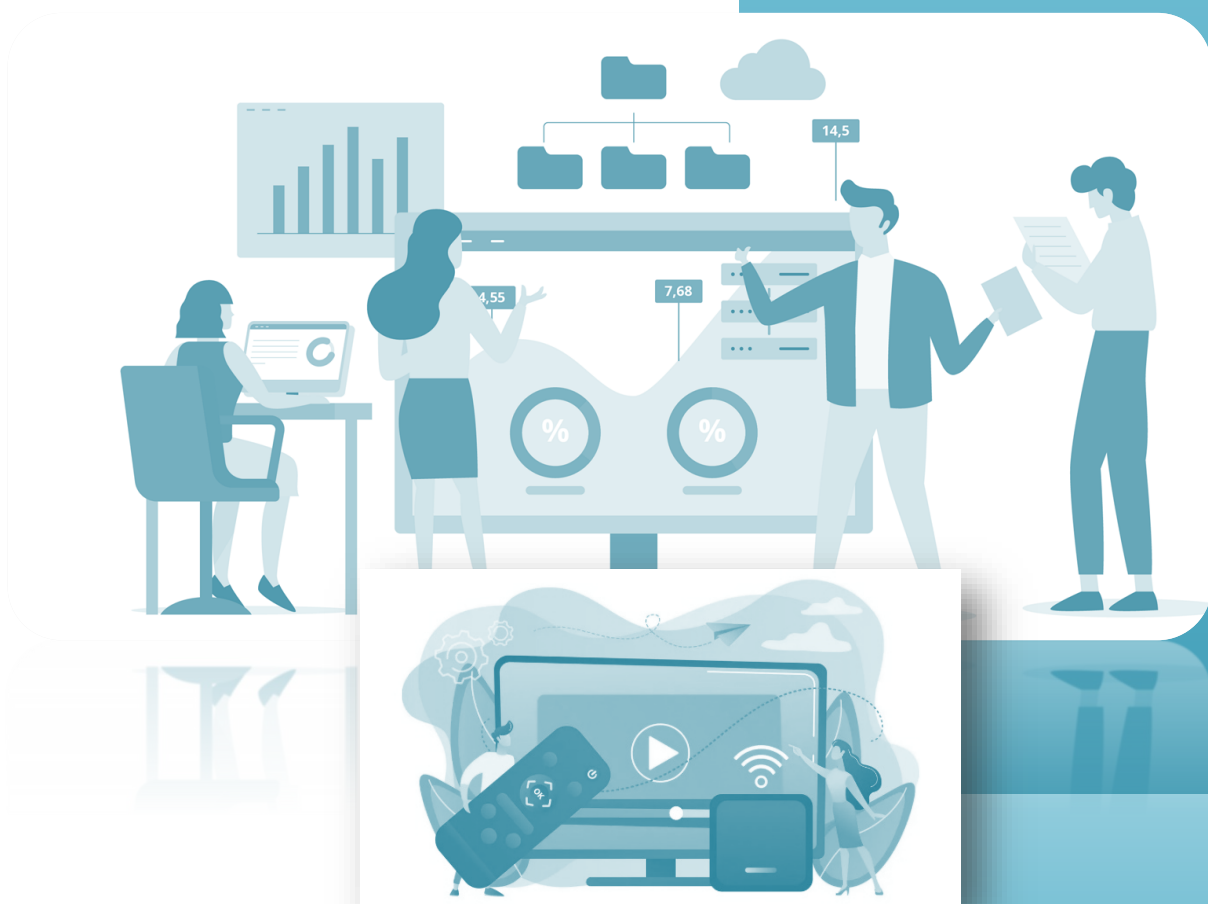


# 2024

## CASO DE ESTUDIO



Diomedes Igua Hernández   Ricardo  
Patiño López   Daniel Arismendi  
Tibaduiza   Diana Bermúdez Carrillo

12-4-2024



## Contenido

1. CASO DE ESTUDIO .....	2
2. GRÁFICA PARA IDENTIFICAR LA TENDENCIA EN LAS VENTAS .....	3
a. ANÁLISIS DE MACROTENDENCIAS DE SUSCRIPCIONES: .....	3
b. ANÁLISIS DE TENDENCIA GENERAL:.....	5
3. GRÁFICA ARQUITECTURA TECNOLÓGICA .....	6
a. DESCRIPCIÓN CASO DE ESTUDIO:.....	6
b. PROPUESTA DE ARQUITECTURA PARA LA EMPRESA.....	8
c. DIAGRAMA DE ARQUITECTURA PROPUESTO PARA LA EMPRESA .....	9
4. PROCESOS ETL .....	10
5. RECOMENDACIONES .....	11



# 1.CASO DE ESTUDIO

Actualmente una compañía de streaming quiere realizar un análisis sobre la cantidad de suscriptores. El objetivo es identificar si se vienen perdiendo clientes, o si, por el contrario, el número de suscripciones ha aumentado. Usted es contratado para realizar entregar un informe a la junta directiva. Adicionalmente, los datos se vienen guardando en un archivo en formato .CSV, por lo que también se espera que usted les haga una recomendación sobre un modelo de negocio soportado en tecnologías actuales para que la manipulación de los datos sea más práctica. Tenga en cuenta que:

- Se estima que las suscripciones ya llegan al orden de millones.
- Se necesita garantizar que las diferentes sucursales de la empresa tengan acceso a los datos desde diferentes lugares del mundo.
- En cada una de las sucursales de la empresa tienen diferentes desarrolladores, los cuales programan con diferentes lenguajes de programación.
- Actualmente los datos pueden llegar en diferentes formatos y desde diversas fuentes.

Los gerentes desean implementar técnicas de Inteligencia Artificial para tener predicciones. ¿Qué recomendaciones le haría a la junta directiva con respecto a las técnicas y a la calidad de los datos?

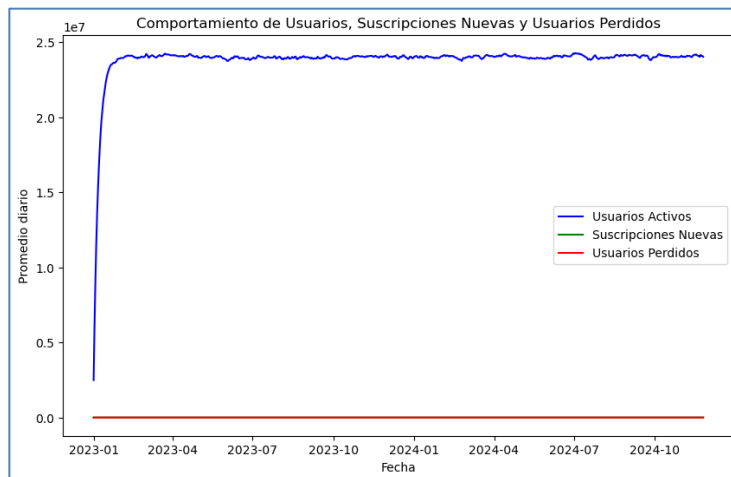
El informe debe tener:

1. Una gráfica para identificar la tendencia en las ventas.
2. Una gráfica para mostrar la arquitectura tecnológica (API's, bases de datos, tecnologías).
3. Listado de procesos ETL que se deben realizar.
4. Recomendaciones sobre las técnicas de aprendizaje supervisado a utilizar.

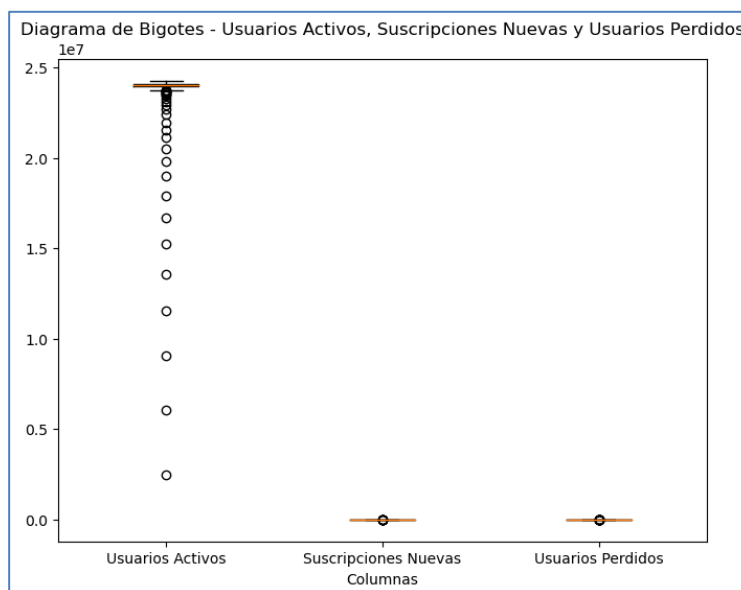
## 2. GRÁFICA PARA IDENTIFICAR LA TENDENCIA EN LAS VENTAS

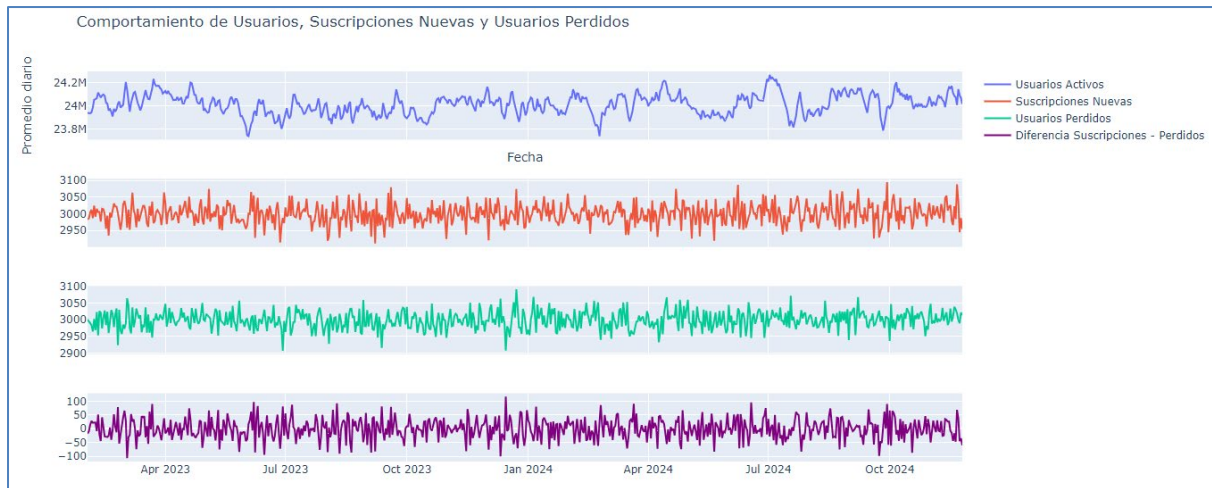
### a. ANÁLISIS DE MACROTENDENCIAS DE SUSCRIPCIONES:

Para visualizar la tendencia en las ventas de suscripciones, se puede utilizar una gráfica de líneas que muestre la cantidad de suscriptores a lo largo del tiempo. Esto permitirá identificar si hay una tendencia creciente, decreciente o estable en el número de suscripciones.

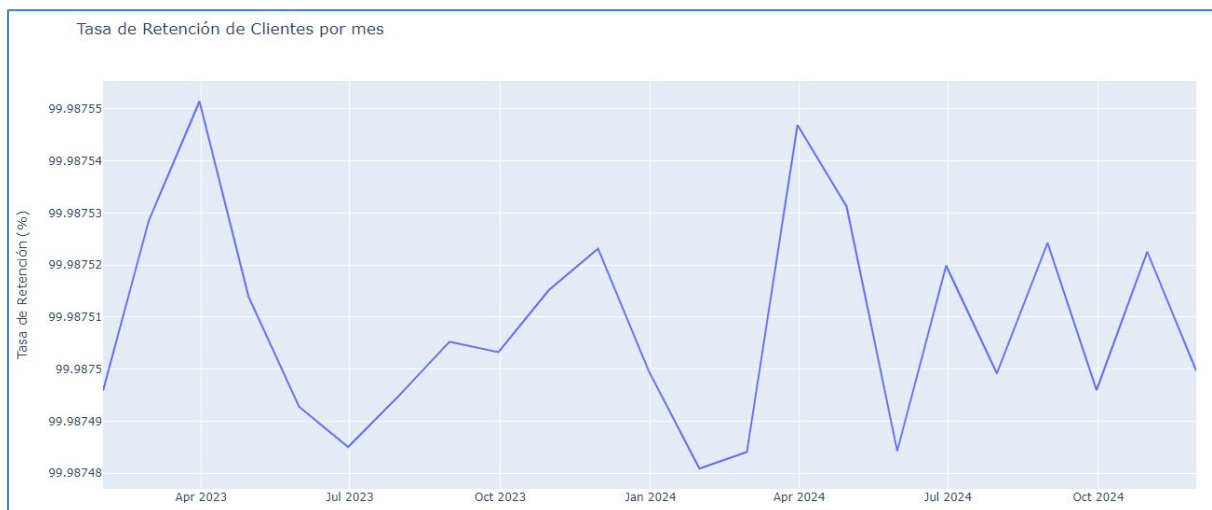


un análisis preliminar de los datos muestra una anomalía en los primeros meses de los datos por tanto se procede a eliminar los primeros valores para generar análisis más precisos

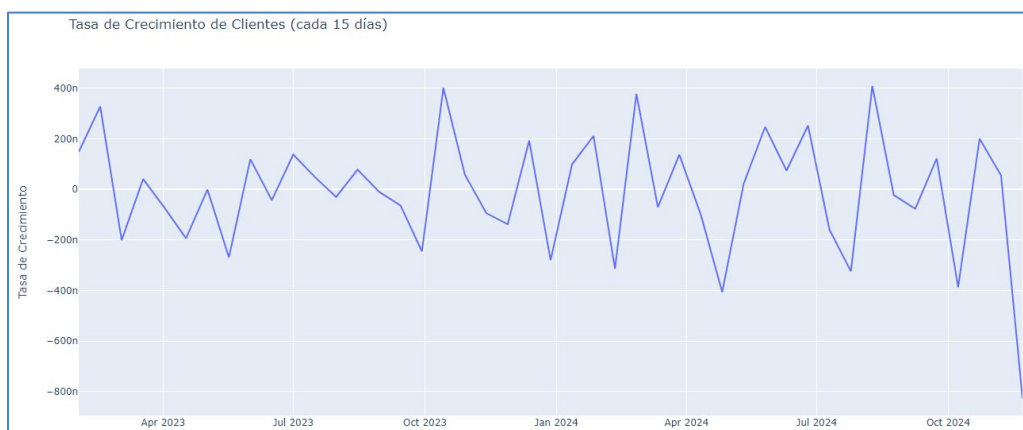




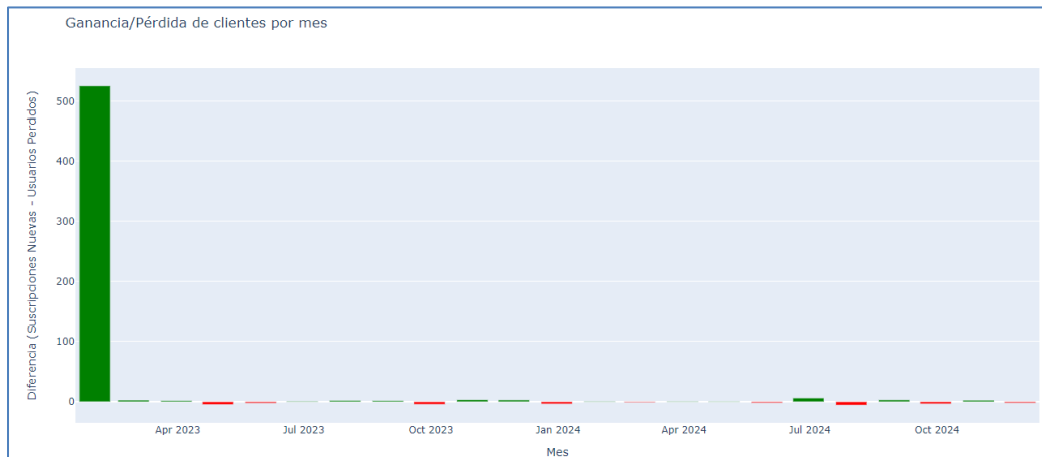
De la gráfica anterior se concluye que la macro tendencia es estable se pierden clientes en el mismo rango en el que se ganan y pese a las variaciones diarias de usuarios en promedio la retención es suscriptores es estable.



la retención de clientes es muy alta superior al 98% sin embargo, se evidencian algunas micro tendencias de resaltar



durante los últimos meses de los registros hay una ligera tendencia a la pérdida de clientes opacada por un alto volumen de usuarios anterior



esto se ve reflejado en la tasa de pérdidas con un ligero declive durante los últimos meses.

## b. ANÁLISIS DE TENDENCIA GENERAL:

En resumen, aunque la macrotendencia muestra estabilidad en el número de suscriptores con una alta retención de clientes, es importante destacar las micro tendencias, especialmente la ligera tendencia a la pérdida de clientes en los últimos meses. Esto sugiere la necesidad de prestar atención a estrategias de retención y fidelización de clientes para contrarrestar esta tendencia y mantener el crecimiento sostenido de la base de usuarios.

## 3. GRÁFICA ARQUITECTURA TECNOLÓGICA

### a. DESCRIPCIÓN CASO DE ESTUDIO:

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Cargar los datos desde el archivo CSV
datos = pd.read_csv("registro_usuarios.csv")

# Convertir la columna "Fecha" a tipo datetime
datos["Fecha"] = pd.to_datetime(datos["Fecha"])

# Agrupar los datos por día y calcular el promedio
datos_diarios = datos.groupby(datos["Fecha"].dt.date).mean()

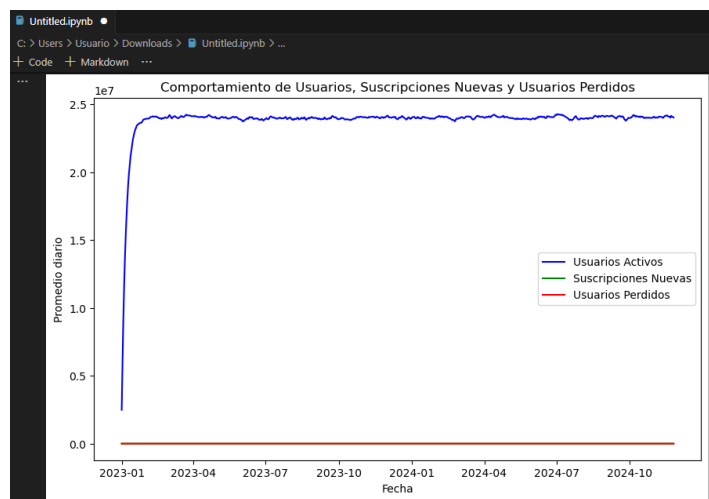
# Graficar el comportamiento de usuarios activos, suscripciones nuevas y usuarios perdidos
plt.figure(figsize=(10, 6))
plt.plot(datos_diarios.index, datos_diarios["Usuarios Activos"], label="Usuarios Activos", color="blue")
plt.plot(datos_diarios.index, datos_diarios["Suscripciones Nuevas"], label="Suscripciones Nuevas", color="green")
plt.plot(datos_diarios.index, datos_diarios["Usuarios Perdidos"], label="Usuarios Perdidos", color="red")
plt.title("Comportamiento de Usuarios, Suscripciones Nuevas y Usuarios Perdidos")
plt.xlabel("Fecha")
plt.ylabel("Promedio diario")
plt.legend()
plt.show()

# Graficar el comportamiento de suscripciones nuevas y usuarios perdidos
plt.figure(figsize=(10, 6))
plt.plot(datos_diarios.index, datos_diarios["Suscripciones Nuevas"], label="Suscripciones Nuevas", color="green")
plt.plot(datos_diarios.index, datos_diarios["Usuarios Perdidos"], label="Usuarios Perdidos", color="red")
plt.title("Comportamiento de Suscripciones Nuevas y Usuarios Perdidos")
plt.xlabel("Fecha")
plt.ylabel("Promedio diario")
plt.legend()
plt.show()

# Diagrama de bigotes para Usuarios Activos
plt.figure(figsize=(8, 6))
plt.boxplot(datos_diarios["Usuarios Activos"], vert=False)
plt.title("Diagrama de Bigotes - Usuarios Activos")
plt.xlabel("Usuarios Activos")
plt.show()

# Diagrama de bigotes para Suscripciones Nuevas
plt.figure(figsize=(8, 6))
plt.boxplot(datos_diarios["Suscripciones Nuevas"], vert=False)

```



## Aplicación en Visual Studio Code:

Esta es la aplicación en la que se desarrolla y ejecuta el código Python para el análisis de datos de suscripciones. Desde aquí, se carga el archivo CSV y se realiza el procesamiento de los datos utilizando bibliotecas específicas.

## Procesamiento y Transformación de Datos (Pandas):

Pandas es una biblioteca de Python que proporciona estructuras de datos y herramientas para el análisis de datos. Se utiliza para cargar el archivo CSV, limpiar y transformar los datos, así como realizar operaciones como agrupaciones y cálculos de métricas.

## Análisis y Visualización de Datos (Matplotlib, Seaborn):

Funcionalidad: Matplotlib y Seaborn son bibliotecas de visualización de datos en Python. Matplotlib proporciona una amplia variedad de gráficos y figuras personalizables, mientras que Seaborn ofrece una interfaz de alto nivel para crear gráficos estadísticos atractivos y informativos.

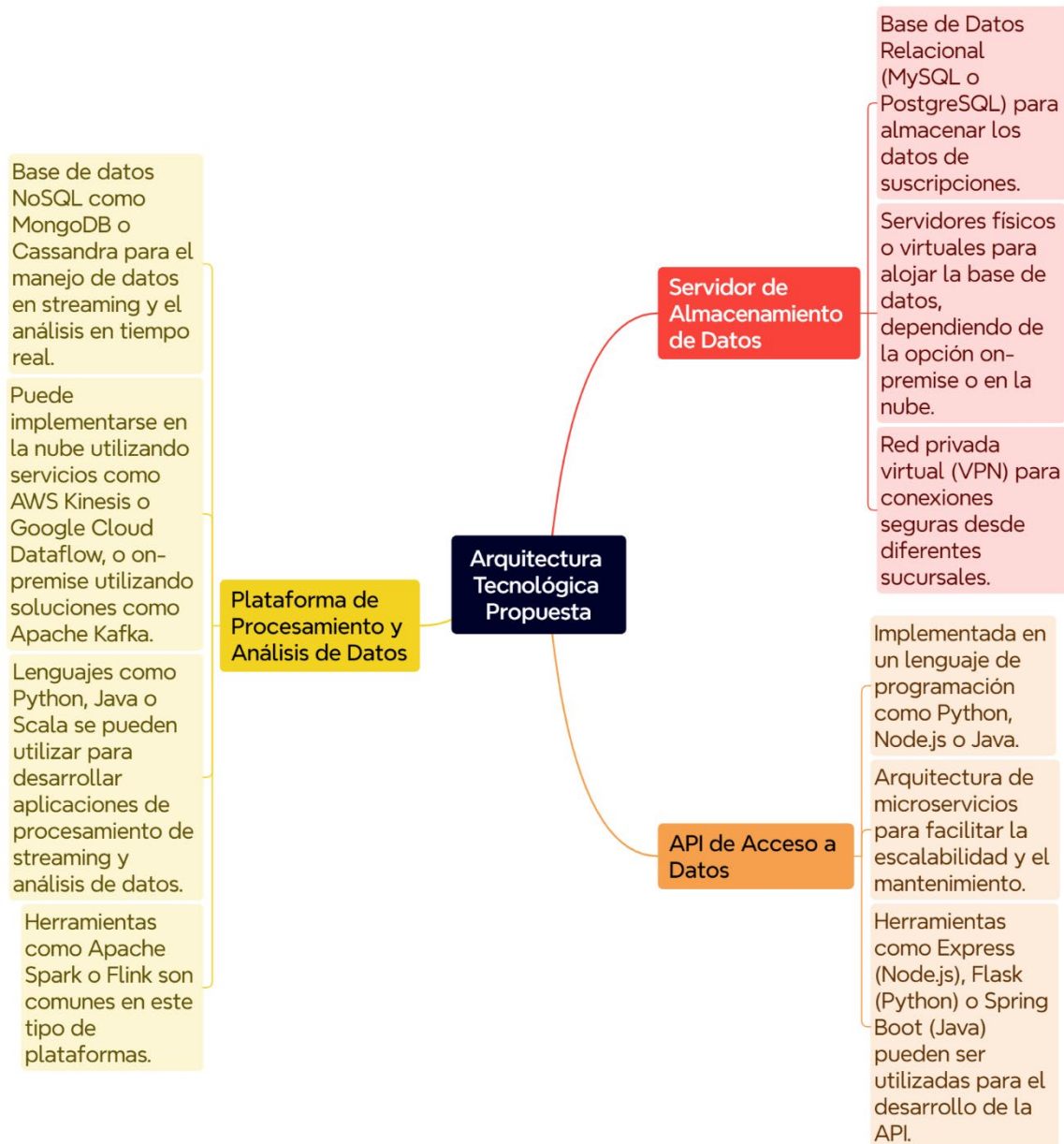
Conexión: Después de procesar los datos con Pandas, la aplicación en Visual Studio Code utiliza Matplotlib y Seaborn para generar visualizaciones que representan tendencias, patrones y relaciones en los datos de suscripciones.

Este enfoque permite cargar, procesar, analizar y visualizar los datos de suscripciones de manera eficiente y flexible, todo dentro del entorno de desarrollo de Visual Studio Code, utilizando las bibliotecas de Python adecuadas para cada paso del proceso.



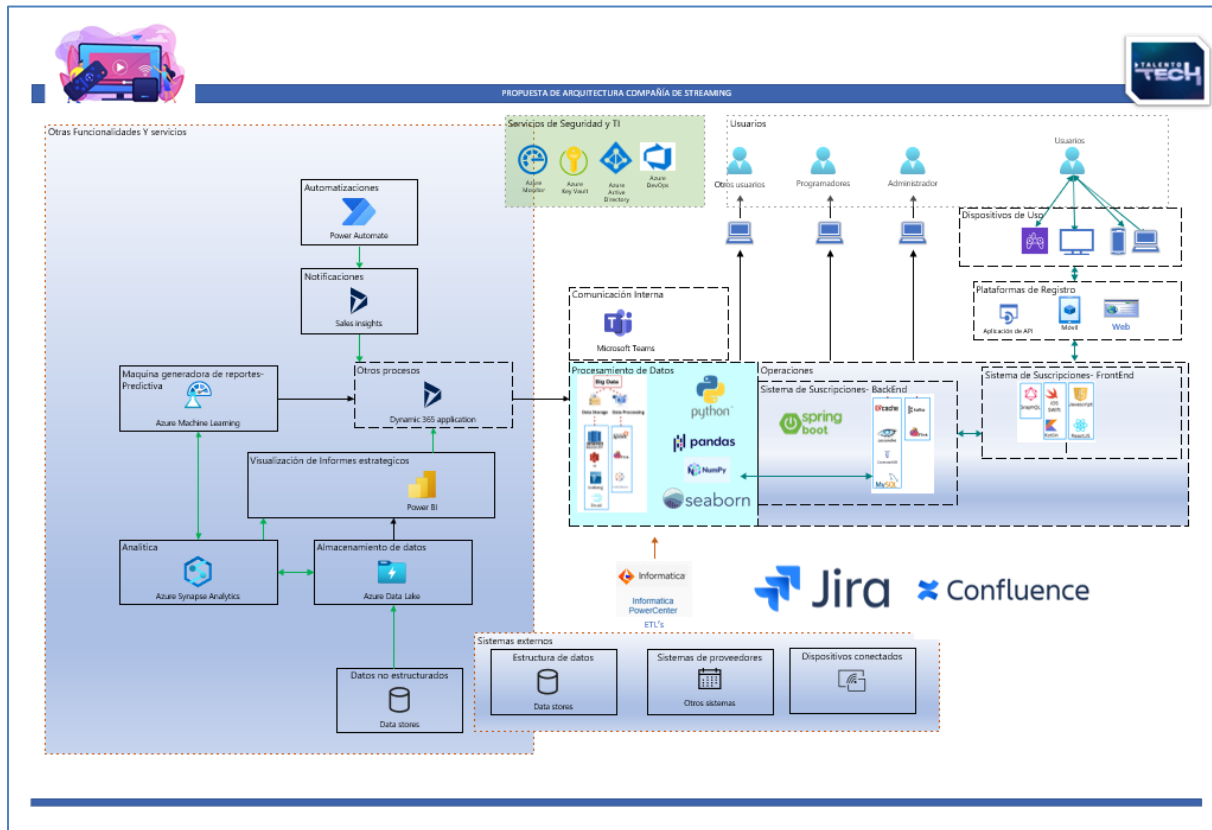


## b. PROPUESTA DE ARQUITECTURA PARA LA EMPRESA



Presented with xmind

## c. DIAGRAMA DE ARQUITECTURA PROPUESTO PARA LA EMPRESA



## 4.PROCESOS ETL

Dado que los datos pueden llegar en diferentes formatos y desde diversas fuentes, se deben realizar los siguientes procesos ETL:

**Extracción:** Recopilar datos de diversas fuentes, como archivos CSV, bases de datos SQL, API's, etc.

**Transformación:** Limpiar y estructurar los datos según sea necesario, para garantizar la coherencia y la calidad.

**Carga:** Insertar los datos transformados en la base de datos para su posterior análisis y utilización.

En este caso particular

	A	B	C	D
1	Fecha	Usuarios Activos	Suscripciones Nuevas	Usuarios Perdidos
2	1/01/2023 0:00	504084	4118	34
3	1/01/2023 0:01	507904	3862	42
4	1/01/2023 0:02	512401	4560	63
5	1/01/2023 0:03	514955	2601	47
6	1/01/2023 0:04	518017	3132	70
7	1/01/2023 0:05	522708	4753	62
8	1/01/2023 0:06	524577	1944	75
9	1/01/2023 0:07	528247	3705	35
10	1/01/2023 0:08	529863	1667	51
11	1/01/2023 0:09	534475	4640	28
12	1/01/2023 0:10	539286	4845	34
13	1/01/2023 0:11	543162	3910	34
14	1/01/2023 0:12	547085	3989	66
15	1/01/2023 0:13	549299	2318	104
16	1/01/2023 0:14	553330	4115	84
17	1/01/2023 0:15	557600	4332	62
18	1/01/2023 0:16	558948	1445	97
19	1/01/2023 0:17	563567	4655	36
20	1/01/2023 0:18	567597	4121	91
21	1/01/2023 0:19	570074	2581	104
22	1/01/2023 0:20	571592	1598	80

## 5.RECOMENDACIONES

Para que el proceso marche de acuerdo a la operación de la empresa y que la arquitectura de negocio y tecnológica puedan soportar un buen nivel business intelligence para una buen análisis, visualización y presentación de datos y transformarlo en información que permita optimizar la toma de decisiones se generan las siguientes recomendaciones:

1. Para predecir las tendencias futuras de suscriptores, se pueden emplear técnicas de aprendizaje supervisado como regresión lineal, regresión logística o modelos de series temporales.
2. Recomendando explorar modelos de series temporales avanzados como ARIMA (Autoregressive Integrated Moving Average) o modelos de redes neuronales recurrentes (RNN) para capturar patrones complejos en los datos de suscripción.
3. Es importante evaluar y comparar varios modelos utilizando métricas de rendimiento como el error cuadrático medio (MSE) o el coeficiente de determinación ( $R^2$ ) para seleccionar el modelo más adecuado.
4. Recomendando implementar una arquitectura basada en microservicios y contenedores, lo que permitirá a las diferentes sucursales acceder a los datos de manera eficiente y escalable. Esto se puede lograr utilizando tecnologías como Docker y Kubernetes.
5. Para la gestión de datos, se puede optar por una base de datos distribuida y escalable como Cassandra o MongoDB, que pueden manejar grandes volúmenes de datos y ofrecer replicación multirregión para garantizar la disponibilidad y la consistencia de los datos en todo el mundo.
6. Se puede establecer una API RESTful para la comunicación entre los diferentes servicios y sucursales, lo que facilitará la integración de sistemas y el intercambio de datos de manera segura y eficiente.
7. En el transcurso del documento se han mencionado diferentes herramientas para poder llegar a la implementación de las funcionalidades, se requiere un análisis de tecnología y la inclusión del PETI- Plan estratégico de tecnología donde se evalúen las posibilidades de implementación y de integración.