

# PROYECTO FINAL: TELECOMUNICACIONES



Identificar operadores ineficaces en la empresa de telefonía virtual  
Call Me Maybe

# ☎️ INTRODUCCIÓN ☎️

La empresa tiene interés en implementar un sistema que permita al área de supervisión identificar a los operadores menos eficaces. Para este caso, se plantea realizar un análisis de datos para identificar patrones y tendencias para facilitar la identificación de dichos operadores.

Se considera que un operador es ineficaz si tiene una gran cantidad de llamadas entrantes perdidas (internas y externas) y un tiempo de espera prolongado para las llamadas entrantes.

Se realizará un análisis exploratorio de datos que permita identificar tendencias y realizar hipótesis que logren facilitar la función para los supervisores identificando más rápido a dichos operadores ineficaces.



# DATOS

El primer dataset (telecom\_dataset\_us.csv) contiene las siguientes columnas:

- user\_id: ID de la cuenta de cliente
- date: fecha en la que se recuperaron las estadísticas
- direction: "dirección" de llamada (out para saliente, in para entrante)
- internal: si la llamada fue interna (entre los operadores de un cliente o clienta)
- operator\_id: identificador del operador
- is\_missed\_call: si fue una llamada perdida
- calls\_count: número de llamadas
- call\_duration: duración de la llamada (sin incluir el tiempo de espera)
- total\_call\_duration: duración de la llamada (incluido el tiempo de espera)

El segundo dataset (telecom\_clients\_us.csv) contiene las siguientes columnas:

- user\_id: ID de usuario/a
- tariff\_plan: tarifa actual de la clientela
- date\_start: fecha de registro de la clientela



# STAKEHOLDERS

Los interesados directos en este análisis son el área de supervisión, y en menor medida tal vez el área de gerencia, por lo que se debe considerar generar un informe acorde a estos interesados, para este caso se plantea generar una presentación ejecutiva que permita comunicar información descubierta de la manera más concisa posible, sin adentrarse en detalles técnicos sobre las técnicas de análisis o código, si no más bien en los resultados y hallazgos obtenidos, así como algunas recomendaciones.





# 📞 OBJETIVOS 📞

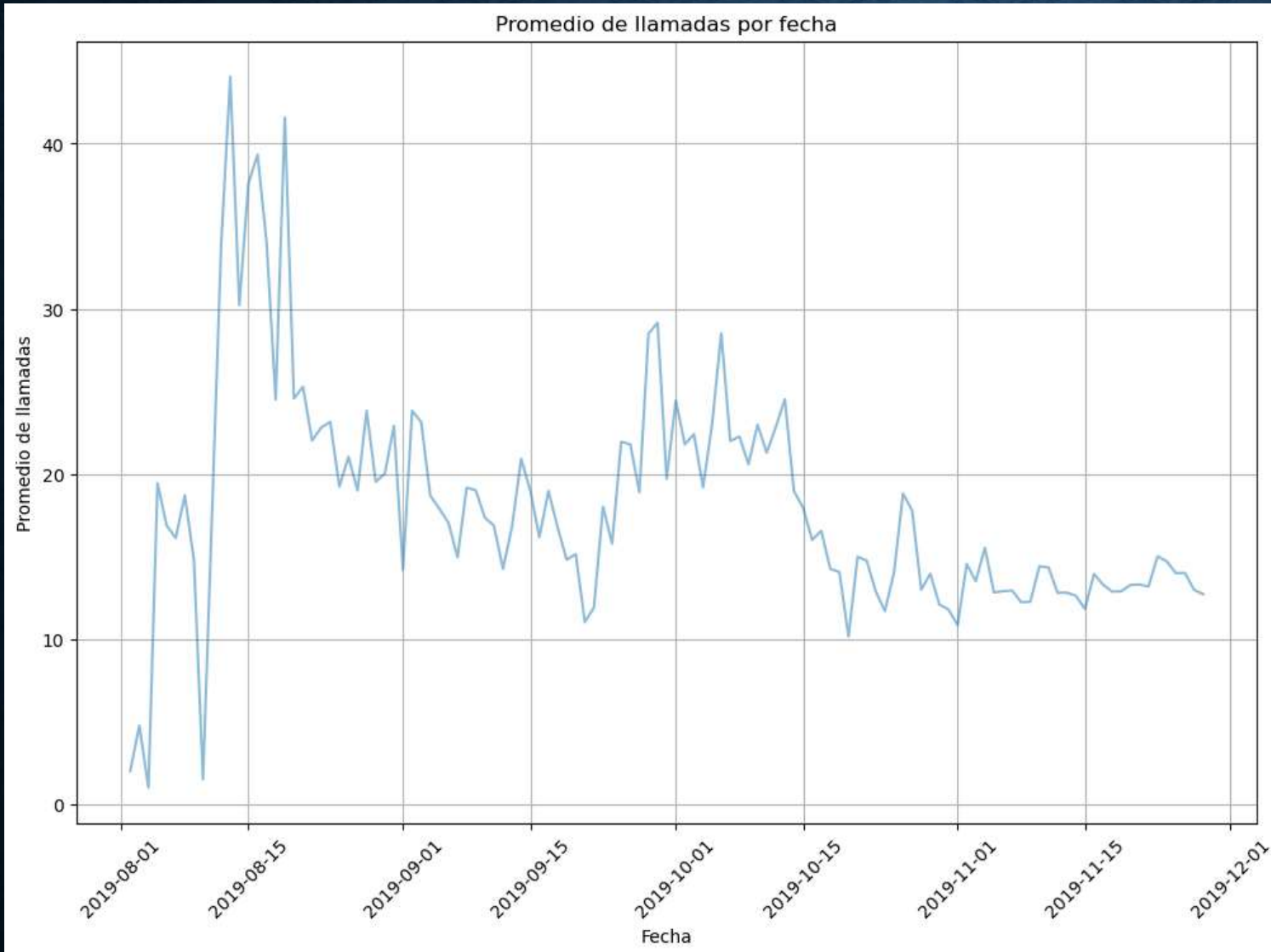
El objetivo es la identificación de patrones para operadores más ineficaces, ya que así podemos saber más fácilmente que estrategias deben implementarse para mejorar el rendimiento, ya que sería inadecuado solo reemplazar operadores con mayor cantidad de llamadas al final de cierto periodo pues se estaría incurriendo en una rotación alta de personal, lo cual implicaría un impacto negativo en la empresa.

Si logramos obtener tendencias de comportamiento, permitiremos a los supervisores tomar acciones preventivas para una vez identificados dichos operadores, se les pueda dar seguimiento y analizar otros factores que permitan dar con las causas del bajo rendimiento.

Así pues las decisiones clave a tomar en el proyecto serán identificar que factores influyen en el rendimiento bajo de los operadores, es decir si existe relación entre la ineffectividad como clientes, y fechas de llamadas.

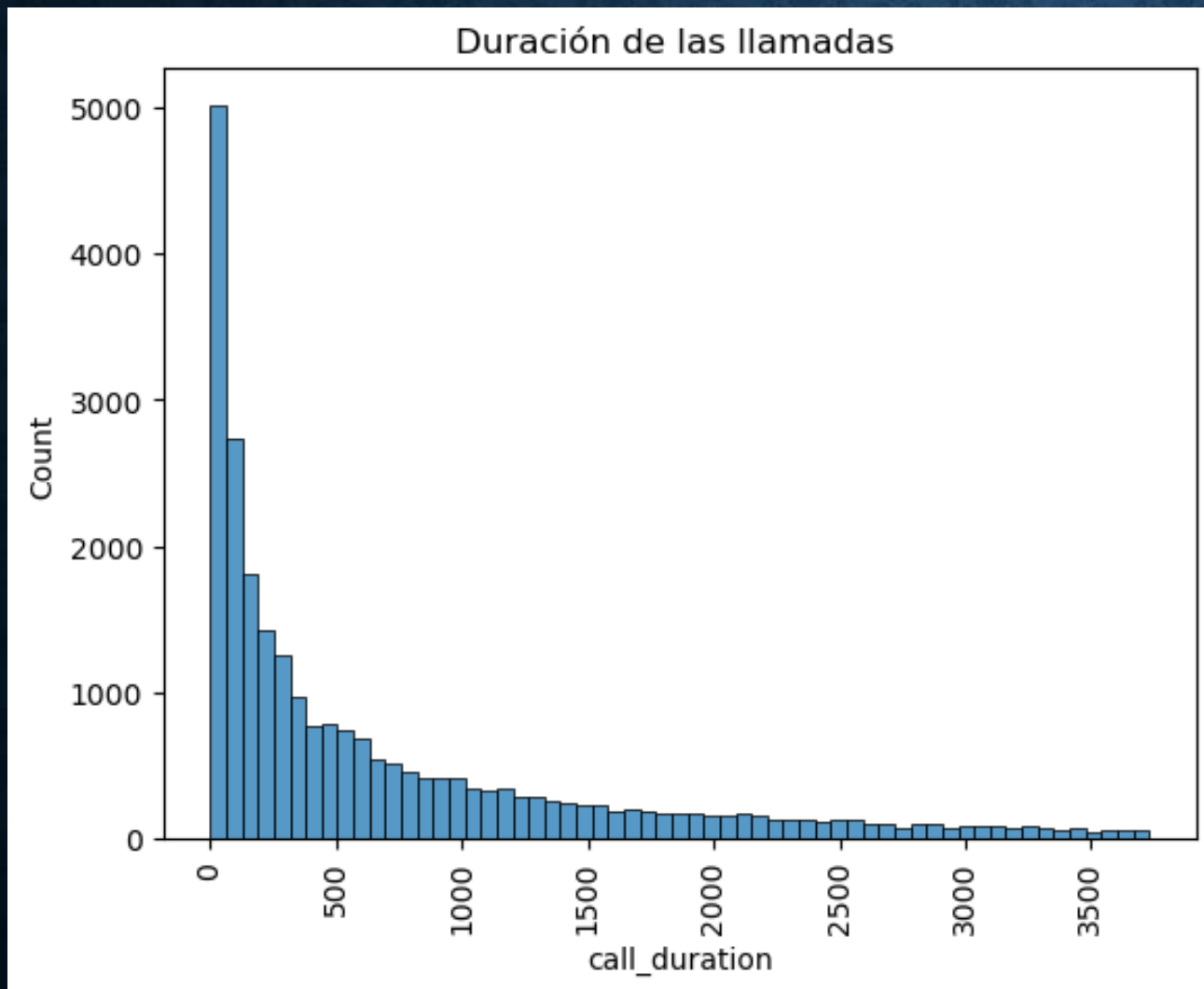
Se interesa saber además si los operadores ineficaces obedecen a estas variables como horarios de las llamadas o atención a ciertos clientes, o bien son un grupo aislado de operadores que no siguen un patrón específico.

# LLAMADAS POR FECHA



Observamos que no hay un patrón claro en el promedio de llamadas realizadas en las fechas, esto podría deberse a la variedad de llamadas realizadas.

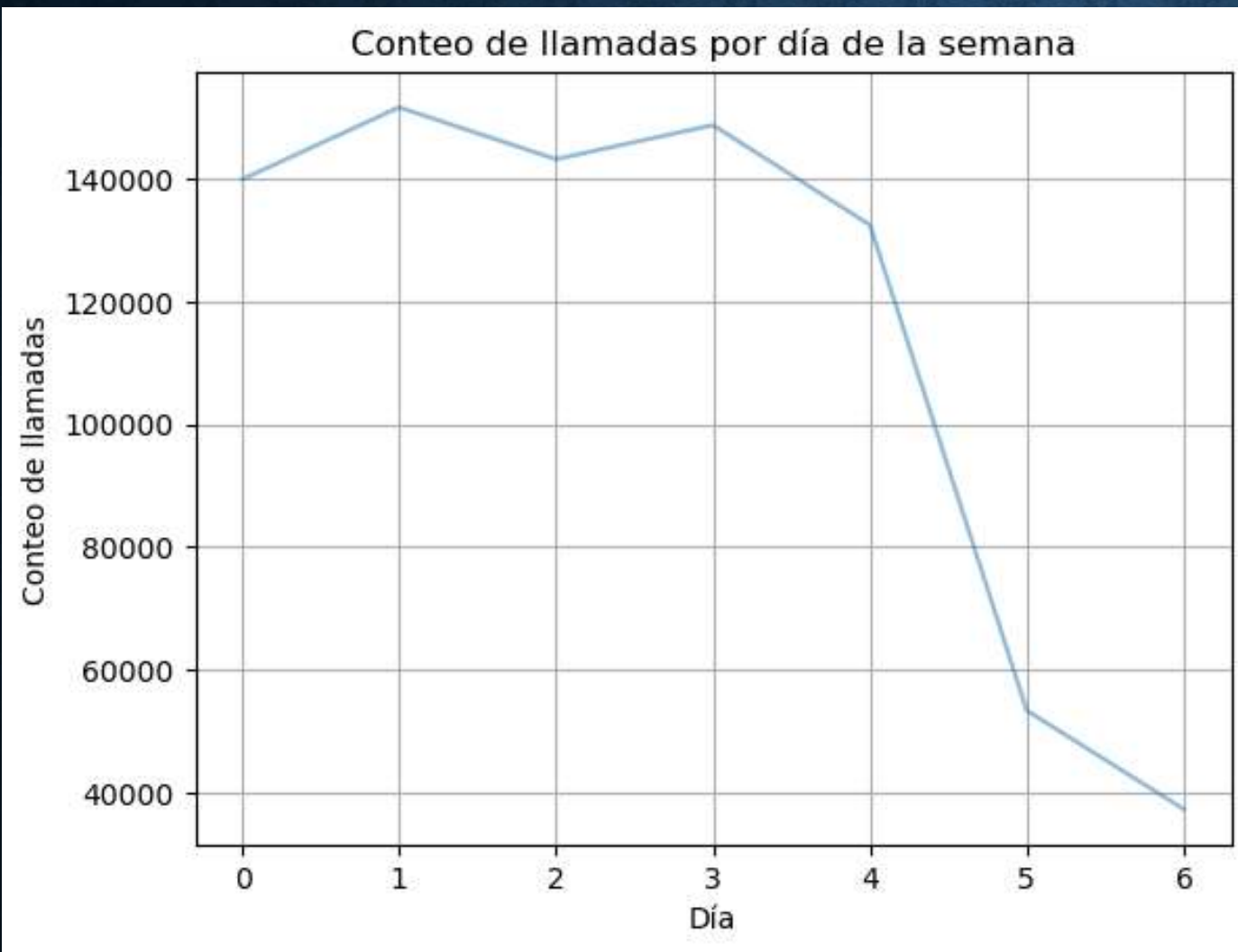
# DURACIÓN DE LLAMADAS



Vemos una distribución sesgada a la derecha lo que indica que la media y moda de valores se encuentran en llamadas menores a 100 segundos, siendo es menos probable hallar llamadas más largas



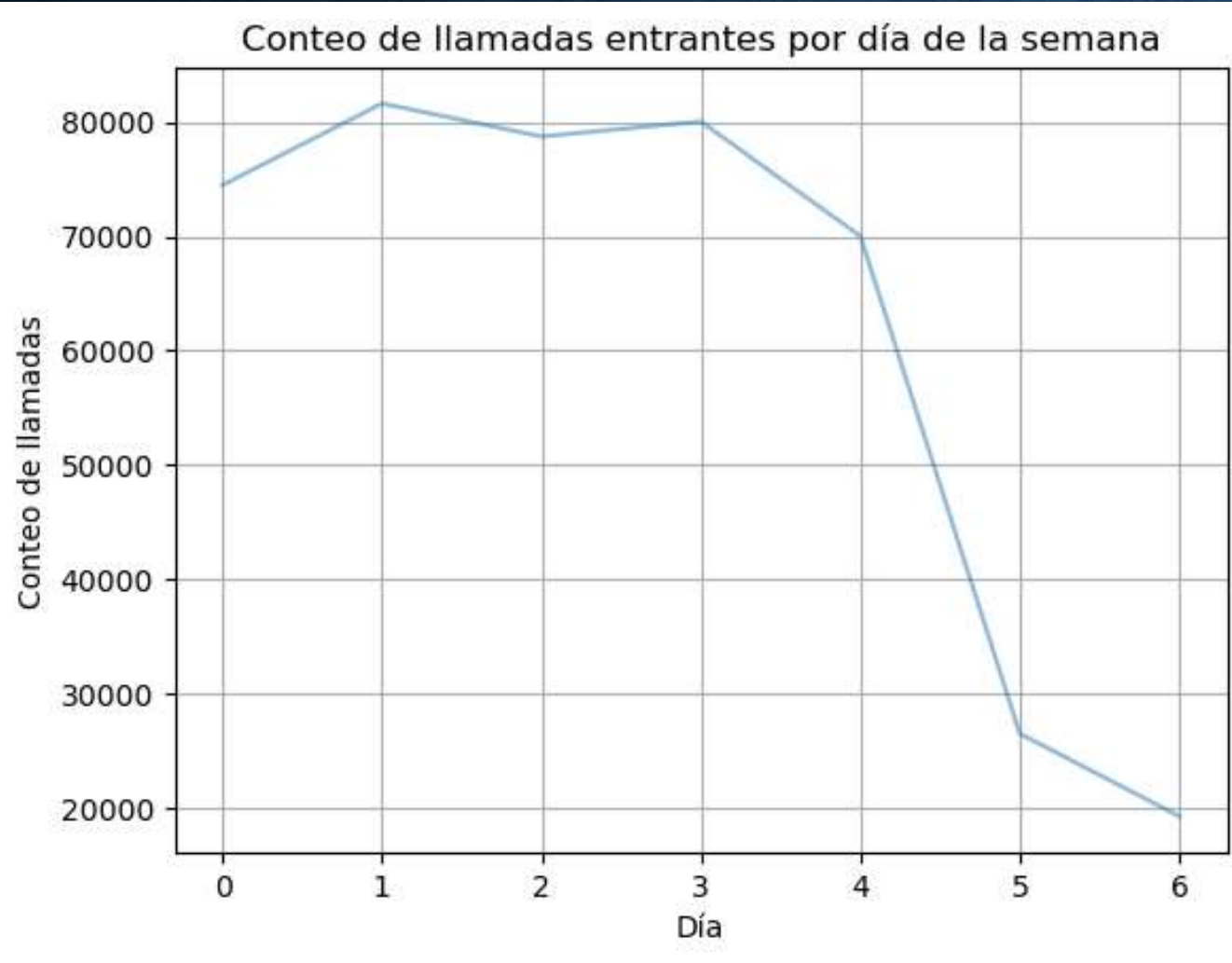
# CONTEO DE LLAMADAS POR DÍA



Observamos que los días con mayor demanda son los primeros días de la semana, bajando significativamente de viernes a domingo, siendo que los picos de llamadas en la gráfica por fechas, en realidad son debido a las cargas de trabajo a lo largo de la semana.

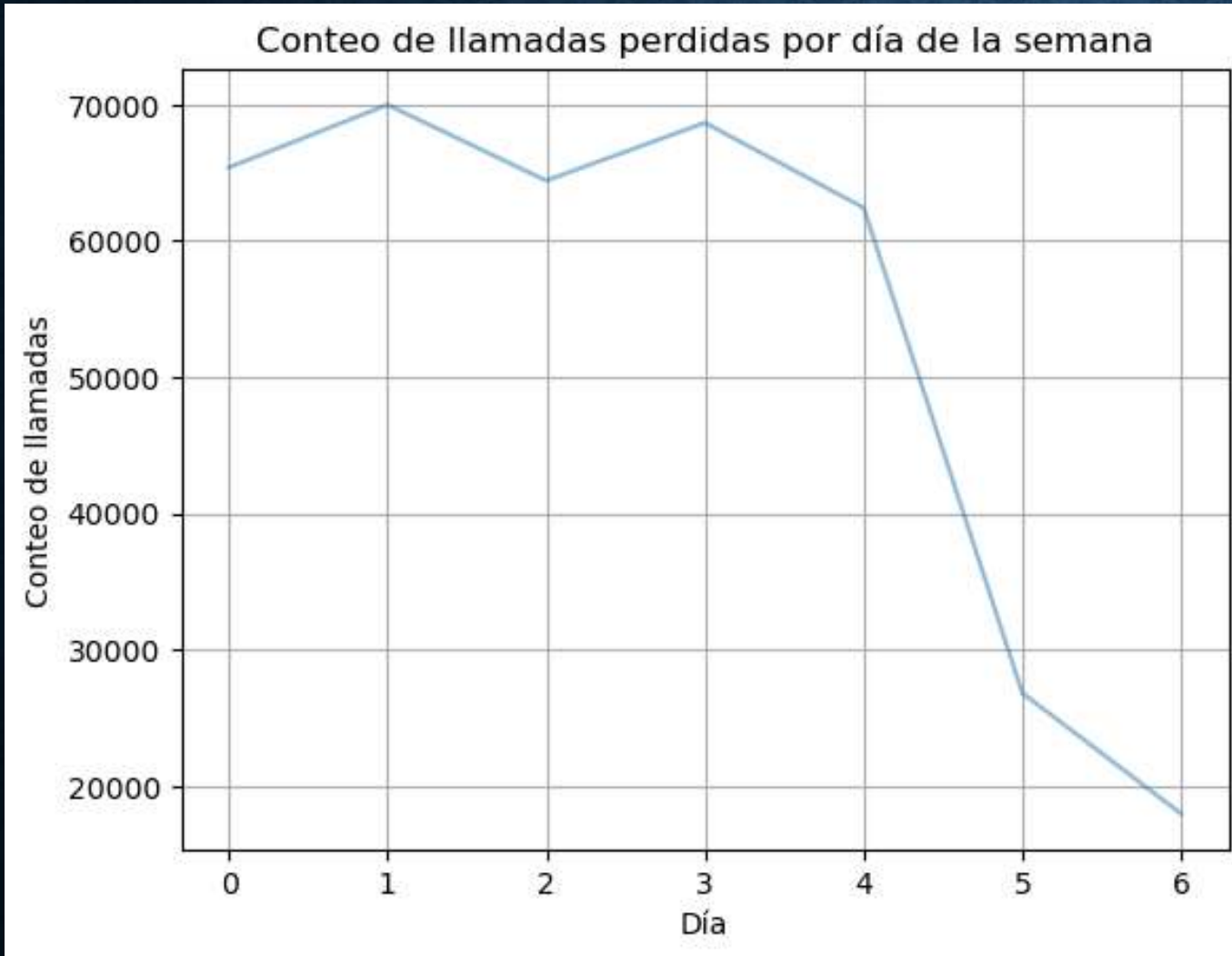


# LLAMADAS ENTRANTES POR DÍA



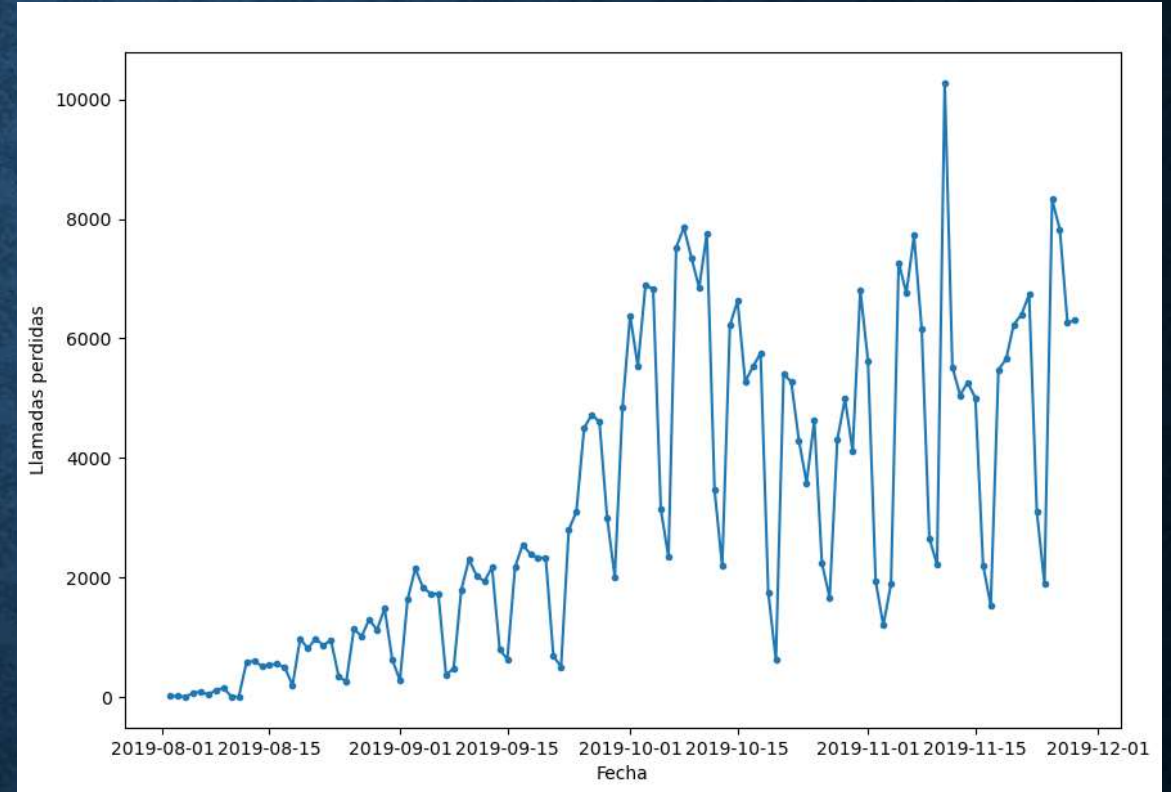
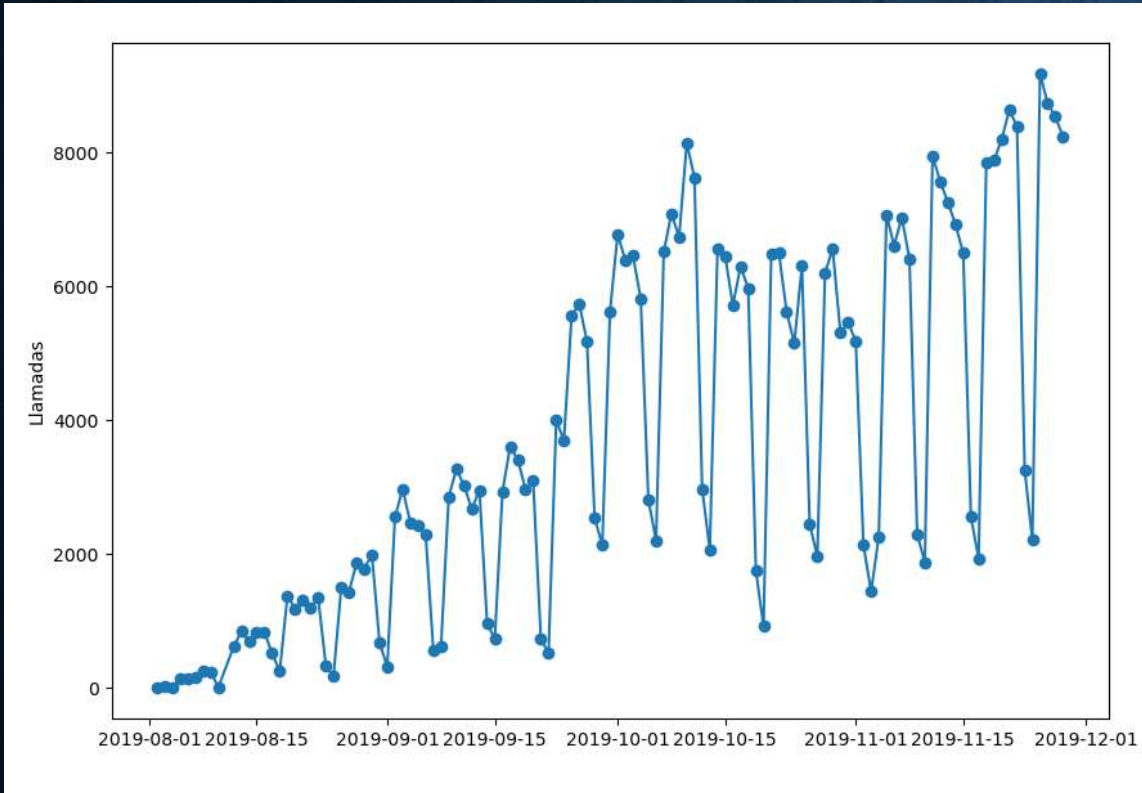
Observamos un patrón similar a la gráficas anterior, con los primeros días de la semana en auge siendo el día con mayor carga de trabajo el martes y bajando significativamente de viernes a domingo

# LLAMADAS PERDIDAS POR DÍA DE LA SEMANA



Observamos un patrón similar a las gráficas anteriores, con primeros días de la semana en auge y bajando significativamente de viernes a domingo

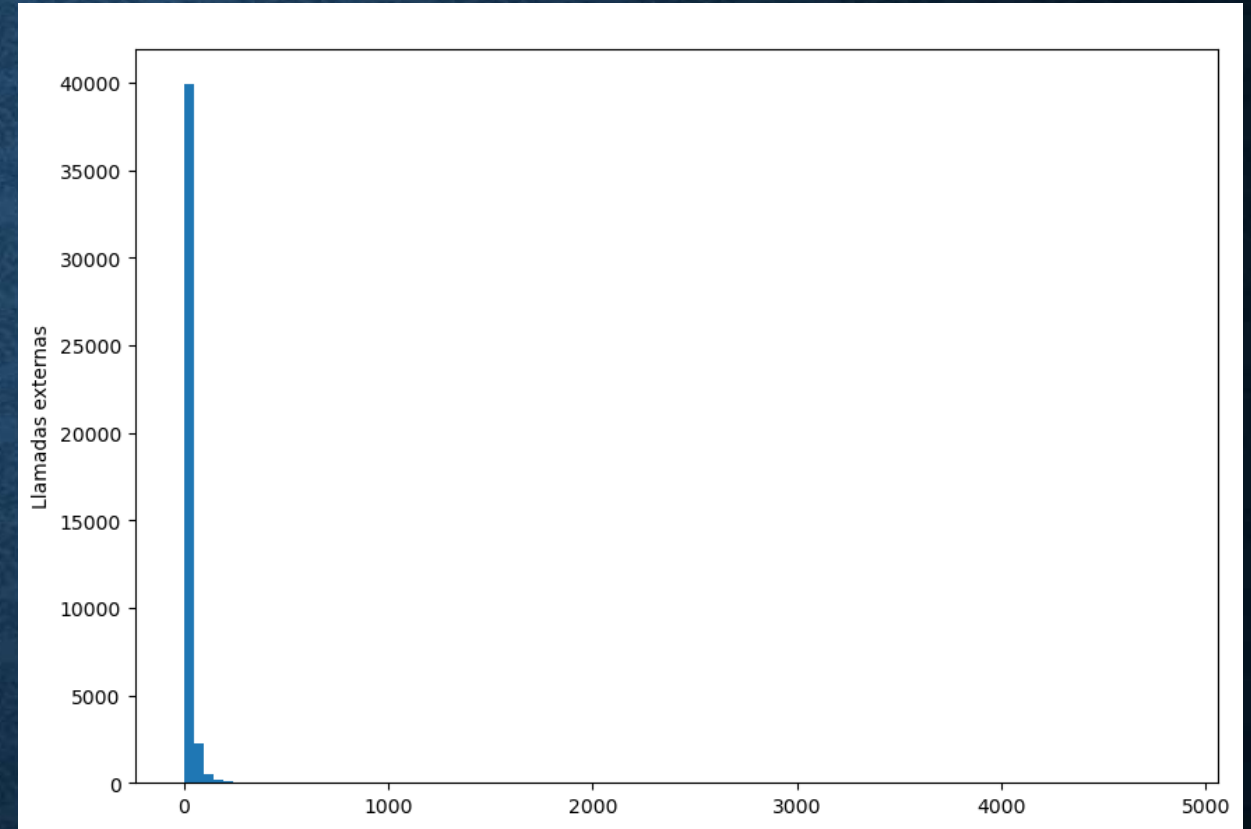
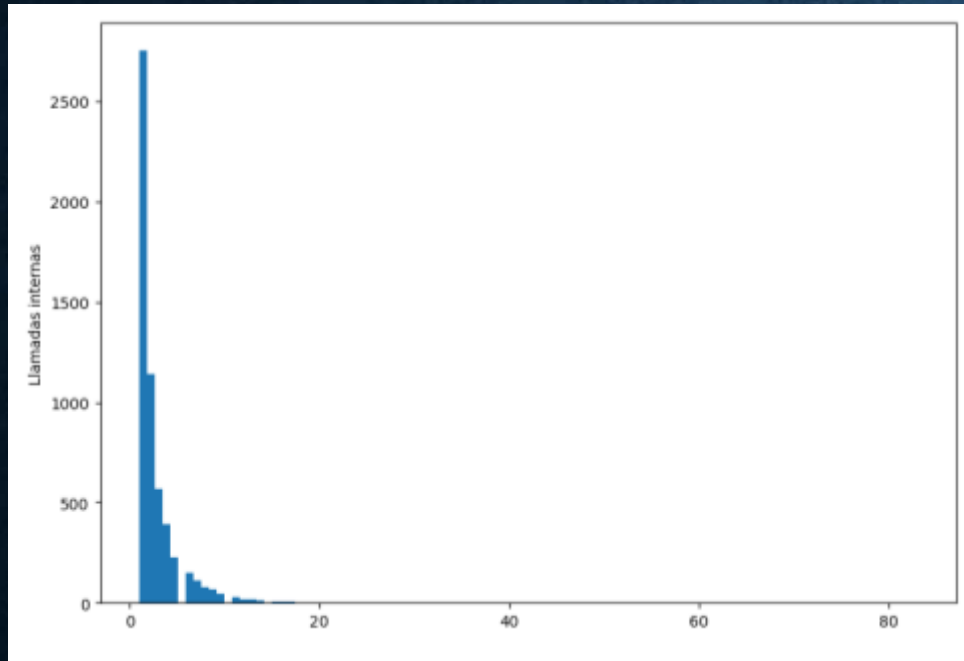
# COMPARACIÓN ENTRE LLAMADAS ENTRANTES Y PERDIDAS



Observamos un patrón similar entre la cantidad y distribución de llamadas perdidas y llamadas entrantes, por lo que es probable que las diferencias se deriven a partir de las fechas como vimos en las gráficas anteriores.

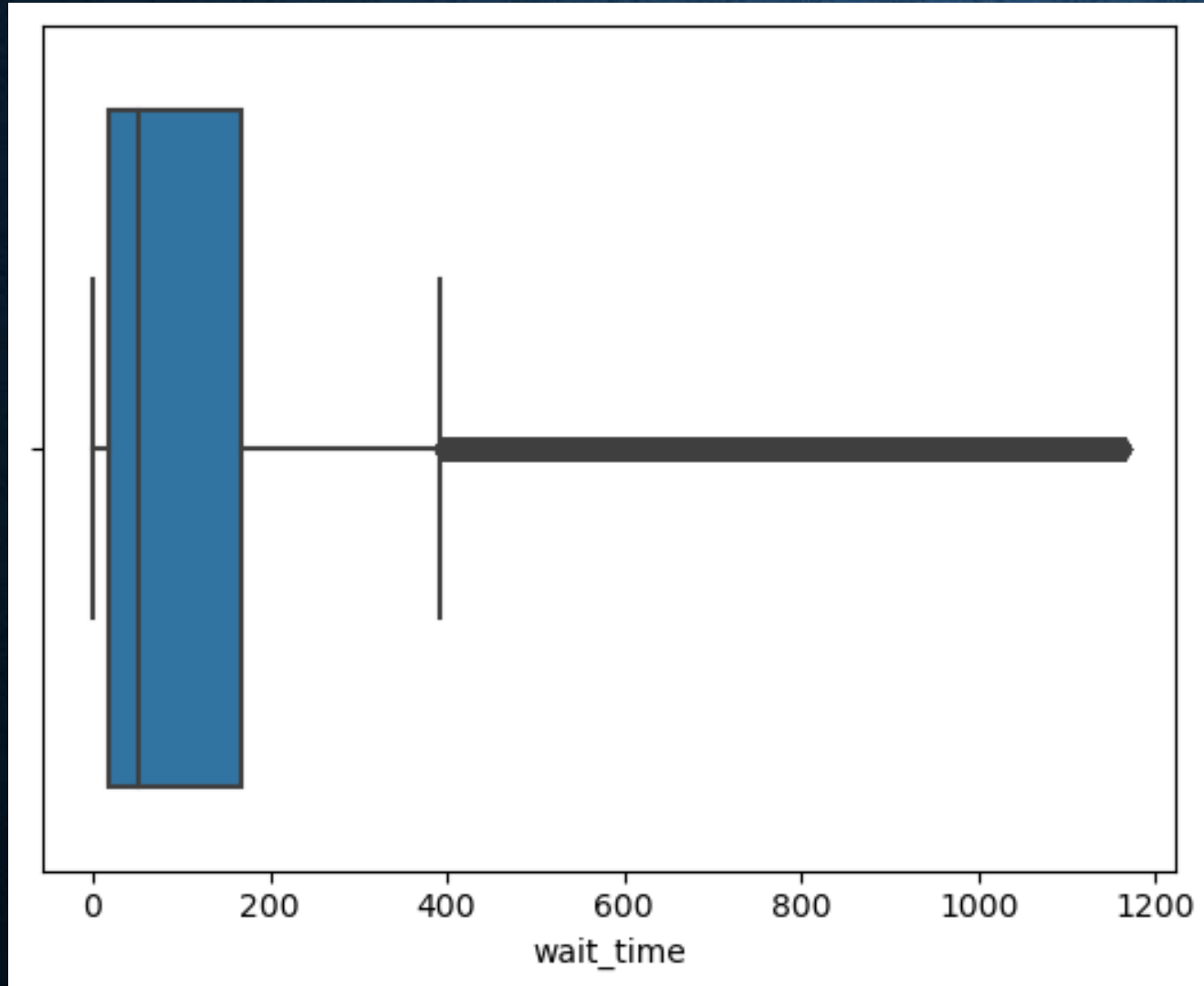


# COMPARACIÓN DE LLAMADAS INTERNAS Y EXTERNAS



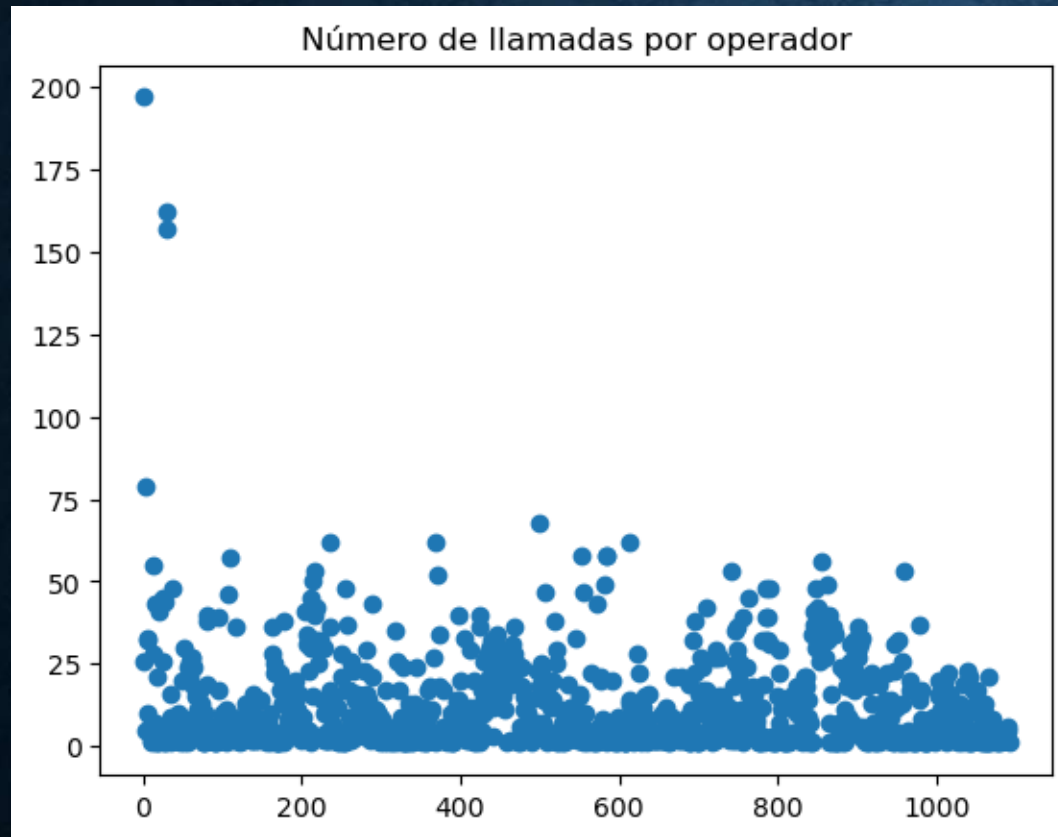
La distribución de ambas llamadas nos indica que son pocas las llamadas tanto internas como externas que superen las 100 llamadas, sin embargo revisando los ejes de las gráficas vemos que hay mayor cantidad de llamadas externas que de llamadas internas

# TIEMPOS DE ESPERA LARGOS (SIN VALORES ATÍPICOS)



Observamos muchos valores por encima de la media de tiempo de espera lo cual podría indicar un alto número de operadores poco eficaces, sin embargo las gráficas de llamadas perdidas no arrojaron tales resultados, por lo que estos valores dependerán de múltiples factores, siendo que se tendrá más claridad tras el análisis estadístico.

# DISTRIBUCIÓN DE LLAMADAS POR OPERADOR



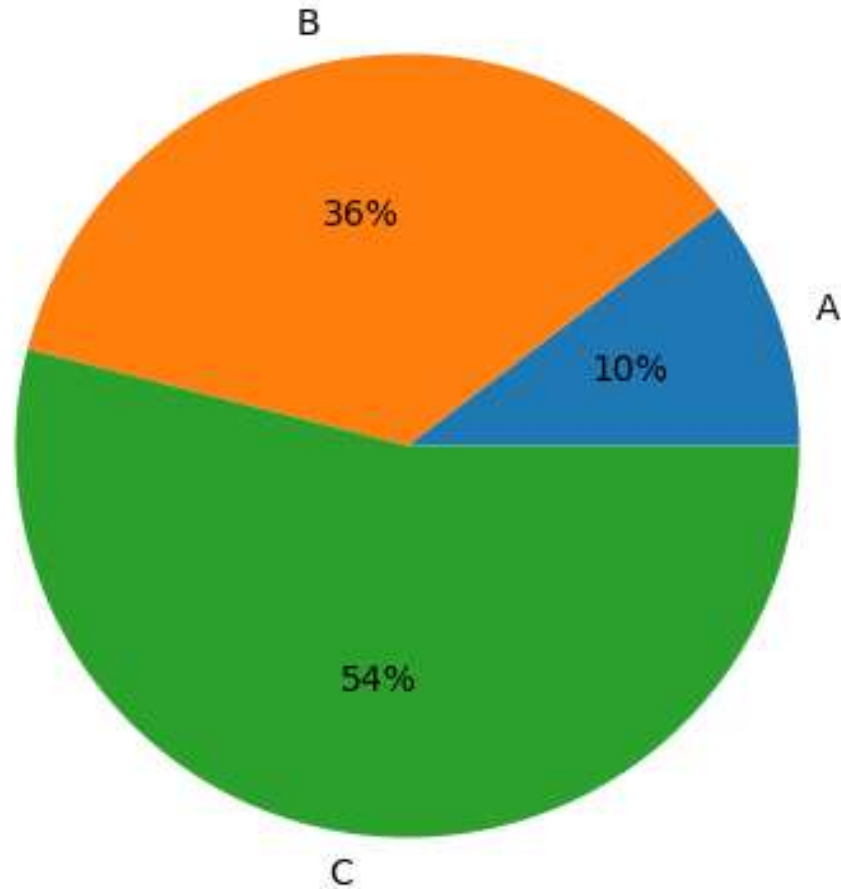
Observamos que una parte importante de los operadores con más llamadas tienen identificación 0, es decir los que no tienen número de identificación.

Por otro lado hay usuarios que tienen tanto llamadas perdidas como llamadas entrantes exitosas, por lo que tampoco hay un patrón en temas de llamadas perdidas, siendo que los operadores con más llamadas perdidas son aquellos con más carga de trabajo, como observamos en la gráfica general donde la mayoría de operadores no supera las 50 llamadas.



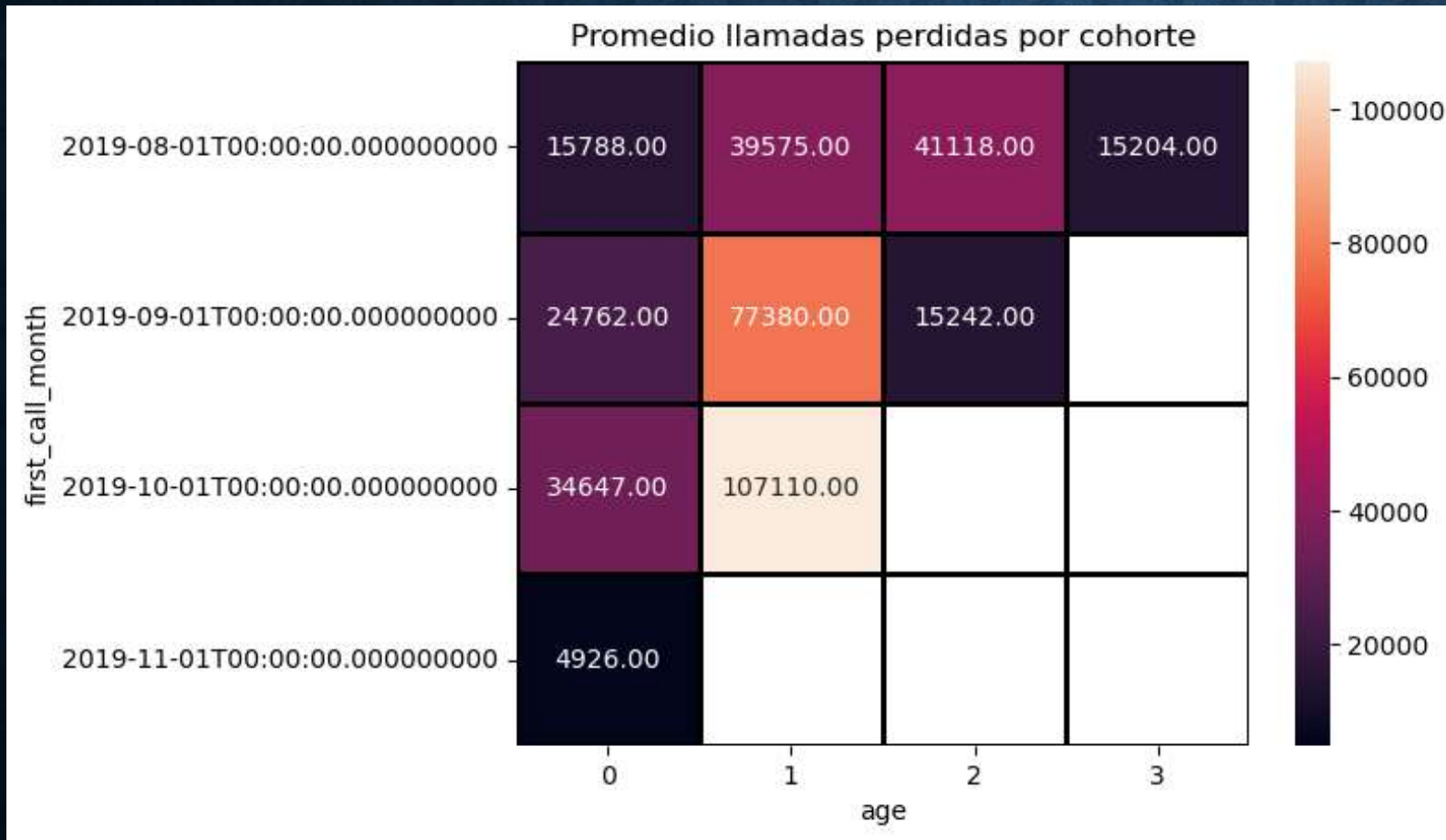
# DISTRIBUCIÓN DE TIPOS DE CLIENTE

Distribución de planes de tarifa de clientes



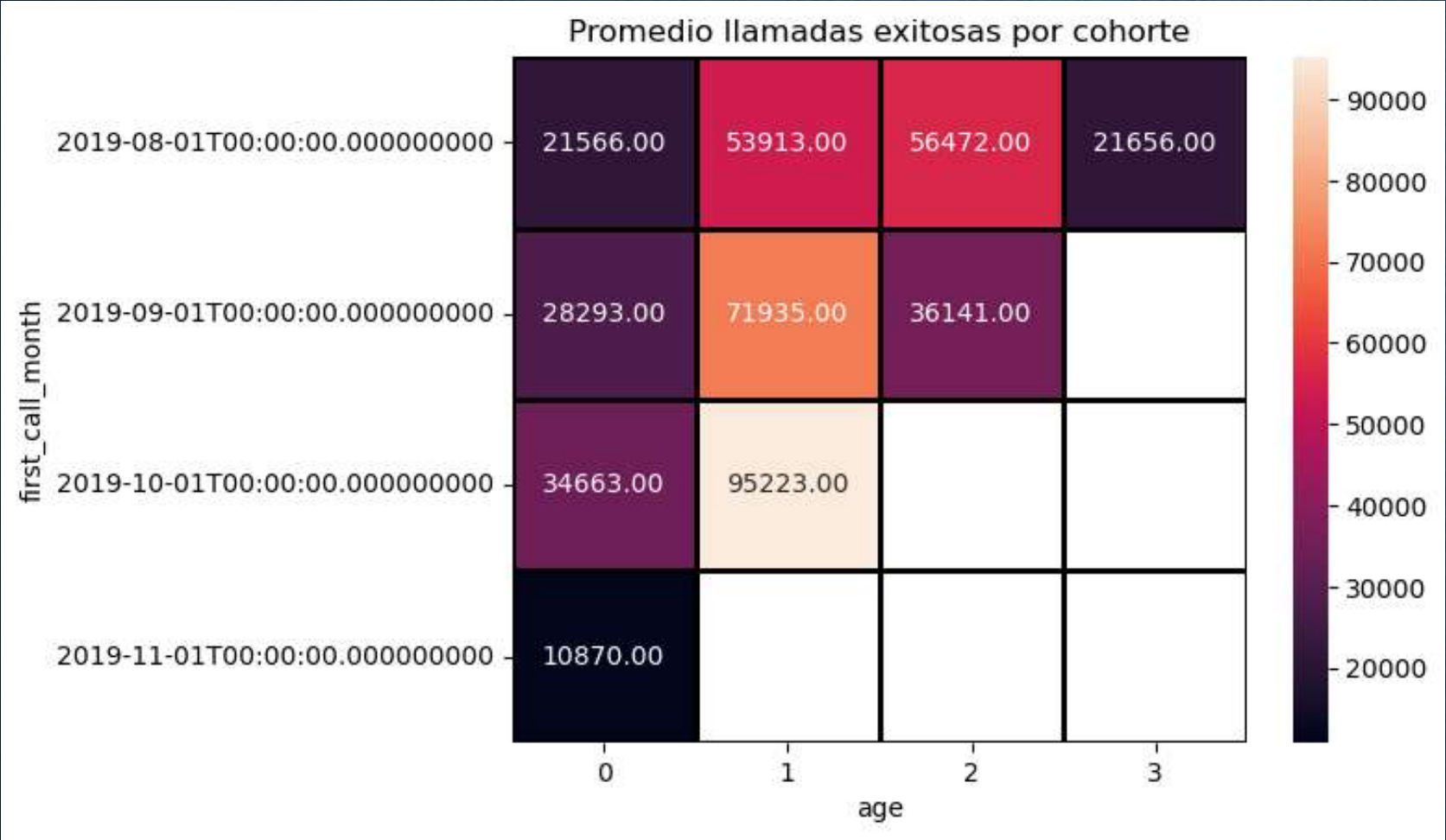
Observamos que la mayoría de clientes se encuentran dentro del grupo C, es decir el más bajo seguidos por el grupo B y A. Así mismo el grupo C contiene la mayor cantidad de llamadas entrantes y perdidas, por la proporción de datos que se maneja.

# LLAMADAS PERDIDAS POR COHORTE DE CLIENTES



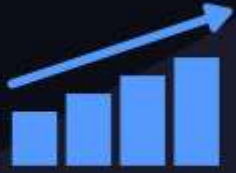
Observamos que el cohorte de octubre es el que tiene mayor cantidad de llamadas tanto perdidas, seguido del cohorte de septiembre, así que la antigüedad de los cohortes no influye necesariamente en la cantidad de llamadas o eficiencia de los trabajadores

# LLAMADAS EXITOSAS POR COHORTE DE CLIENTES

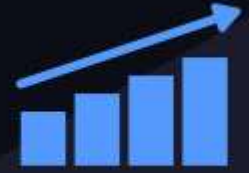


Observamos que el cohorte de octubre es el que tiene mayor cantidad de llamadas exitosas, seguido del cohorte de septiembre, confirmando que no habría razones para pensar en una relación directa entre la eficiencia y la antigüedad del cliente.





# PRUEBA DE HIPÓTESIS



Al realizar la prueba de hipótesis, que hay operadores que tenían tiempos de espera largos o llamadas perdidas y aquellos que cumplen con al menos una característica tienen tiempos de llamada promedio más cortos, aunque se vio que no había operadores con ambas características, (tiempos de espera largos y varias llamadas perdidas).

No podemos decir que hay correlación directa entre el tiempo de las llamadas o llamadas perdidas, pero sí es un parámetro a considerar para obtener mejores resultados, por lo al final de los periodos dados por la dirección o supervisores podremos dar un parámetro para identificar operadores poco eficaces y estimar que tendrán menor rendimiento en cuanto a duración de llamadas.

Para mejorar la información obtenida se crea un algoritmo de datos para intentar predecir las llamadas perdidas, así como la agrupación por clústeres con variables como operadores, tiempos de espera largos y varias llamadas perdidas

# SELECCIÓN ALGORITMO PARA LLAMADAS PERDIDAS

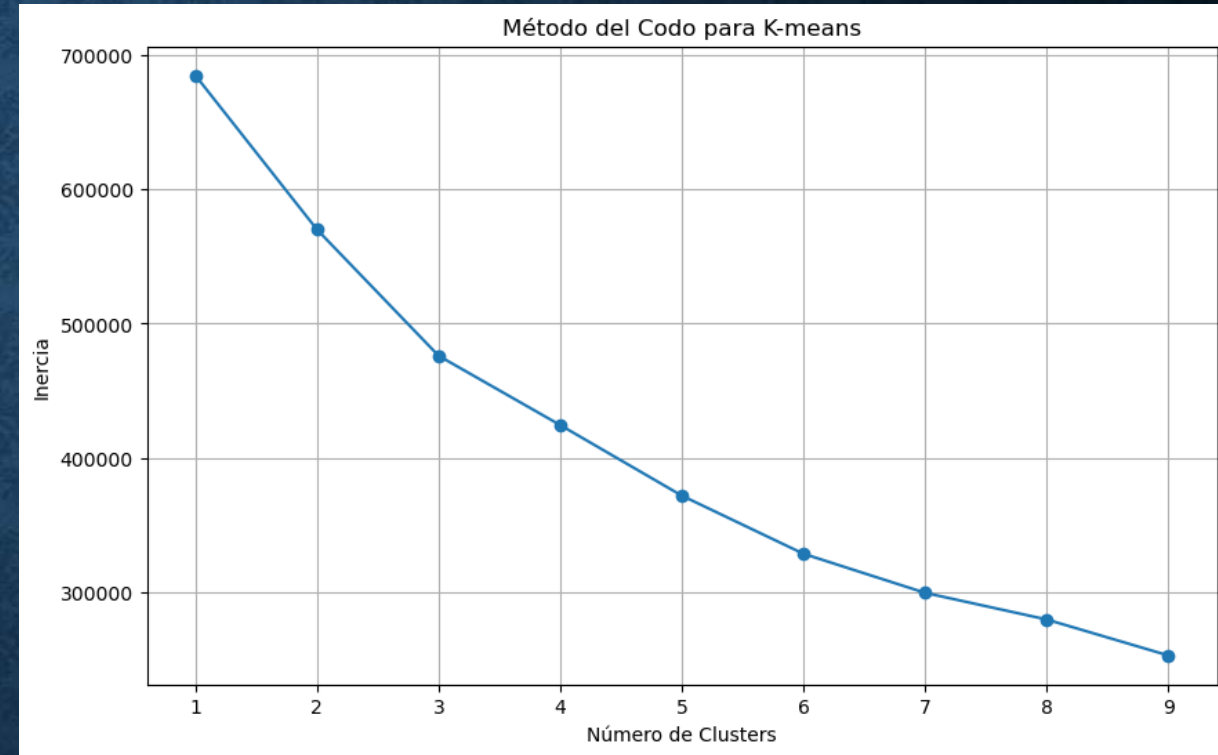
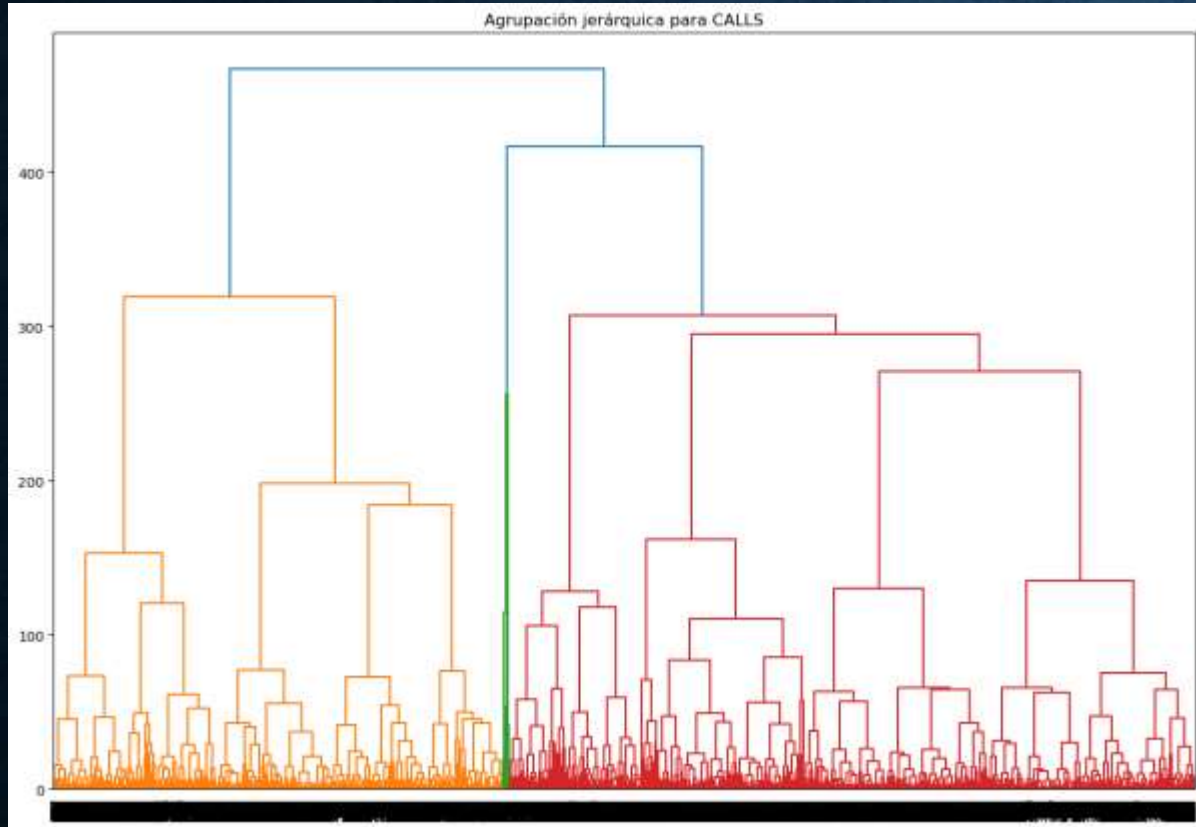
- `LogisticRegression()`
- Accuracy: 0.95
- Precision: 0.91
- Recall: 0.97

- 
- `RandomForestClassifier()`
  - Accuracy: 0.99
  - Precision: 1.00
  - Recall: 0.99



Al observar los valores de exactitud (accuracy), precisión (precision) y sensibilidad (recall) observamos que el mejor modelo es `RandomForestClassifier()`, al obtener una precisión de 100% y exactitud y sensibilidad de 99%, así mismo se recomienda revisar este modelo y su mejor implementación en un equipo especializado en ciencia de datos.

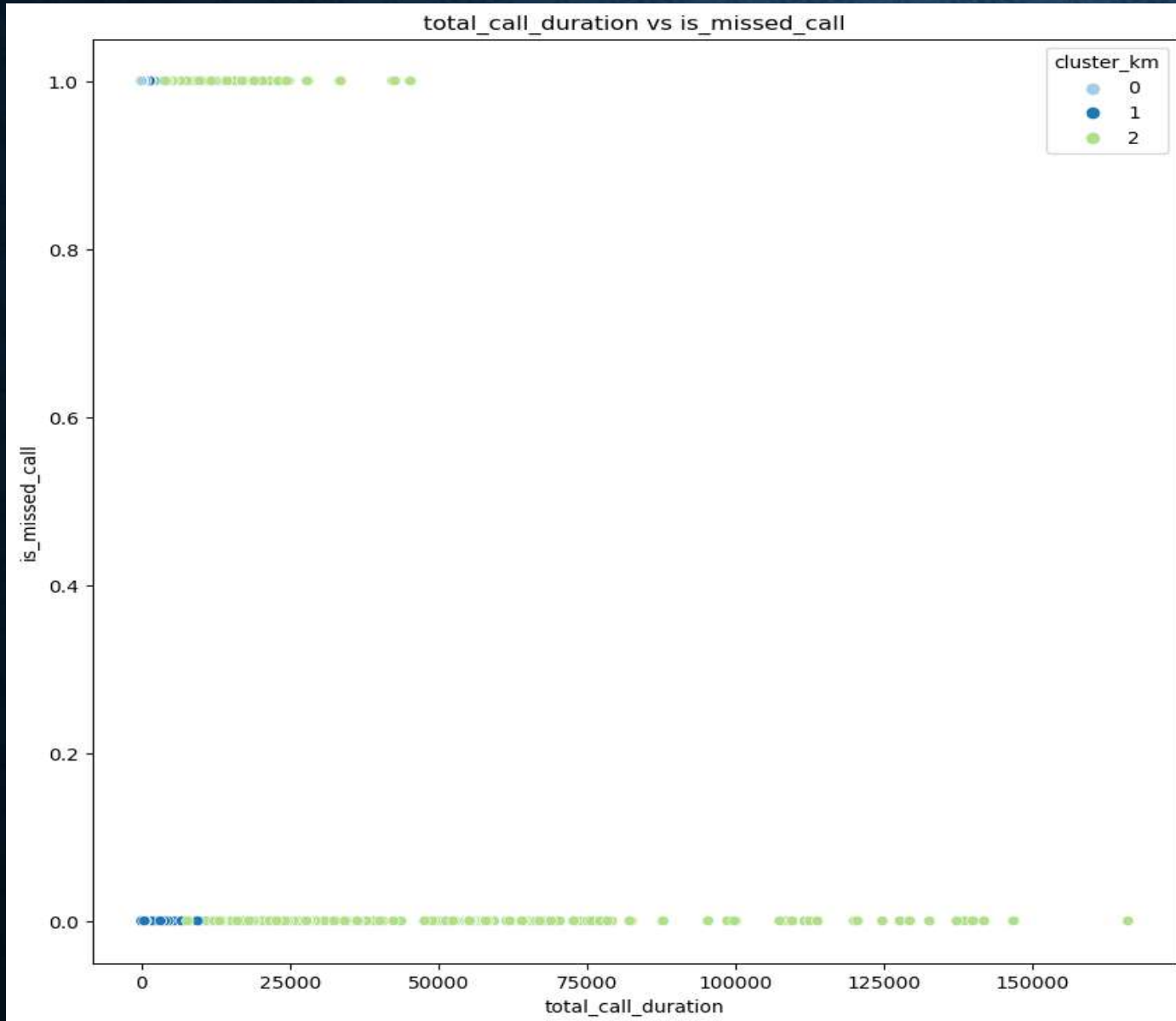
# ANÁLISIS DE CLÚSTERES



Observamos que tanto en el método del codo como en el dendograma el número ideal de clústeres es tres

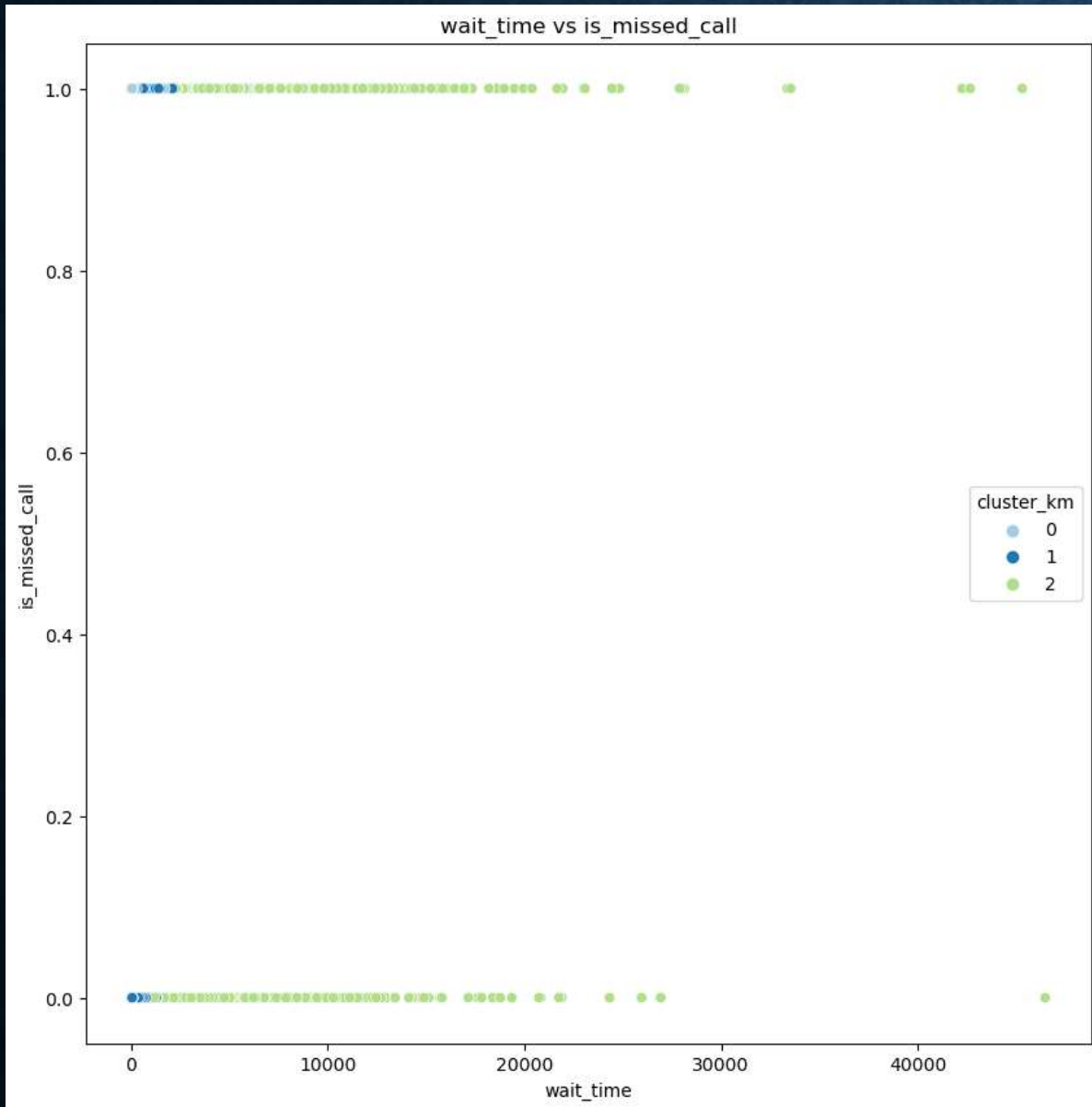


# RELACIÓN LLAMADAS PERDIDAS Y DURACIÓN



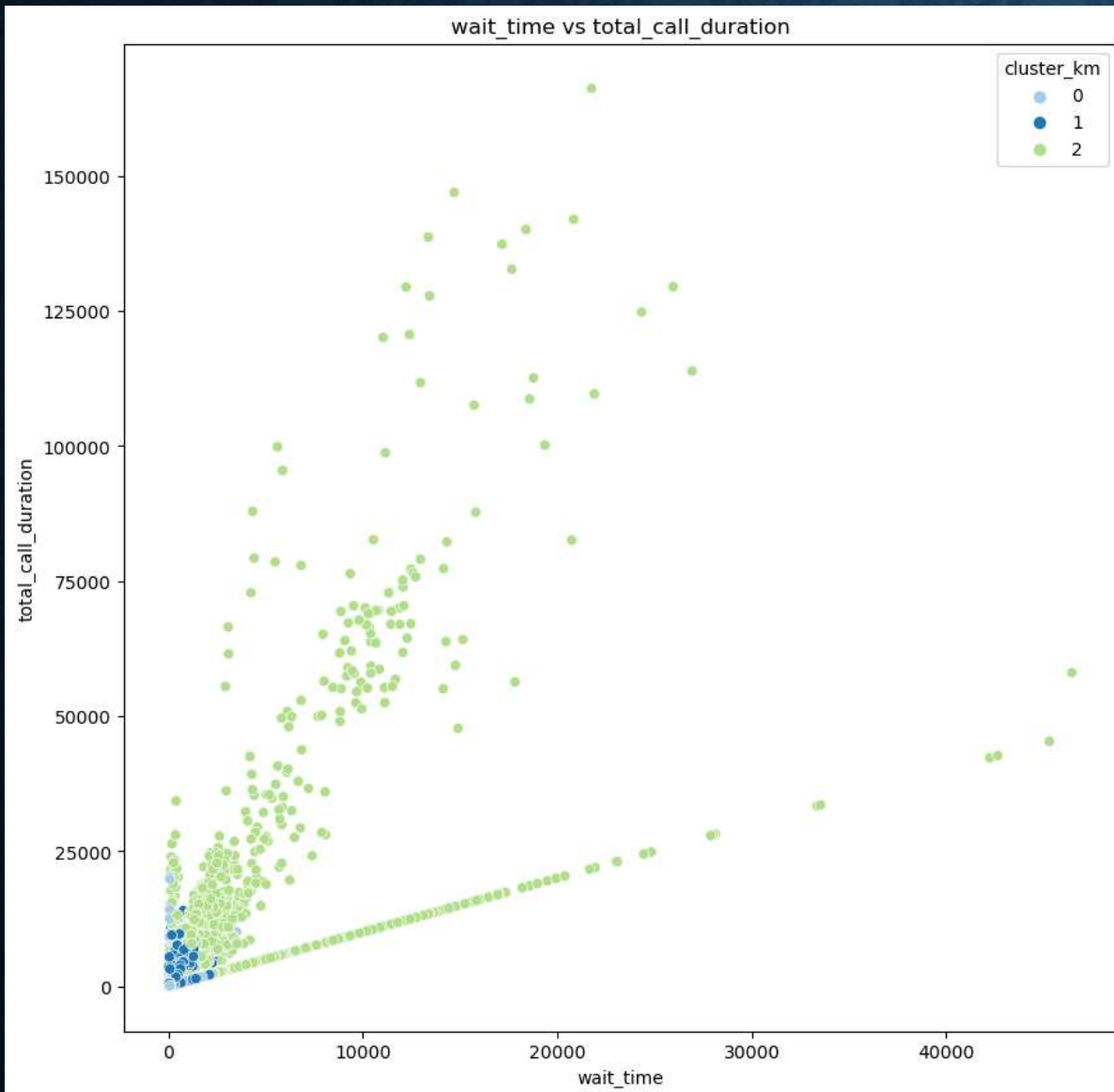
Observamos que la duración de llamadas para los clústeres que tienen mayor cantidad de llamadas perdidas también tienen la mayor duración de llamadas

# RELACIÓN LLAMADAS PERDIDAS Y TIEMPO DE ESPERA



Observamos que los clústeres se agrupan con tiempos de espera más cortos tienden a obtener menor cantidad de llamadas perdidas

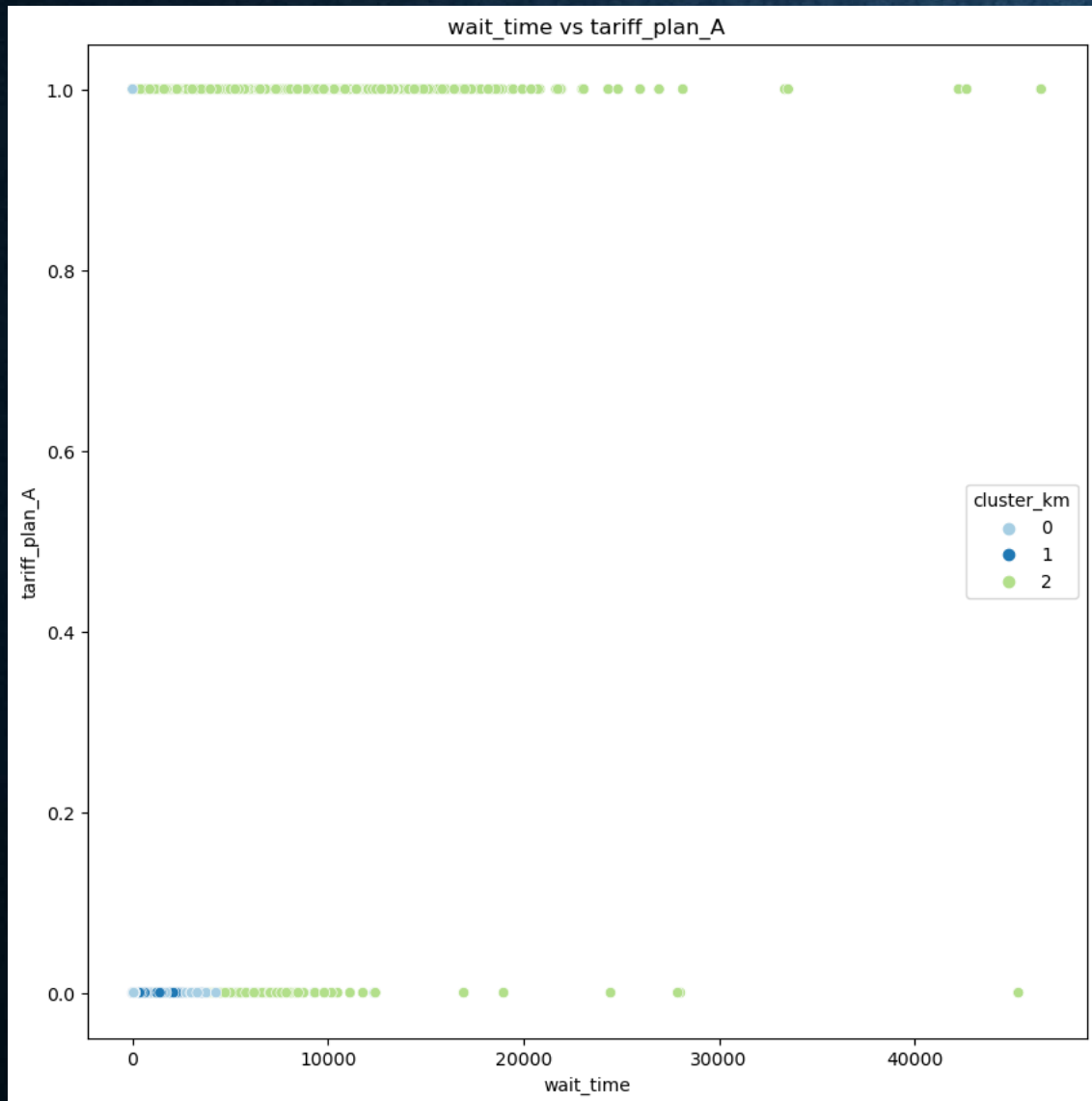
# RELACIÓN TIEMPO DE ESPERA Y DURACIÓN



Igualmente podemos ver que los clústeres con mayor cantidad de duración de llamadas son aquellos con el tiempo de espera más largo

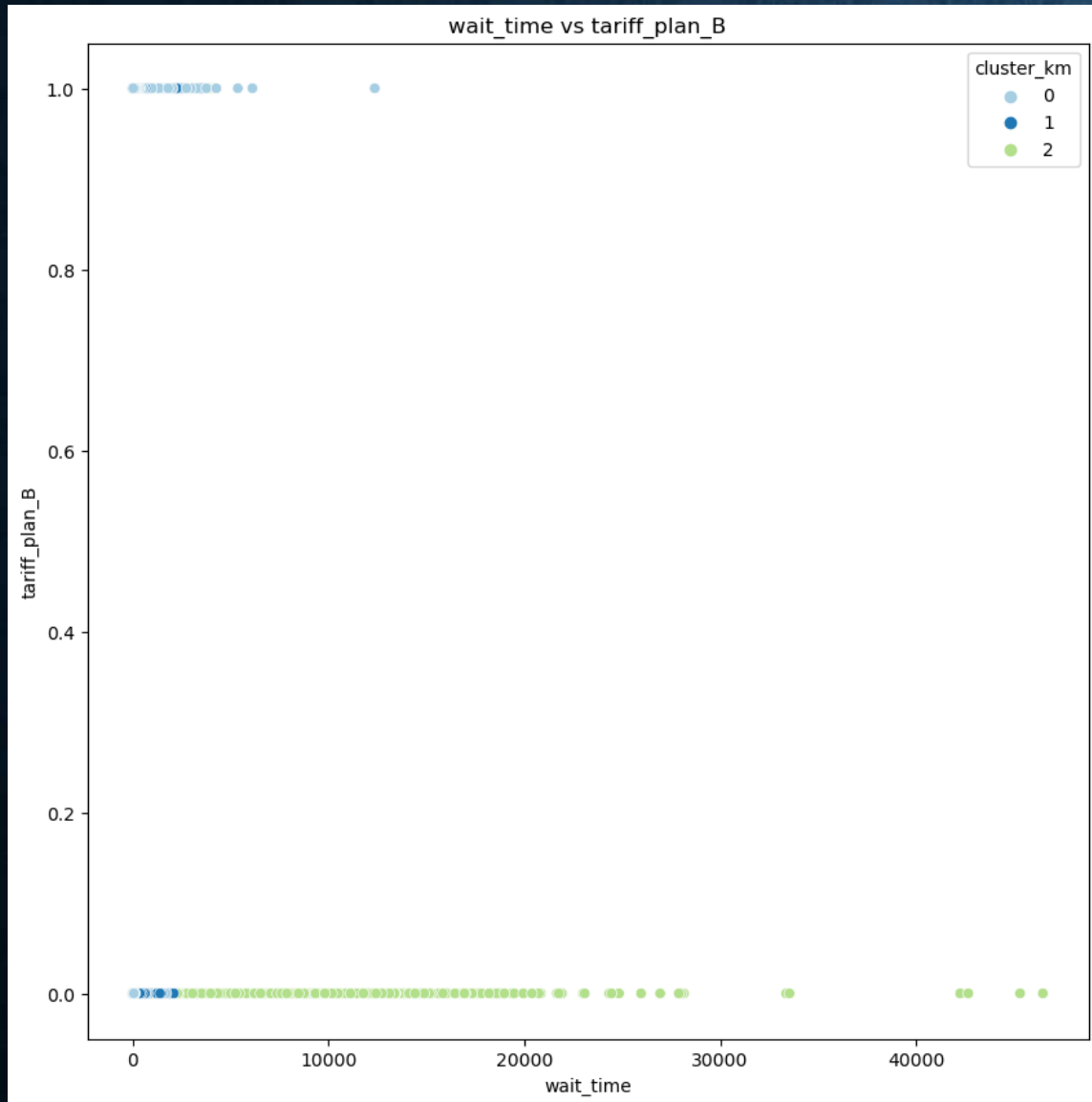


# RELACIÓN TIEMPO DE ESPERA Y CLIENTES A



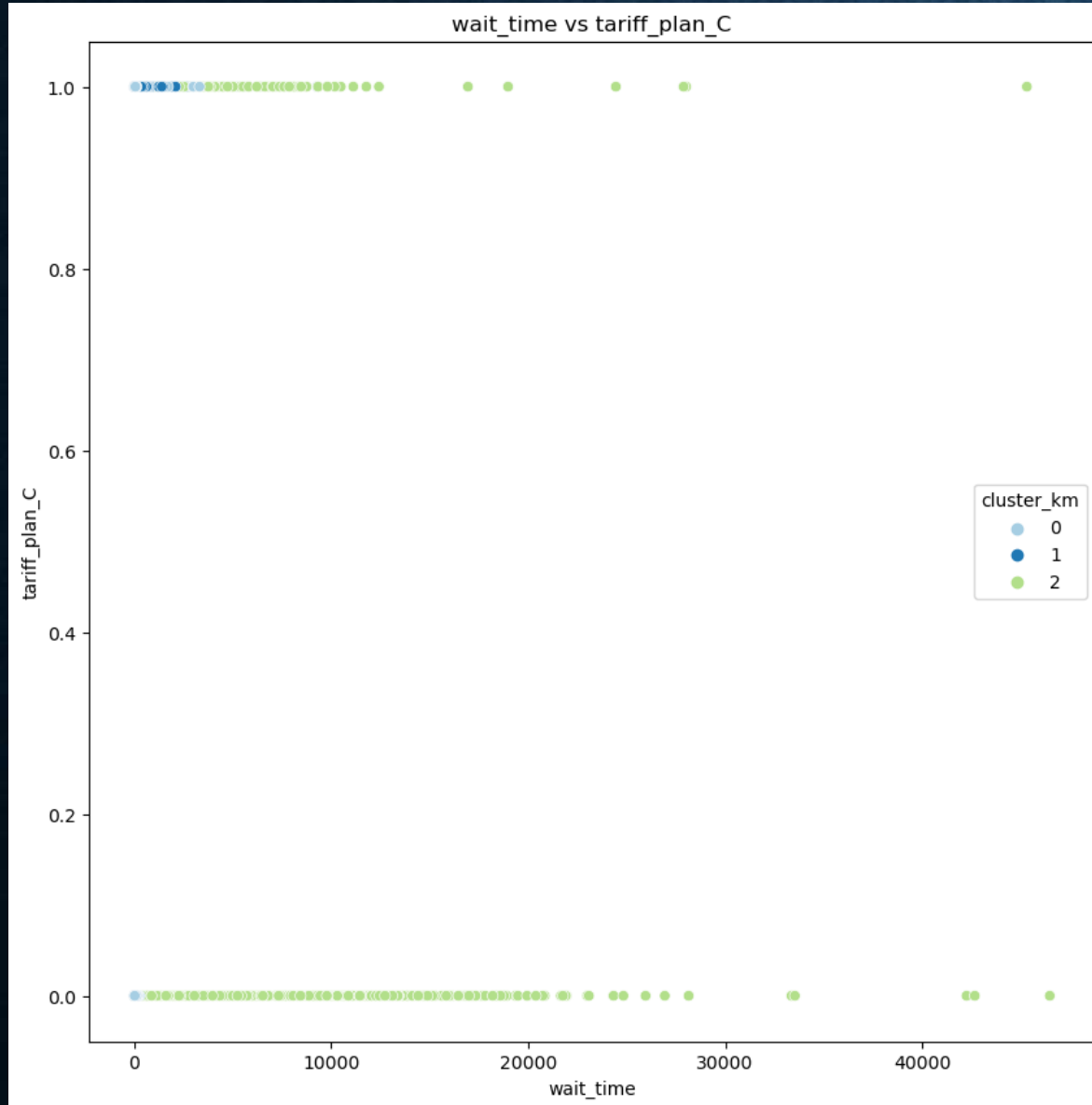
Observamos que el mayor tiempo de espera pertenece al plan A

# RELACIÓN TIEMPO DE ESPERA Y CLIENTES B



El plan B tiene  
el menor tiempo  
de espera

# RELACIÓN TIEMPO DE ESPERA Y CLIENTES C



Observamos que el plan C tiene los tiempos de espera más variados



# 📞 CONCLUSIONES 📞

Se identificaron dos directrices donde la primera fue hallar operadores con varias llamadas perdidas o bien operadores con tiempos de espera muy largos.

Primero se realizó un análisis exploratorio donde hallamos que no existía una relación o patrón directo aparente para dictar la ineficiencia de los operadores, siendo que una de las hipótesis era la fecha se observó que esta no influía, sino que las métricas de ineficiencia eran más altas entre los primeros días de la semana por la carga de trabajo.

Más adelantes se tenía la hipótesis que los tipos de clientes influían en el número de llamadas perdidas o tiempos de espera largos, sin embargo vimos que no había una relación evidente entre la categoría de los clientes o el número de llamadas perdidas, sin embargo para confirmar esta distribución se generó un análisis de clústeres para identificar mejor las distribuciones.

Para el análisis estadístico se realizaron pruebas de hipótesis dando como resultado que las diferencias entre la duración de llamadas entre los usuarios eficaces e ineficaces si era significativa.

Finalmente se obtuvieron los clústeres para identificar las características de operadores ineficaces, obteniendo como resultados 3 clústeres obteniendo que las llamadas con mayor duración pertenecían los operadores más ineficaces, es decir con mayor tiempo de espera y número de llamadas perdidas

# ☎️ RECOMENDACIONES ☎️

- Observamos que los operadores más ineficaces son aquellos con mayor carga de trabajo como se mostró en análisis exploratorio y de clústeres, por lo que se recomienda generar estrategias para lograr repartir de mejor manera la carga de trabajo de los operadores.
- Realizar futuros análisis para revisar el perfil de operadores, como por ejemplo el tiempo que llevan dentro de la empresa.
- Mejorar los procesos de capacitación de los colaboradores para una mejora continua.
- Complementar este análisis con equipos especializados en ciencia de datos que permitan mejorar e implementar modelos de predicción de llamadas perdidas así como segmentar mejor los clientes o tipos de llamadas dando otros parámetros para generar mejoras o incluso la probabilidad de un operador de volverse ineficaz en función de las llamadas y los clientes que atiende para mejorar las estrategias de capacitación de los mismos.
- Para esto se recomienda implementar los siguiente en futuras capacitaciones:
  1. Flujograma de atención al cliente
  2. Inteligencia emocional
  3. Estrategias de comunicación no violenta
  4. Protocolos de etiqueta telefónica
  5. Técnicas de persuasión
  6. Habilidades de negociación
  7. Uso de herramientas y sistemas



# GRACIAS

