

Análisis de Operadores de Telecomunicaciones

Un estudio integral para identificar operadores ineficaces mediante métricas cuantificables y análisis estadístico robusto. Este proyecto transforma datos operacionales en insights accionables para supervisores.

Autor: Octavio Landa Verde

Fecha: 27 de octubre del 2025



Planificación y Gestión del Proyecto

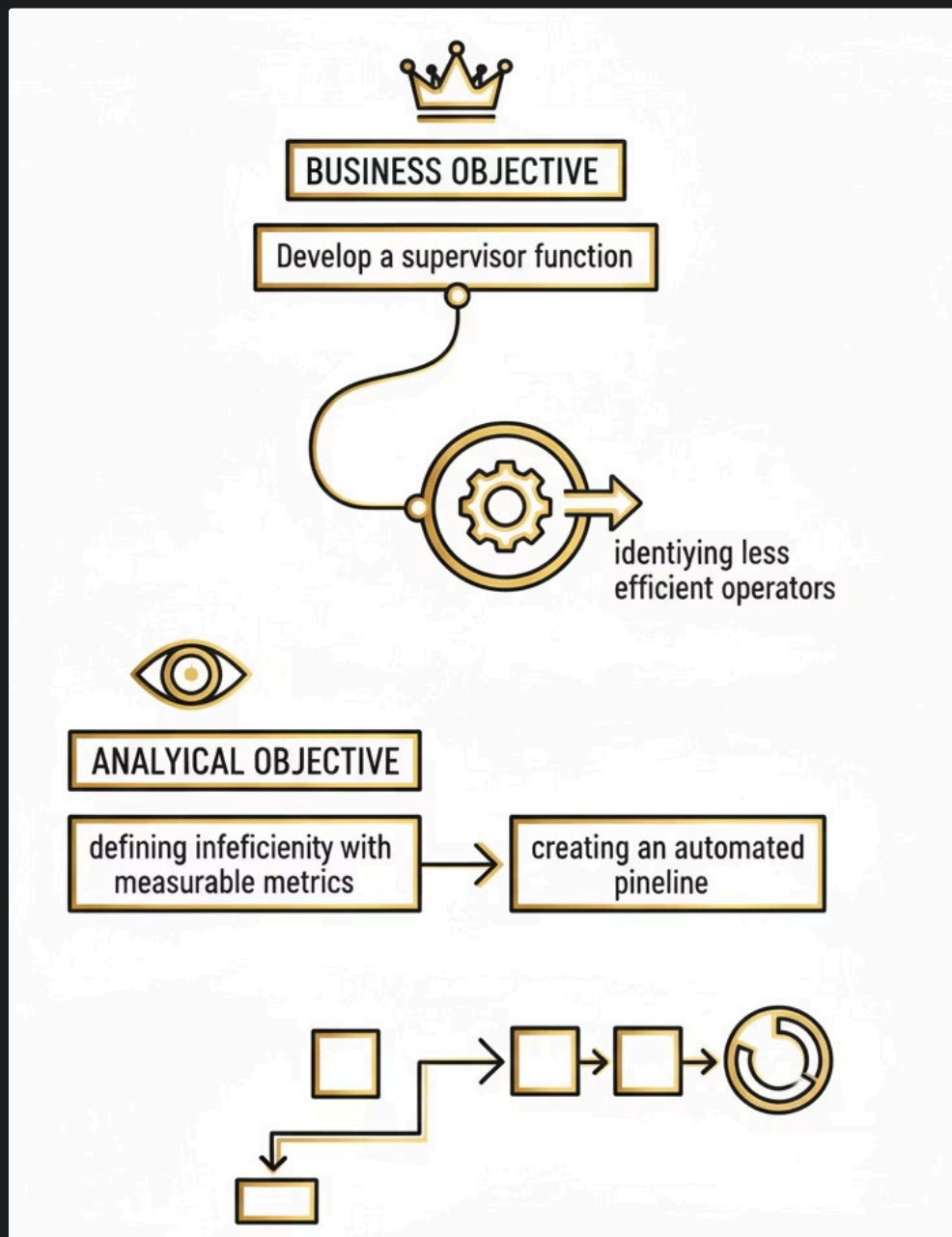
Objetivo de Negocio

Desarrollar una función para supervisores que identifique operadores menos eficaces mediante un sistema de métricas cuantificables y análisis predictivo.

Objetivo Analítico

Definir "ineficacia" con métricas medibles y crear un pipeline automatizado para identificar operadores que requieren intervención o reentrenamiento.

Esta fase inicial traduce un problema de negocio ambiguo en un conjunto específico de métricas analíticas, estableciendo las bases para un análisis riguroso y orientado a resultados.



Definición de Ineficacia Operacional

1

Llamadas Perdidas

Alto porcentaje de llamadas entrantes perdidas, tanto internas como externas, indicando problemas de disponibilidad.

2

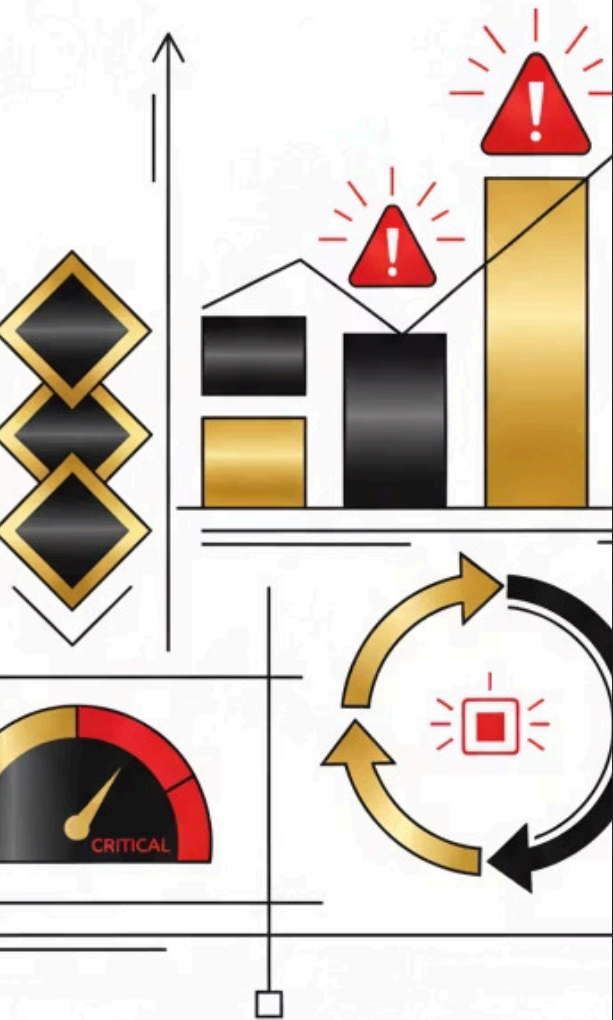
Tiempo de Espera

Tiempo de espera promedio elevado para llamadas entrantes, reflejando ineficiencia en la gestión del flujo de trabajo.

3

Llamadas Salientes

Bajo número de llamadas salientes cuando es parte del rol esperado, señalando falta de proactividad.



Preguntas Clave del Análisis

Calidad de Datos

¿Cuál es la integridad de los datos?
Identificación de valores nulos y tipos incorrectos.

Métricas Derivadas

¿Cómo crear la métrica clave
tiempo_de_espera a partir de variables
existentes?

Distribuciones

¿Cómo se distribuyen las llamadas por
duración, conteo y tipos?

Análisis por Operador

¿Cómo varían estas métricas entre diferentes
operadores?

Umbrales Estadísticos

¿Cómo definir "alto" y "bajo"
usando cuantiles como
enfoque robusto?

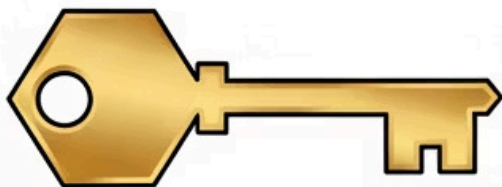
Identificación

¿Qué operadores caen en
categorías de ineficacia
según los criterios
establecidos?

Significancia Estadística

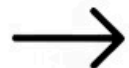
¿Es la diferencia entre
grupos estadísticamente
significativa?

Calidad de Datos



¿Cuál es la integridad de los
datos? Identificación de
de valores nulos y tipos
incorrectos.

Métricas Derivadas



¿Cómo crear la métrica clave
tiempo_de_espera a
partir de varibales existentes?

Adquisición y Preparación de Datos

La fase de ingeniería de datos transforma información cruda en un formato analizable. Se cargaron dos datasets principales: **telecom_dataset_us.csv** con registros de llamadas y **telecom_clients_us.csv** con información de clientes.

01

Importación de Bibliotecas

Pandas para manipulación de datos, NumPy para cálculos numéricos, Matplotlib y Seaborn para visualización, SciPy para análisis estadístico.

02

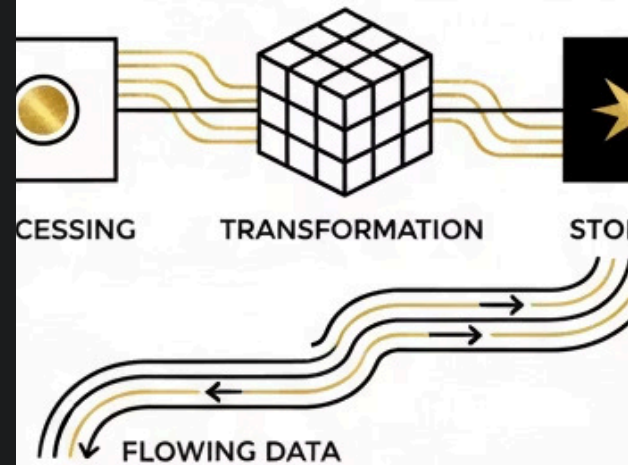
Carga de Datasets

Verificación exitosa de archivos y validación de dimensiones de los conjuntos de datos.

03

Inspección Inicial

EDA defensivo para verificar tipos de datos, valores nulos y valores únicos antes de la limpieza.



Hallazgos de la Inspección de Datos

Dataset Telecom

- Campo `date` requiere conversión a `datetime`
- Campos `internal` e `is_missed_call` correctamente como booleanos
- `operator_id` contiene valores nulos críticos

Dataset Clientes

- Campo `date_start` necesita conversión a `datetime`
- `user_id` es la clave común para unir tablas

❏ Decisión Crítica

Las llamadas sin `operator_id` no pueden usarse para evaluar operadores. Se eliminaron estos registros para mantener la integridad del análisis.

HALLAZGOS DE LA INSPENCIÓN DE DATOS

DATASET TELECOM



date requires conversion to `datetime`



Campos `internal` e `is_missed_call` correctamente como booleanos



`operator_id` contiene valores nulos críticos

DATASET CLIENTES



Campo `date_start` necesita conversión a `datetime`



`user_id` es la clave común a `datetime`



`user_id` es la clave común para unir tablas

DECISIÓN CRÍTICA




Las llamadas sin `operator_id` no pueden usarse para evaluar operadores



Se eliminaron estos registros para mantener

Ingeniería de Características

El paso más crítico del proyecto: crear las métricas que definen la ineficacia operacional. Esta transformación convierte datos brutos en indicadores accionables de rendimiento.

	$\frac{f}{dx}$	
Tiempo de Espera Total	Tiempo de Espera Promedio	Banderas de Dirección
Diferencia entre duración total y duración efectiva de llamada.	Tiempo de espera dividido por número de llamadas, con protección contra división por cero.	Columnas booleanas para identificar llamadas entrantes y salientes.

Estas características derivadas permiten análisis granular del comportamiento de cada operador, facilitando la identificación de patrones de ineficacia.

INGENIERÍA DE CARACTERÍSTICAS





Tiempo de Espera Total



Tiempo de Espera Promedio



Banderas de Dirección



Análisis Granular

These derived features enable granural analysis of each operator's behavior, facilitatieg the identificación de paterns of operational ineficienci.

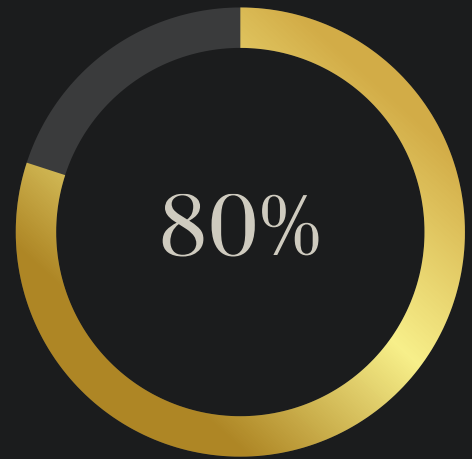
Análisis Exploratorio de Datos

Distribución de Llamadas

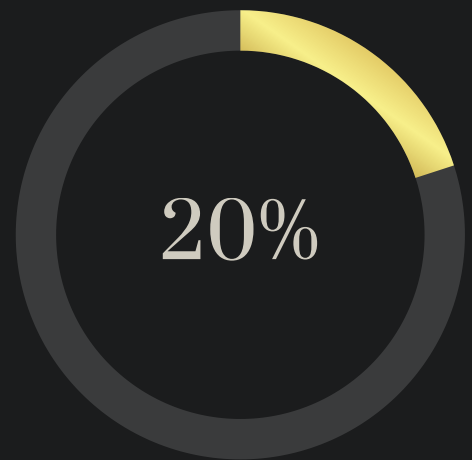
La duración de llamadas muestra un sesgo pronunciado hacia la derecha, típico en telecomunicaciones. La mayoría son muy cortas, con algunas llamadas extremadamente largas que afectan el promedio.

Tipos de Llamadas

- **80%** son llamadas entrantes
- **20%** son llamadas perdidas (problema crítico)
- Mayoría son externas (no entre operadores)



Entrantes



Perdidas

El manejo de llamadas entrantes es la función principal, y el 20% de llamadas perdidas representa el problema central a abordar.

Métricas por Operador

Transformación pivotal: agregación de datos por operador para crear un perfil de rendimiento individual. Cada fila representa un operador único con sus métricas consolidadas.



Métricas por Operador



Identificación de Operadores Ineficaces

Aplicación de umbrales estadísticos usando percentiles como método robusto. "Alto" se define como percentil 75 (25% superior), "Bajo" como percentil 25 (25% inferior).

Q3

Umbral Llamadas Perdidas

Percentil 75 para identificar alto porcentaje de pérdidas.

Q3

Umbral Tiempo Espera

Percentil 75 para identificar tiempos de espera elevados.

Q1

Umbral Salientes

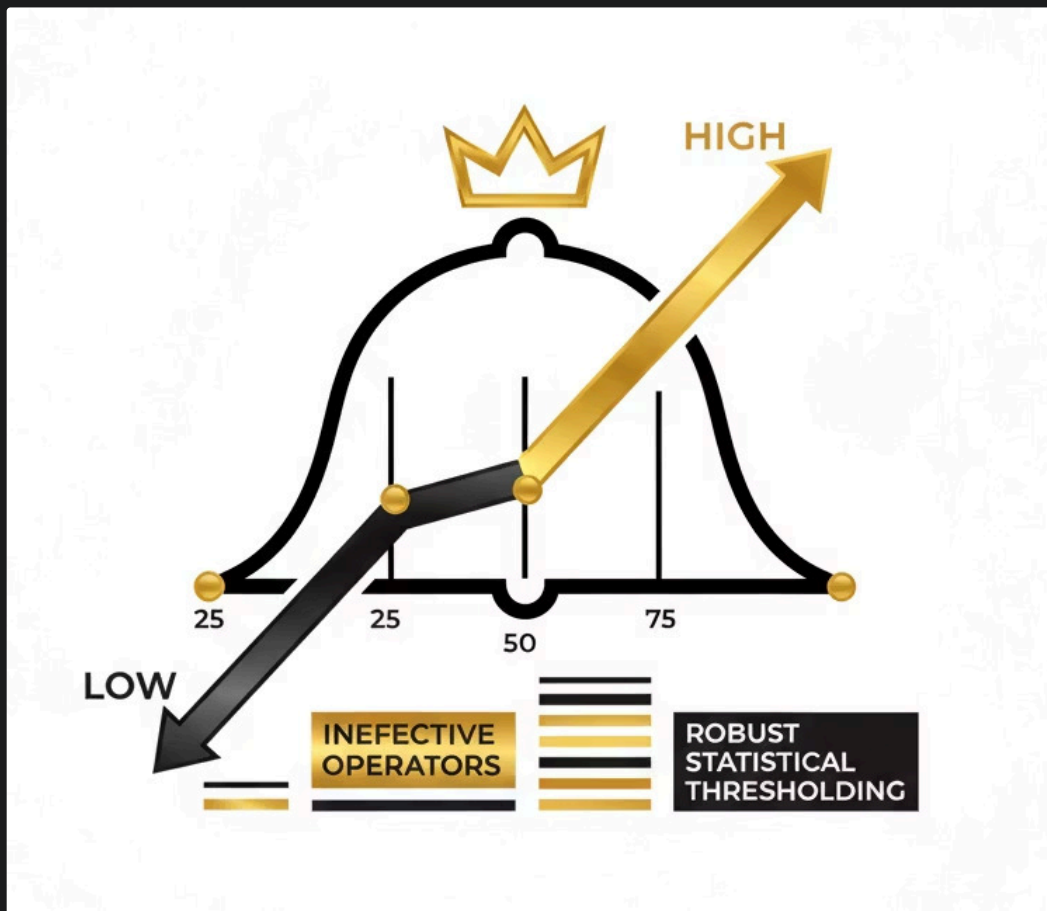
Percentil 25 para identificar baja producción saliente.

Criterio 1: Gestión de Entrada

Operadores con alto % de perdidas Y alto tiempo de espera simultáneamente.

Criterio 2: Producción de Salida

Operadores en el 25% inferior de llamadas salientes entre quienes realizan esta actividad.



Pruebas de Hipótesis Estadísticas

Validación de que las diferencias observadas no son producto del azar. Se utilizó la Prueba U de Mann-Whitney, apropiada para datos no normales.

Hipótesis 1: Tiempo de Espera

H₀: Tiempo de espera del grupo ineficaz = grupo eficaz

H₁: Tiempo de espera del grupo ineficaz > grupo eficaz

Resultado: Se rechaza H₀ ($p < 0.05$)

La diferencia es estadísticamente significativa. El grupo ineficaz tiene tiempo de espera significativamente **MAYOR**.

Hipótesis 2: Llamadas Salientes

H₀: Llamadas salientes del grupo ineficaz = grupo eficaz

H₁: Llamadas salientes del grupo ineficaz < grupo eficaz

Resultado: Se rechaza H₀ ($p < 0.05$)

La diferencia es estadísticamente significativa. El grupo ineficaz tiene número de llamadas significativamente **MENOR**.



MANN-WHITNEY U TEST
VALIDATED DIFFERENCE

Conclusiones y Hallazgos Principales



Métricas Clave Establecidas

Se crearon indicadores de rendimiento: % de llamadas perdidas, tiempo de espera promedio y total de llamadas salientes por operador.



Umbrales Definidos

Criterios de ineficacia usando cuantiles (Q3 para métricas altas, Q1 para bajas), método estándar de la industria.



Operadores Identificados

Dos grupos críticos: ineficaces en gestión de entrada (alto % pérdidas + alto tiempo espera) e ineficaces en producción saliente (bajo volumen).

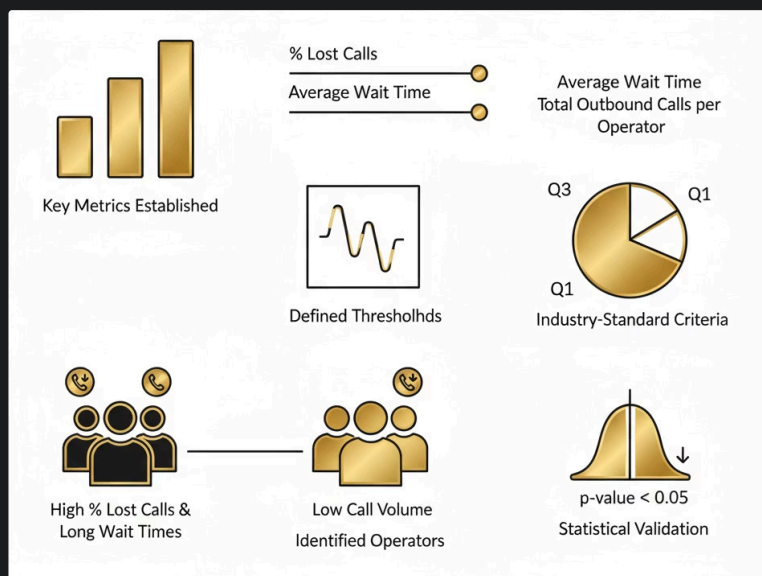


Validación Estadística

Pruebas Mann-Whitney U confirmaron con alta confianza ($p < 0.05$) que las diferencias no se deben al azar.

Recomendación Accionable

Implementar estas métricas y umbrales en el dashboard de supervisores. Los operadores identificados deben ser marcados automáticamente para revisión o reentrenamiento inmediato.



Fuentes y Referencias del Proyecto

Documentación de Pandas

Fundamental para operaciones groupby, agg, merge y fillna en la creación del DataFrame de estadísticas de operadores.

SciPy Stats

Guía para parámetros de la prueba U de Mann-Whitney y configuración de pruebas de una cola.

- **Python for Data Analysis** (Wes McKinney): Mejores prácticas para limpieza y transformación de datos
- **Matplotlib y Seaborn**: Creación de histogramas, gráficos circulares y personalización de visualizaciones
- **Handling Missing Data in Python**: Estrategias para manejo de valores nulos con dropna
- **Feature Engineering for Machine Learning** (Zheng & Casari): Metodología para derivación de características
- **Artículos sobre Pruebas A/B**: Selección entre t-test y pruebas no paramétricas según distribución de datos

Este proyecto demuestra un enfoque integral desde la planificación hasta la validación estadística, proporcionando un sistema robusto y accionable para la identificación de operadores ineficaces en telecomunicaciones.

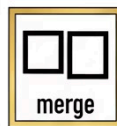
DOCUMENTACIÓN DE PANDAS



groubup



agg



merge

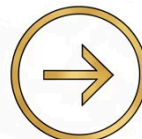


finlla

SCIPY STATS



Mann-Whitney U
test parameteras



one-tailed test
configuración