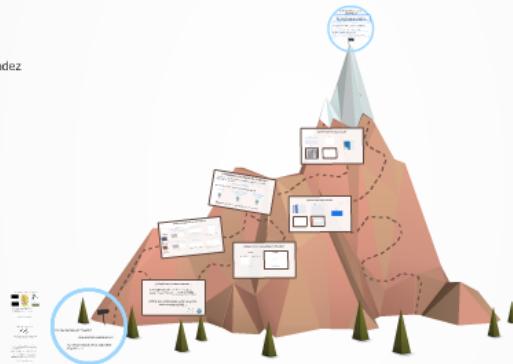


# Análisis de una Red de Telecomunicaciones

*Como encaramos nuestro primer Proyecto de Ciencia de Datos y ... sobrevivimos al intento.*

Integrantes  
• Sergio Sulca  
• Carlos Bacia  
• Soledad Fernández

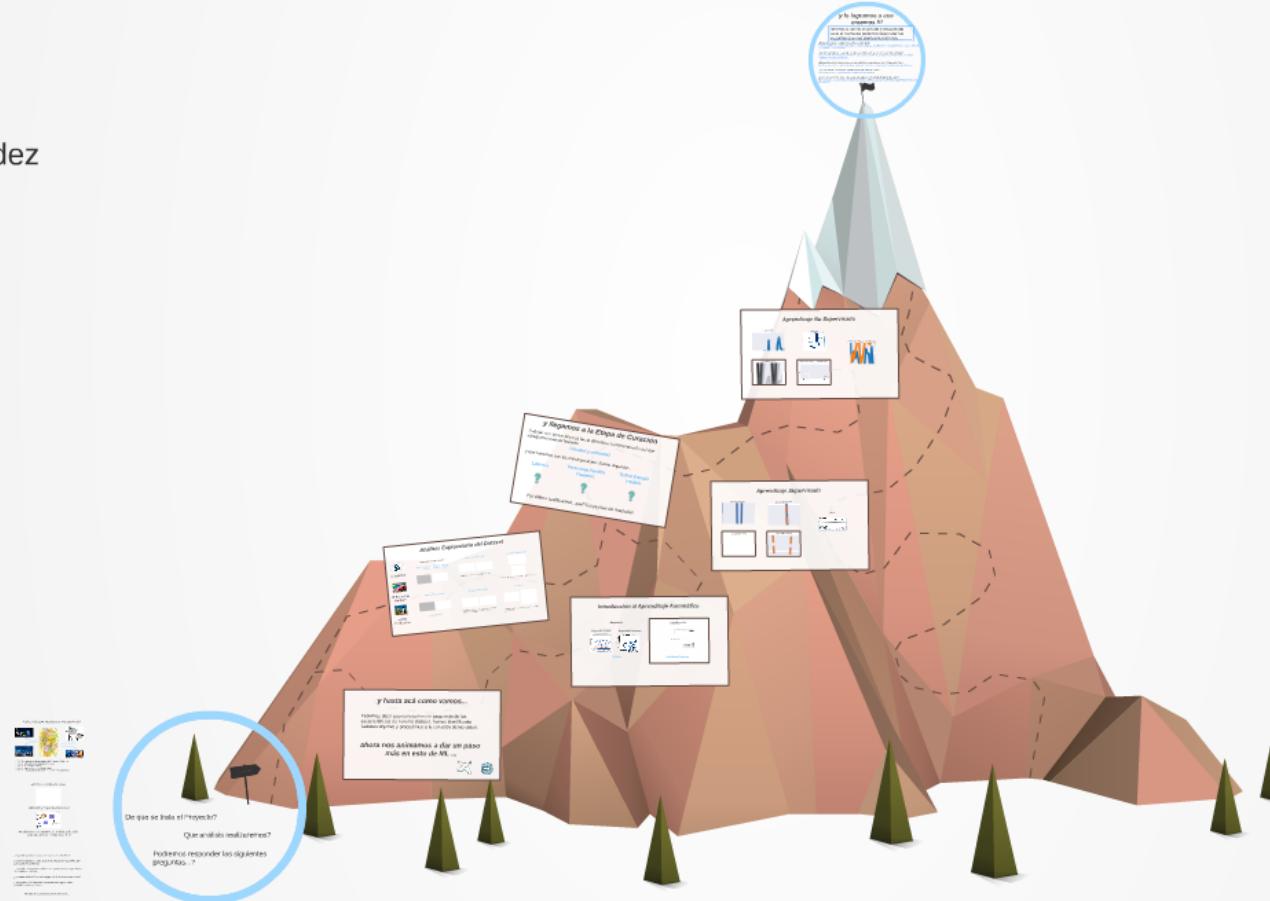


# Análisis de una Red de Telecomunicaciones

Como encaramos nuestro primer Proyecto de Ciencia de Datos y ... sobrevivimos al intento.

## Integrantes

- Sergio Sulca
- Carlos Bacia
- Soledad Fernández



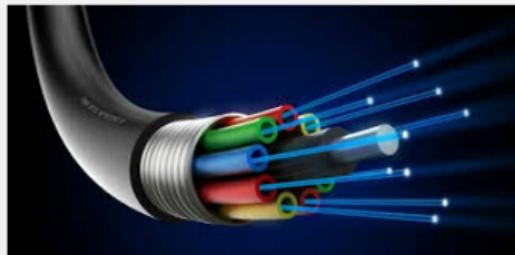


De que se trata el Proyecto?

Que análisis realizaremos?

Podremos responder las siguientes  
preguntas...?

# Red de Telecomunicaciones a nivel provincial



- Red Teleoperación e Interconexión de Estaciones Eléctricas
- Red de Telemedición y Transporte de Datos
- Red de Infraestructura de TI
- Red de Protección del Tendido Eléctrico
- Red para Provisión/Comercialización de Internet Corporativo



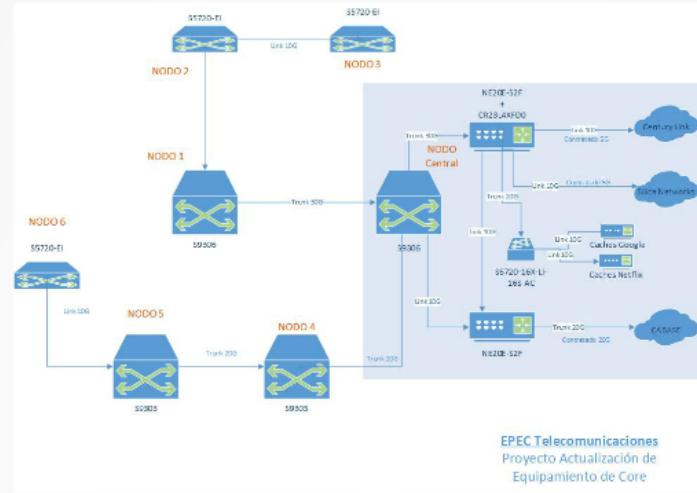


De que se trata el Proyecto?

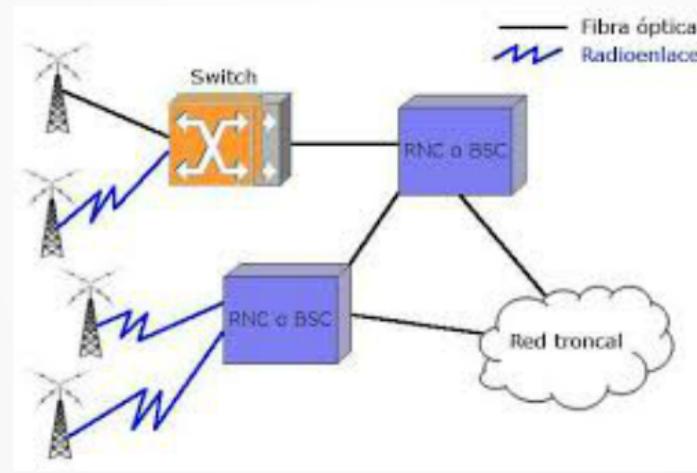
Que análisis realizaremos?

Podremos responder las siguientes  
preguntas...?

# Análisis del Tráfico de la Red



## Utilización y Capacidad de la Red



Predicciones de Consumo, saturación, calidad de la información que transporta la Red



De que se trata el Proyecto?

Que análisis realizaremos?

Podremos responder las siguientes  
preguntas...?

¿Es posible predecir cortes o desperfectos en la Red?

¿Podemos predecir el uso de la Red en las próximas 24 horas?

¿En la próxima semana?

¿Es posible detectar anomalías o comportamientos sospechosos en el tráfico de la Red?

¿La actividad de la Red varía según el día de la semana o mes?

¿Los cortes en la Red están asociados con alguna otra/s variable/s del set de datos?

*Al final de la presentación lo veremos....*

# Análisis Exploratorio del Dataset



11 Equipos



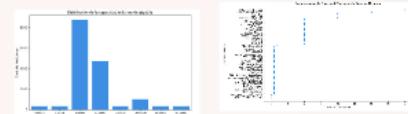
48 Puntos de Medición



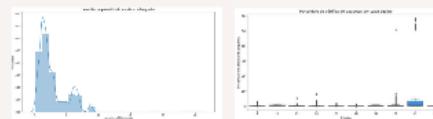
61548  
Mediciones

Cuales son las features objetivo?

Latencia Pérdida de Paquetes Tráfico de Entrada Tráfico de Salida

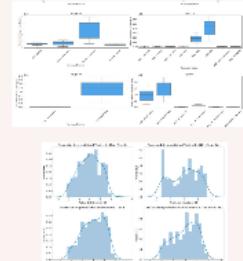


Analizamos la Feature Latencia



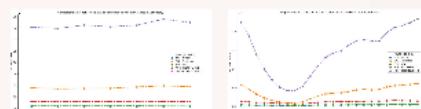
Detectamos anomalías en ciertos punto de medición...

Analizamos la Feature Tráfico



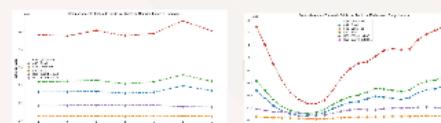
Empezamos por el set de datos completo pero no fue buena elección....

Analisis del Tráfico de Entrada



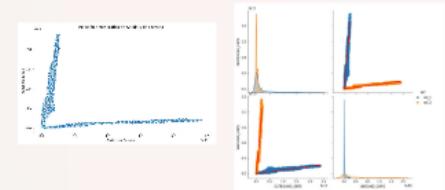
Como se comporta?

Analisis del Tráfico de Salida



y descubrimos básicamente que tiene un comportamiento similar al tráfico de entrada.... era un poco obvio!

Correlaciones



Podemos decir que hipotéticamente si pero... encontramos algo inesperado



Prezi

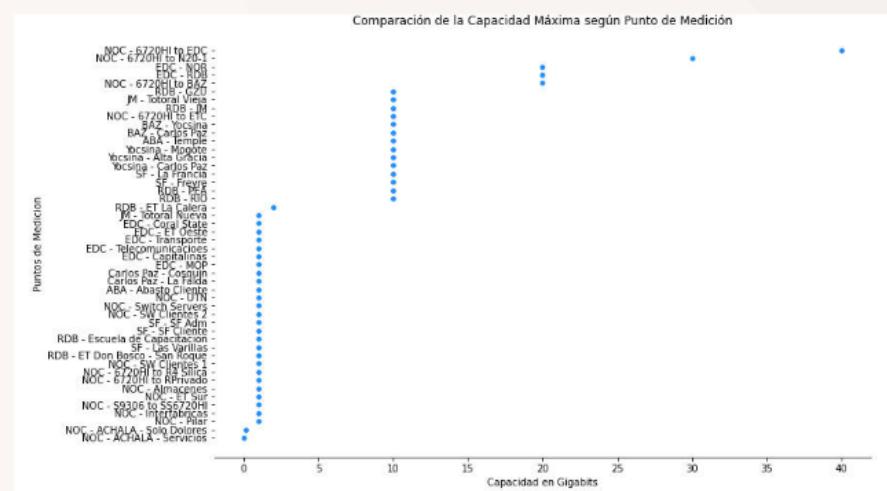
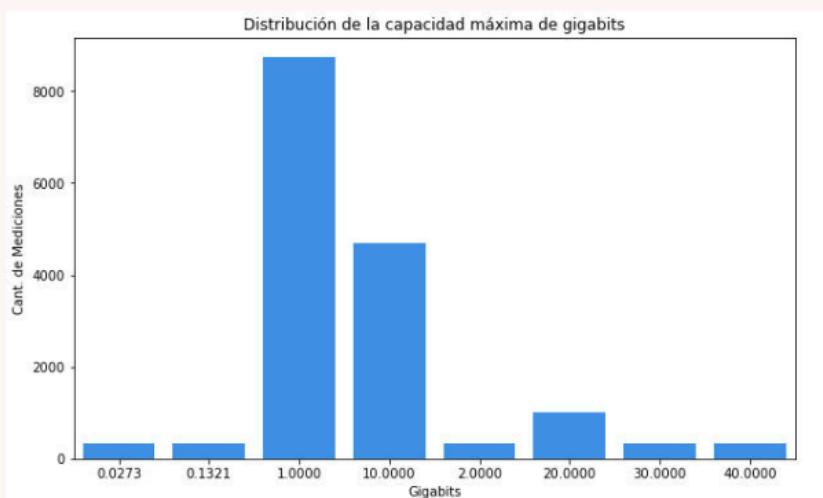
# Cuales son las features objetivo?

Latencia

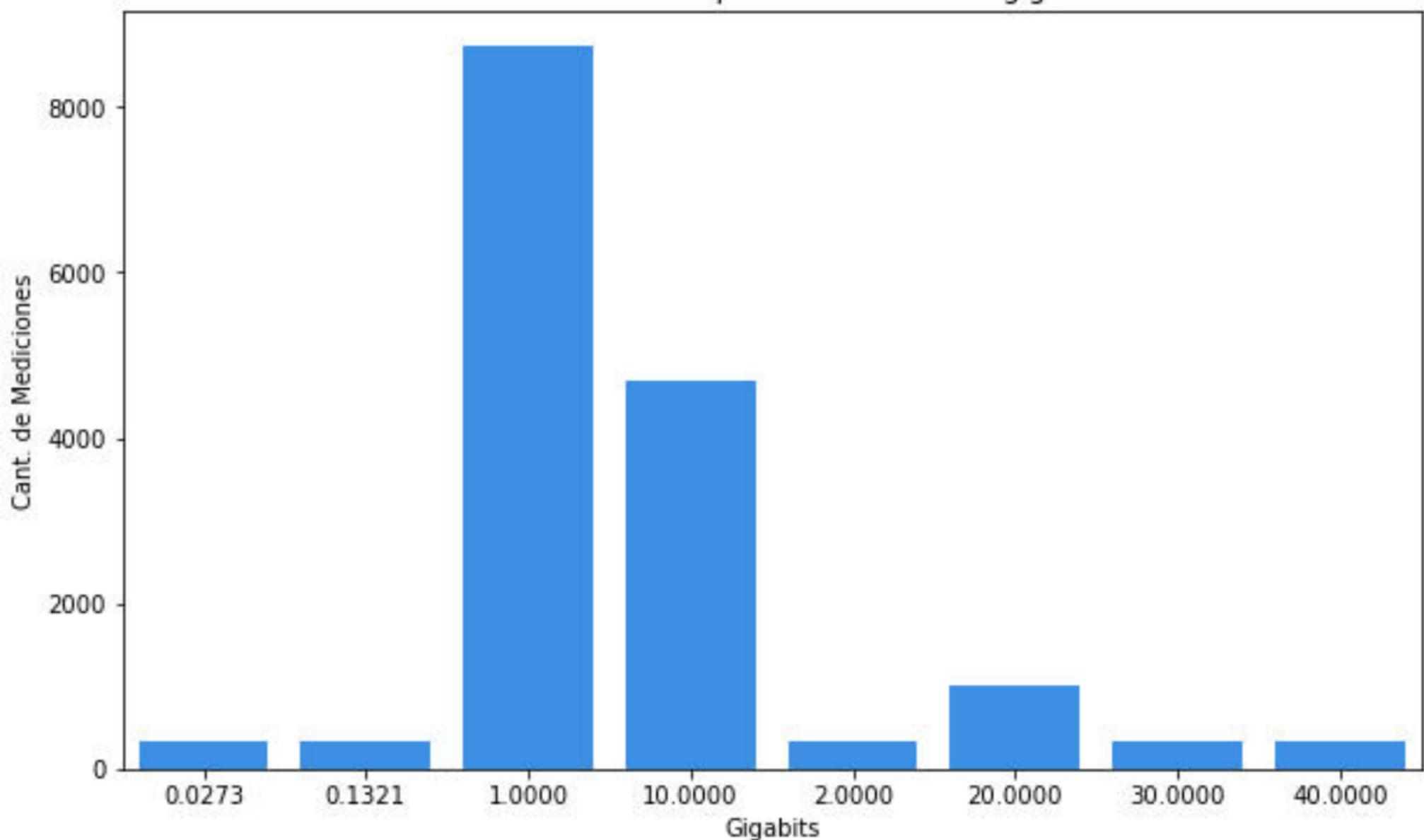
Pérdida de Paquetes

Tráfico de Entrada

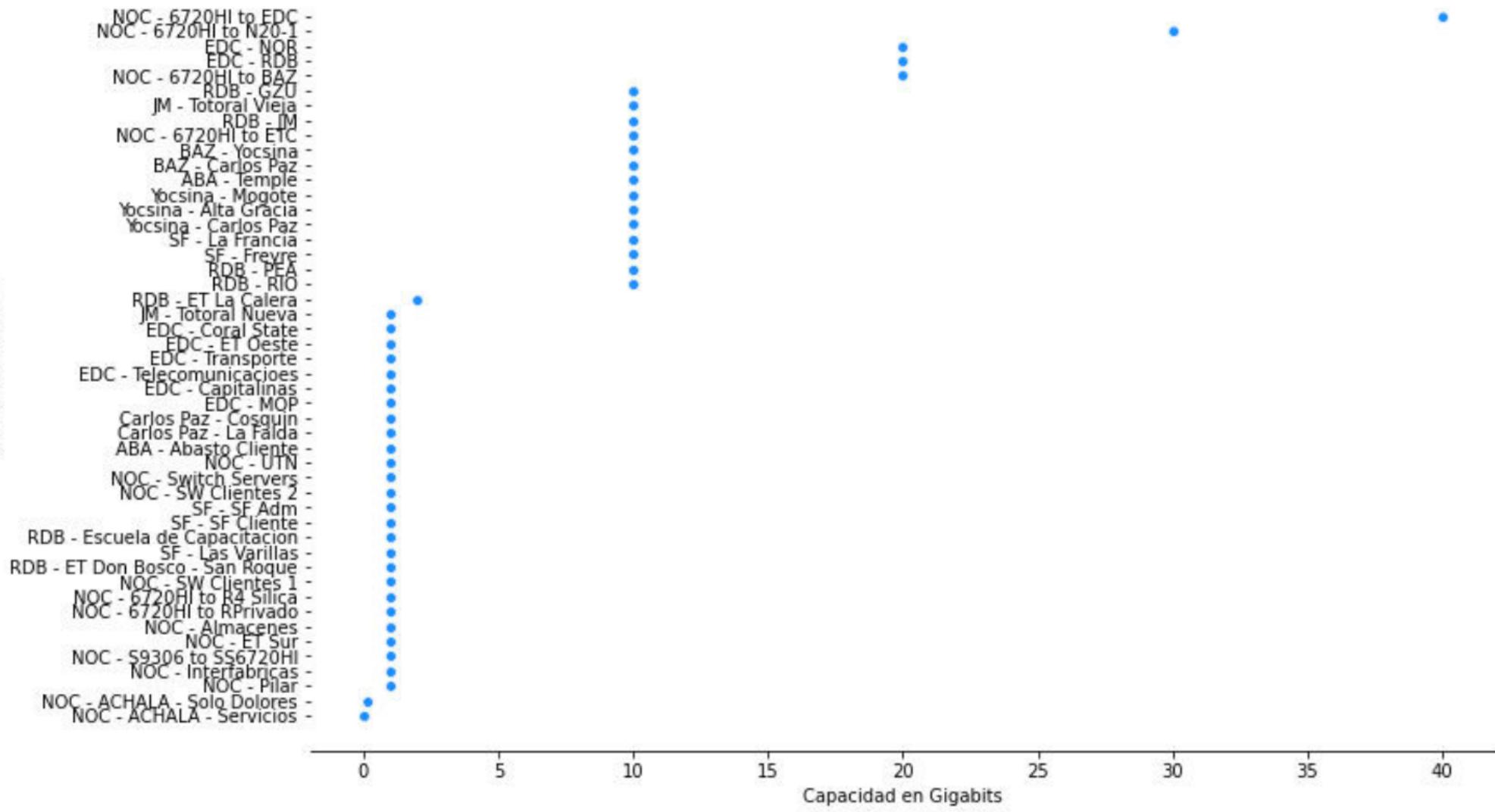
Tráfico de Salida



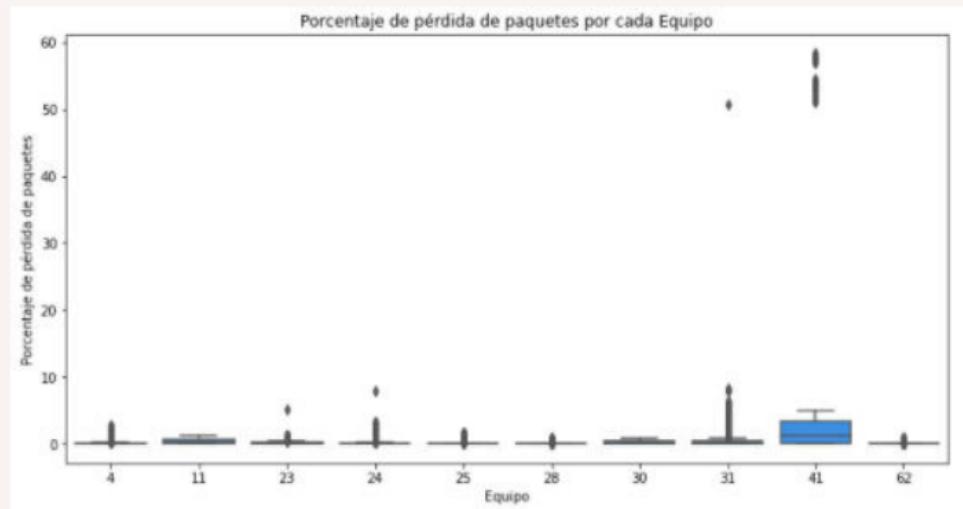
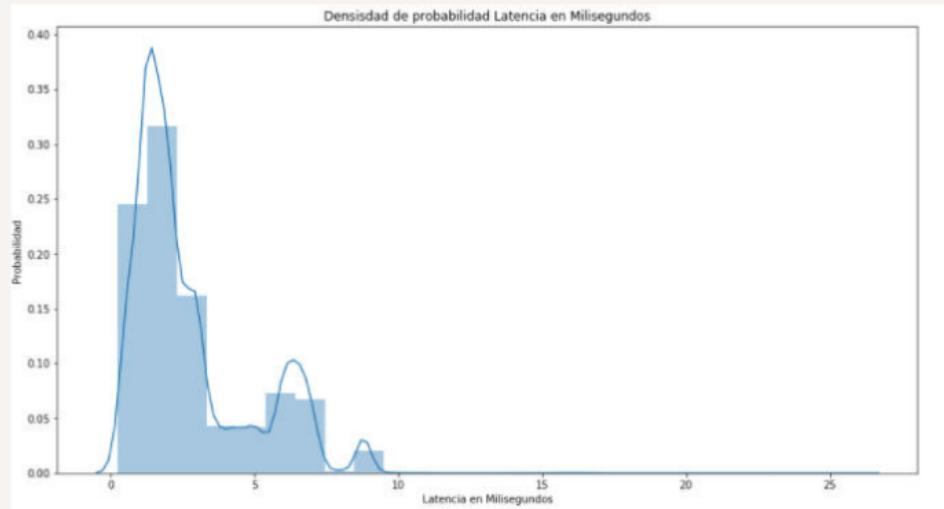
Distribución de la capacidad máxima de gigabits



Comparación de la Capacidad Máxima según Punto de Medición

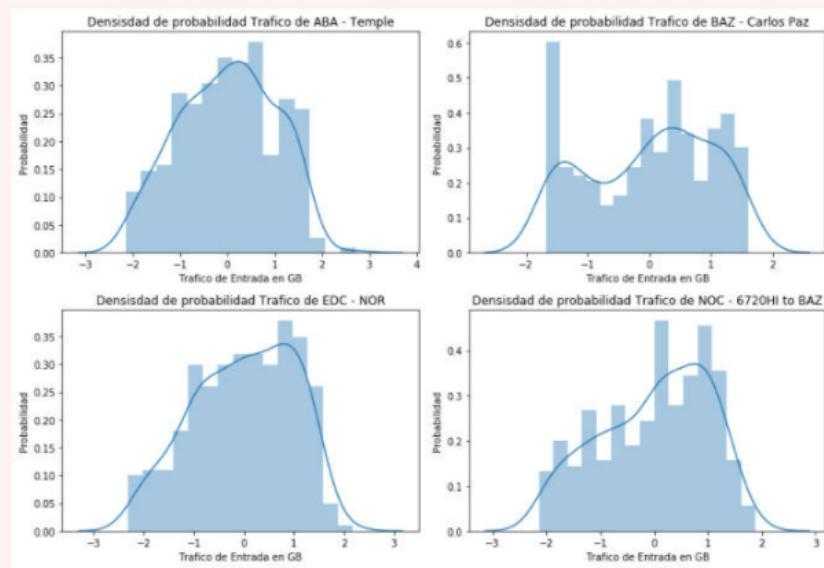
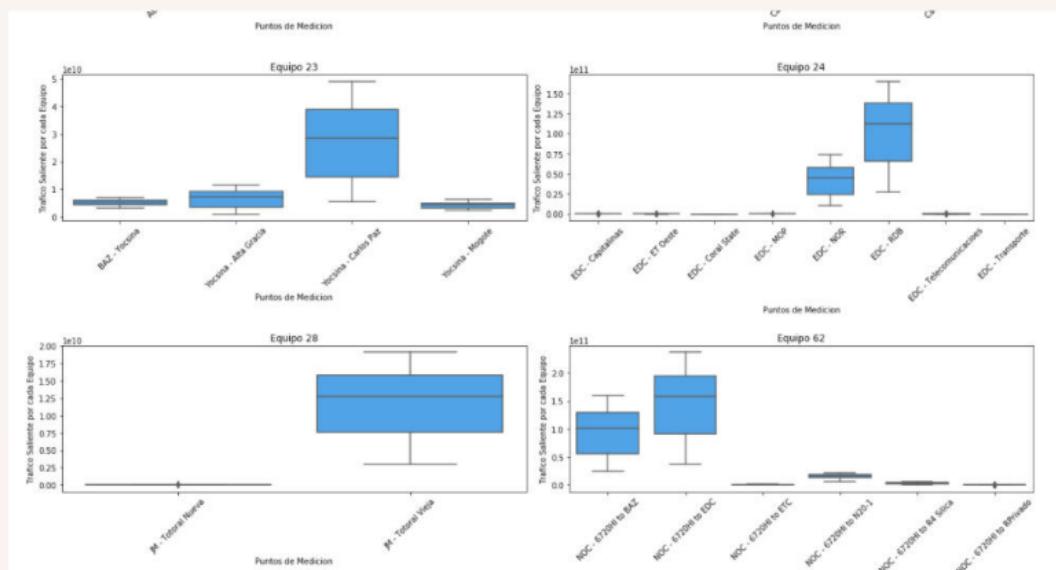


# Analizamos la Feature Latencia



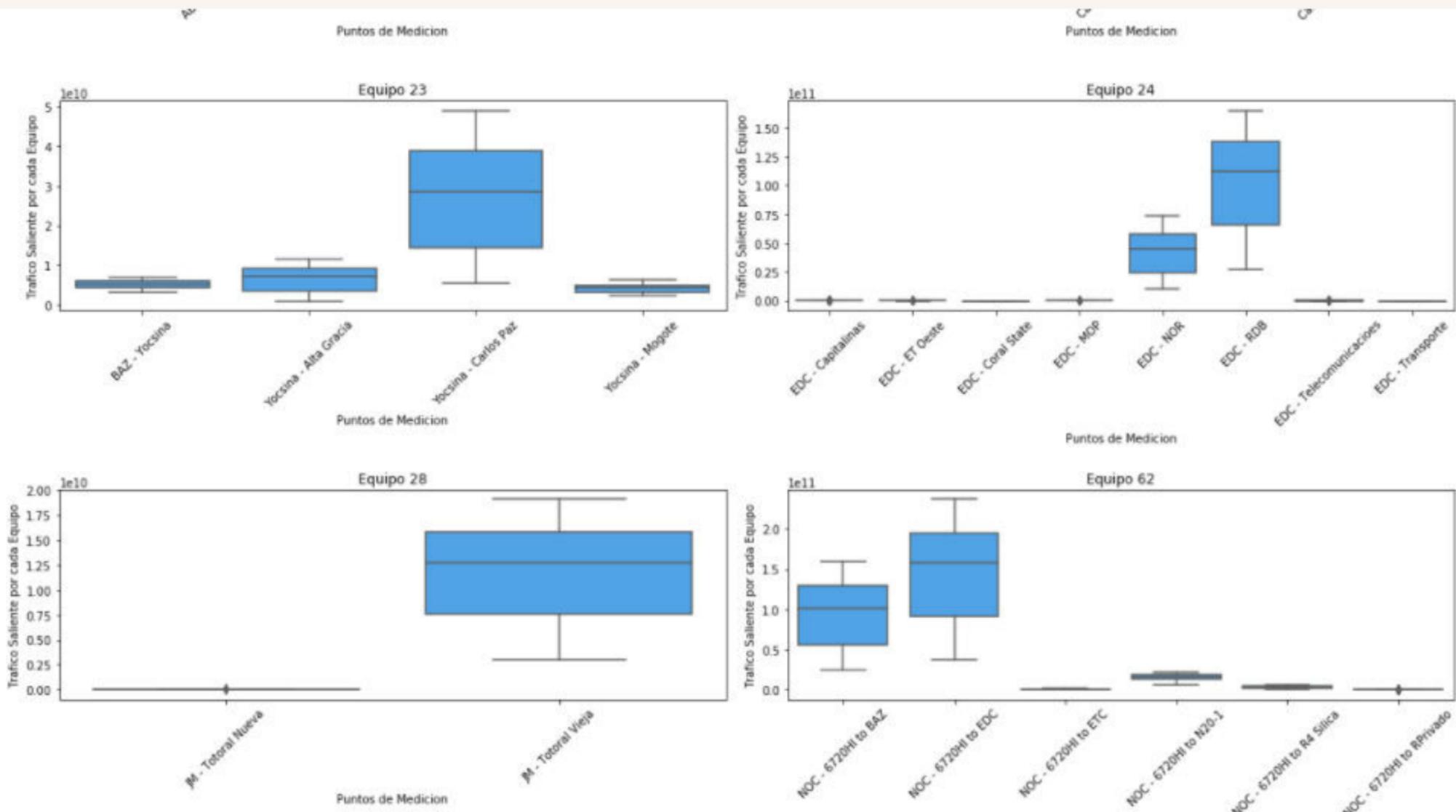
Detectamos anomalías en ciertos punto de medición...

# Analizamos la Feature Tráfico

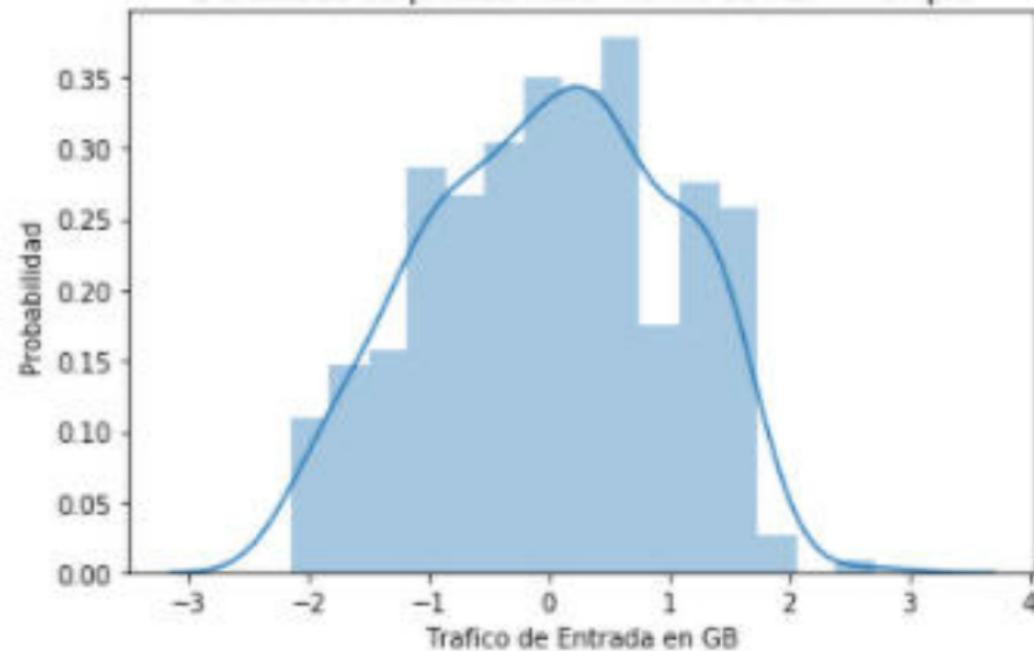


Empezamos por el set de datos completo pero no fue buena elección....

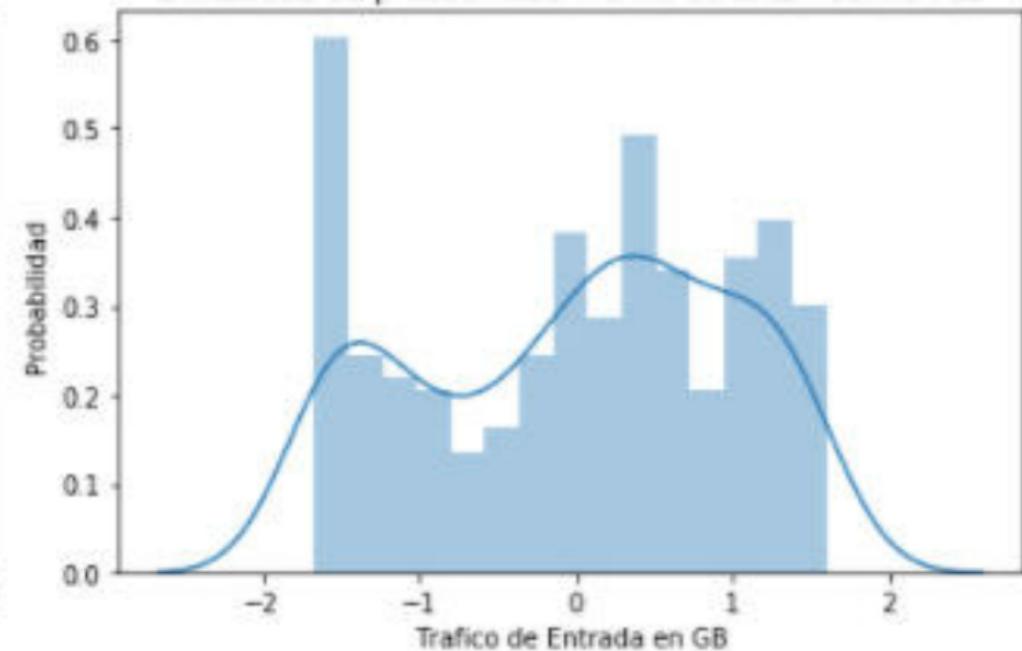
# Analizamos la Feature Tráfico



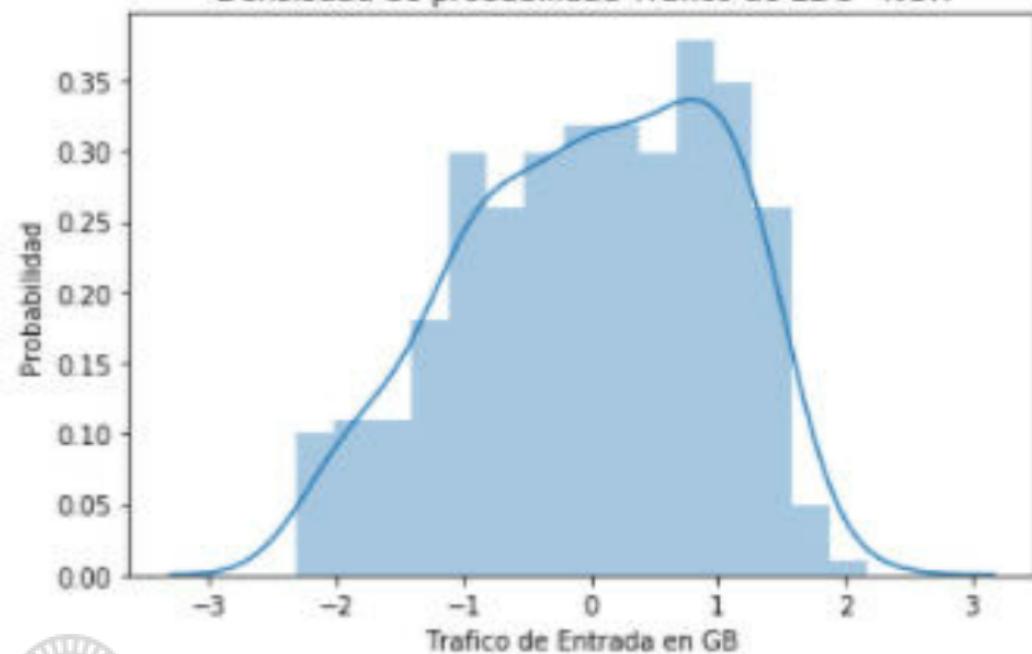
Densidad de probabilidad Trafico de ABA - Temple



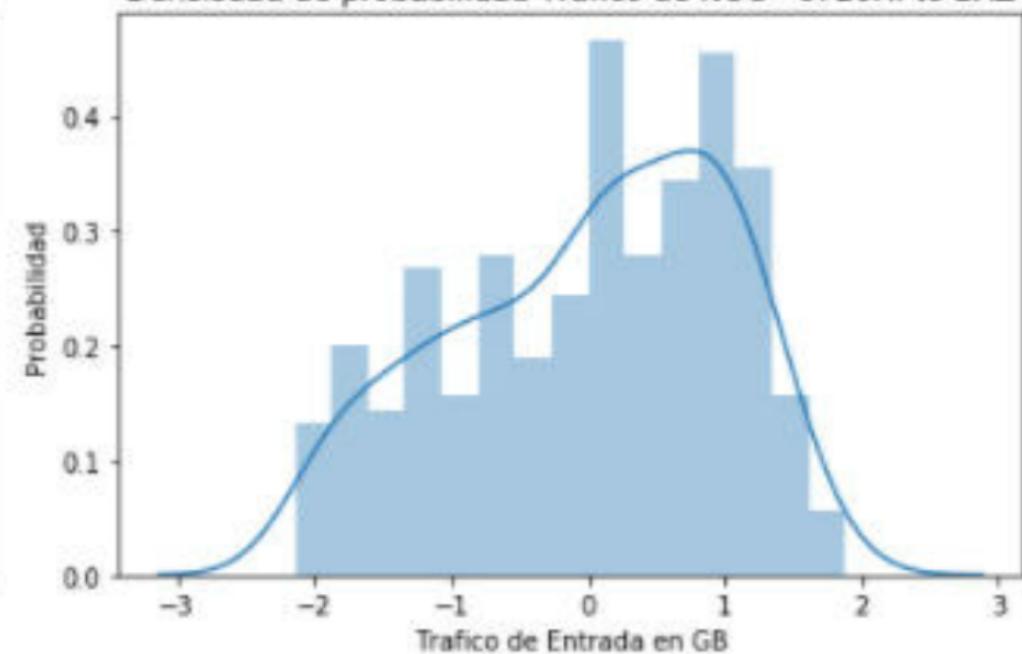
Densidad de probabilidad Trafico de BAZ - Carlos Paz



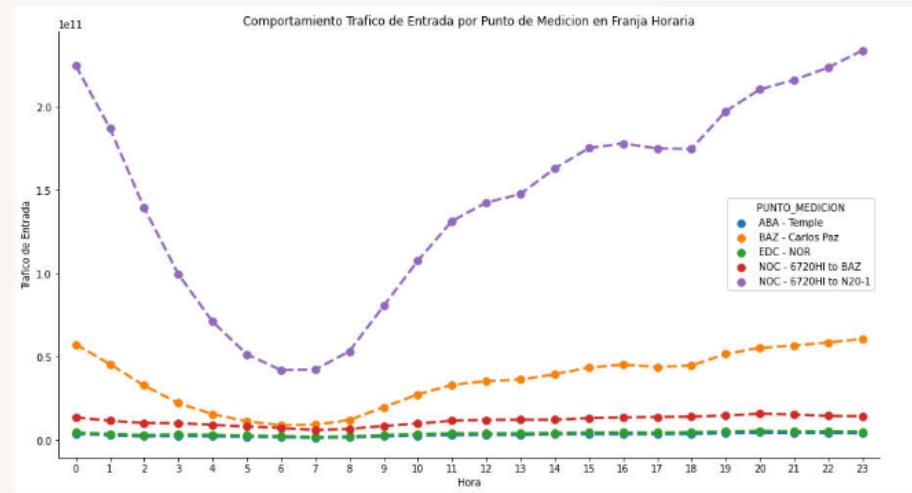
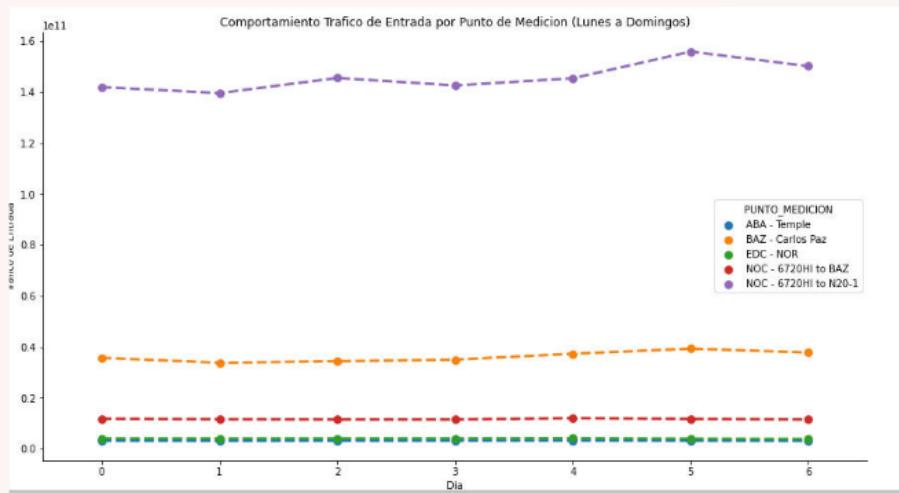
Densidad de probabilidad Trafico de EDC - NOR



Densidad de probabilidad Trafico de NOC - 6720HI to BAZ

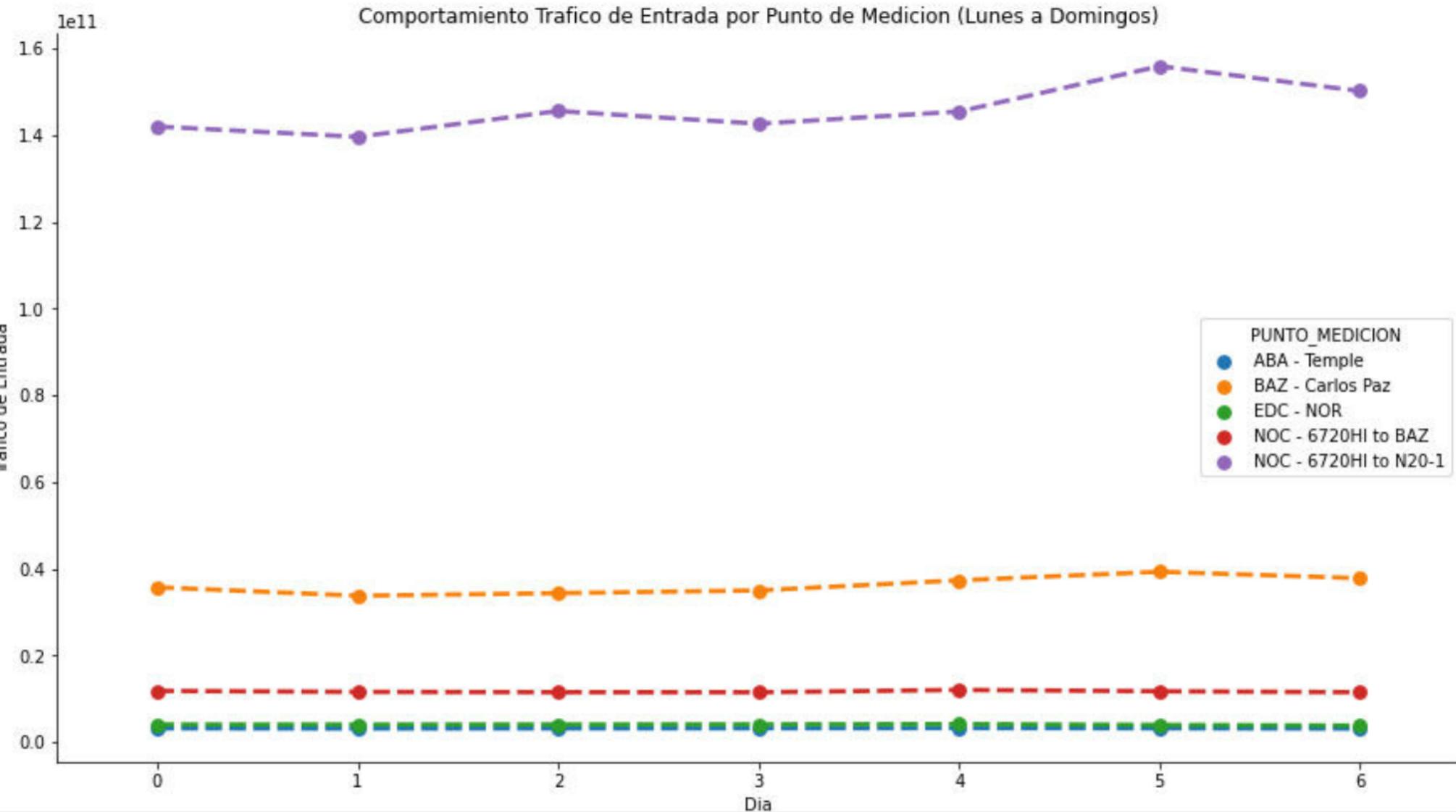


# Análisis del Tráfico de Entrada

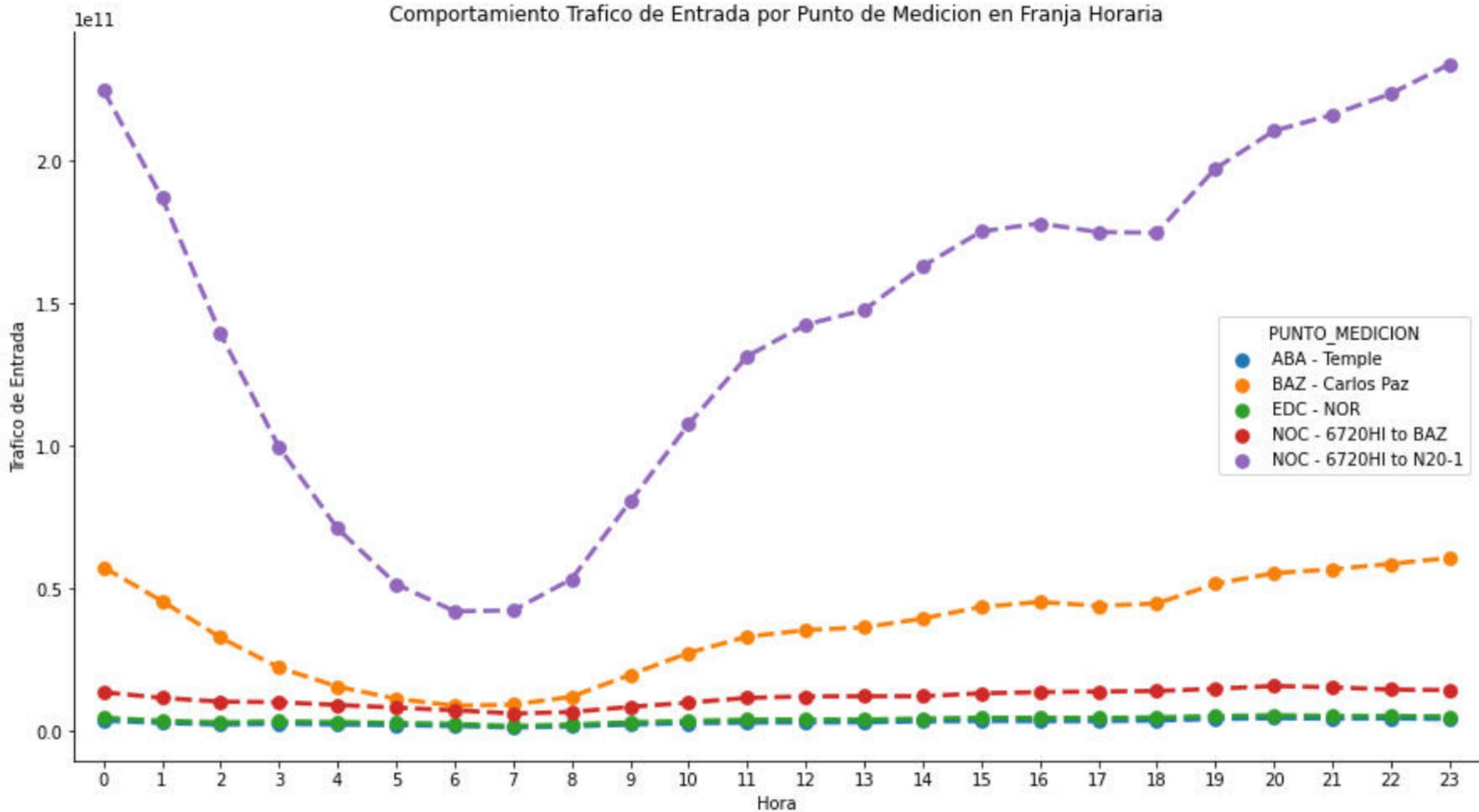


Como se comporta?

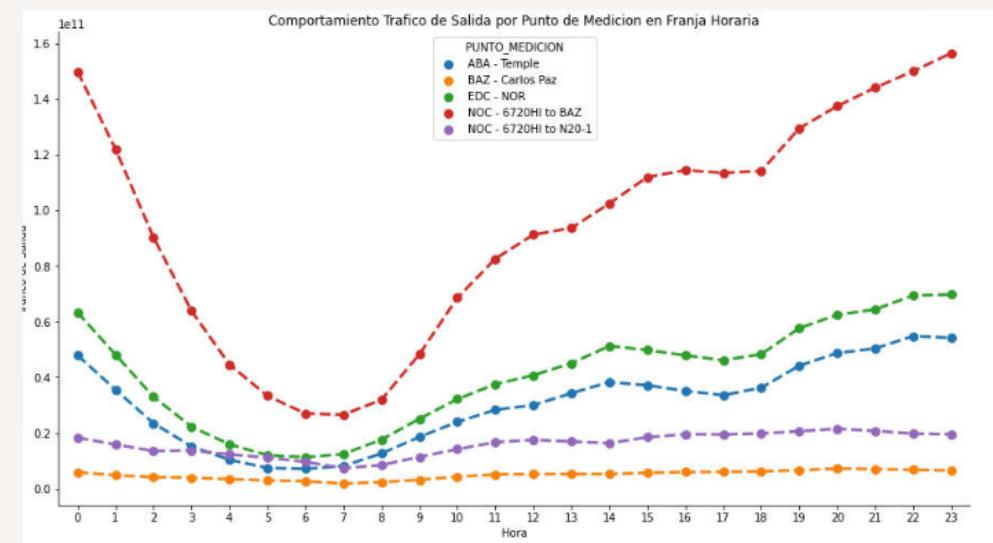
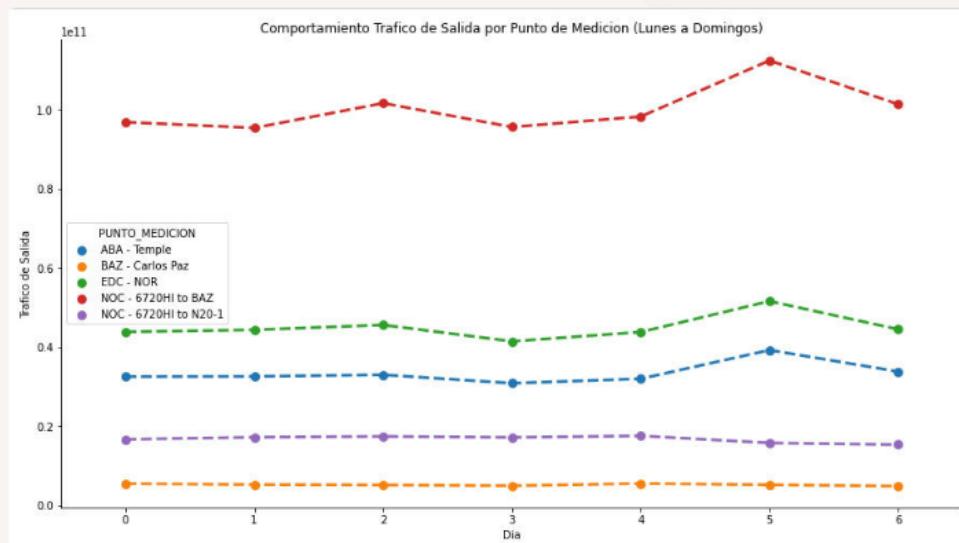
Comportamiento Trafico de Entrada por Punto de Medicion (Lunes a Domingos)



### Comportamiento Trafico de Entrada por Punto de Medicion en Franja Horaria

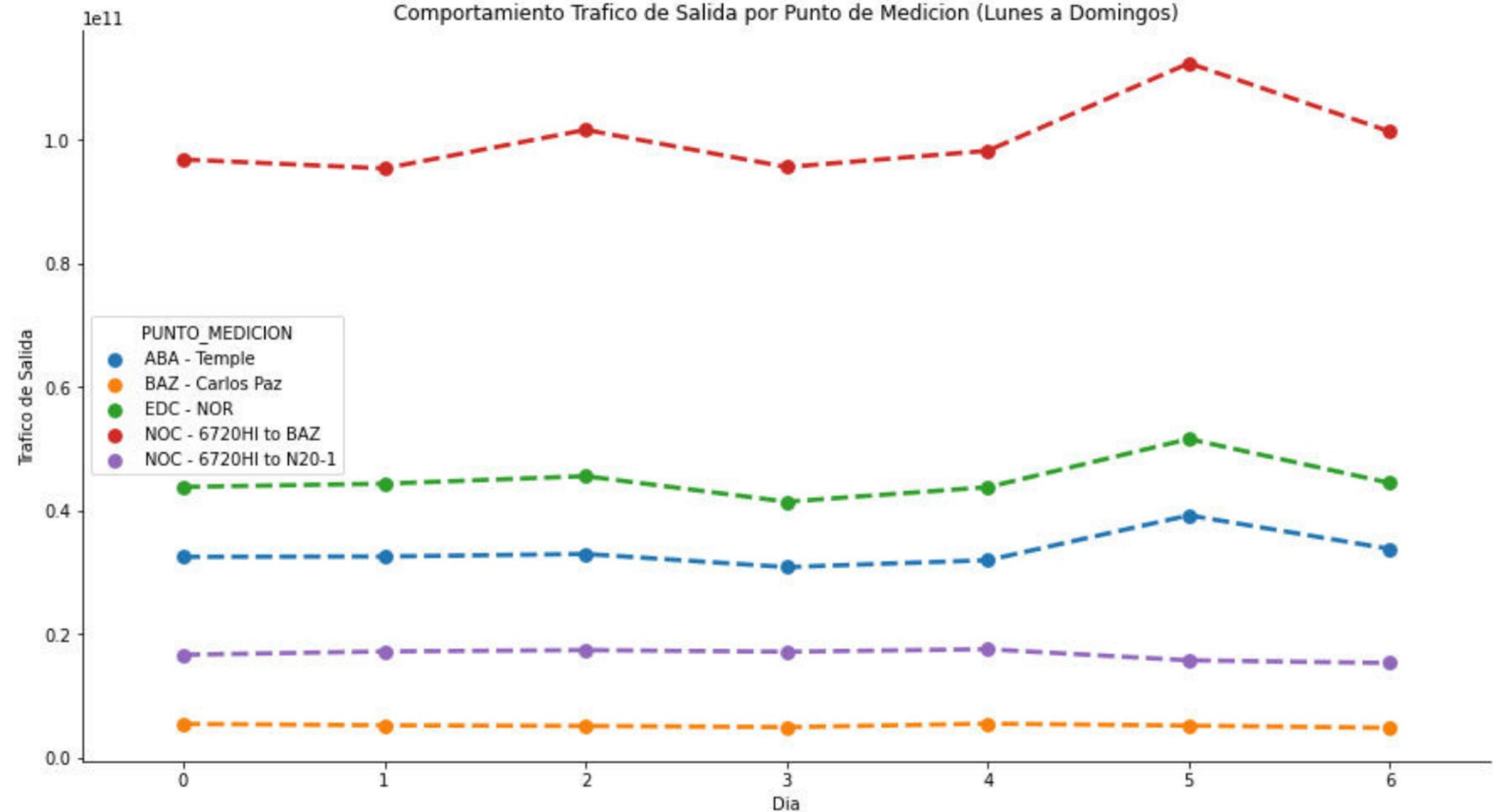


# Análisis del Tráfico de Salida

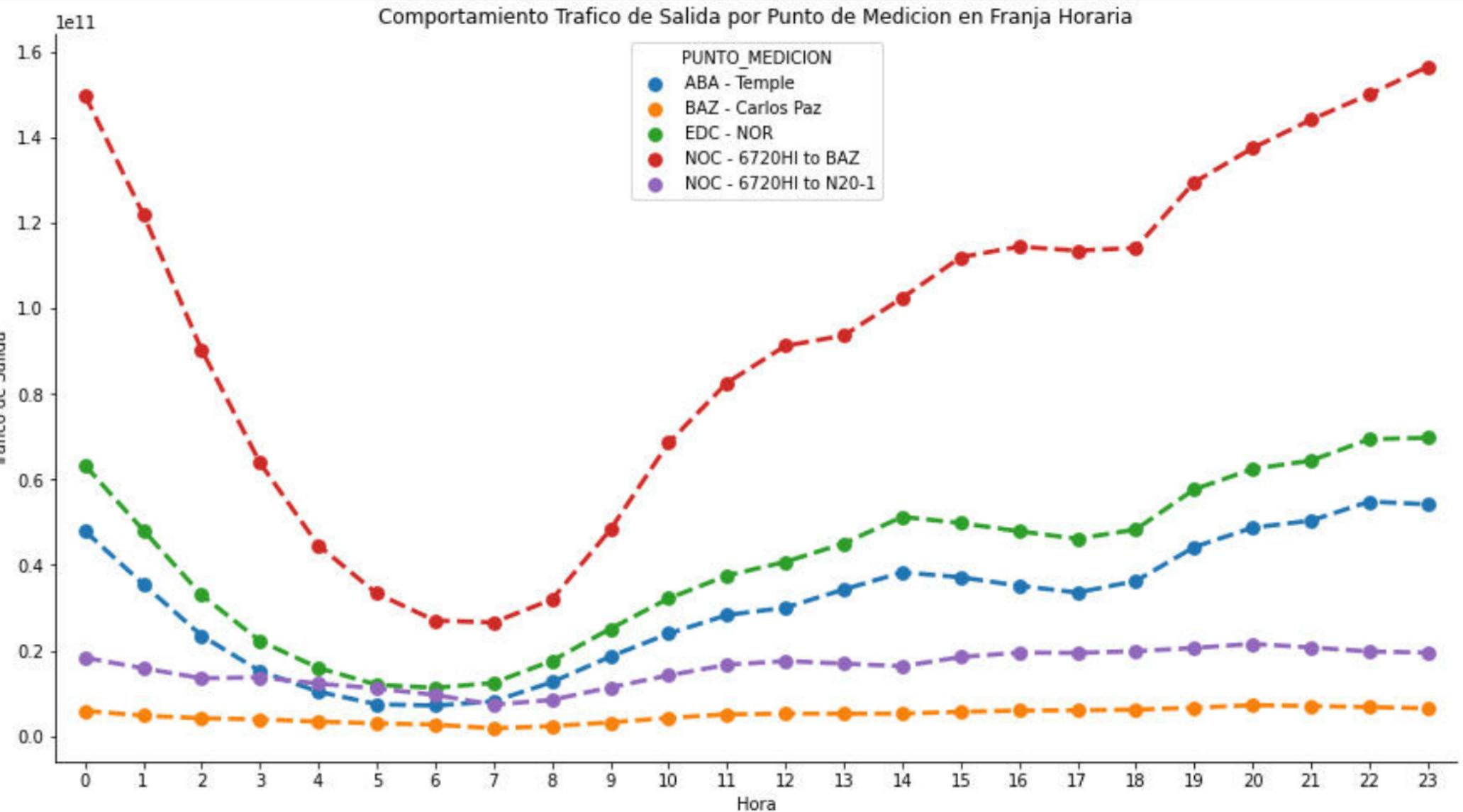


y descubrimos básicamente que tiene un comportamiento similar al tráfico de entrada.... era un poco obvio!

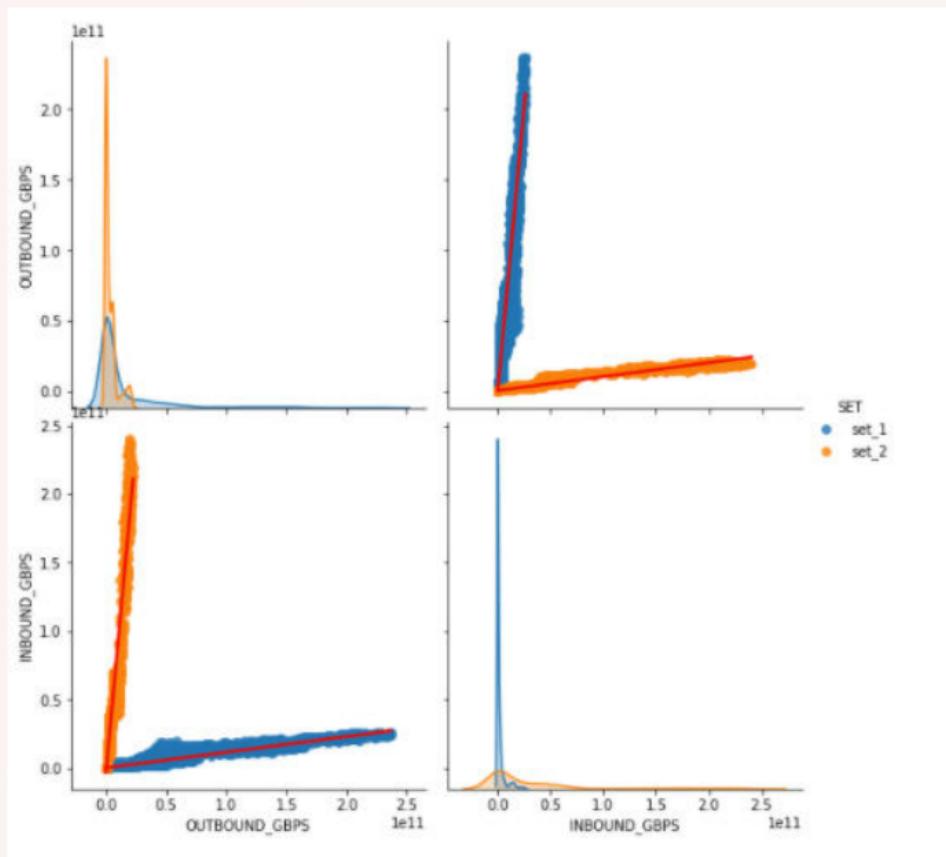
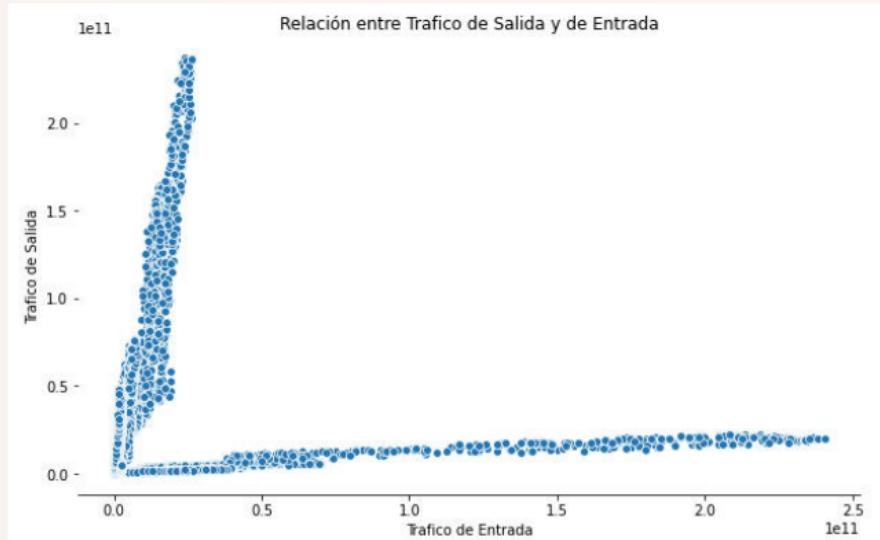
### Comportamiento Trafico de Salida por Punto de Medicion (Lunes a Domingos)



### Comportamiento Trafico de Salida por Punto de Medicion en Franja Horaria



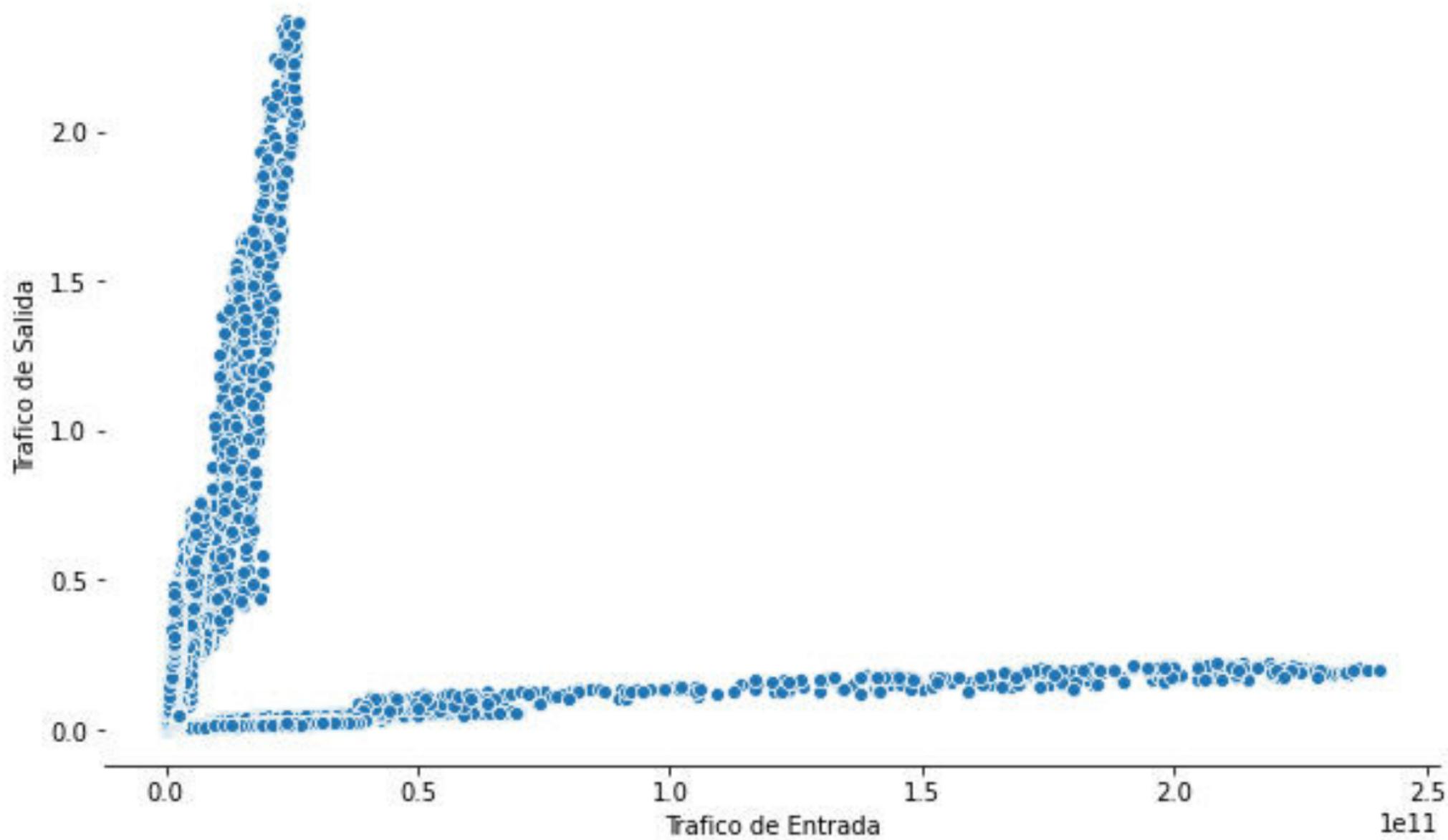
# Correlaciones

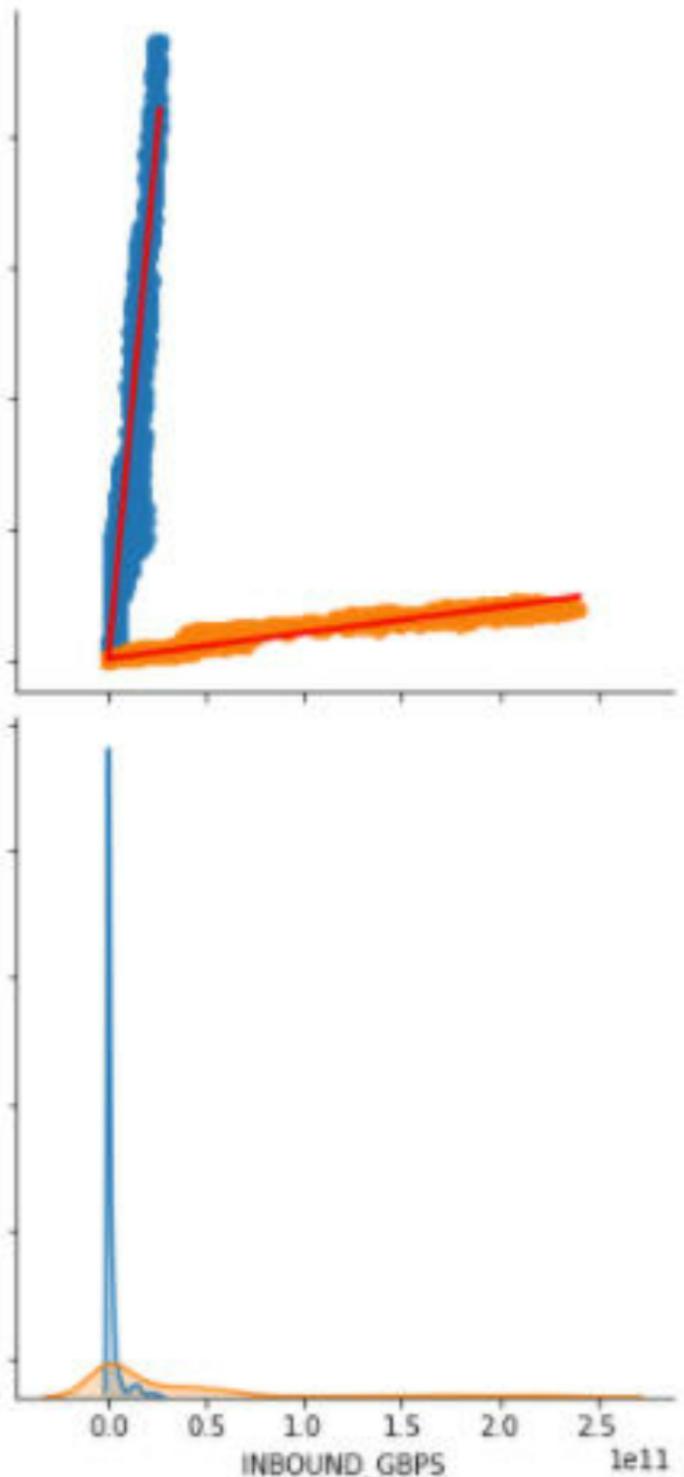
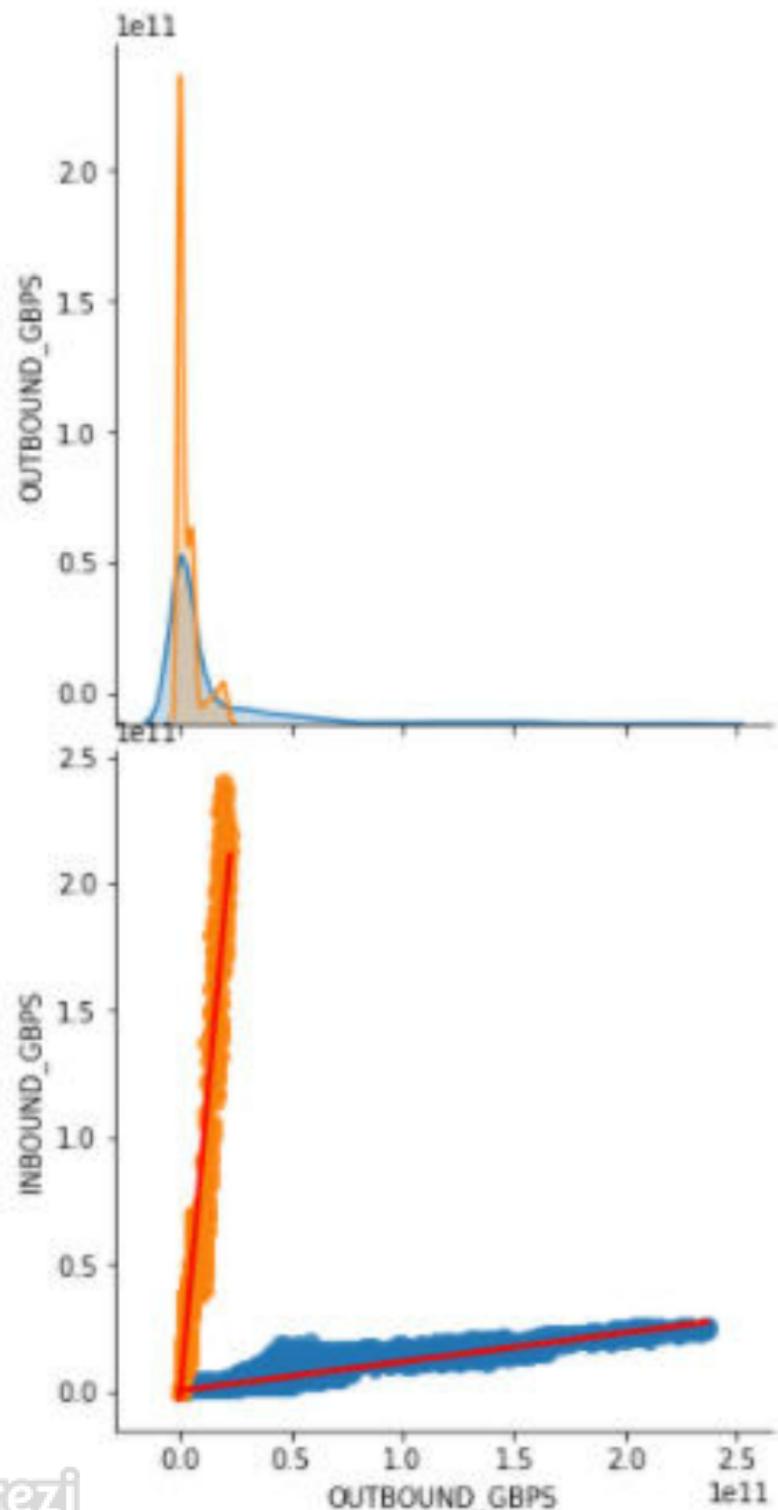


Podemos decir que hipotéticamente si pero... encontramos algo inesperado

le11

### Relación entre Trafico de Salida y de Entrada





## *y llegamos a la Etapa de Curación*

Trabajar con tantos bits nos hacía dificultosa la interpretación así que agregamos nuevas features

**inbound y outbound**

y que hacemos con los missing values? Bueno depende...

**Latencia**



**Porcentaje Perdida  
Paquetes**



**Tráfico Entrada  
y Salida**



Por último codificamos, que? los puntos de medición!



Prezi

*y hasta acá como vamos...*

Podemos decir que conocemos un poco más de las características de nuestro dataset, hemos identificado features objetivo y procedimos a la curación de los datos.

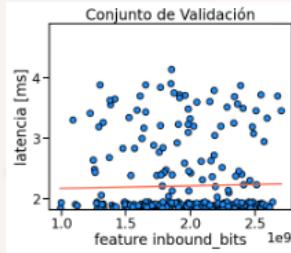
*ahora nos animamos a dar un paso  
más en esto de ML ...*



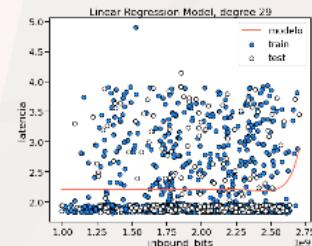
# Introducción al Aprendizaje Automático

## Regresión

### Regresión Lineal



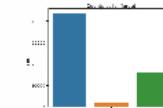
### Regresión Polinomial



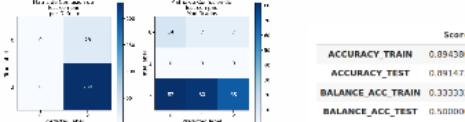
Latencia

## Clasificación

- SI PORCENTAJE\_PACK\_LOSS  $\in [0, 0.05] \Rightarrow 0$
- SI PORCENTAJE\_PACK\_LOSS  $\in [0.05, 0.1) \Rightarrow 1$
- SI PORCENTAJE\_PACK\_LOSS  $\in [0.1, \infty) \Rightarrow 2$



### Regresión Logística

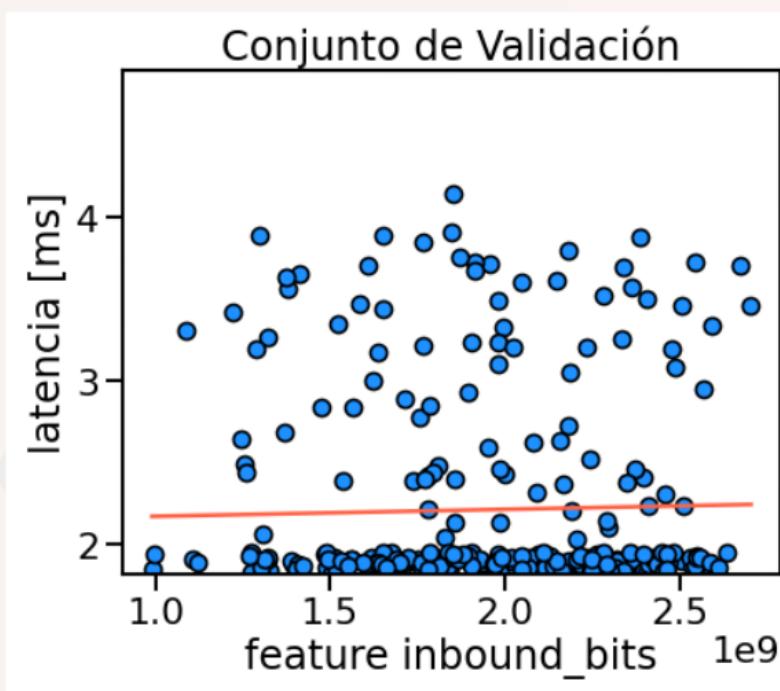


Score
ACCURACY_TRAIN 0.094200
ACCURACY_TEST 0.091473
BALANCE ACC TRAIN 0.333333
BALANCE ACC TEST 0.500000

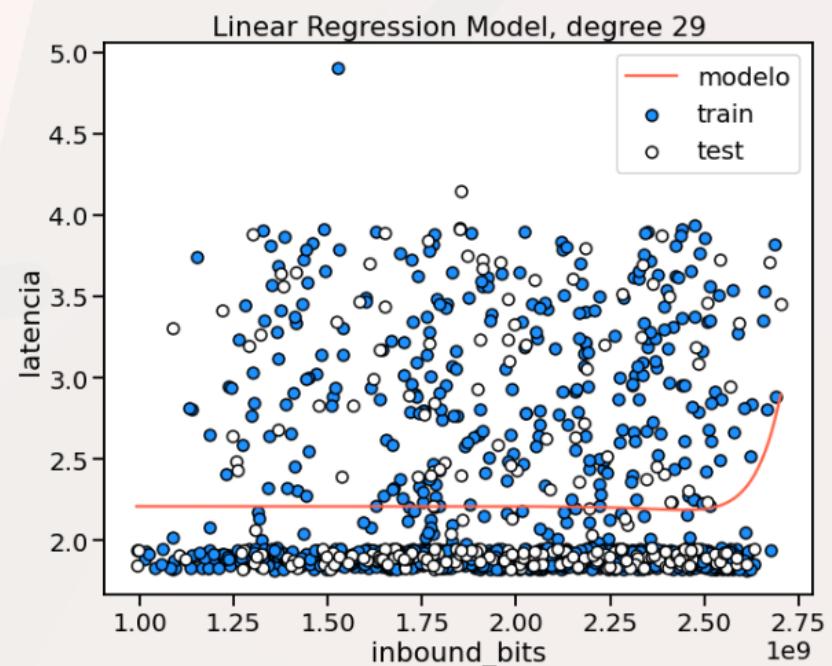
Pérdida de Paquetes

# *Regresión*

## *Regresión Lineal*

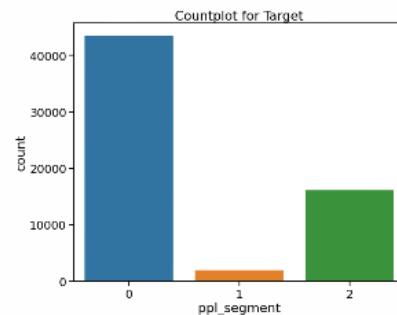


## *Regresión Polinomial*



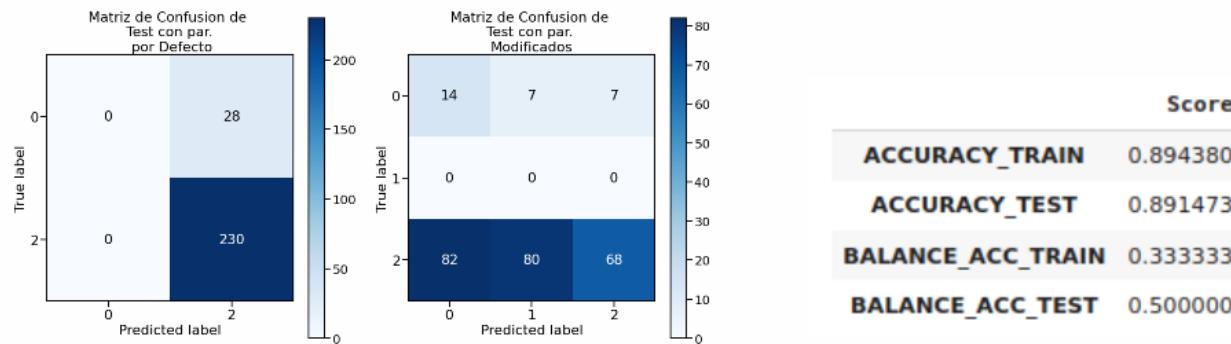
Latencia

# Clasificación



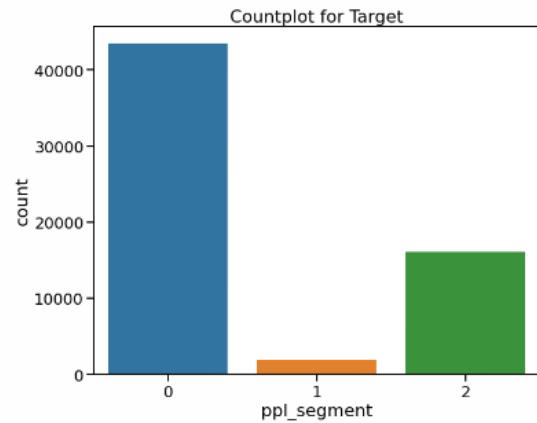
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0, 0.05) \Rightarrow 0$
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0.05, 0.1) \Rightarrow 1$
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0.1, \infty) \Rightarrow 2$

## Regresión Logistica



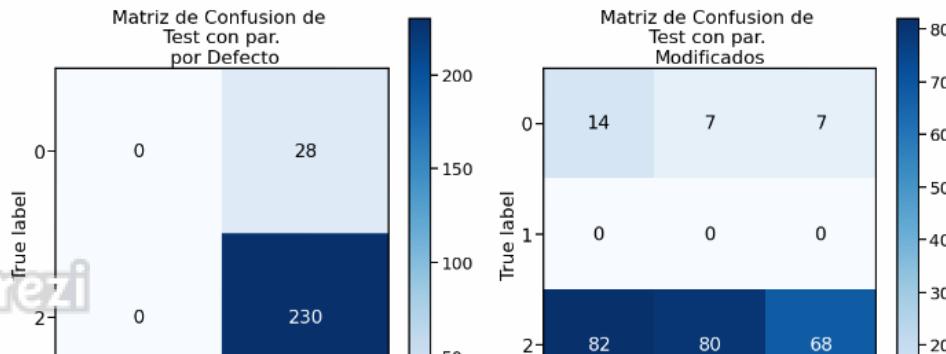
# Pérdida de Paquetes

# Clasificación

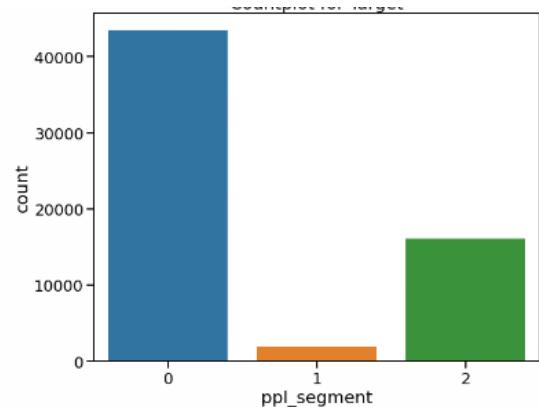


- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0, 0.05) \Rightarrow 0$
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0.05, 0.1) \Rightarrow 1$
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0.1, \infty) \Rightarrow 2$

## Regresión Logistica

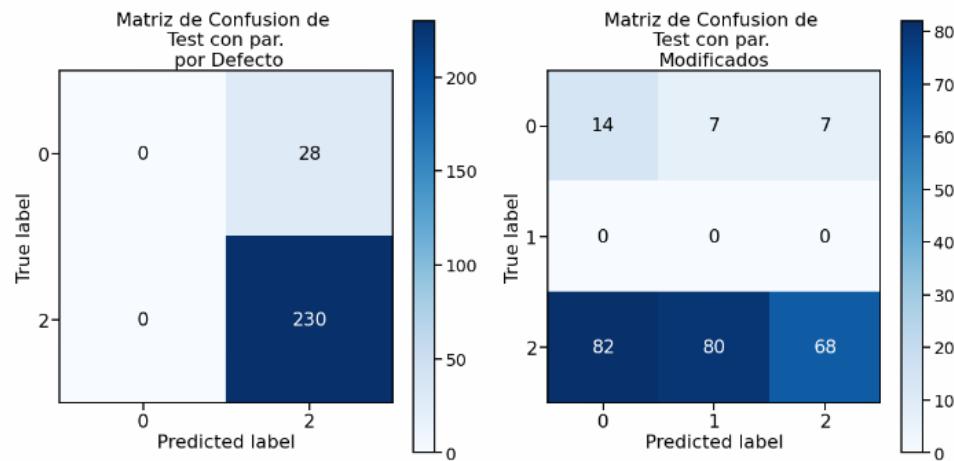


Score
ACCURACY_TRAIN 0.894380
ACCURACY_TEST 0.891473
BALANCE_ACC_TRAIN 0.333333



- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0, 0.05) \Rightarrow 0$
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0.05, 0.1) \Rightarrow 1$
- Si  $\text{PORCENTAJE\_PACK\_LOSS} \in [0.1, \infty) \Rightarrow 2$

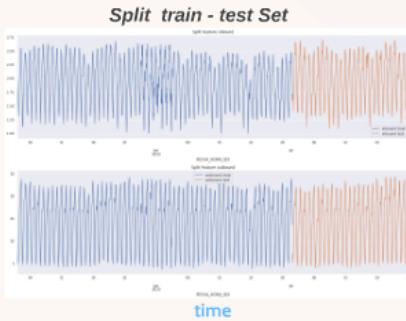
## *Regresión Logistica*



	Score
<b>ACCURACY_TRAIN</b>	0.894380
<b>ACCURACY_TEST</b>	0.891473
<b>BALANCE_ACC_TRAIN</b>	0.333333
<b>BALANCE_ACC_TEST</b>	0.500000

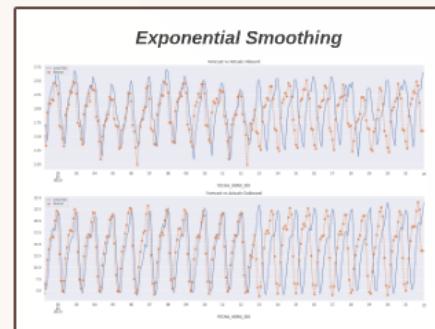
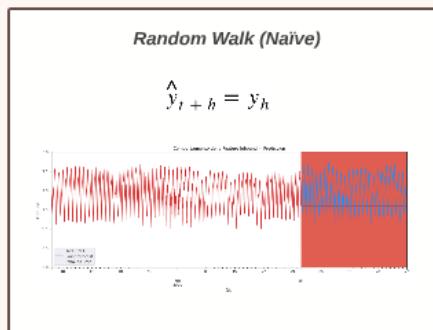
# Pérdida de Paquetes

# Aprendizaje Supervisado

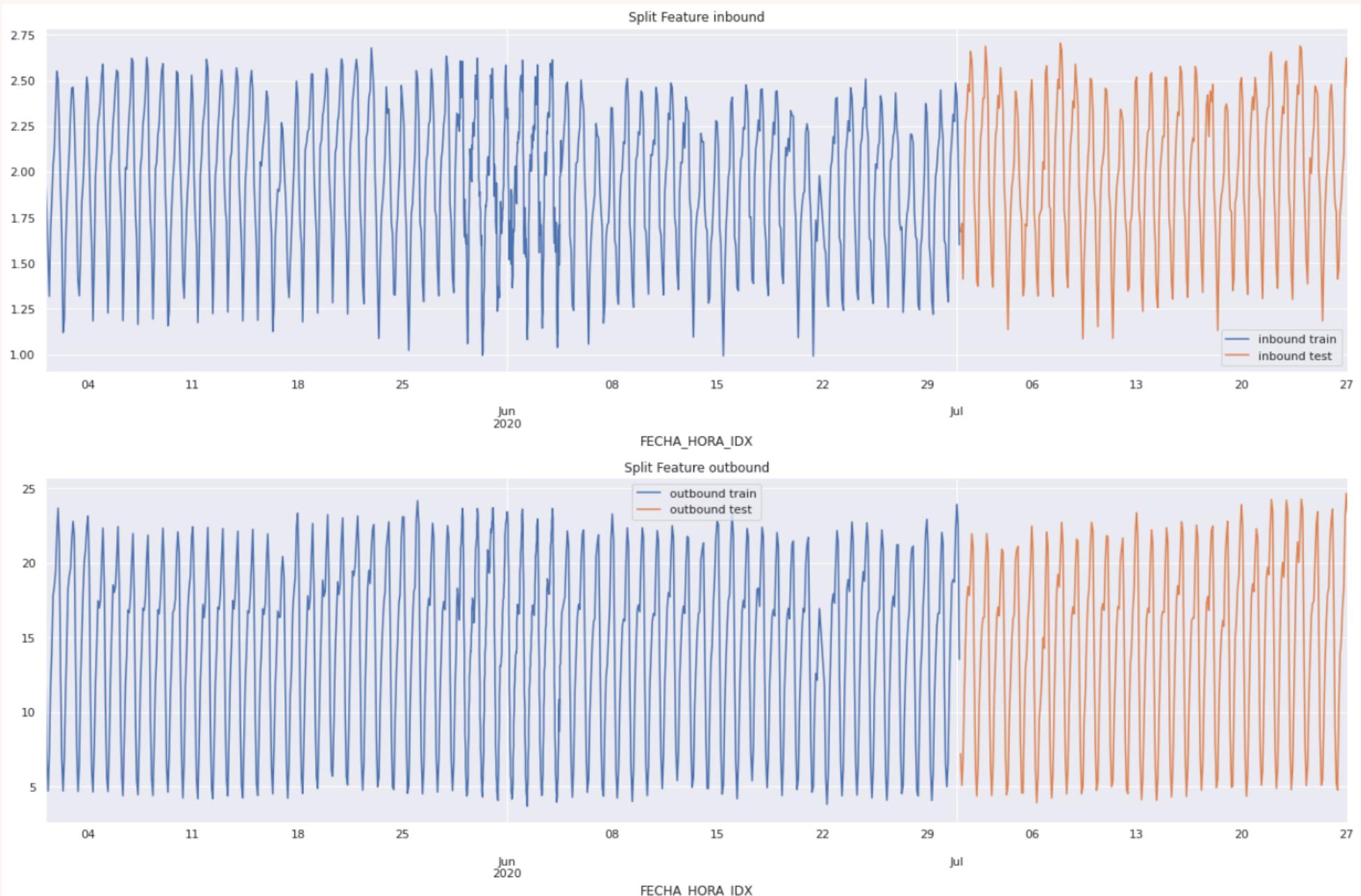


**Metrics**

Inbound Traffic	RMSE	SE
Seasonal RW	0.2	0.011
Exponential Smoothing	0.309	0.156
Outbound Traffic		
Seasonal RW	1.727	0.08
Exponential Smoothing	7.375	0.419

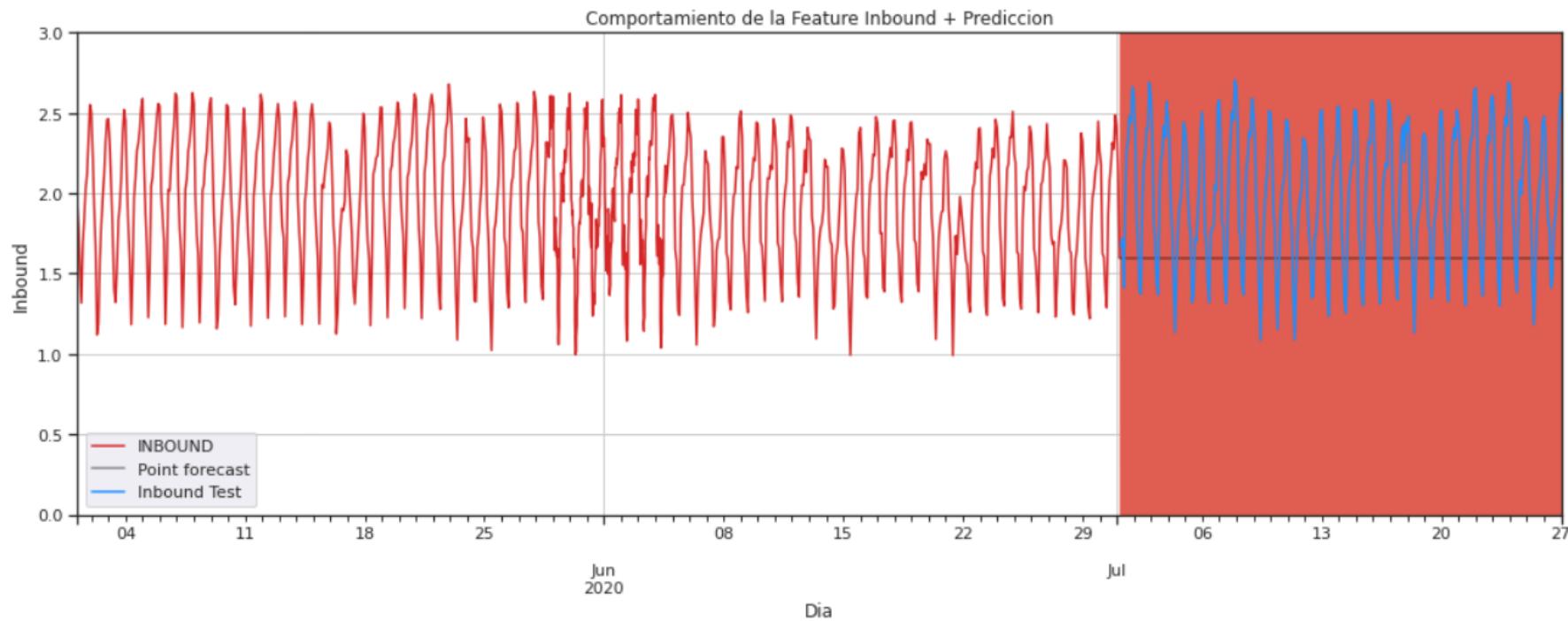


# *Split train - test Set*

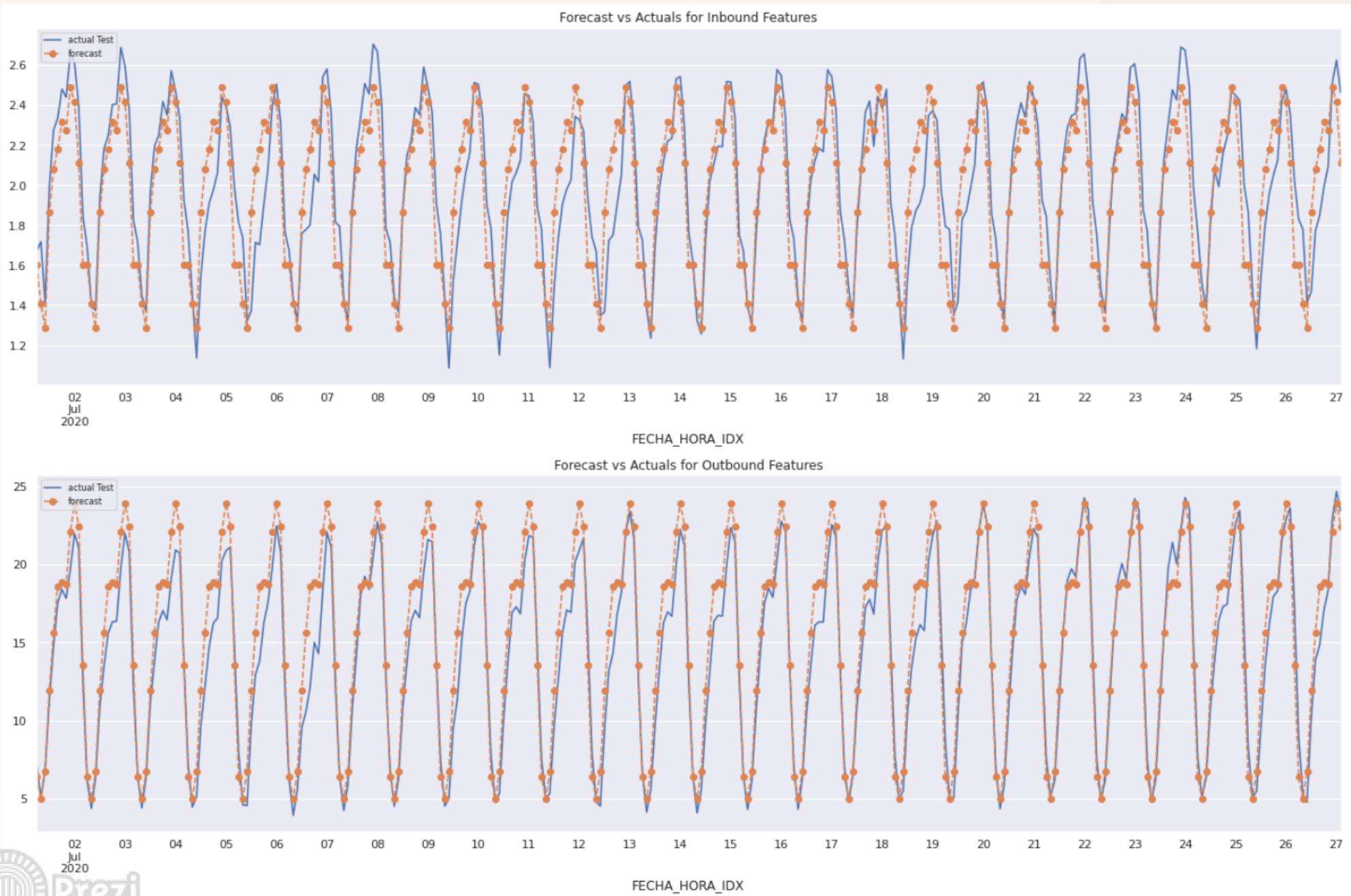


# *Random Walk (Naïve)*

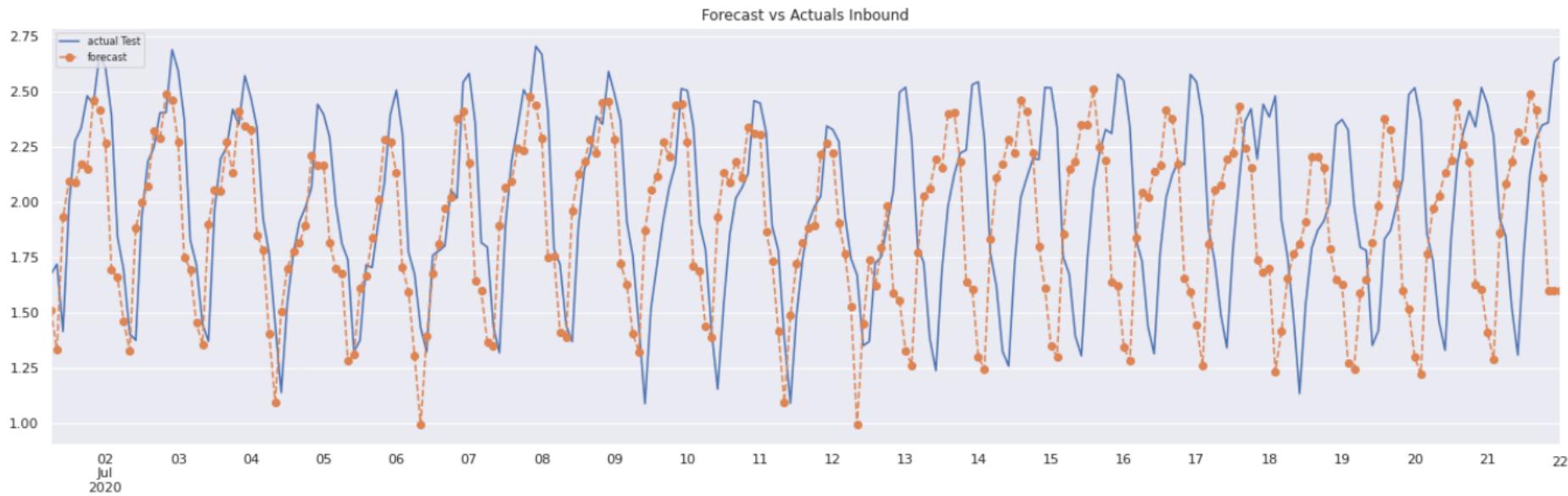
$$\hat{y}_{t+h} = y_h$$



# *Seasonal Random Walk*

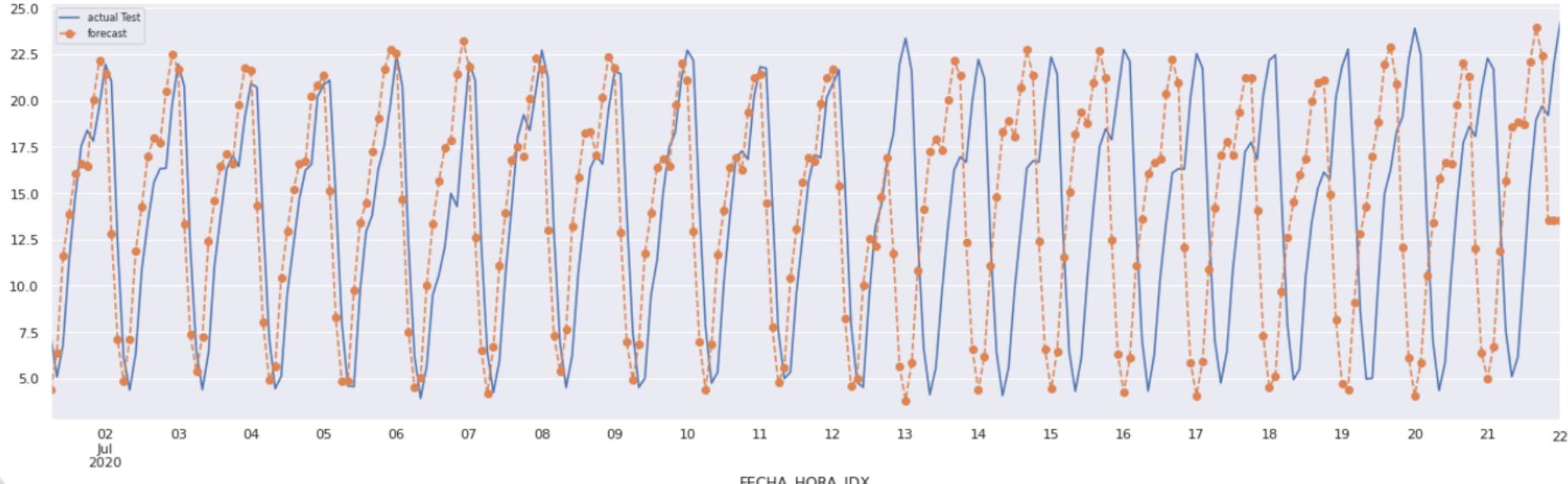


# *Exponential Smoothing*



FECHA\_HORA\_IDX

Forecast vs Actuals Outbound



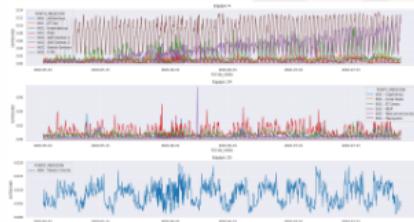
FECHA\_HORA\_IDX

# *Metrics*

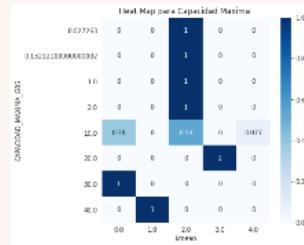
Inbound Traffic	RMSE	SE
Seasonal RW	0.2	0.011
Exponential Smoothing	0.309	0.156
Outbound Traffic		
Seasonal RW	1.727	0.08
Exponential Smoothing	7.375	0.419

# Aprendizaje No Supervisado

Análisis Previo



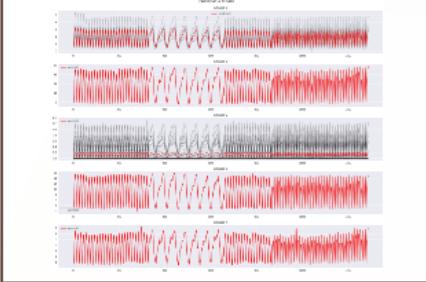
HEAT MAP



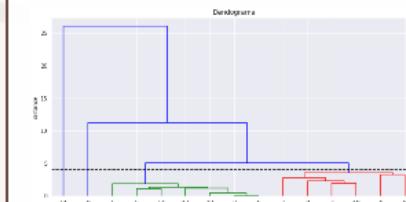
Clustering para dispositivos con tráfico irregular



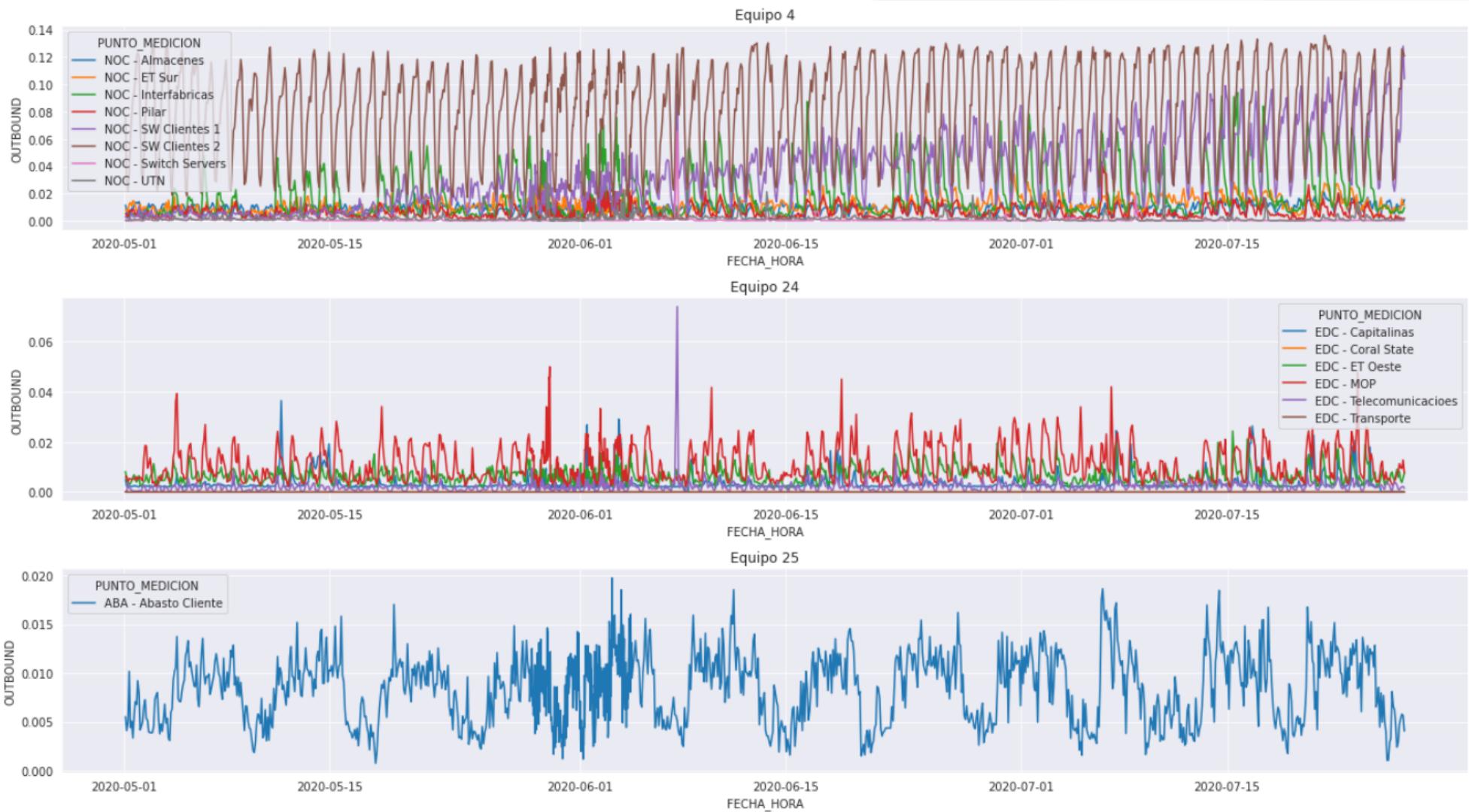
Time Series KMeans



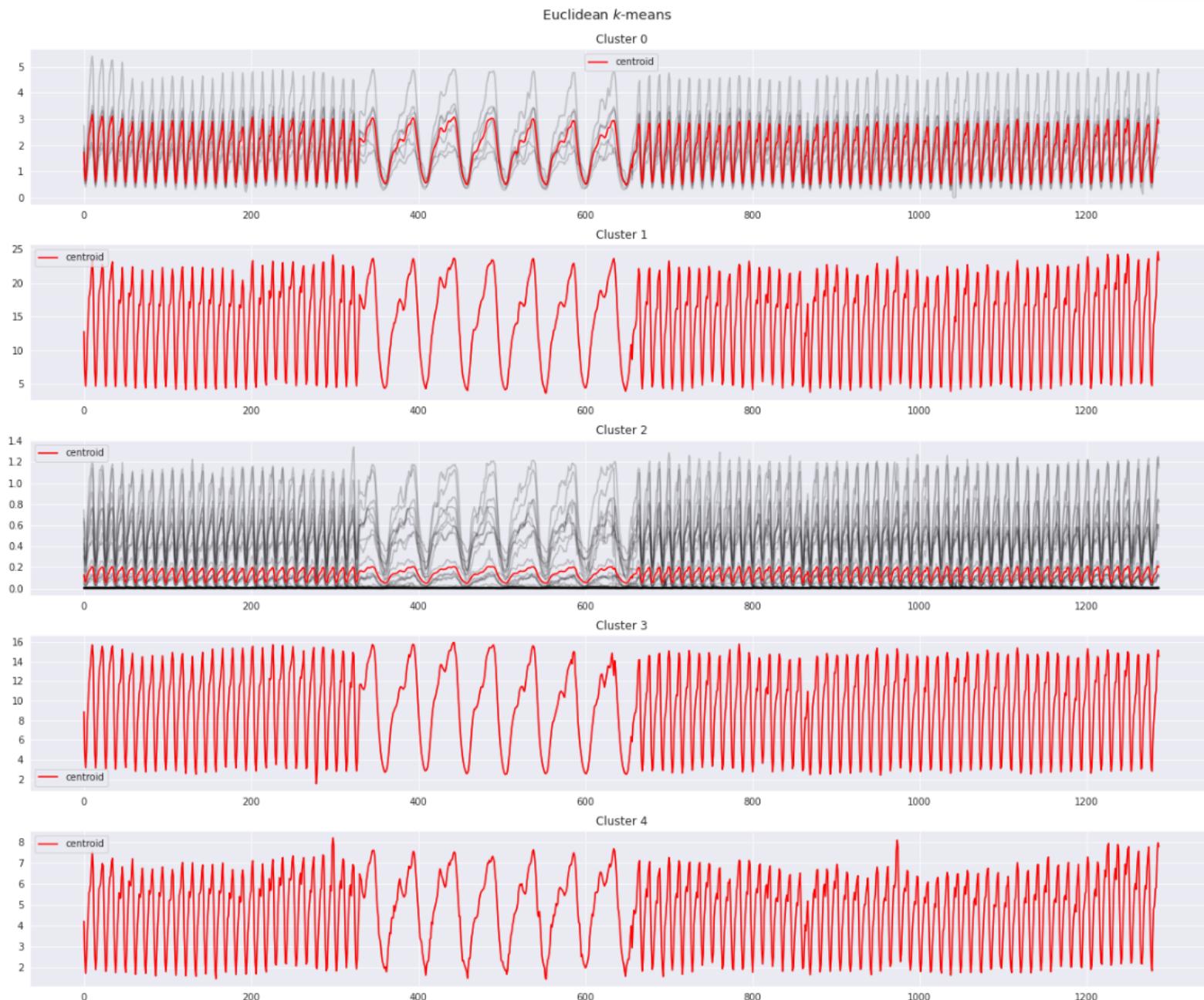
Clustering para dispositivos con tráfico irregular



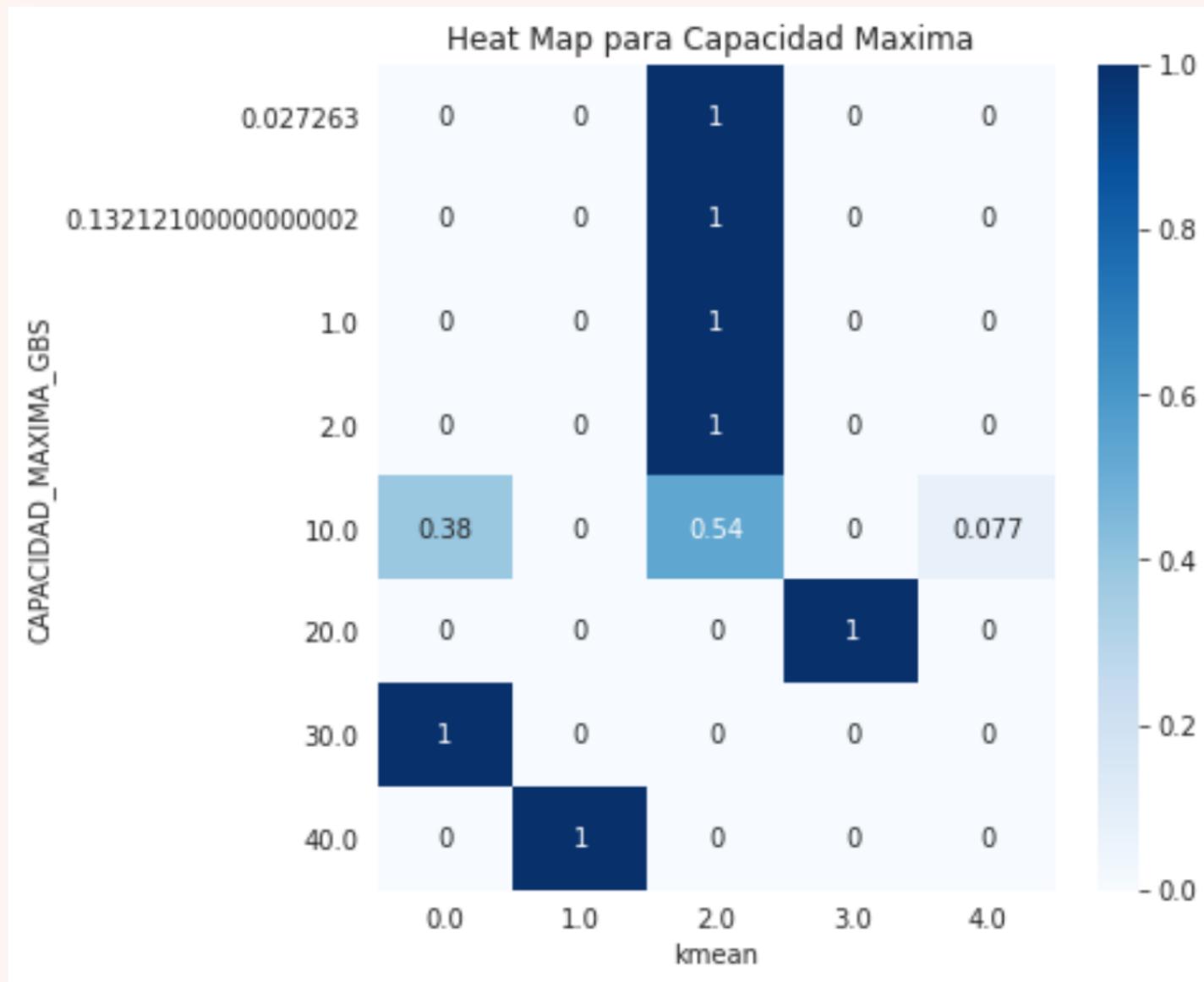
# Análisis Previo



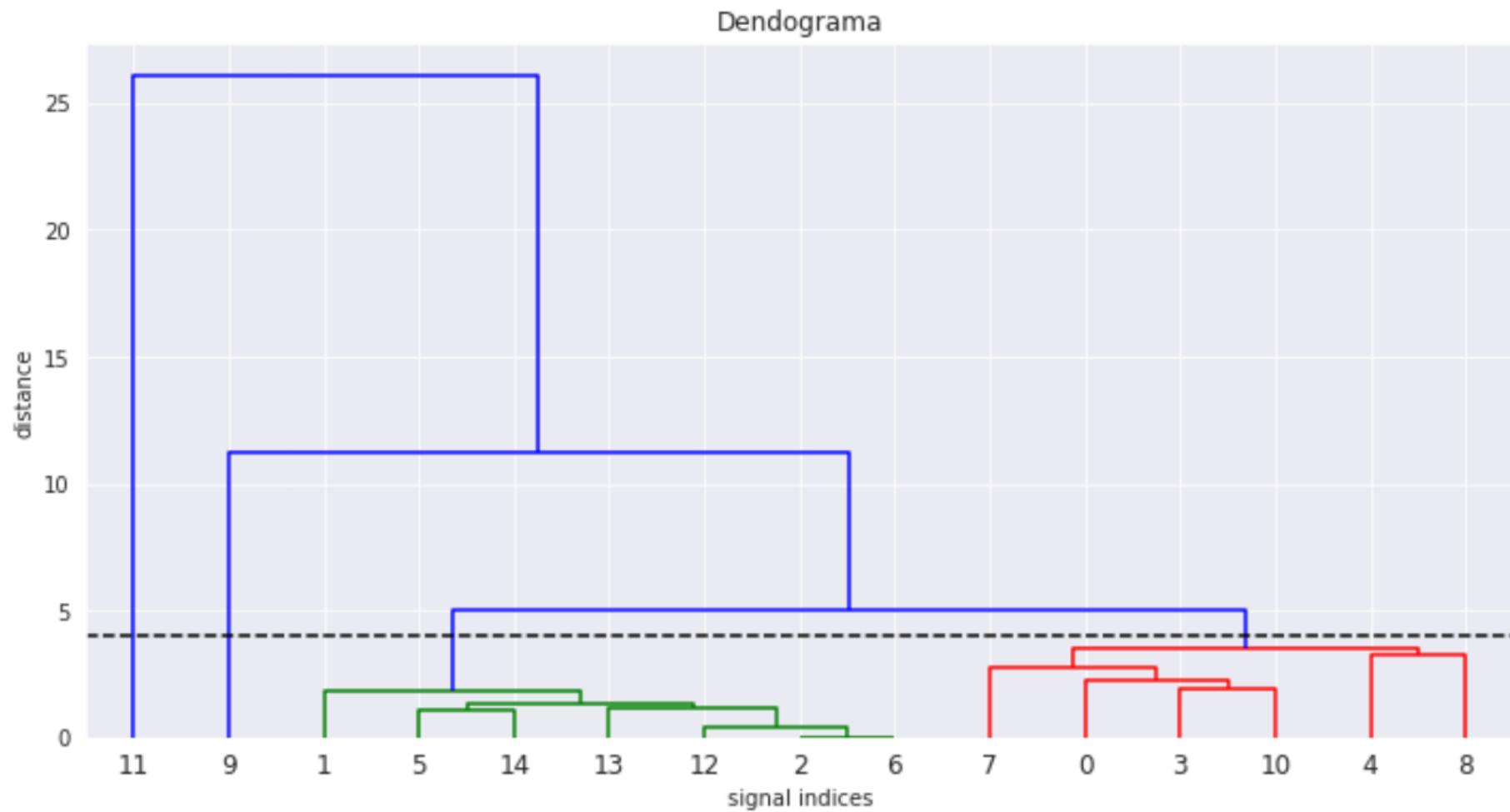
# Time Series KMeans



# HEAT MAP



# *Clustering para dispositivos con tráfico irregular*



# *Clustering para dispositivos con tráfico irregular*



# *y lo logramos o eso creemos !!!*

Veremos si con lo analizado y descubierto hasta el momento podemos responder las preguntas que nos planteamos al inicio.....

¿Es posible predecir cortes o desperfectos en la Red?

*Si bien analizamos la predicción sobre la feature latencia y tráfico aún no descubrimos un modelo eficiente para aplicar a esta situación*

¿Podemos predecir el uso de la Red en las próximas 24 horas? ¿En la próxima semana?

*Sobre este punto obtuvimos resultados muy buenos, podemos concluir en que más tiempo a predecir requiere más datos del dataset*

¿Es posible detectar anomalías o comportamientos sospechosos en el tráfico de la Red?

*Si, de hecho varios análisis y gráficos abordados durante el Proyecto detectaron estos problemas.*

¿La actividad de la Red varía según el día de la semana o mes?

*Definitivamente si, como lo hemos comentado previamente.*

¿Los cortes en la Red están asociados con alguna otra/s variable/s del set de datos?

*Este análisis era mucho más ambicioso de lo que pudimos abarcar en el Proyecto por lo tanto lo dejamos en estudio....*



# **creemos !!!**

Veremos si con lo analizado y descubierto hasta el momento podemos responder las preguntas que nos planteamos al inicio.....

¿Es posible predecir cortes o desperfectos en la Red?

*Si bien analizamos la predicción sobre la feature latencia y tráfico aún no descubrimos un modelo eficiente para aplicar a esta situación*

¿Podemos predecir el uso de la Red en las próximas 24 horas? ¿En la próxima semana?

*Sobre este punto obtuvimos resultados muy buenos, podemos concluir en que más tiempo a predecir requiere más datos del dataset*

¿Es posible detectar anomalías o comportamientos sospechosos en el tráfico de la Red?

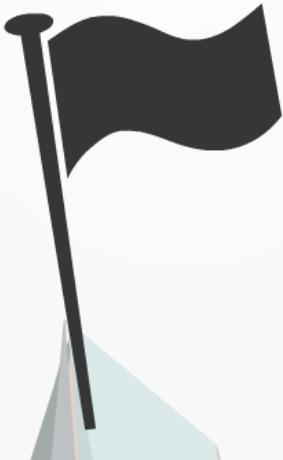
*Si, de hecho varios análisis y gráficos abordados durante el Proyecto detectaron estos problemas.*

¿La actividad de la Red varía según el día de la semana o mes?

*Definitivamente si, como lo hemos comentado previamente.*

¿Los cortes en la Red están asociados con alguna otra/s variable/s del set de datos?

*Este análisis era mucho más ambicioso de lo que pudimos abarcar en el Proyecto por lo tanto lo dejamos en estudio....*



# Análisis de una Red de Telecomunicaciones

Como encaramos nuestro primer Proyecto de Ciencia de Datos y ... sobrevivimos al intento.

## Integrantes

- Sergio Sulca
- Carlos Bacia
- Soledad Fernández

