



# Resultado proyecto servicio de telefonía virtual CallMeMaybe

# INTRODUCCIÓN Y OBJETIVOS

El servicio de telefonía virtual CallMeMaybe está desarrollando una nueva función que brindará a los supervisores y las supervisoras información sobre los operadores menos eficaces. 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. Además, si se supone que un operador debe realizar llamadas salientes, un número reducido de ellas también será un signo de ineficacia.

- ▶ Lleva a cabo el análisis exploratorio de datos
- ▶ Identificar operadores ineficaces
- ▶ Prueba las hipótesis estadísticas

# DESCRIPCIÓN DE LOS DATOS

telecom\_dataset\_us.csv contiene las siguientes columnas:

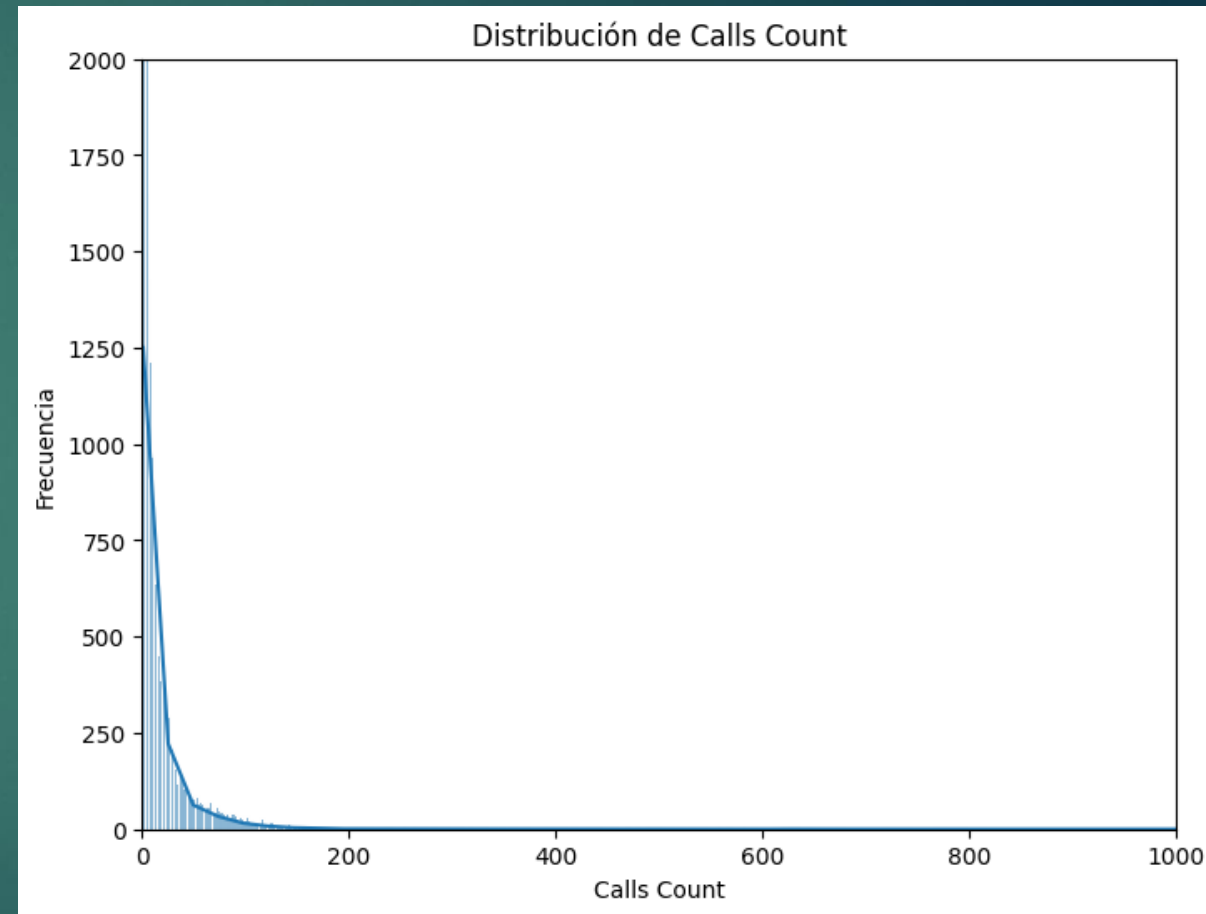
- ▶ user\_id: ID de la cuenta de cliente
- ▶ date: fecha en la que se recuperaron las estadísticas
- ▶ direction: "dirección" de llamada (out para saliente, in para entrante)
- ▶ internal: si la llamada fue interna (entre los operadores de un cliente o clienta)
- ▶ operator\_id: identificador del operador
- ▶ 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)

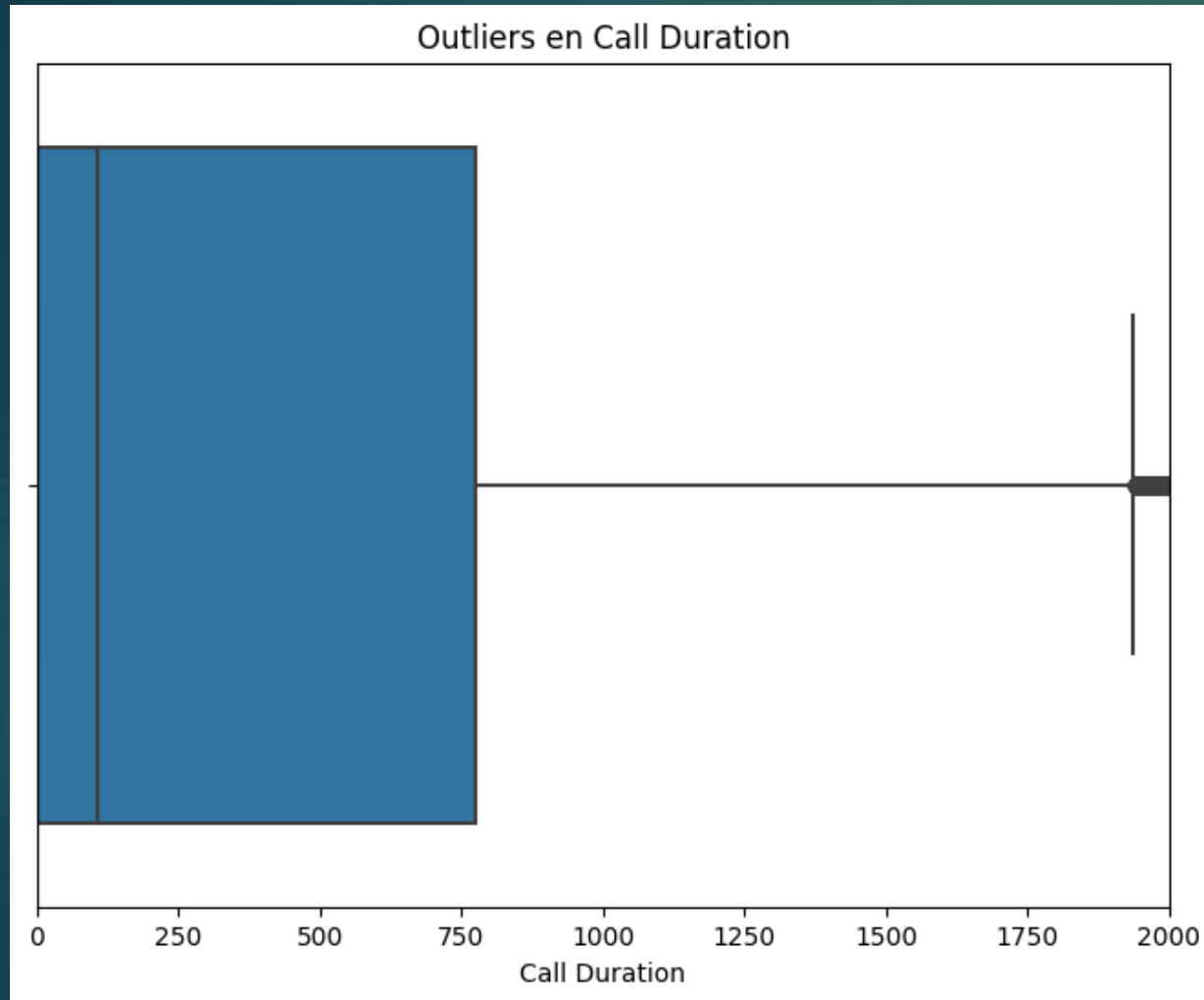
telecom\_clients\_us.csv tiene las siguientes columnas:

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

# ANÁLISIS EXPLORATORIO DE DATOS

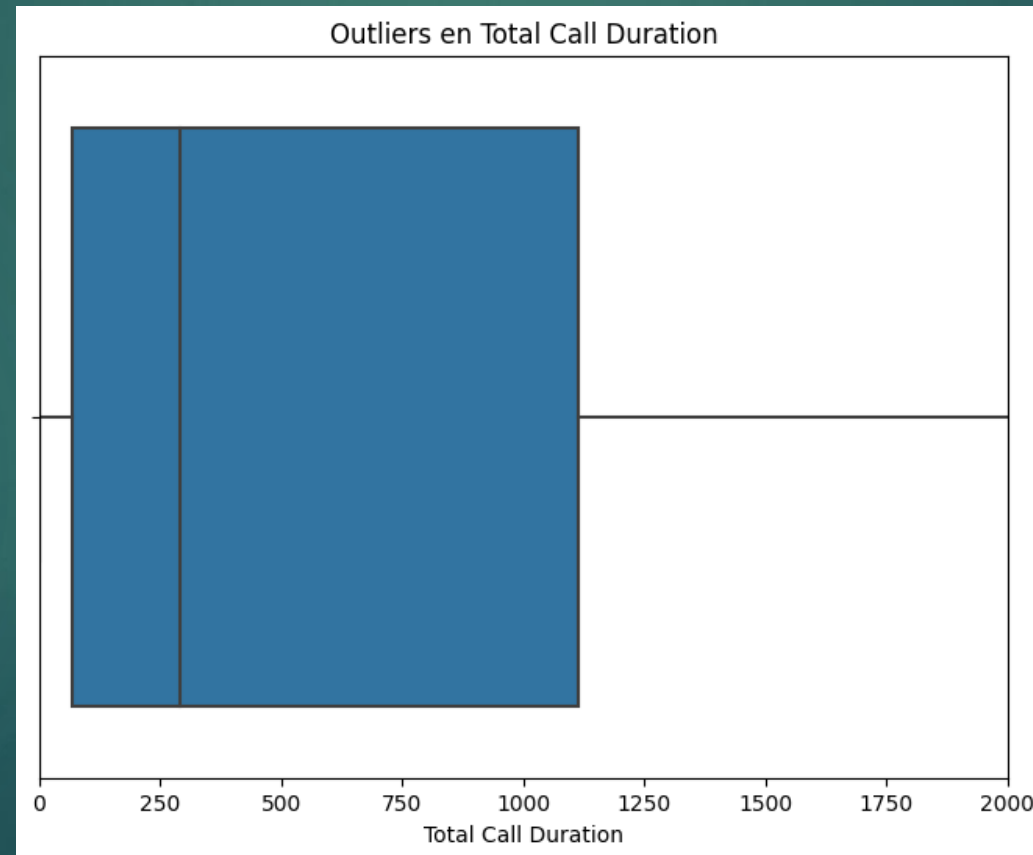
- Podemos observar que el gráfico tiene un sesgo de datos hacia la derecha, lo que nos indica que hay mayor cantidad de datos atípicos partiendo aproximadamente de las 100 o 150 de llamadas.
- Concluimos también que tenemos un sesgo de frecuencia de llamadas después de 1250.





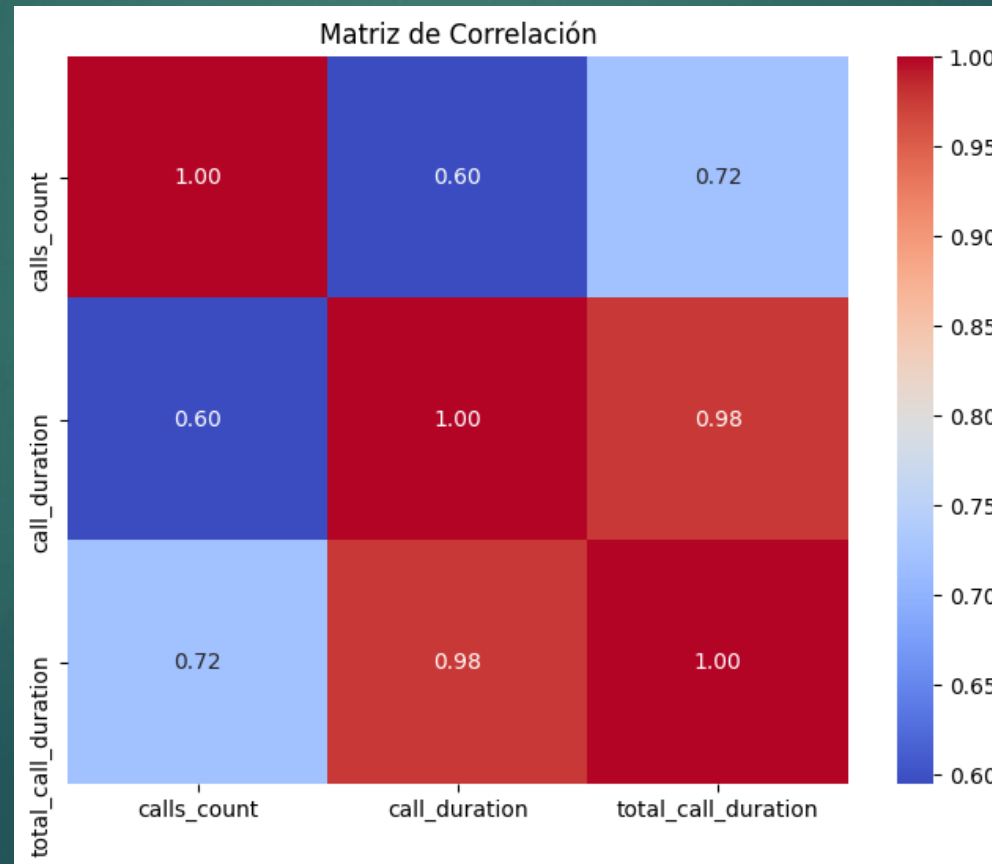
- ▶ Con los resultados del gráfico podemos observar que la duración promedio de las llamadas está alrededor de 100 minutos, además, que a partir de los 700 minutos encontramos ya datos atípicos.

- Con los resultados del gráfico concluimos que la media total de las duraciones de la llamadas esta alrededor de 300 minutos. Partiendo de dicha media podríamos identificar datos atípicos en la duración total de algunas llamadas, esto da pie a aundar en detalle los motivos de las duraciones de dichas llamadas e identificar casos complejos o razones por las cuales se presentan estas duraciones.

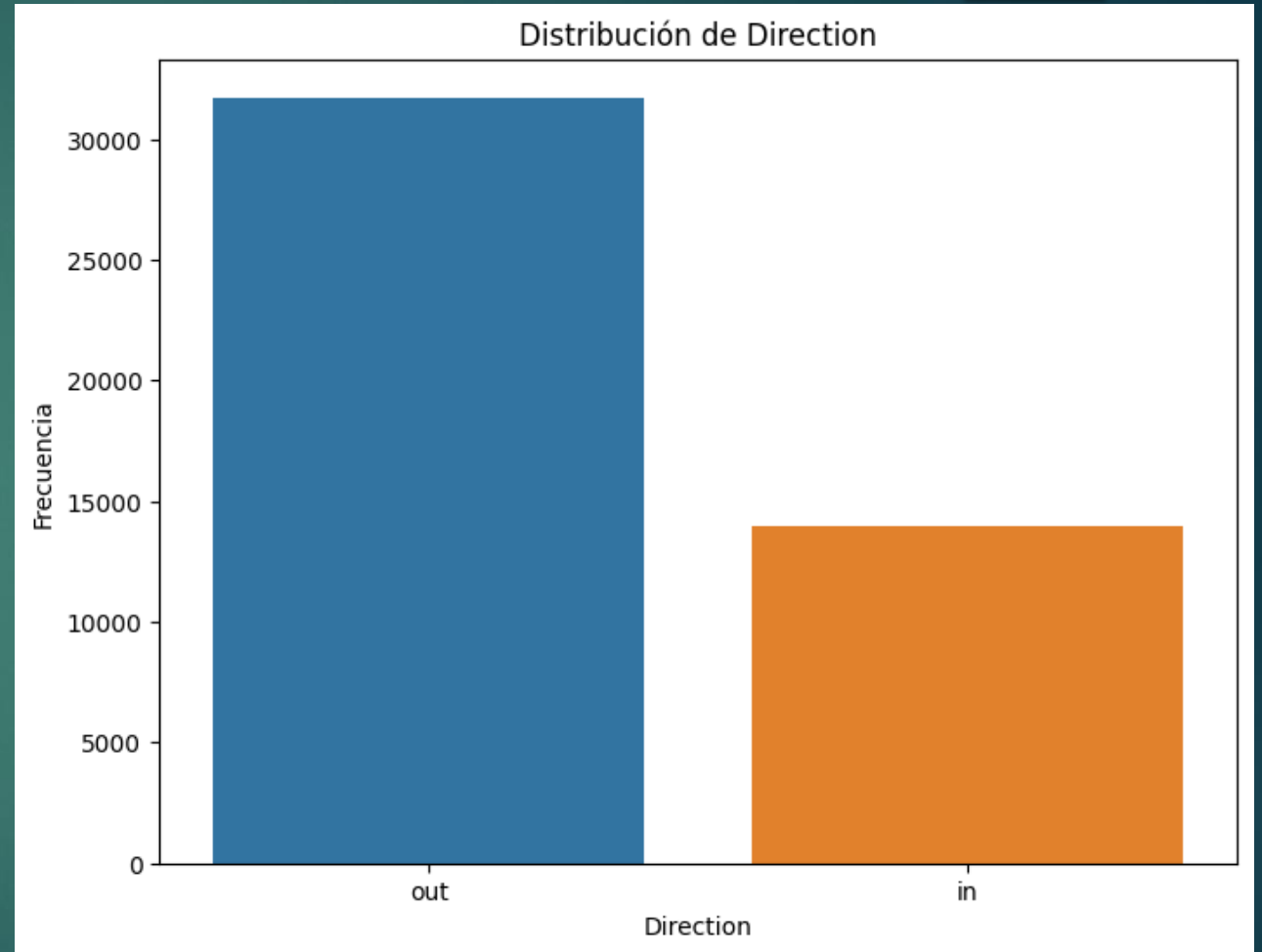


- Para nuestro análisis consideraremos las mismas variables anteriores: `calls_count`, `call_duration`, `total_call_duration`.

Podemos observar que aunque las variables tienen buena relación entre sí en términos generales con más del 60%, las variables que mayor relación tienen son '`total_call_duration`' y '`call_duration`' con una correlación del 98%.



Se observa que se realizan más llamadas hacia afuera 'out' que llamadas internas 'in', en proporción podría concluirse que están distribuidas en un 60% 'out' y 40% 'in'.





# IDENTIFICAR OPERADORES INEFICIENTES

Según lo planteado en la descomposición del proyecto previamente, se realizará el análisis de los siguientes aspectos:

- Qué servicio es más requerido entre llamada entrante o llamada saliente (out para saliente, in para entrante)

out	31717
in	13953

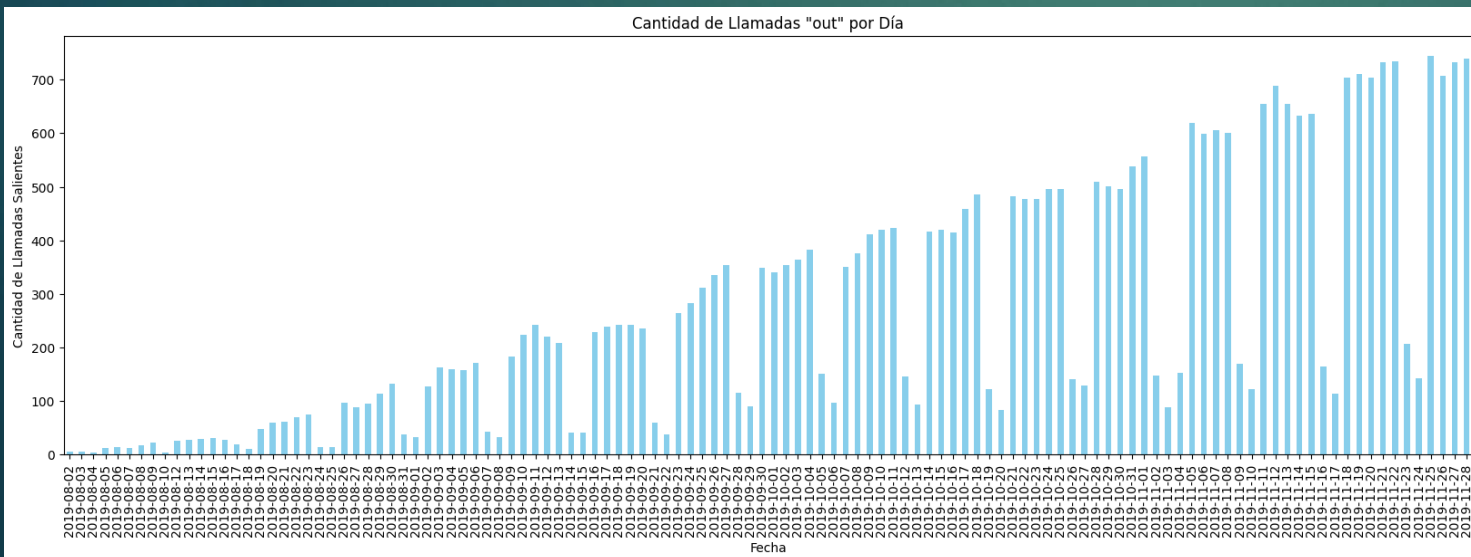
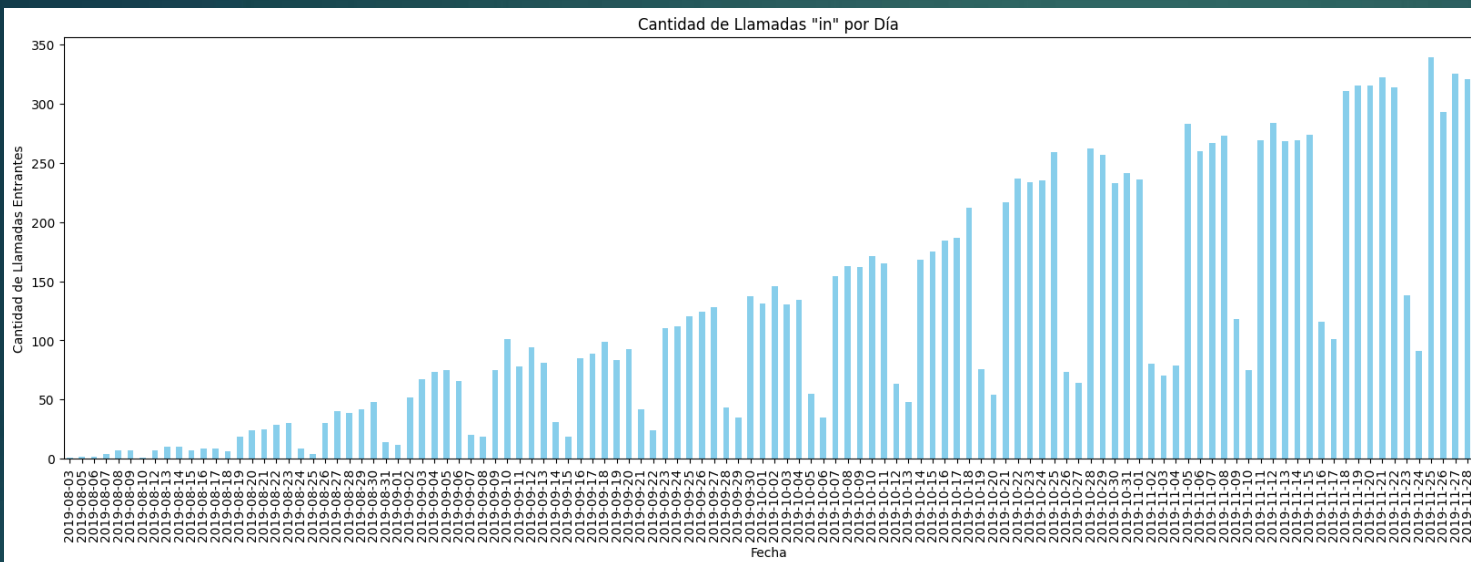
Como se había identificado en el gráfico de distribución, efectivamente las llamadas saliente 'out' son mayores que las llamadas 'in', y que en proporción está en un 60% 'out' - 40% 'in'

- Que servicio puede tardar más en la llamada (duración total de la llamada).

Duración total de las llamadas 'out'	48582322
Duración total de las llamadas 'in'	11870920

El servicio de llamadas 'out' es el que puede tardar más en total en comparación con las llamadas 'in'.

## - Cantidad de llamadas realizadas o entrantes en un día.



- En ambos gráficos vemos que hay una tendencia creciente exponencial a medida que pasan los días y que también en fechas específicas de los meses en ambos casos es que se presentan mayor cantidad de llamadas 'in' y 'out'.
- Se observa que las llamadas salientes 'out' son mayores y constantes que las llamadas entrantes 'in'.

- Obtener en orden descendente el ranking de los operadores ineficaces. Se considerará partiendo del número de llamadas perdidas cada operador.

operator_id	missed_calls_operator	total_duration_operator
885876.0	135.0	6200335
891410.0	131.0	407064
893804.0	130.0	394536
901880.0	128.0	530866
891414.0	126.0	329745
901884.0	115.0	363821
885890.0	110.0	5253549
887276.0	101.0	155132
879898.0	100.0	391841
905538.0	99.0	249782
880026.0	94.0	193666
891416.0	94.0	93212
880028.0	91.0	216250
890404.0	90.0	421164
882686.0	88.0	469596
905566.0	87.0	60208
905542.0	85.0	167938
921818.0	84.0	278168
890410.0	84.0	238964

Según el criterio establecido para identificar ineficiencia en los operadores (llamadas perdidas), se puede concluir que los 10 primeros operadores que encabezan la lista tienen más de 100 llamadas perdidas y aquellos que los siguen están por debajo de las 100 llamadas.

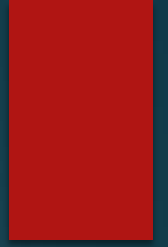


- Del total de llamadas cuál es el % de llamadas realizadas entre operadores internamente.

12.7195 %

Como resultado del análisis realizado incluyendo todas las llamadas sin importar su clasificación de entrante o saliente, se identifica que casi el 13% de estas llamadas son realizadas de forma interna entre los operadores.

# HIPÓTESIS ESTADÍSTICA



Hipótesis nula ( $H_0$ ): Hay diferencia en el total de llamadas entrantes (in) con respecto a las llamadas saliente (out).

Hipótesis alternativa ( $H_1$ ): No existe diferencia entre el total de llamadas entrantes (in) con respecto a las llamadas saliente (out).

# HIPÓTESIS ESTADÍSTICA

## Resultado:

T-Statistic: -22.917474993914173

P-Value: 1.4002711319676033e-115

## Conclusión:

1. Al obtener un valor negativo en T-Statistic concluimos que las llamadas 'in' y 'out' son completamente diferentes.
2. El valor P-Value es mucho menor que el valor alfa establecido, por dicha razón se rechaza la hipótesis nula corroborando que hay una diferencia significativa entre el total de llamadas entrantes y llamadas salientes.
3. En análisis previos se había identificado que el total de llamadas salientes 'out' y llamadas entrantes 'in' existía gran diferencia, siendo esta de 60% del total para llamadas 'out' y del 40% del total para llamadas 'in'.



# CONCLUSIONES GENERALES

- ▶ Se logró identificar aquellos operadores menos eficientes en términos de cantidad de llamadas pérdidas, concluyendo que los 10 primeros operadores ineficientes tienen más de 100 llamadas perdidas.
- ▶ Se identificó que hay un valor superior en total de llamadas salientes 'out' que de entrantes 'in' para definir si este factor altera ineficiencia en los operadores, se recomienda evaluar específicamente para los operadores ineficientes identificados (la cantidad que se considere pertinente) calcular la diferencia en llamadas saliente 'out' y llamadas entrantes 'in' para saber si este factor también aplica como determinante para que un operador sea clasificado como ineficiente.