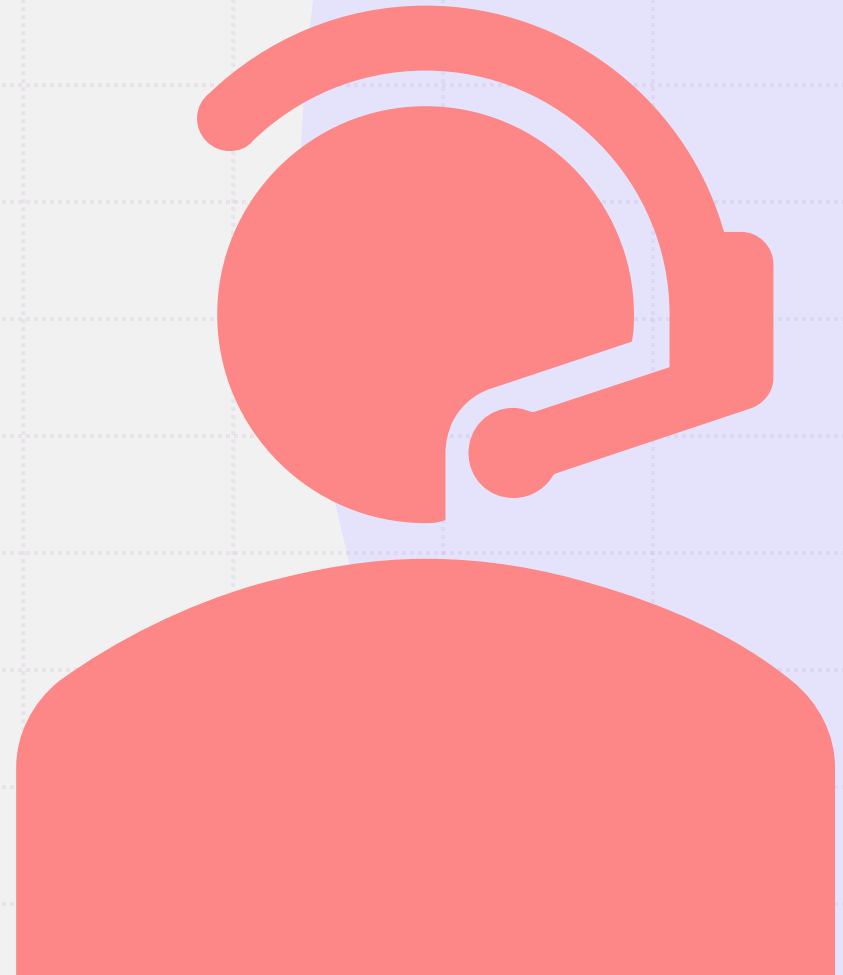


Christian Ortega Ochoa

Proyecto Final:
Telecomunicaciones: Identificar
operadores ineficaces



Descripción del Proyecto

El servicio de telefonía virtual CallMeMaybe está desarrollando una nueva función que brindará a los supervisores 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

Los datasets contienen información sobre el uso del servicio de telefonía virtual CallMeMaybe. Sus clientes son organizaciones que necesitan distribuir gran cantidad de llamadas entrantes entre varios operadores, o realizar llamadas salientes a través de sus operadores. Los operadores también pueden realizar llamadas internas para comunicarse entre ellos. Estas llamadas se realizan a través de la red de CallMeMaybe.

El dataset comprimido `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 conjunto de datos `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

Etapa de Preprocesamiento

Fuente: <https://medium.com/@dossieranalysis/python-for-data-cleaning-best-practices-and-efficient-techniques-3072ed3935af>

Durante esta etapa se definen las librerías a usar, la carga de datos, revisión y limpieza de los datasets a desarrollar.

Con métodos como import, read_csv, .info(), head(), describe() se percibe la estructura de los datos y se llega a una mejor comprensión de los mismos.

Se define el tratamiento que se le darán a los datos nulos, datos duplicados y a los tipos de datos que cada columna tendrá con el fin de ser lo más limpio, coherente y eficiente posible.

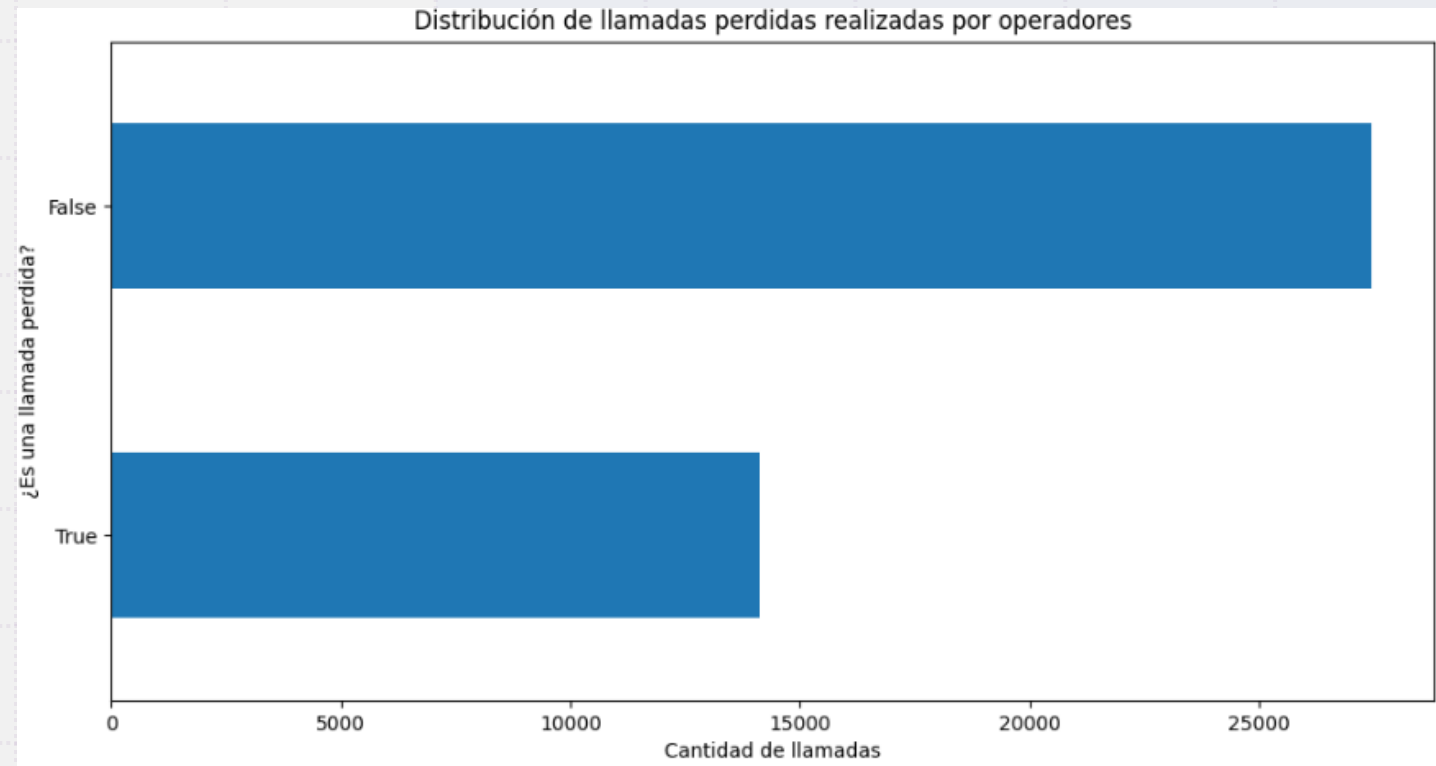
Etapa de Análisis Exploratorio de Datos (EDA):

Fuente: <https://medium.com/data-and-beyond/mastering-exploratory-data-analysis-eda-everything-you-need-to-know-7e3b48d63a95>

Aquí se espera realizar una revisión estadística, visual, que a grandes rasgos nos haga entender el comportamiento de los operadores, clientes en función de cómo se comportan las diversas variables a determinar en el análisis.

Análisis de datos (EDA):

1. La gran mayoría de llamadas NO fueron llamadas perdidas. Sin embargo, aproximadamente un 34% de las llamadas son perdidas, sigue siendo una proporción alta (1 de cada 3 llamadas es una llamada perdida).



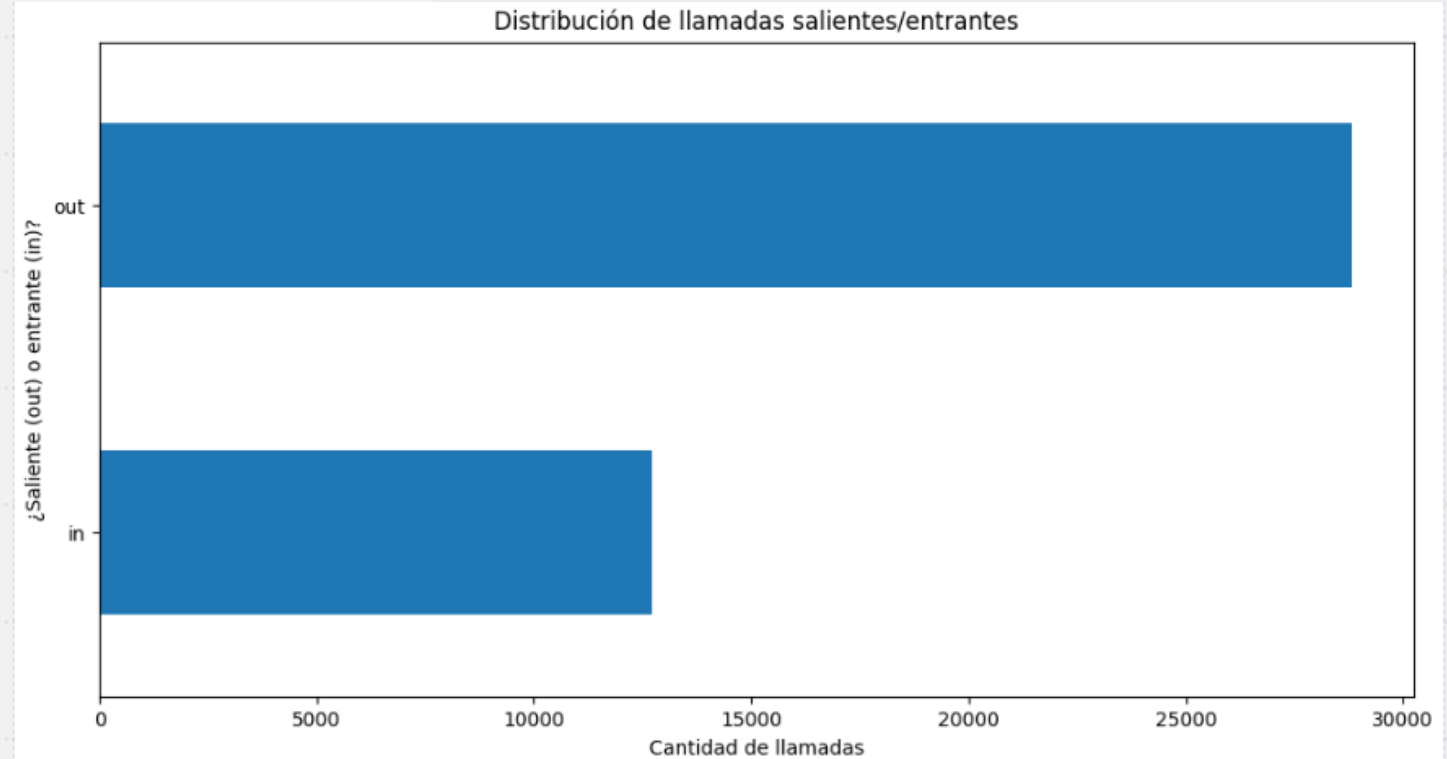
Análisis de datos (EDA):

1. El 90% de los tiempos de espera son menores a 658 segundos (casi 11 minutos), en promedio el tiempo de espera es de 311.82 segundos (poco menos de 5 minutos).



Análisis de datos (EDA):

1. La gran mayoría de llamadas fueron salientes. Sin embargo aproximadamente un 30% de las llamadas son entrantes.



Identificación de correlaciones

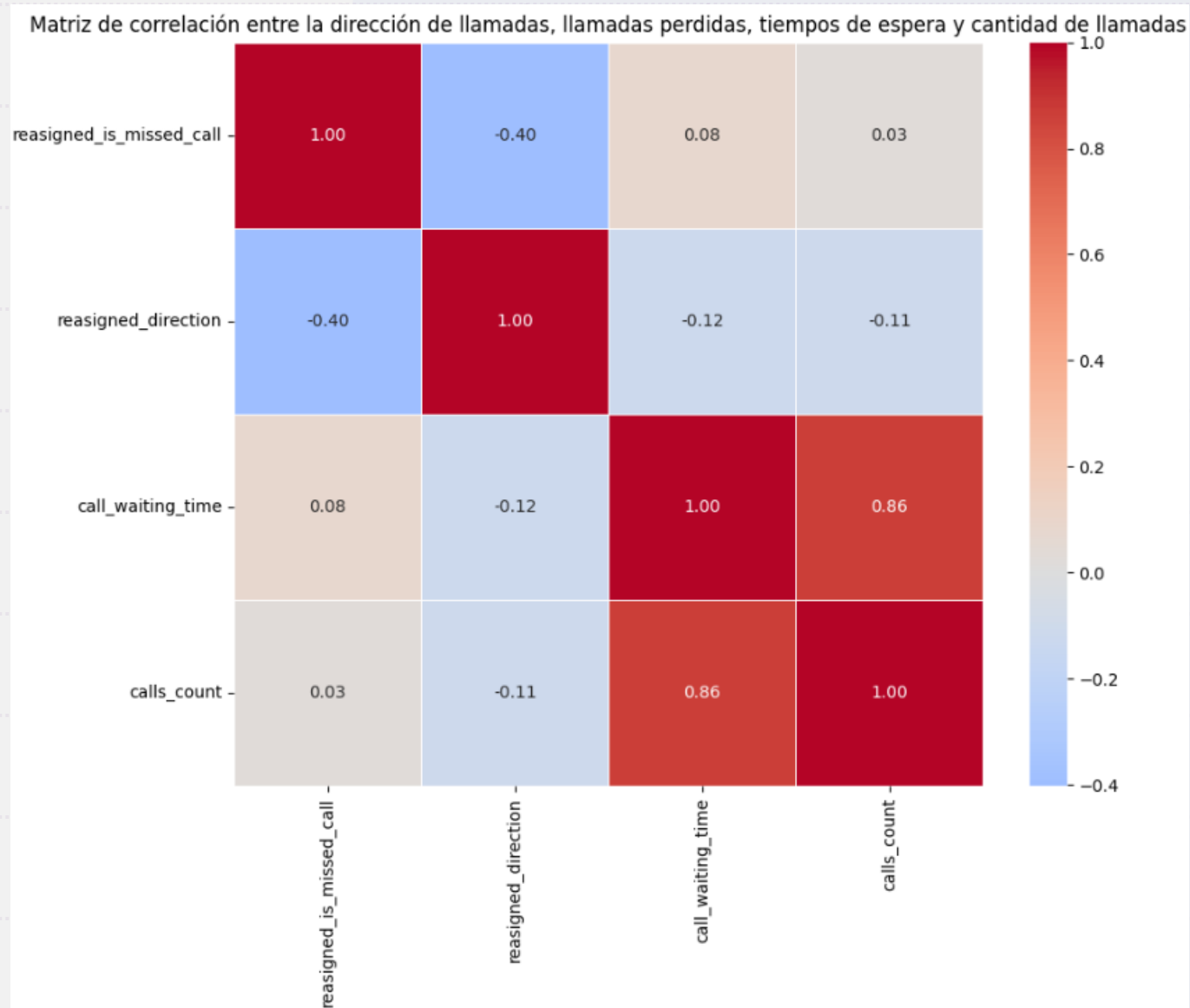
Fuente: <https://realpython.com/numpy-scipy-pandas-correlation-python/>

Conclusiones Generales de las pruebas de correlación:

1. Primera Prueba: Indica una relación inversa moderada entre las llamadas perdidas y la dirección de las llamadas. Es posible que estas variables tengan un impacto significativo en la identificación de ineficiencia.

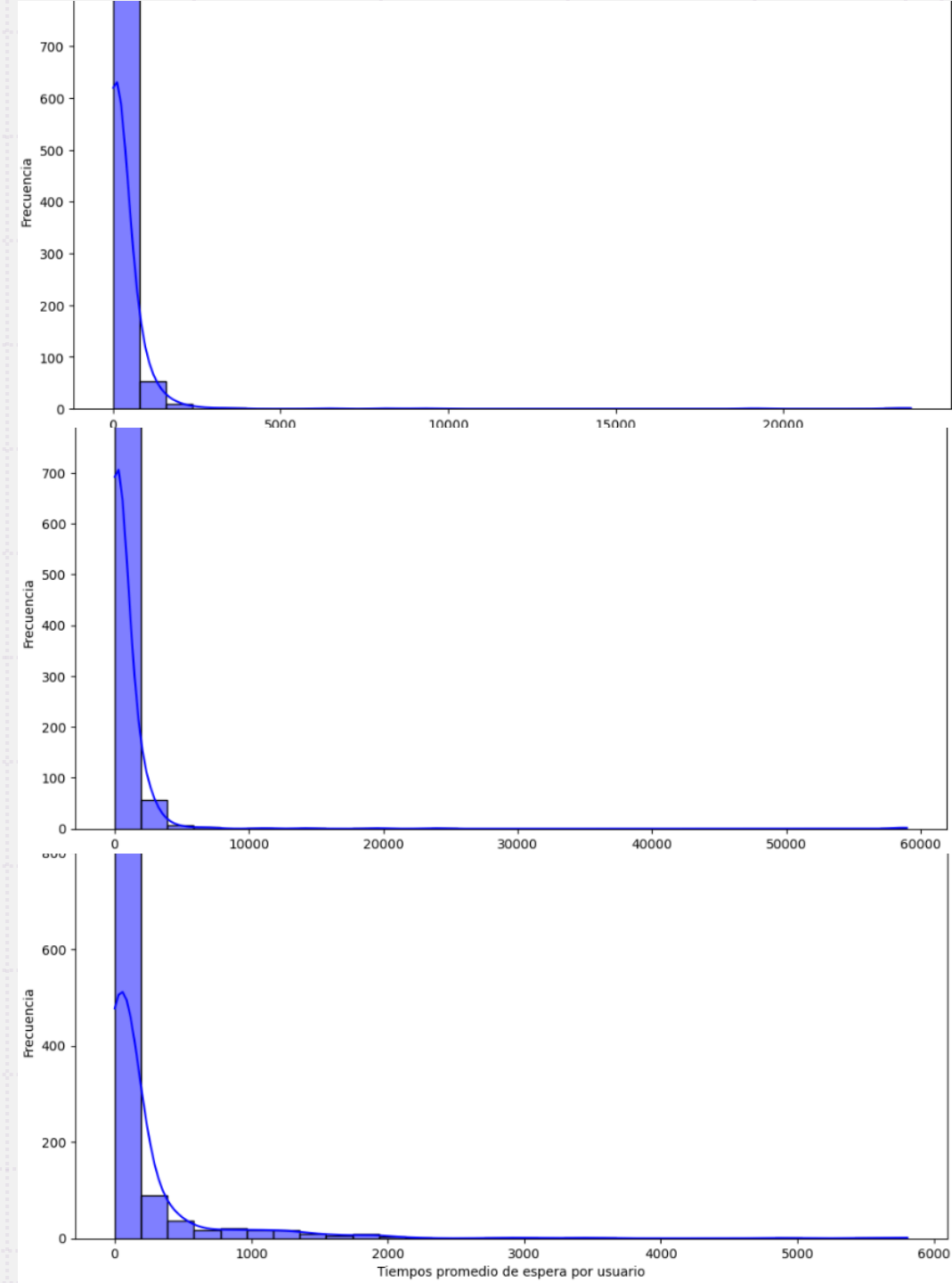
2. Segunda Prueba: No muestra una relación significativa entre las variables, sugiriendo que podrían no estar directamente relacionadas en el contexto de ineficiencia operativa.

3. Tercera Prueba: Muestra una fuerte relación positiva, lo que sugiere que las variables están estrechamente vinculadas y podrían ser críticas para identificar o comprender la eficacia de los operadores.



Identificación de Operadores Ineficaces:

1. El 75% de los operadores tienen 229 llamadas perdidas o menos, sin embargo los que están fuera de ese rango contribuyen demasiado a que la gráfica tenga un sesgo alto hacia la derecha.
2. Los tiempos promedios de espera están en 216 segundos, sin embargo los operadores que lideran el Top están muy por encima de ese promedio, sesgando la gráfica a la derecha.



Etapa de Hipótesis:

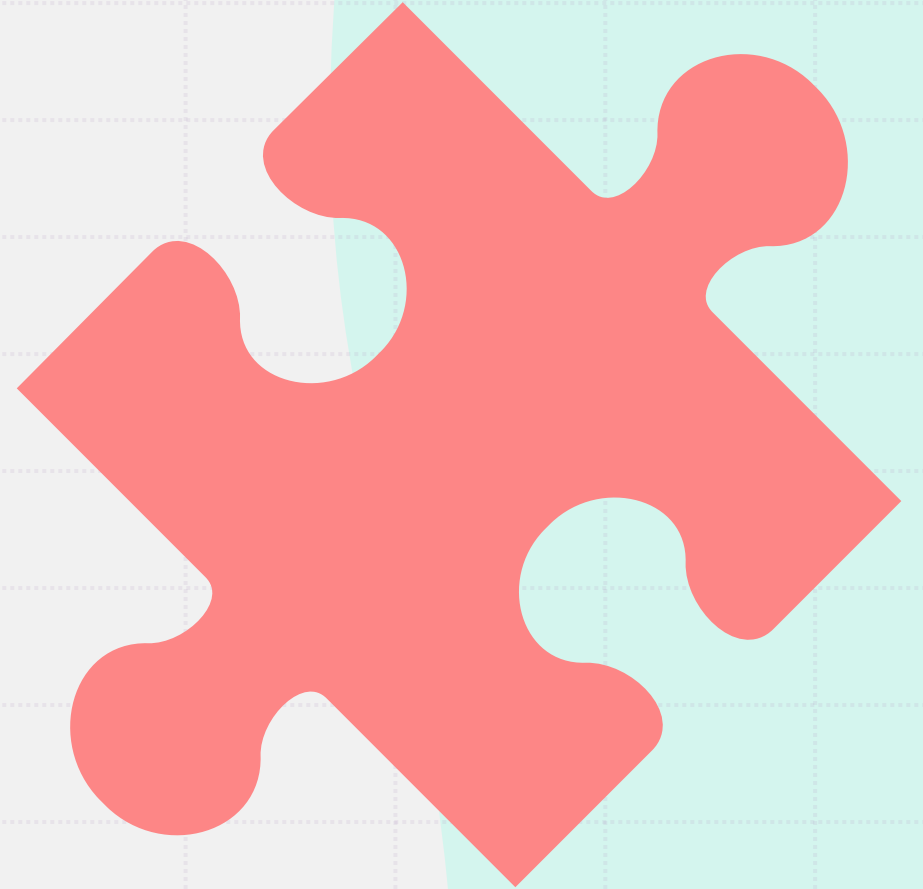
1. Fuentes:

https://www.statsmodels.org/dev/generated/statsmodels.stats.proportion.proportions_ztest.html

2. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

3. En esta etapa se definen las hipótesis a partir de una noción previa producto del análisis. ****Por ejemplo**** se podría establecer que:

4. 1. Hipótesis 1: Los operadores que realizan pocas llamadas salientes, cuando se supone que deben hacerlas, son ineficaces.
5. 2. Hipótesis 2: Los operadores con un alto porcentaje de llamadas perdidas son ineficaces.
6. 3. Hipótesis 3: Los operadores con tiempos de espera prolongados para las llamadas entrantes son ineficaces.
7. 4. Hipótesis 4: Los operadores que contactan con clientes con fechas de registros más recientes suelen ser ineficaces.



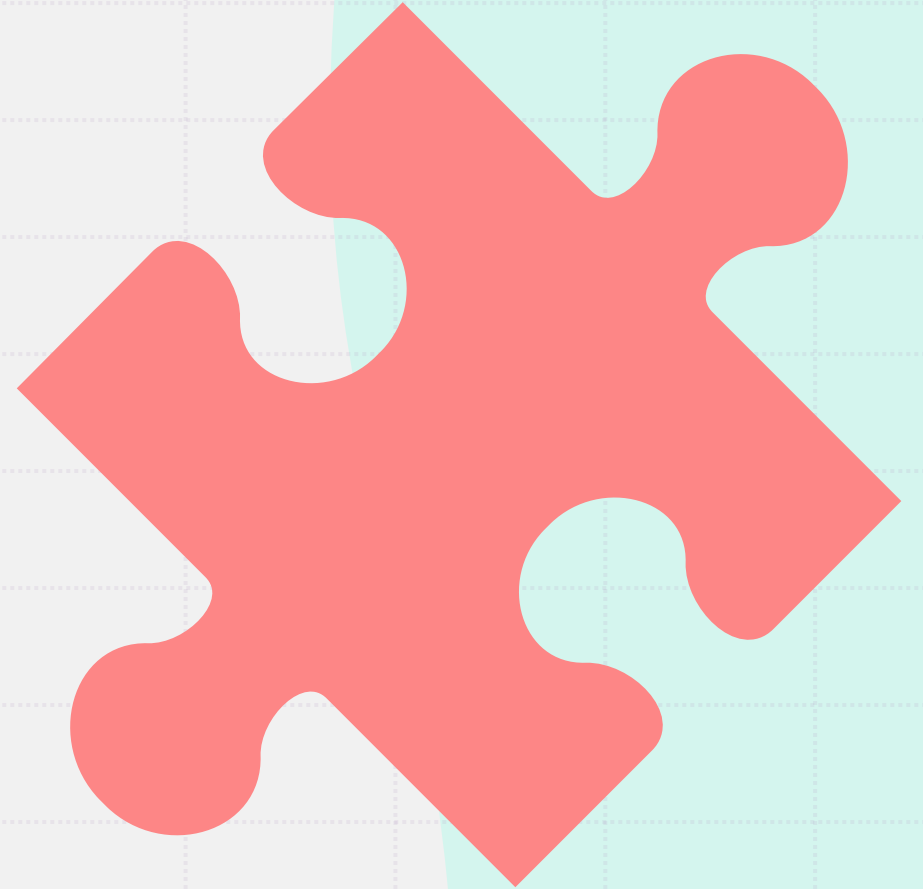
Etapa de Hipótesis:

1. **Conclusión sobre la Hipótesis 1:**

2. Dado el T-statistic negativo y el P-value tan bajo, rechazamos la hipótesis nula. La hipótesis nula probablemente plantea que no hay diferencia significativa en la eficacia de los operadores basada en la cantidad de llamadas salientes realizadas. Al rechazar la hipótesis nula, apoyamos la hipótesis alternativa, que establece que los operadores que realizan pocas llamadas salientes, cuando se supone que deben hacerlas, son ineficaces.

3. **Recomendaciones basadas en los resultados:**

4. 1. Los trabajadores deben tener un script guía claro que especifique como actuar en caso de interactuar con un usuario. Tener una especie de 'TroubleShooting' que estudien constantemente y se retroalimenten de ellos para evitar ser ineficaces cuando interactúen con los clientes.
5. 2. Definir una meta en un periodo de tiempo concreto. Por ejemplo: Número mínimo de llamadas diarias, para poder establecer más fácilmente si el operador, al menos a la hora de realizar llamadas externas, es eficiente o no.



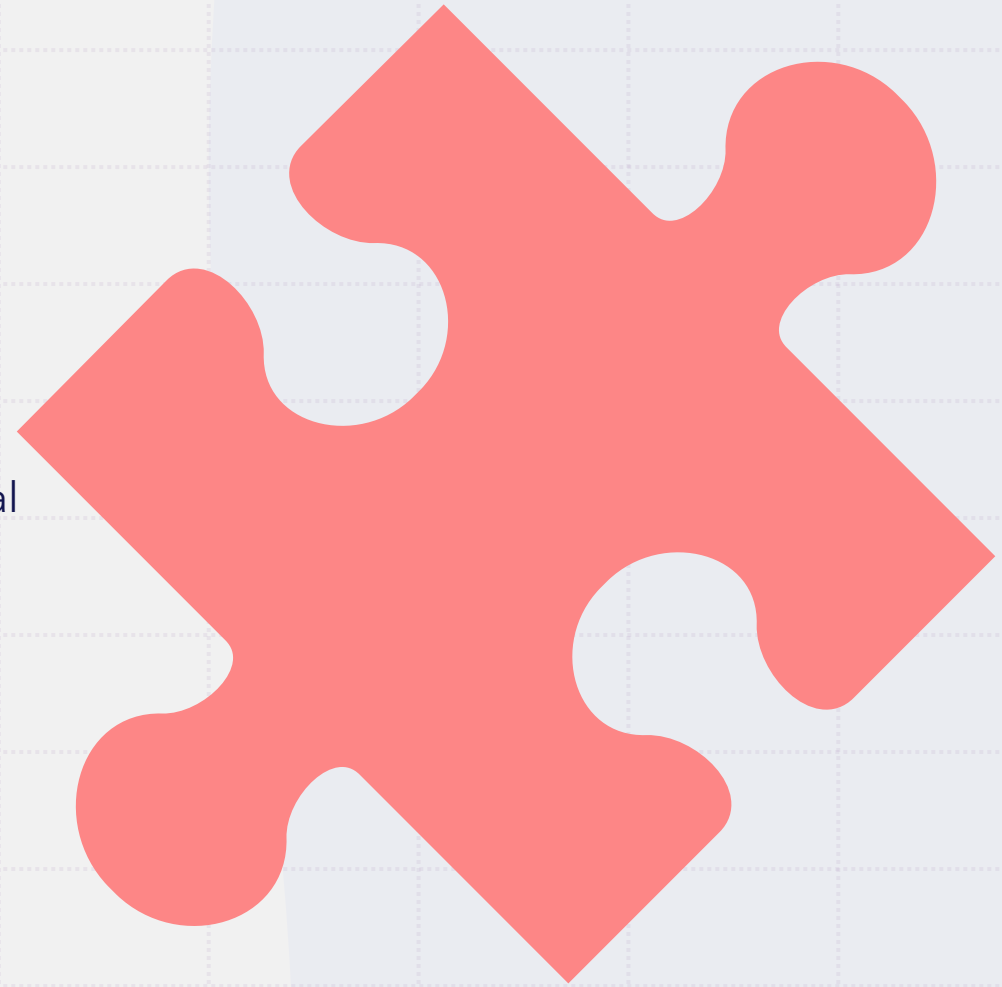
Etapa de Hipótesis:

1. Conclusión sobre la Hipótesis 2:

2. Dado el Z-statistic tan extremo y alejado de 0; y el P-value extremadamente bajo, rechazamos la hipótesis nula. En este caso, la hipótesis nula sostendría que no hay diferencia significativa en la eficacia de los operadores basada en el porcentaje de llamadas perdidas. Al rechazar la hipótesis nula, apoyamos la hipótesis alternativa, que sugiere que los ****operadores con un alto porcentaje de llamadas perdidas son ineficaces.****

3. Recomendaciones basadas en los resultados:

1. Revisar la carga laboral de los operadores. Es probable que al tener un volumen alto de asignaciones no puedan atender llamadas y por ende impactar en la eficacia de recepción de llamadas.
2. Revisión de procedimientos: Implementar cambios en los procedimientos de manejo de llamadas para reducir las llamadas perdidas.
3. Monitorear a través de seguimiento estadístico a los operadores que actualmente estén incurriendo en alto porcentaje de llamadas perdidas.



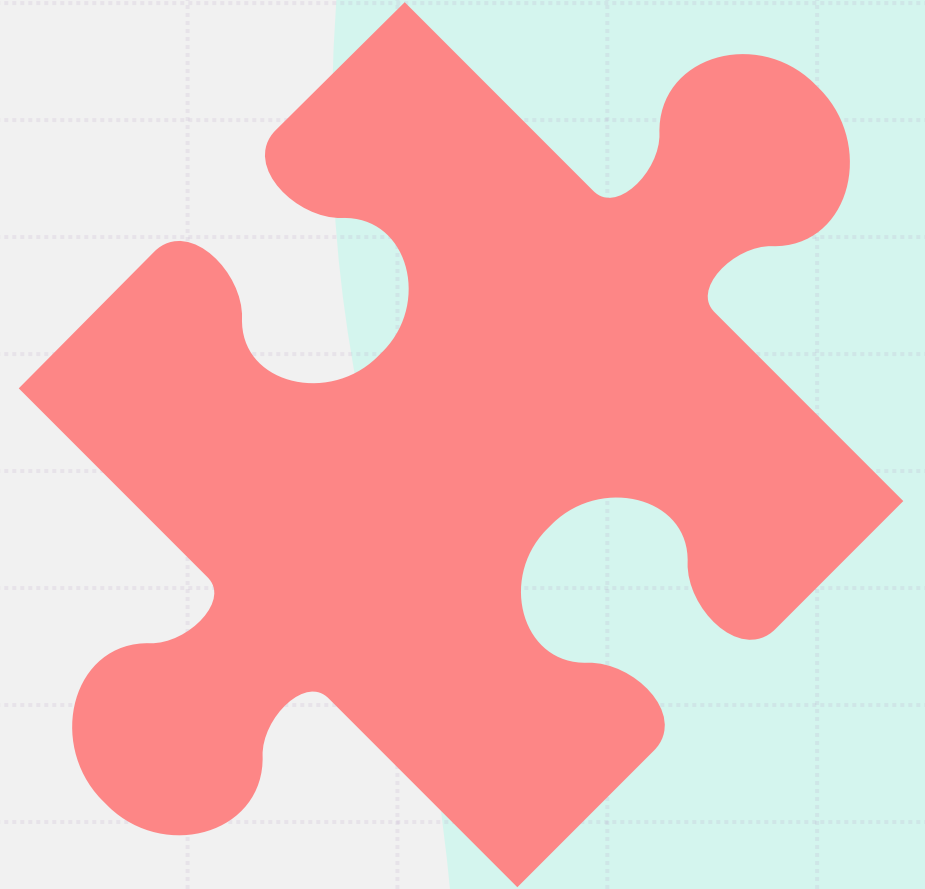
1. **Conclusión sobre la Hipótesis 3:**

2. Dado el T-statistic tan alto y el P-value extremadamente bajo, rechazamos la hipótesis nula. En este caso, la hipótesis nula probablemente sostiene que no hay diferencia significativa en la eficacia de los operadores basada en el tiempo de espera para las llamadas entrantes. Al rechazar la hipótesis nula, apoyamos la hipótesis alternativa, que sugiere que **los operadores con tiempos de espera prolongados para las llamadas entrantes son ineficaces.

3. **Recomendaciones basadas en los resultados:**

4. 1. Revisar la carga laboral de los operadores. Es probable que al tener un volumen alto de asignaciones tienen unos altos tiempos de espera a sus usuarios.
5. 2. Mejorar los sistemas de gestión de llamadas.
6. 3. Capacitar a los operadores sobre en diferentes habilidades: Gestión del tiempo, definición de prioridades, ser más eficientes a la hora de dar instrucciones a los usuarios, entre otros.

Etapa de Hipótesis:



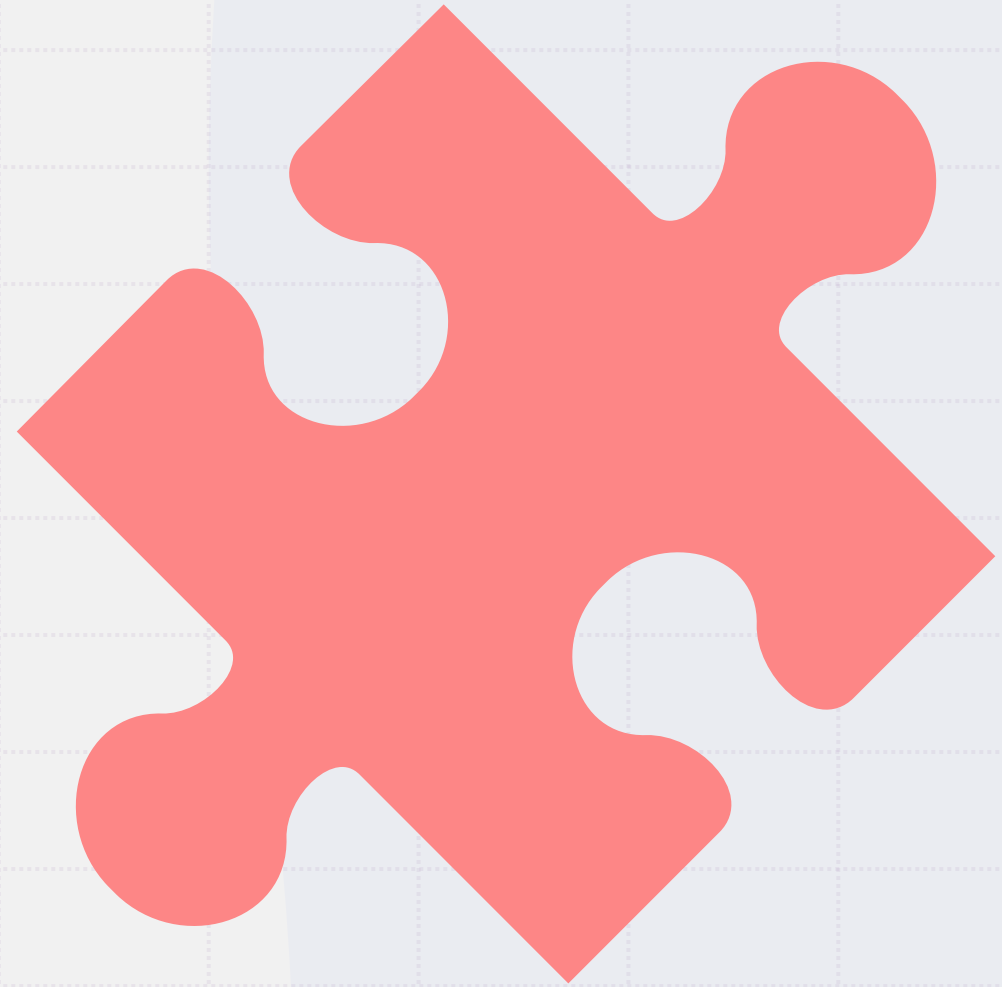
1. **Conclusión sobre la Hipótesis 4:**

2. Resultado: Dado que la correlación es prácticamente nula, no hay evidencia práctica que respalde la hipótesis declarada. No se puede concluir que contactar a clientes con fechas de registro más recientes influya de manera significativa en la ineficacia de los operadores.

3. **Recomendaciones basadas en los resultados:**

4. En este caso se recomienda explorar otras variables que puedan estar afectando la eficacia de los operadores:
5. 1. Considerar Contexto Operativo: Es importante contextualizar los resultados dentro de las operaciones diarias y procedimientos de la empresa para interpretar adecuadamente los hallazgos.
6. 2. Estrategia de Contacto: Los resultados indican que la fecha de registro de los clientes no es un factor determinante en la eficacia de los operadores. Por lo tanto, no sería necesario ajustar las estrategias de contacto basándose únicamente en la antigüedad del registro de los clientes.
7. 3. Enfoque en Otros Factores: Sería más productivo centrar la atención en otros factores que puedan tener una influencia más significativa en la eficacia de los operadores, como la capacitación, la experiencia, la satisfacción del cliente, entre otros.

Etapa de Hipótesis:



Conclusiones de los análisis realizados y proposiciones de acciones de mejora



Posibles soluciones luego del análisis y las pruebas de hipótesis:



- Capacitación de operadores: Identificar necesidades de capacitación para mejorar el desempeño de operadores ineficaces y de aquellos que estén haciendo bien su trabajo actualmente (ejemplos a seguir).



- Redistribución de tareas: Ajustar la carga de trabajo y responsabilidades de los operadores para equilibrar la eficiencia operativa. Además de eliminar tareas que puedan resultar innecesarias.



- Revisión de procedimientos: Implementar cambios en los procedimientos de manejo de llamadas para reducir los tiempos de espera y las llamadas perdidas.



- Evaluación de rendimiento: Realizar evaluaciones periódicas del rendimiento de los operadores utilizando métricas clave para garantizar la mejora continua.



- Optimización del servicio: Ajustar las estrategias operativas y los planes de acción para mejorar la eficiencia general del servicio de telefonía virtual.



- Retroalimentación del proceso: Al establecer acciones de retroalimentación, se garantiza la vinculación de todos aquellos actores implicados en el proceso y que tengan la capacidad de establecer acciones propositivas, documentadas y de mejora continua que alimenten el proceso constantemente, indiferente de su posición jerárquica en la empresa.