

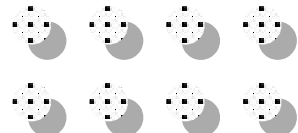


Proyeto final

03/04/25

Obetivo de proyecto:

- Llevar a cabo el análisis exploratorio de datos
- Identificar operadores ineficaces
- Prueba las hipótesis (a reserva que el area interesada tenga necesidades de informacion particulares) se sugieren las siguientes hipotesis y analisis
 - Hay una mayor cantidad de llamadas internas que externas.
 - La mayoría de las llamadas se conectaron (es decir no son llamadas perdidas)
 - Las llamadas internas tienen una duracion mayor con respecto a las externas
- Esto utilizando dos datasets sobre el uso del servicio de telefonía virtual CallMeMaybe.



Datos:

- 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)

Datos:

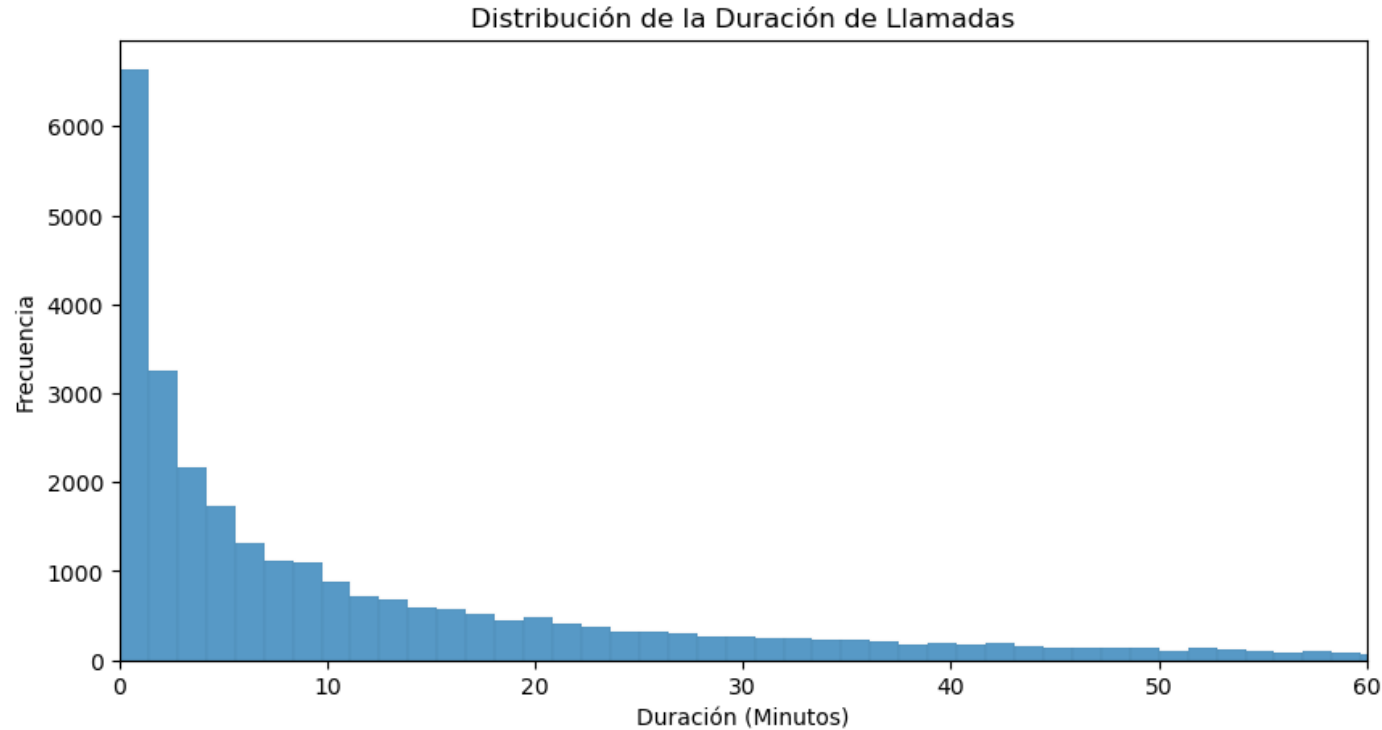
- El conjunto de datos telecom_clients_us.csv tiene las siguientes columnas:
 - user_id: ID de usuario/a
 - tariff_plan: tarifa actual de la clientela
 - date_start: fecha de registro de la clientela

EDA

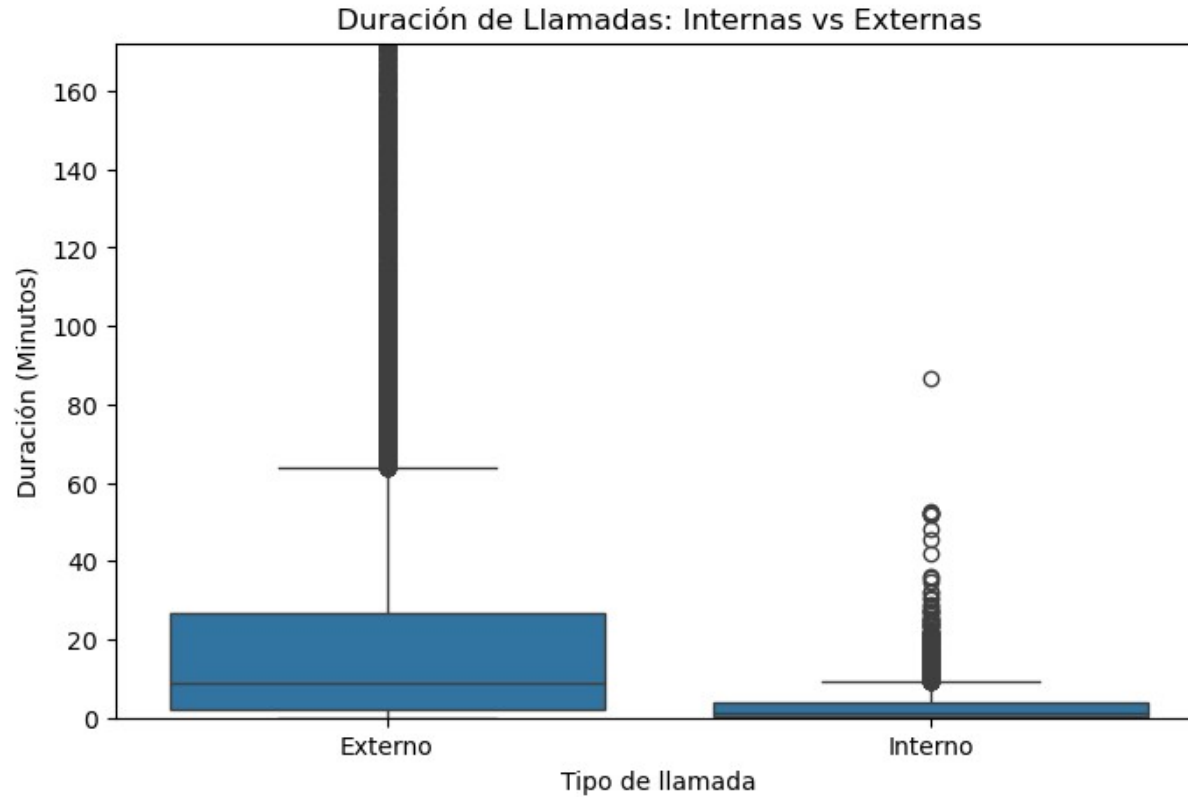
- No se encuentran particularidades especialmente relevantes en los datos.
- Se hacen correcciones de valores vacíos y formatos.

The background of the slide is composed of numerous overlapping rectangular shapes in various shades of gray, from light to dark, and some white. These shapes are scattered across the upper and middle portions of the slide, creating a modern, minimalist aesthetic. The word "Resultados" is prominently displayed in the upper right area.

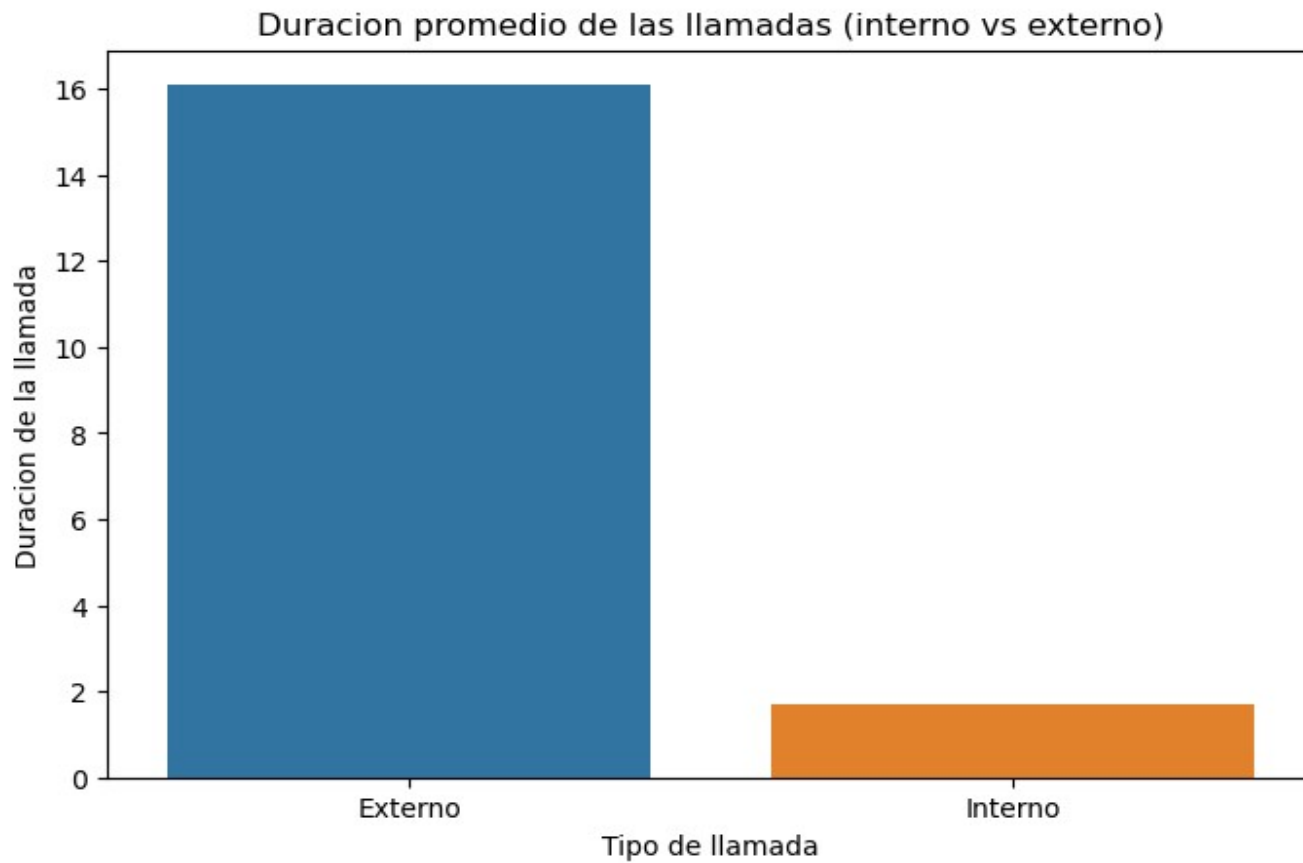
Resultados



Se observa que la mayoría de las llamadas tienen una duración de aproximadamente menor a 10 minutos.



- Con base en la grafica se sugiere que las llamadas externas duran mas que las internas
- Tambien observamos que existen algunos valores outlier que denotan una duracion mayor de las llamadas.



- Observamos que la duración promedio de las llamadas externas es mayor que las internas. (Confirmado con prueba estadística)

T-statistic: -46.96548101210464, P-value: 0.0

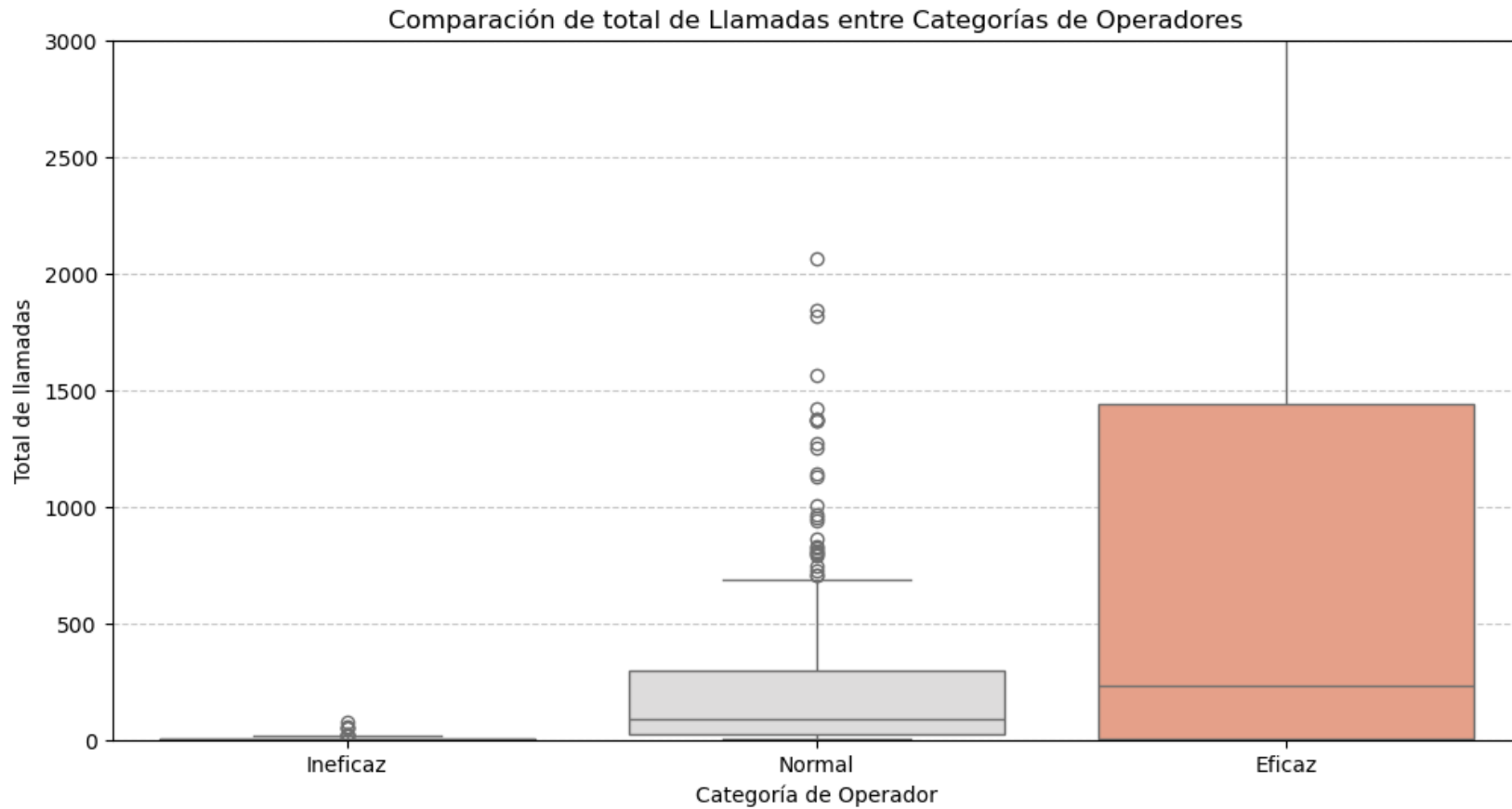
Existe una diferencia estadísticamente significativa: las llamadas internas y externas tienen duraciones diferentes.

📌 Las llamadas externas tienen una duración mayor en promedio.

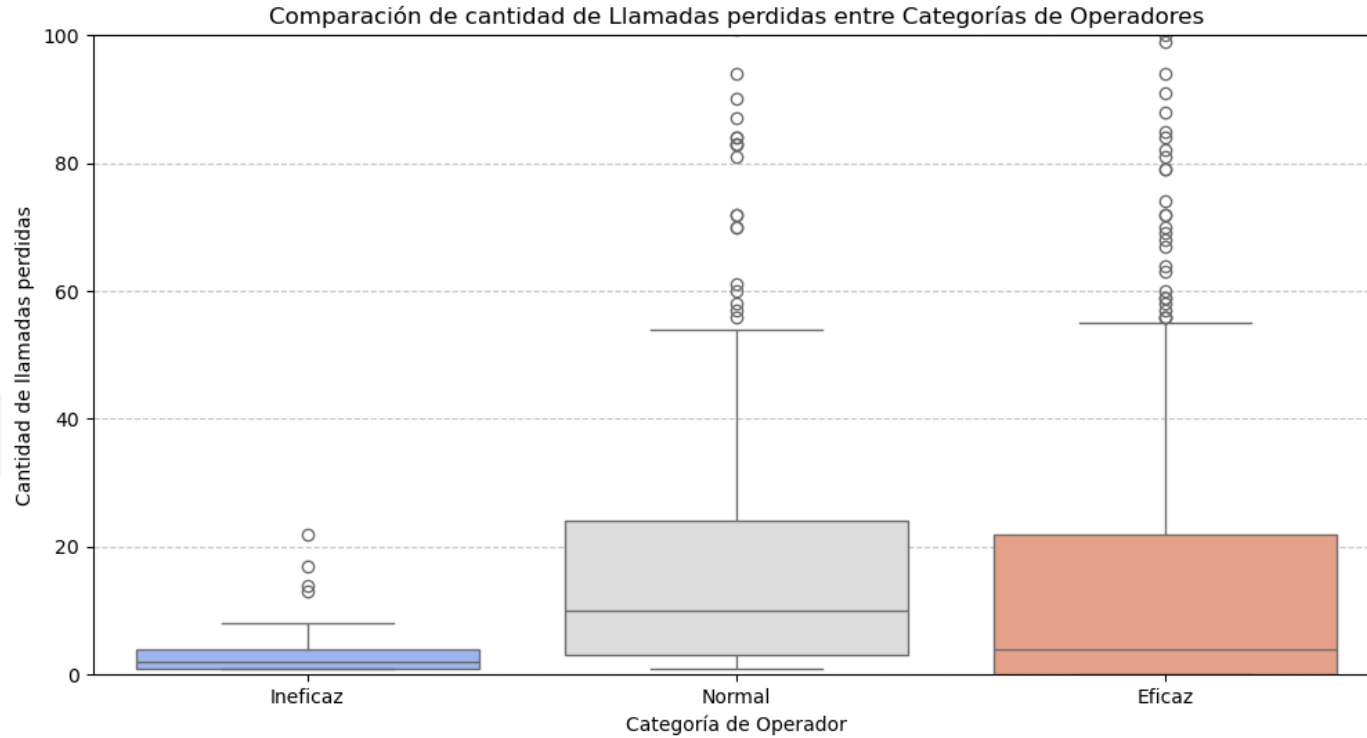
Extracto de la tabla de operadores "Ineficaces "

# operator_id	# total_llamadas	# llamadas_perdidas	# duracion_promedio	# porcentaje_perdidas	📁 categoria
941826.0	1	1	0.0	100.0	Ineficaz
937872.0	1	1	0.0	100.0	Ineficaz
885682.0	3	3	0.0	100.0	Ineficaz
914272.0	1	1	0.0	100.0	Ineficaz
970258.0	1	1	0.0	100.0	Ineficaz
924572.0	1	1	0.0	100.0	Ineficaz
914626.0	1	1	0.0	100.0	Ineficaz
937778.0	1	1	0.0	100.0	Ineficaz
958458.0	1	1	0.0	100.0	Ineficaz
937772.0	1	1	0.0	100.0	Ineficaz

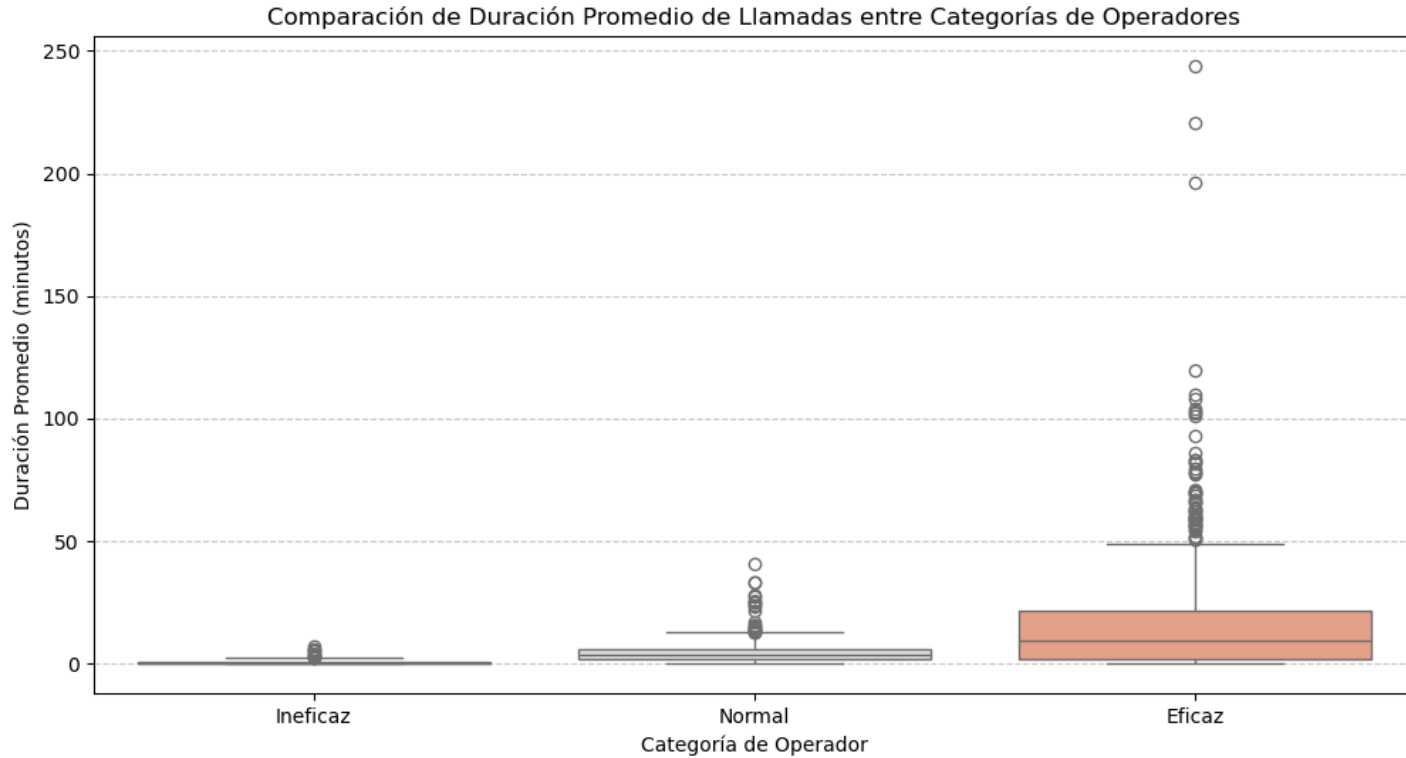
- Para determinar si un operador es ineficaz o no se hace una determinación en función de percentiles.
 - Percentil 50 = Promedio o normales (operadores promedio).
 - Percentil 90 = Operadores en el peor 10% (más ineficaces).
 - Si se encuentran por debajo del percentil 50 son considerados eficaces



- Se observa una mayor cantidad de llamadas para los operadores Eficades que para el resto.



- Observamos que la cantidad de llamadas perdidas es mayor aparentemente para los operadores normales que el resto del grupo, sin embargo los eficaces tienen una cantidad aparentemente cercana



- Con esta ultima grafica se sugiere que los operadores eficaces tienen una mayor duracion promedio de llamadas

- Al separar por percentiles identificamos tres grupos si son ineficientes, promedio o eficientes. Para saber si existen diferencias entre estos 3 grupos es necesario emplear un ANOVA en lugar de una prueba T para evitar un error de comparación múltiple.
 - Se observa que los operadores ineficaces y con desempeño promedio son menores respecto a los eficaces en la duracion de llamadas.
 - La cantidad de llamadas perdidas entre los grupos de operadres eficientes y normales es mayor a los ineficaces.
 - Y finalmente observamos que los operadores eficaces tienen una mayor duracion de llamada.
- Esto señala que los operadores ineficaces son los mas inactivos al atender una menor cantidad de llamadas tener una menor duracion, y eso repercute en un mayor porcentaje de llamadas perdidas ya que, por ejemplo. de una llamada que recibieron no la atendieron, esto significando una proporcion de llamadas perdidas del 100%.

Dif Total de llamadas

- Observamos que existe diferencias entre el total de llamadas por tipo de operador.
- Sin embargo el tamaño del efecto (eta cuadrada es muy pequeño)

```
pg.anova(data=operador_ineficaz, dv='total_llamadas', between='categoria')[['F', 'p-unc', 'np2']]
```

✓ 0.0s

	# F	# p-unc	# np2
0	5.18775065274911	0.005723266195070591	0.009429055166339679

Dif Llamadas perdidas

- Aquí observamos, por el contrario que no existen diferencias entre los grupos.
- De igual manera el tamaño del efecto es muy pequeño

```
pg.anova(data=operador_ineficaz, dv='llamadas_perdidas', between='categoria')[['F', 'p-unc', 'np2']]
```

	# F	# p-unc	# np2
0	1.2359908304395701	0.2909536186266607	0.0022627414728943434

Dif Duracion promedio de llamada

- Aqui vemos que efectivamente existe una diferencia considerable entre grupos para la duracion de la llamada.
- Y aunque el tamaño del efecto es reducido, se observa que es mayor al resto de las comparaciones.

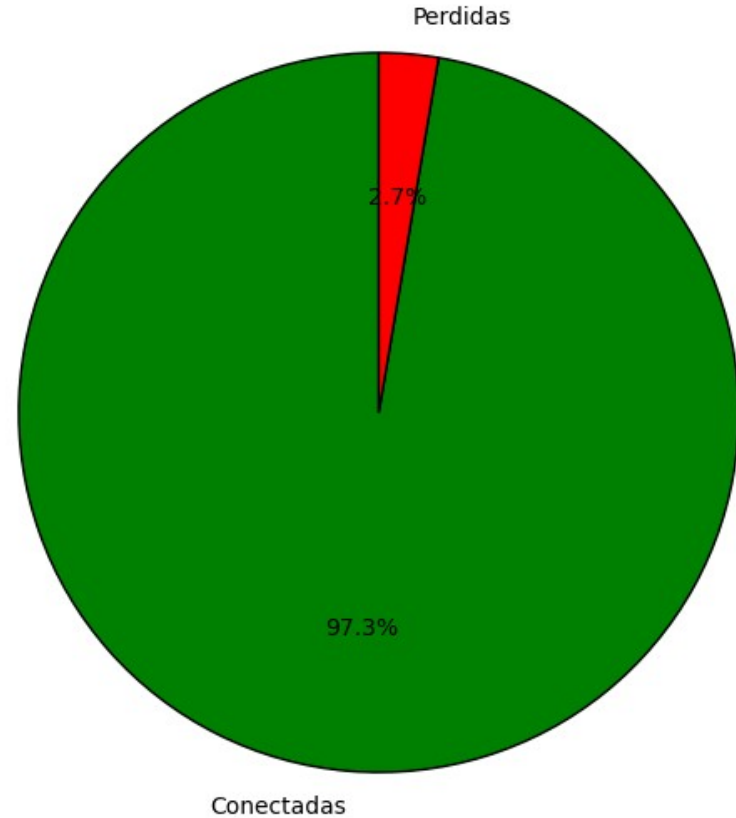
```
pg.anova(data=operador_ineficaz, dv='duracion_promedio', between='categoria')[['F', 'p-unc', 'np2']]
```

0.0s

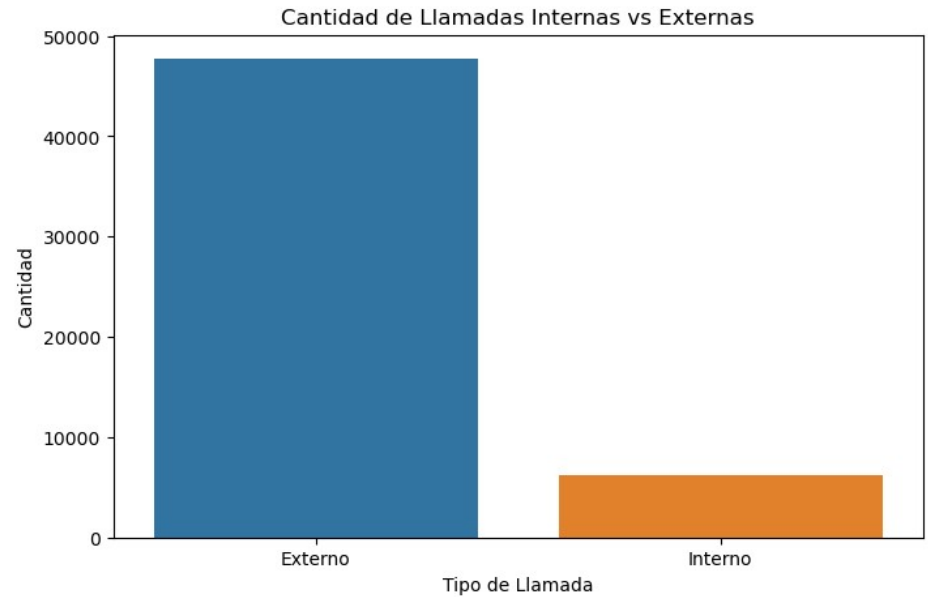
# F	# p-unc	# np2	
0	74.47935571953283	4.802496222749495e-31	0.12022895522163794

- Se observa que el 97.3% de las llamadas son enlazadas lo cual muestra una eficiencia general en ese sentido.

Proporción de Llamadas Conectadas vs. Perdidas



- Confirmamos de manera visual que las llamadas externas son mayores que las internas.



Conclusion

- El presente proyecto tuvo como objetivo proveer informacion que pueda ayudar a tomar desiones empresariales.
 - Con base en esto hemos identificado a los operadores "ineficaces" como aquellos operadores que tienen una menor actividad, se sugiere indagar al respecto con el area encargada de los operadores, y ver porque es que estos operadores estan teniendo una menor actividad respecto a sus compañeros.
 - Se corrobora que las llamadas externas son mayores a las internas. y estas tienen una mayor duracion respecto a las internas.
 - Se observa de igual manera que la mayoria de las llamadas fueron conectadas exitosamente

