DigitalHouse>
Coding School

¿Comemos afuera?

Predicción de visitas a restaurants

Gastón Domínguez

Data Science - 2018

Objetivo

- Predicción de visitas a restaurant
- Caracterización como Serie de Tiempo
- Aplicación de modelos idóneos
- Incorporación de información complementaria

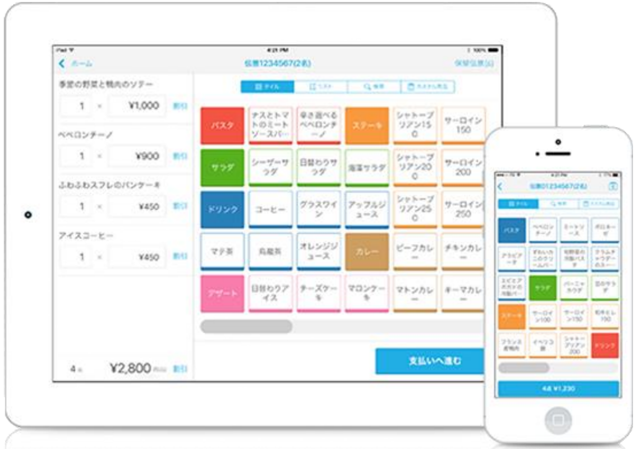
GENERALIDADES

Información disponible

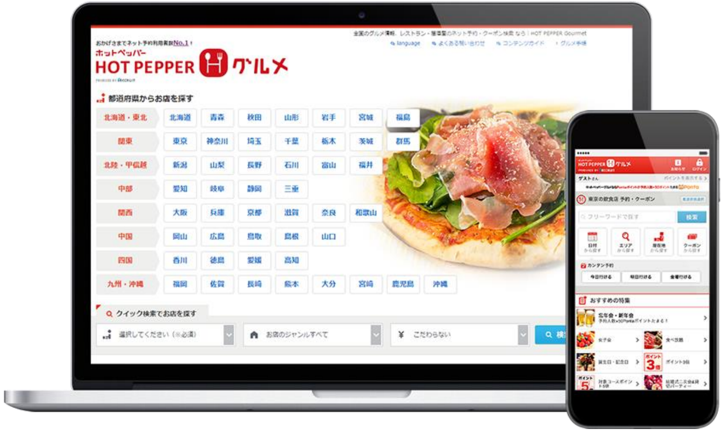
RECRUIT

AirREGI

HOT PEPPER グルメ



RESERVAS Y VISITANTES



RESERVAS

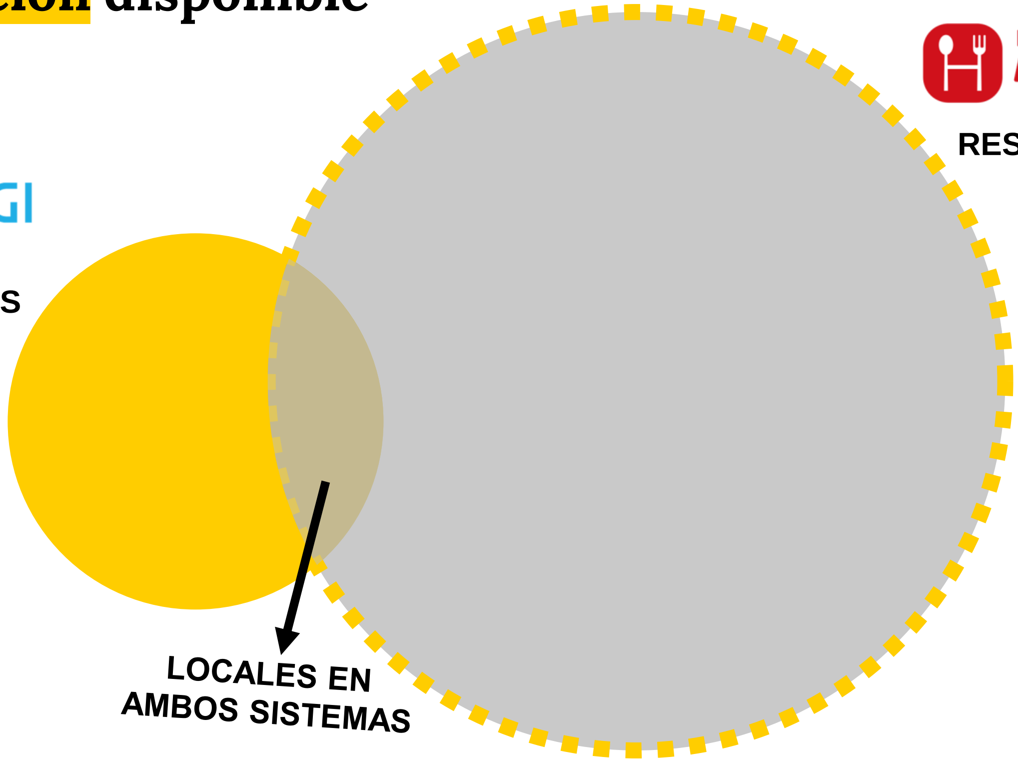
Información disponible

AirREGI

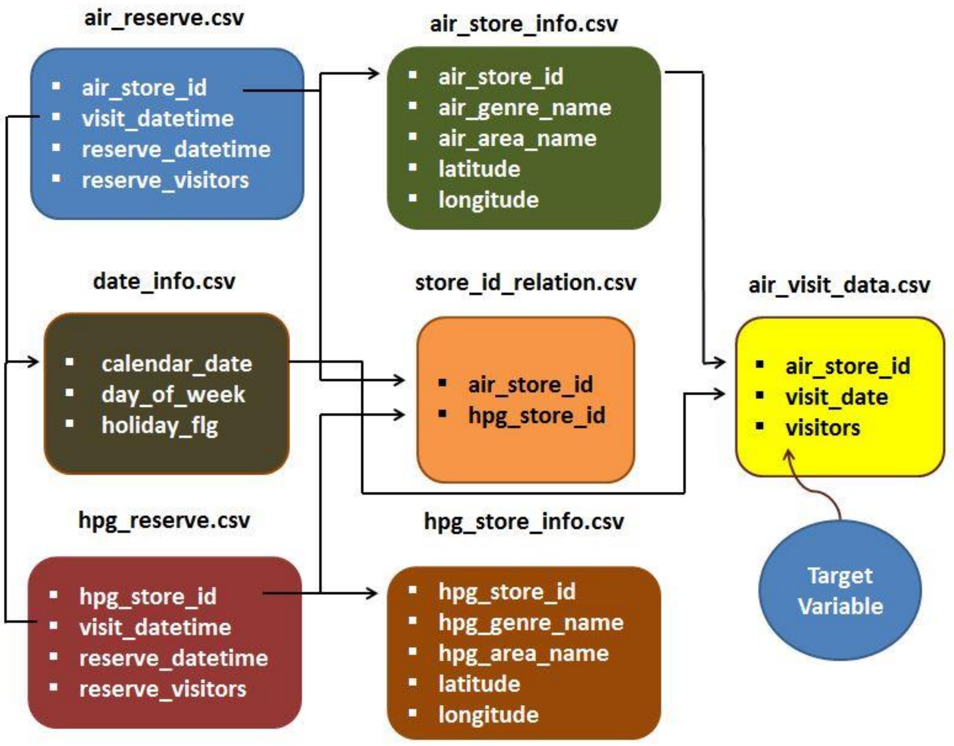
RESERVAS
Y VISITANTES

HOT PEPPER グルメ

RESERVAS

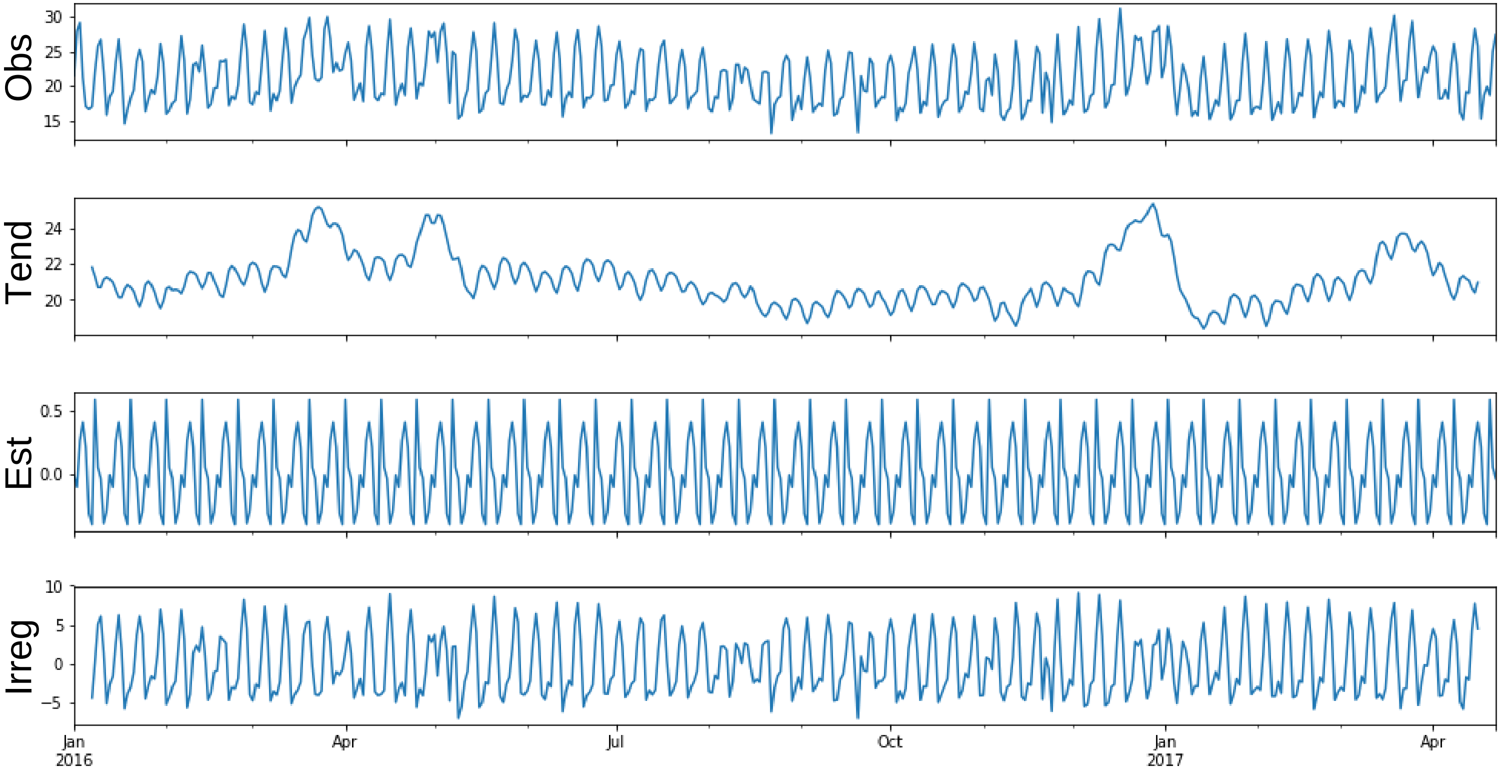


Información disponible



Series de Tiempo

seasonal_decompose(air.visit_mean, freq=12)

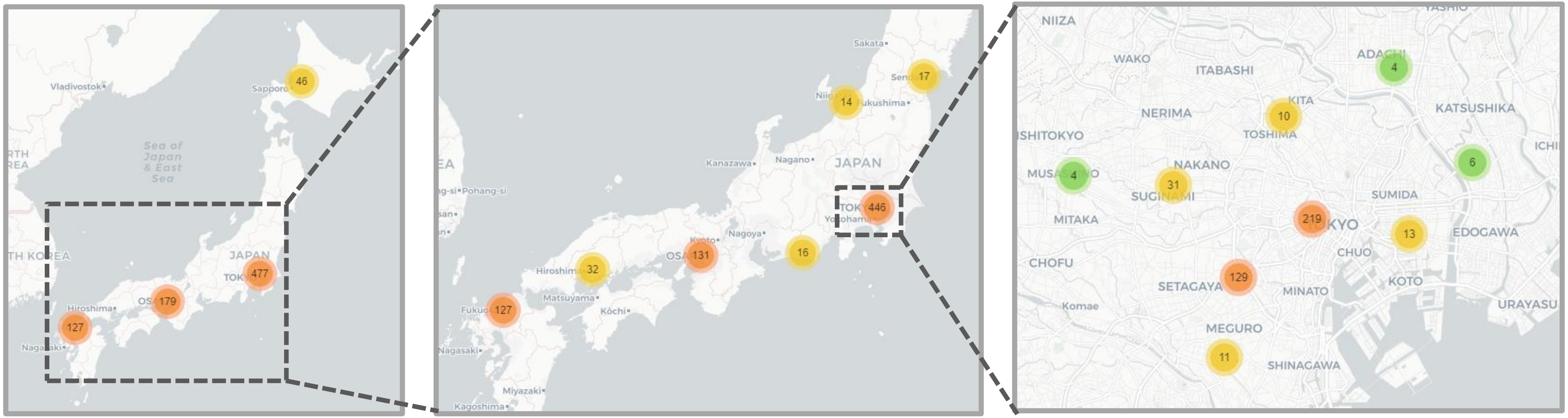


Series de Tiempo

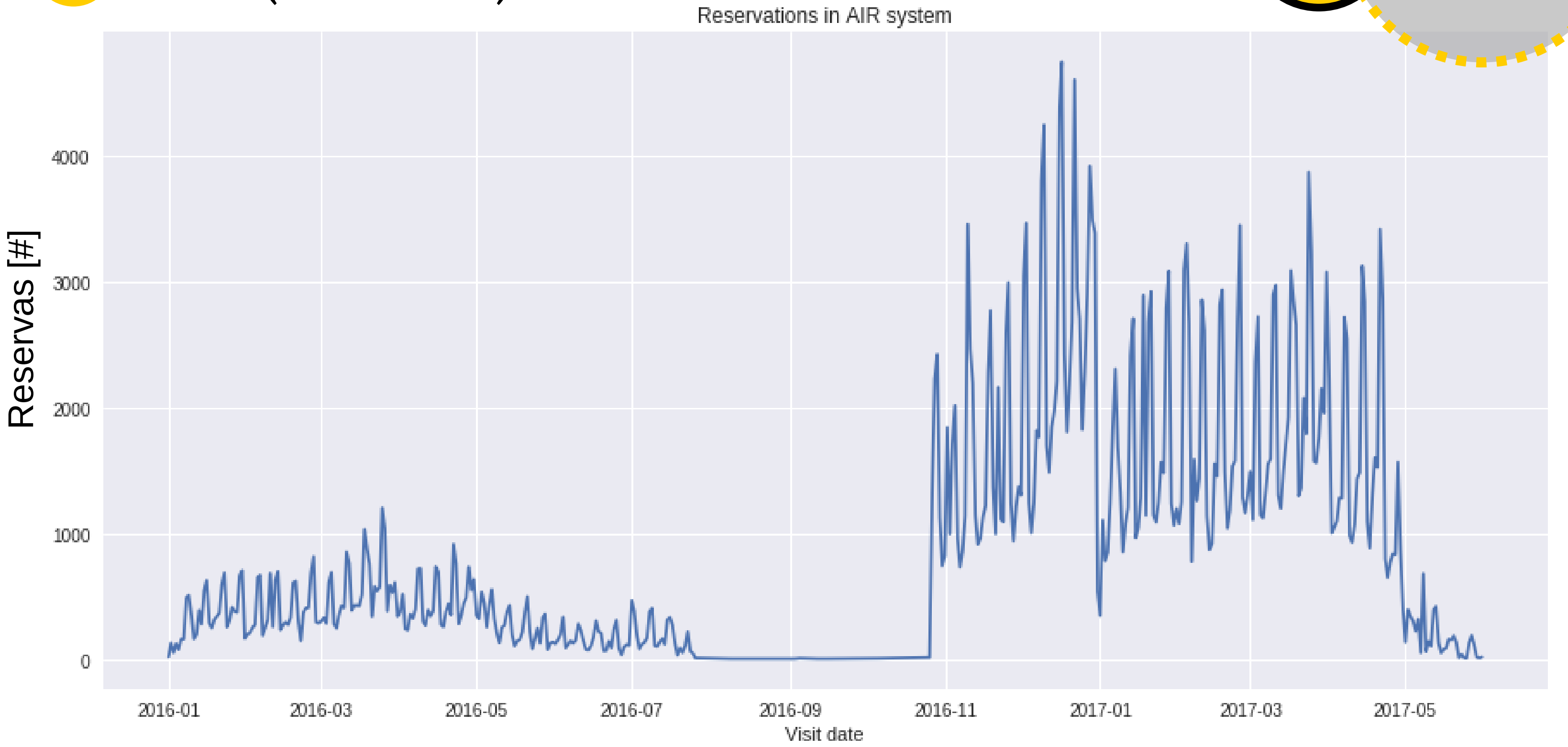
Tendencia	• Incremento/disminución a largo plazo
Estacional	• Serie influenciada por patrones estacionales
Cíclico	• Aumento/Caída no propios al período fijo (>2años)
Irregular	• Variaciones aleatorias o residuales

ANÁLISIS EXPLORATORIO

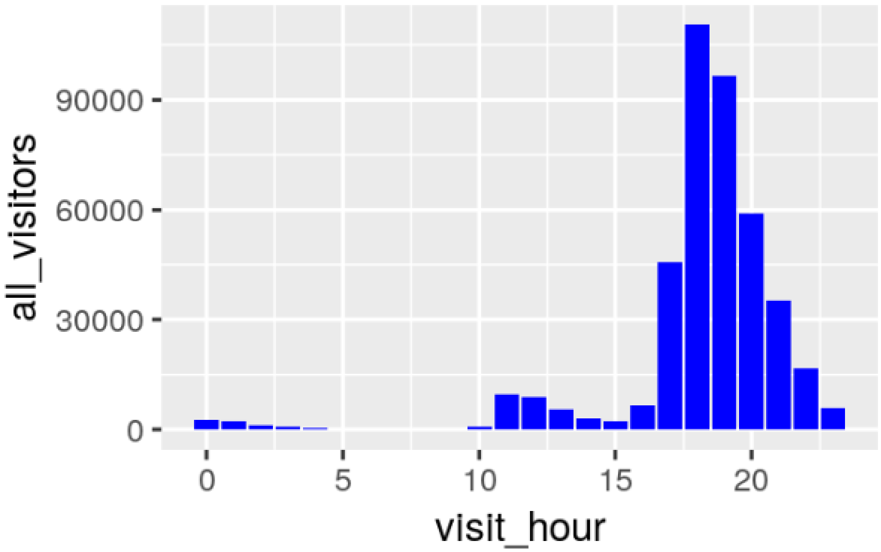
Ubicación de locales

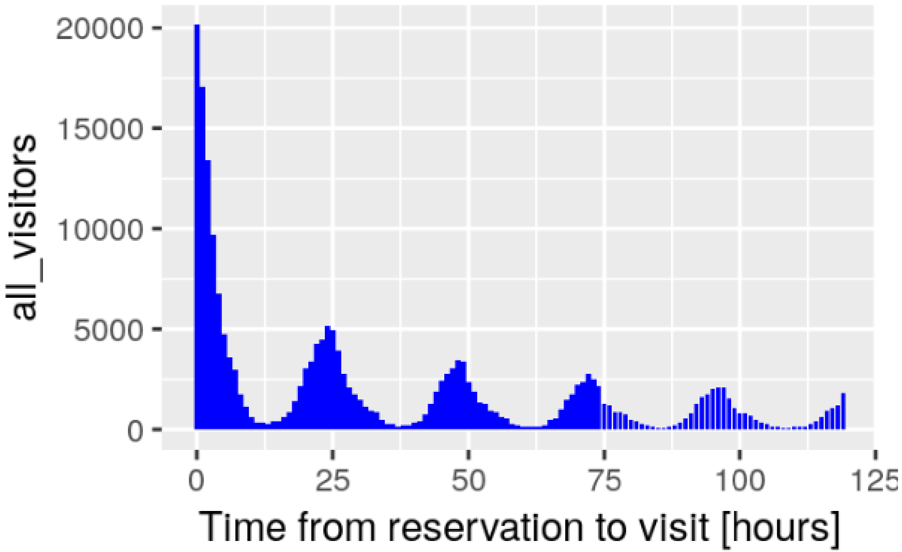


AIR (Reservas)

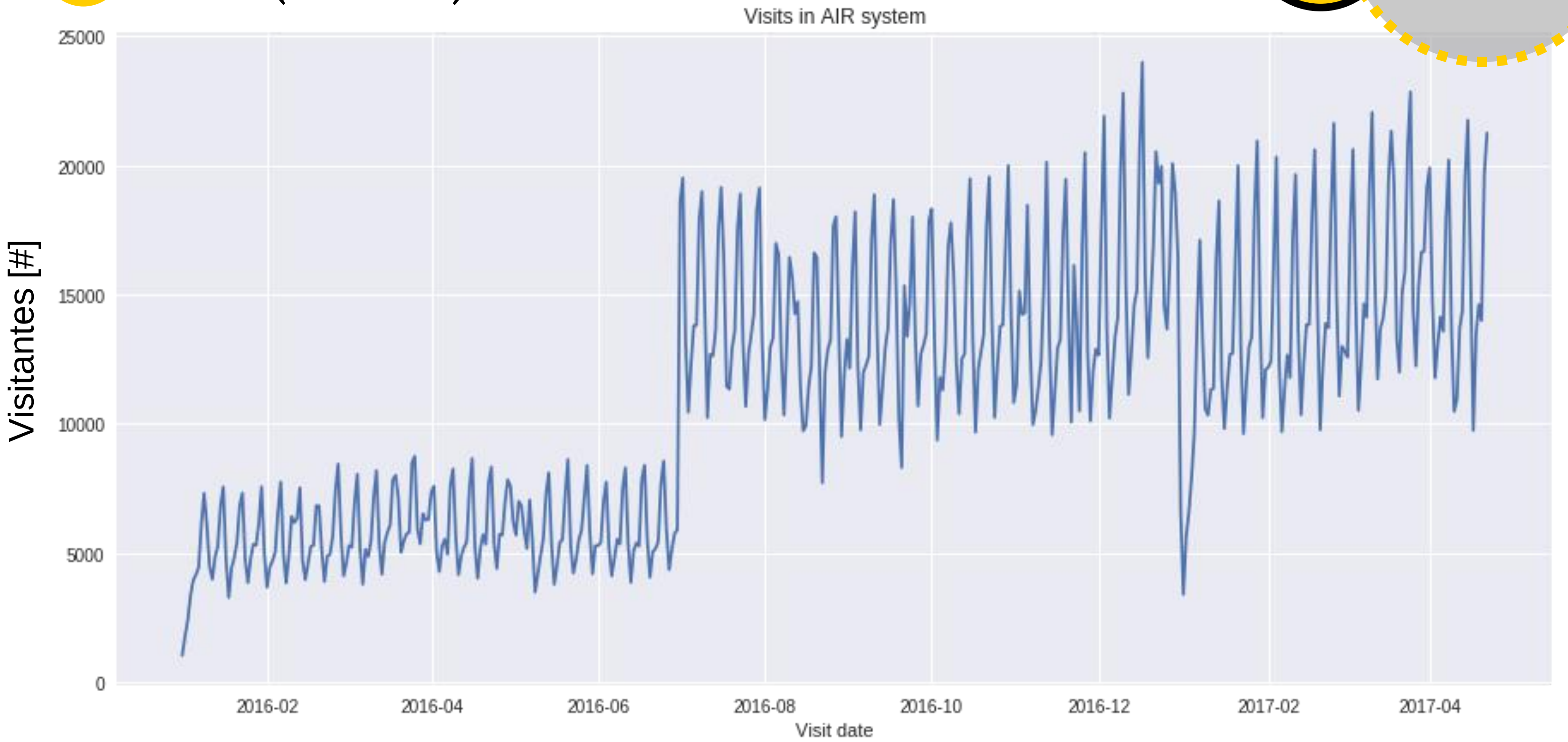


AIR (Reservas)



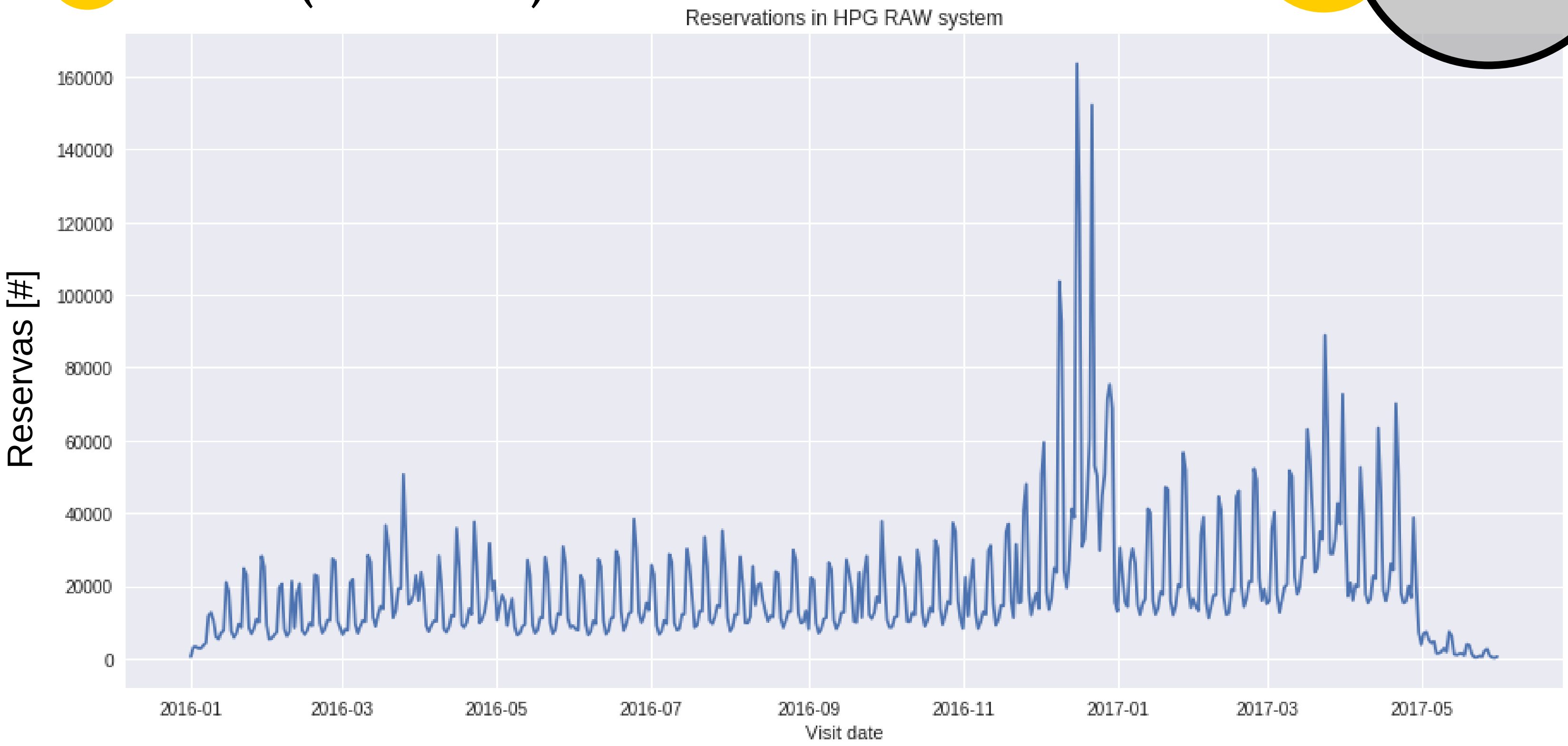


AIR (Visitas)

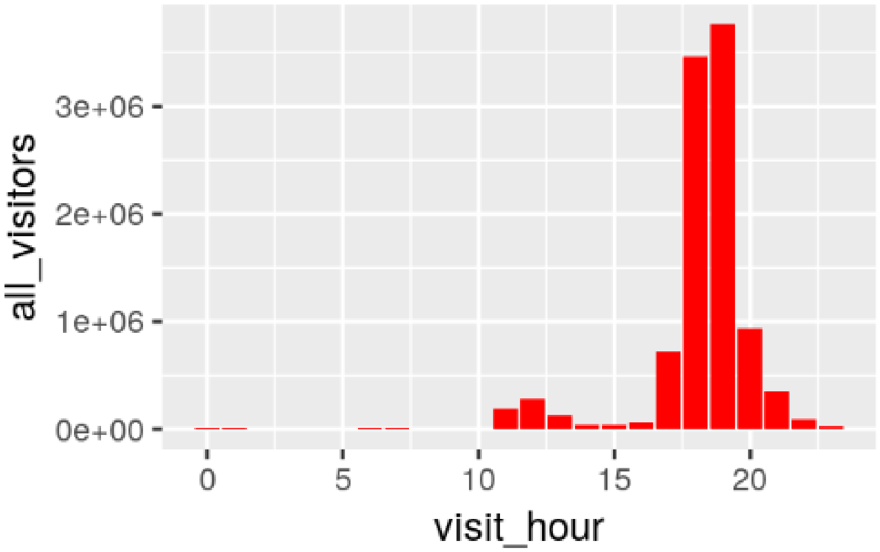


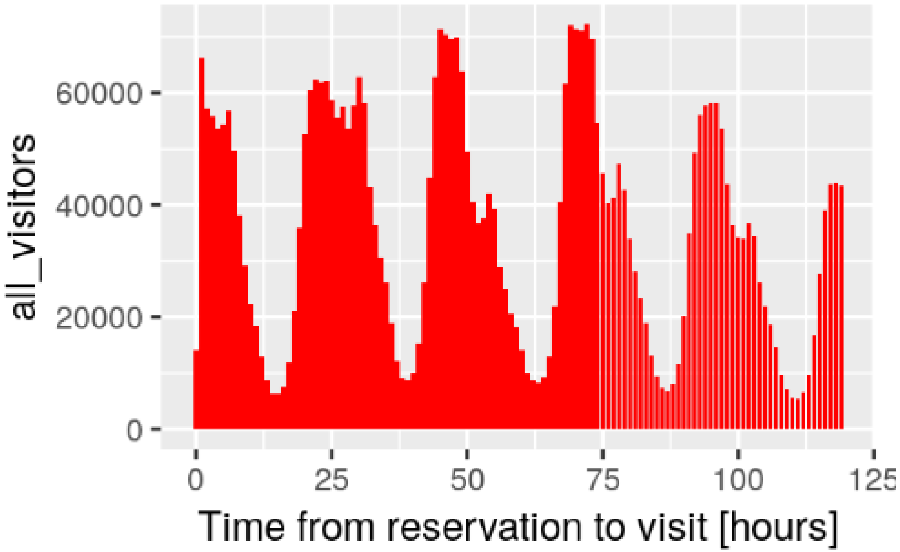
Visits in AIR system
Visitantes [#]
25000
20000
15000
10000
5000
0
2016-02
2016-04
2016-06
2016-08
2016-10
2016-12
2017-02
2017-04
Visit date

HPG (Reservas)



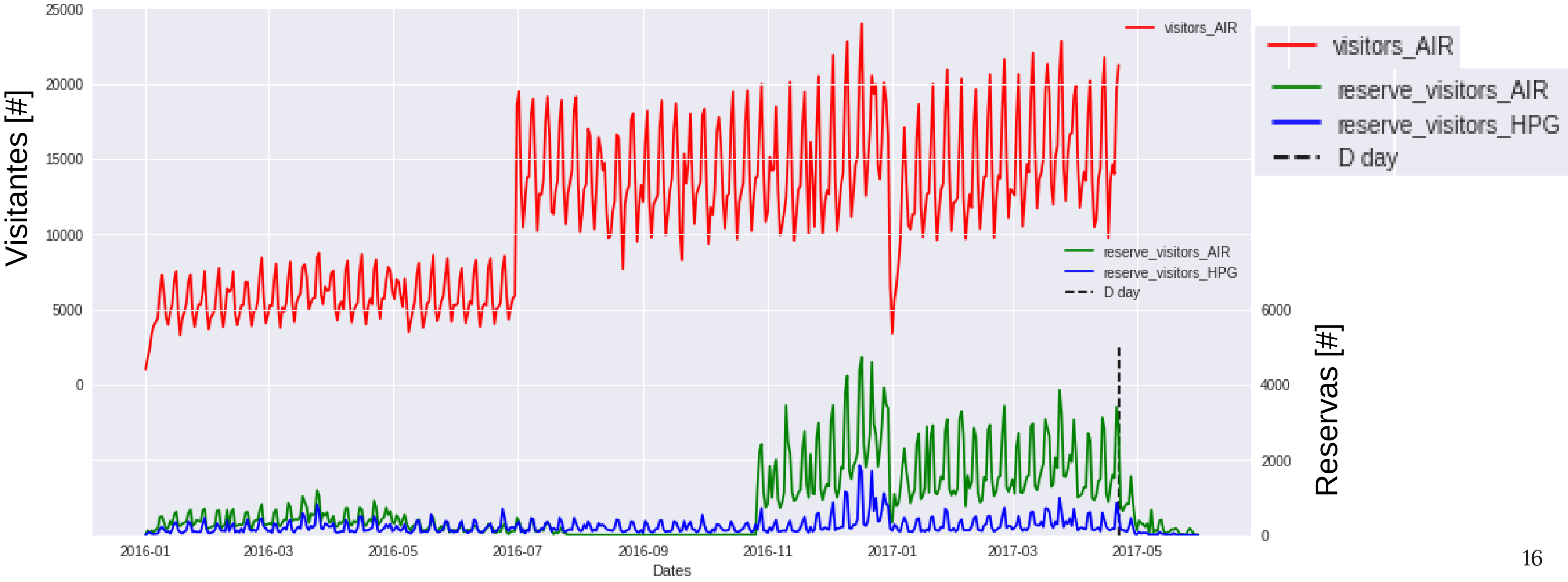
HPG (Reservas)



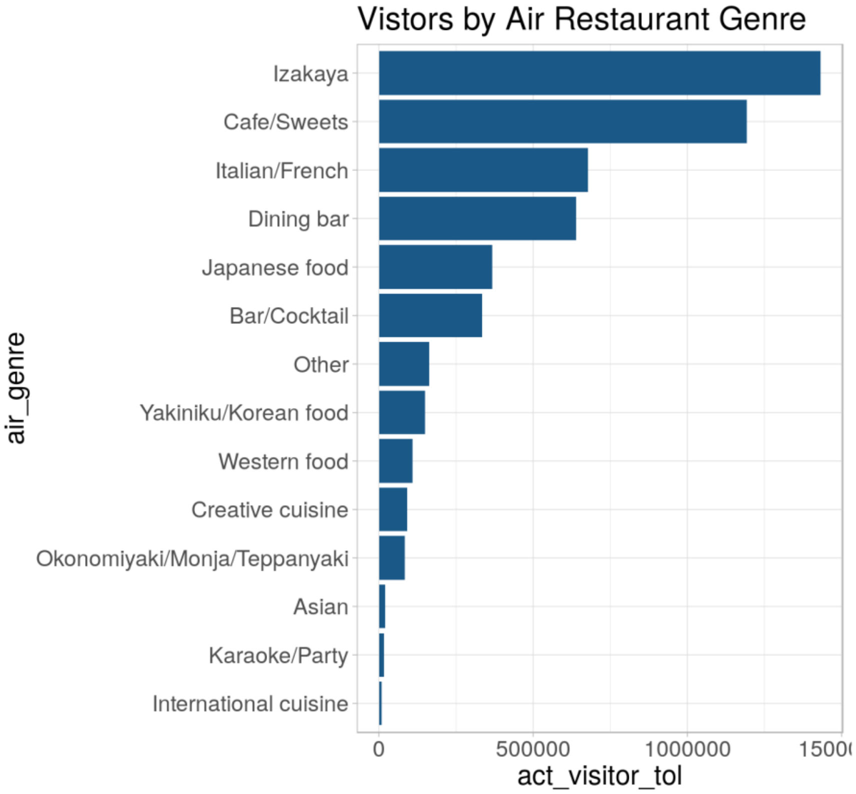


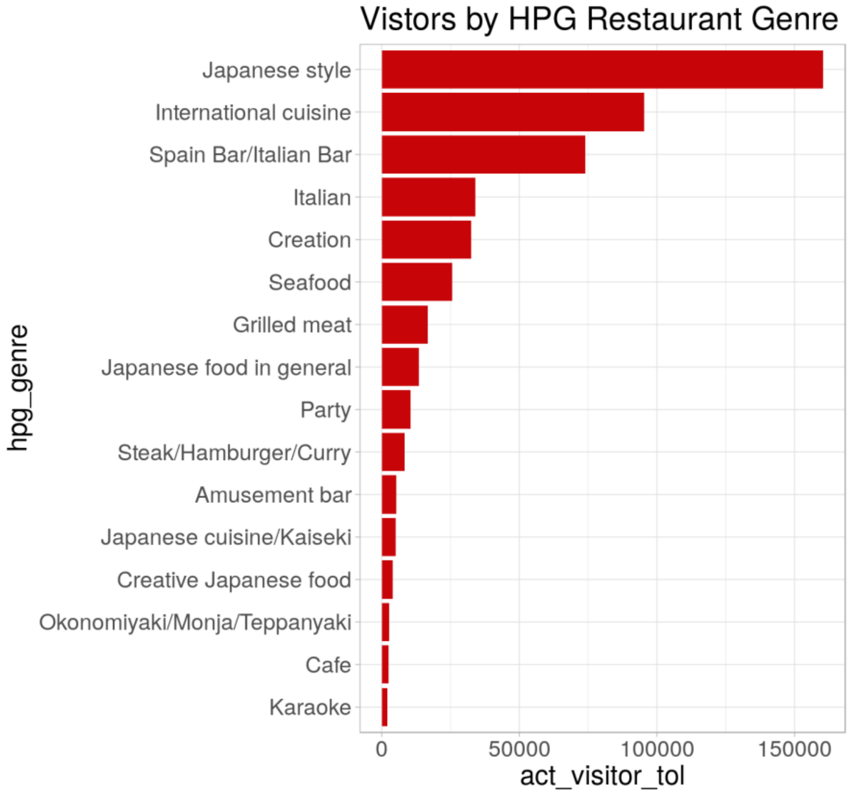
AIR & HPG (Reservas y Visitas)

Visitors against reservations in AIR & HPG systems

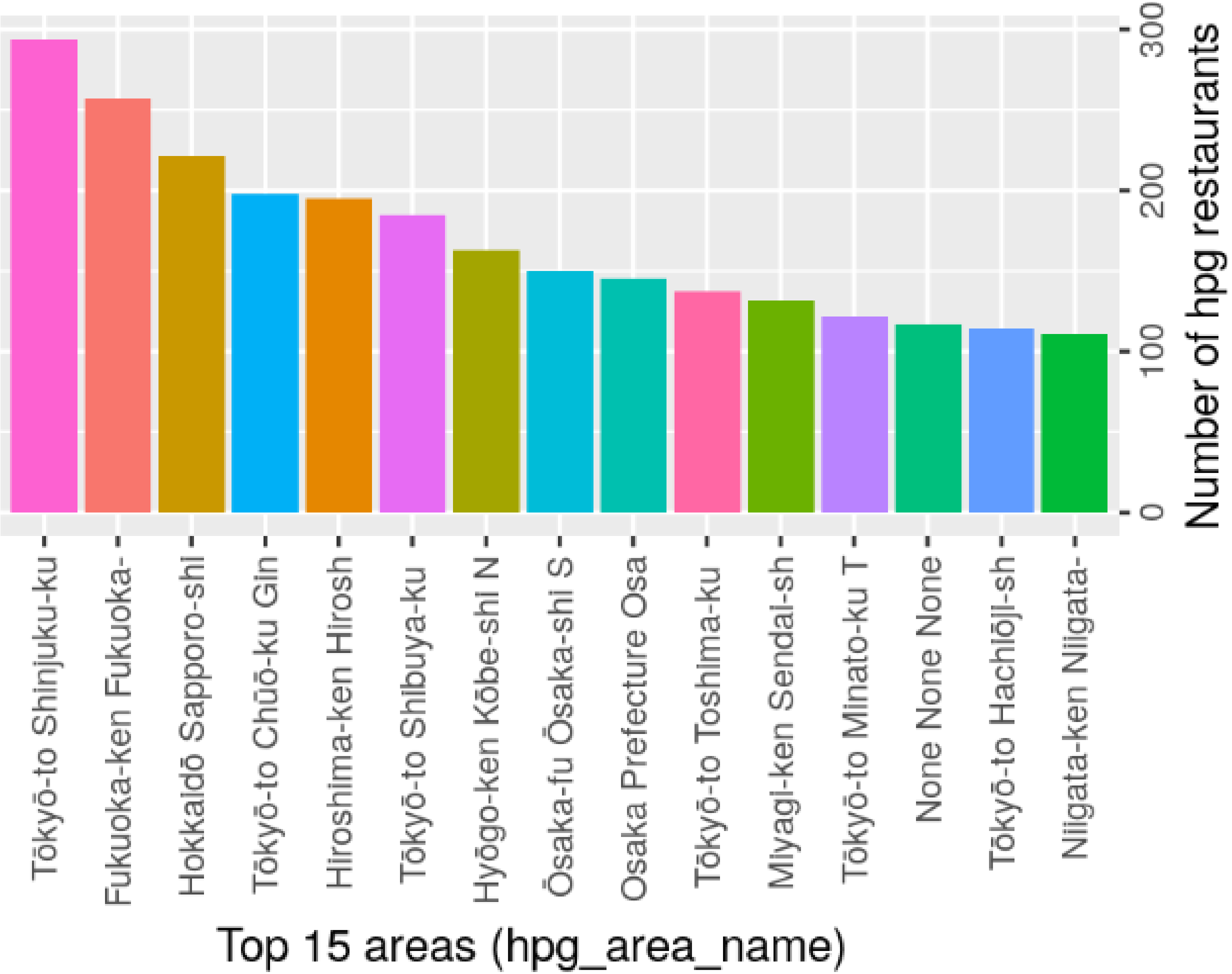


Tipo de cocina



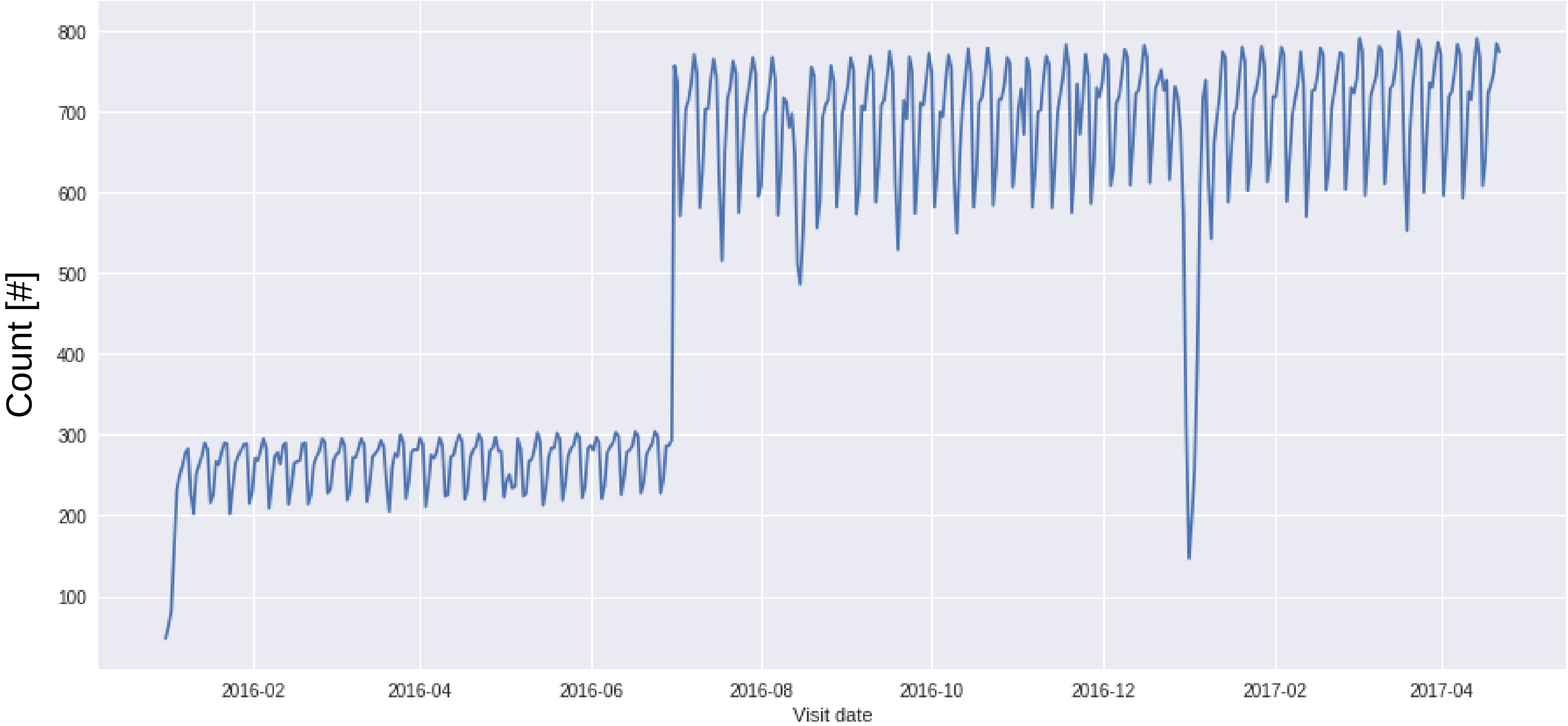


Top Areas (HPG)

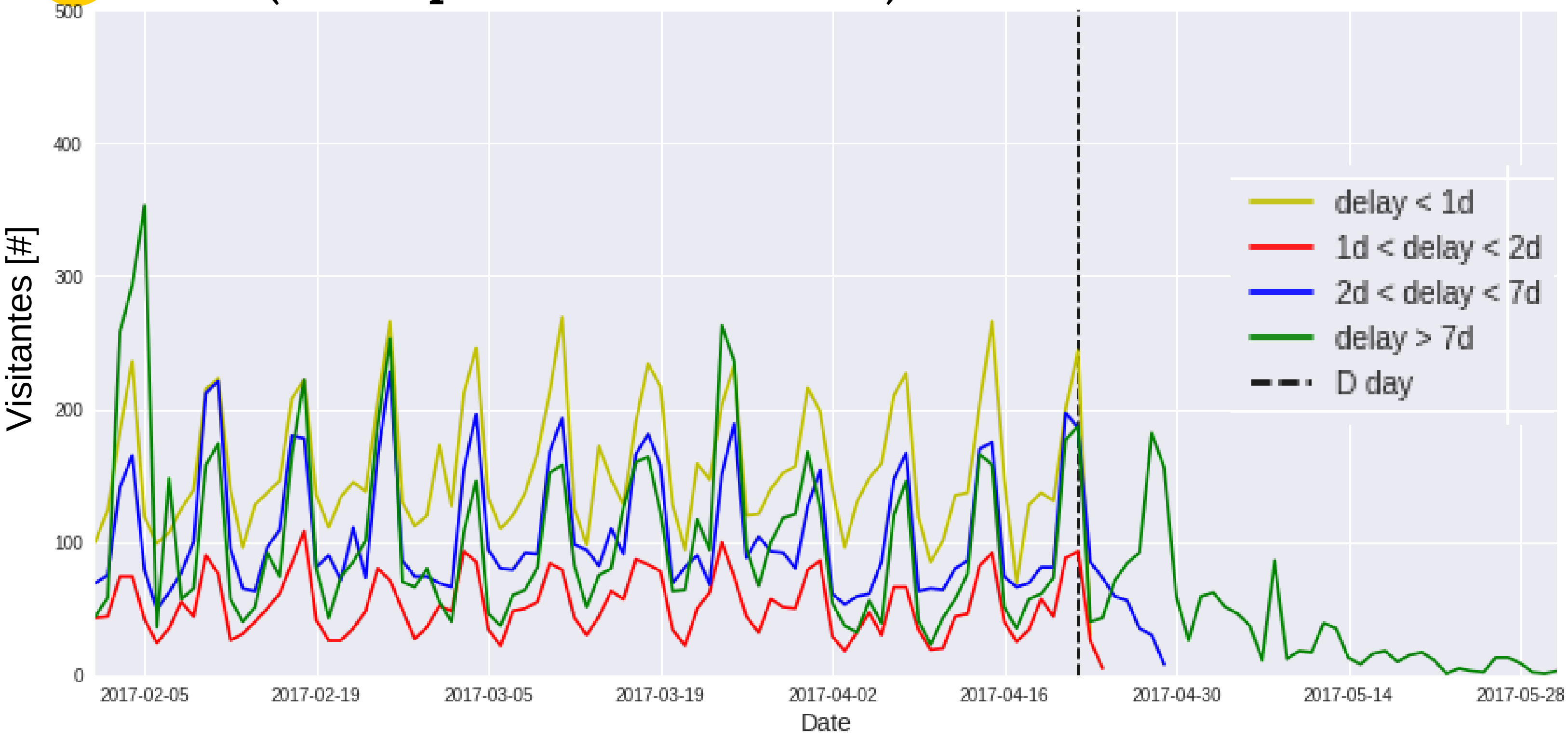


INFORMACIÓN COMPLEMENTARIA

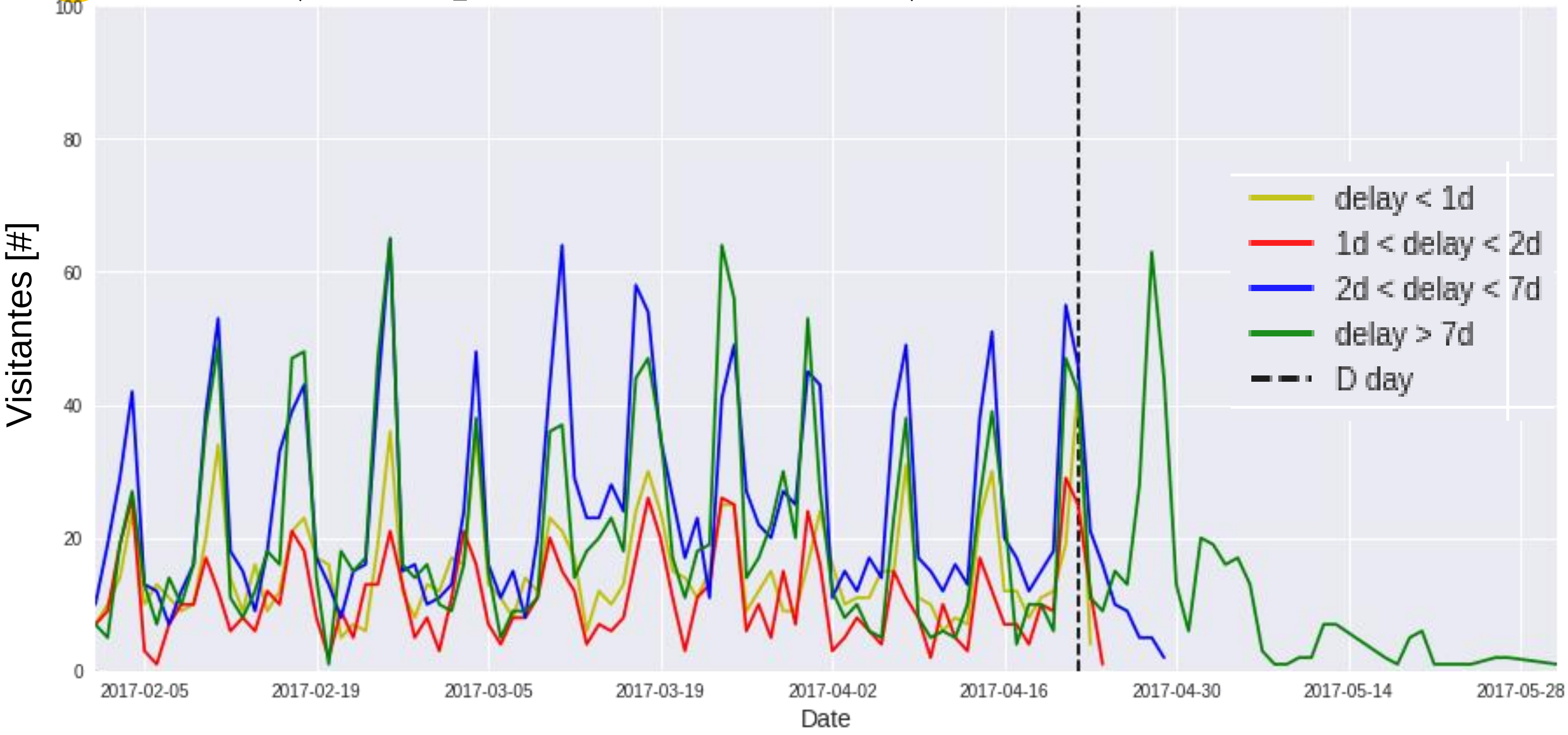
Total de Locales



AIR (Anticipación de reservas)

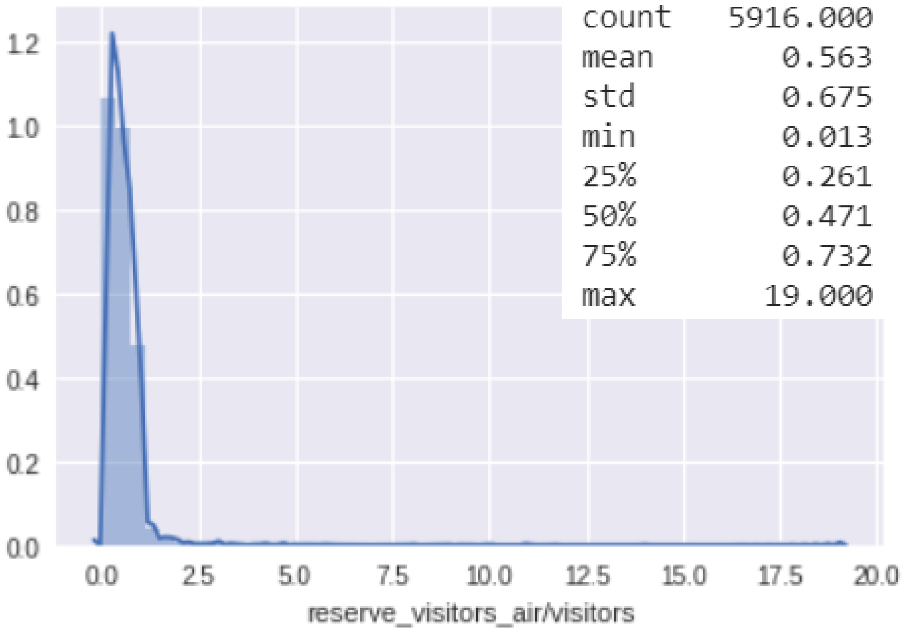


HPG (Anticipación de reservas)



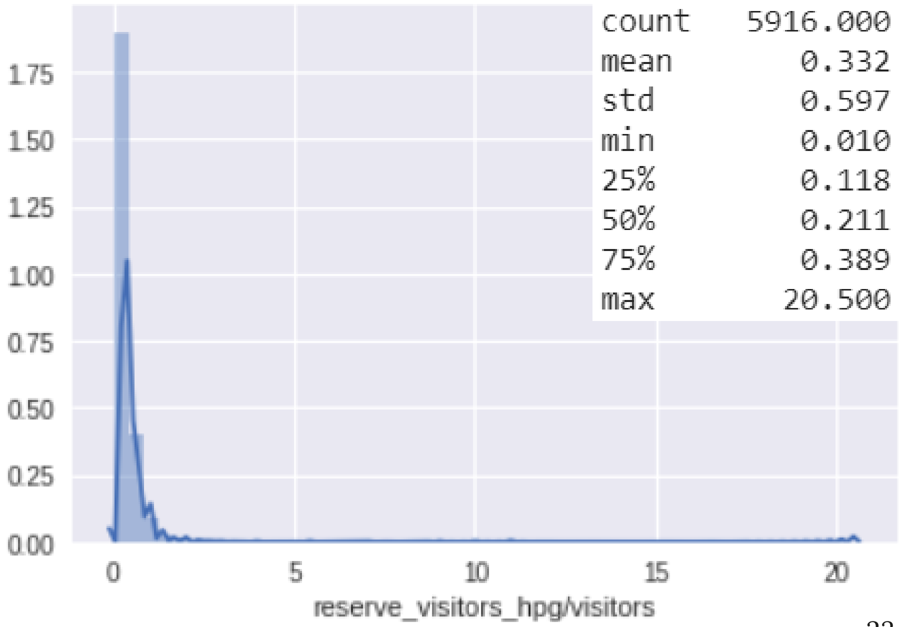
Overbooking

AIR



count	5916.000
mean	0.563
std	0.675
min	0.013
25%	0.261
50%	0.471
75%	0.732
max	19.000

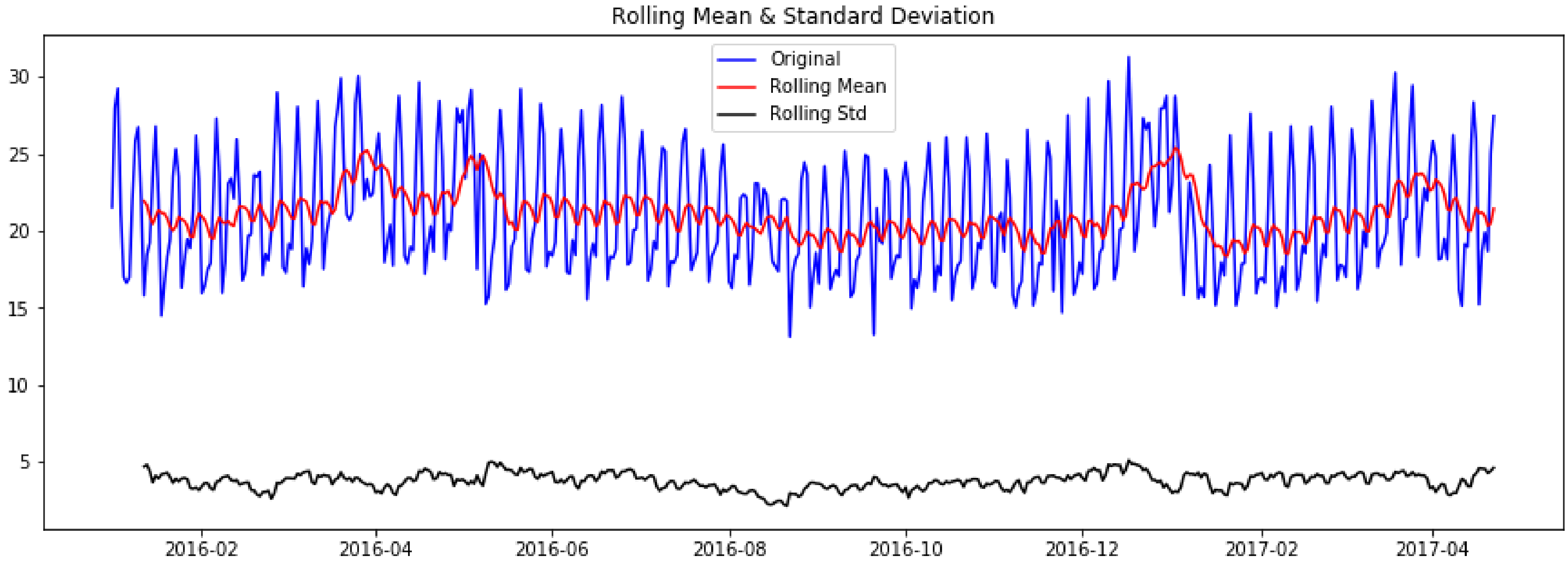
HPG



count	5916.000
mean	0.332
std	0.597
min	0.010
25%	0.118
50%	0.211
75%	0.389
max	20.500

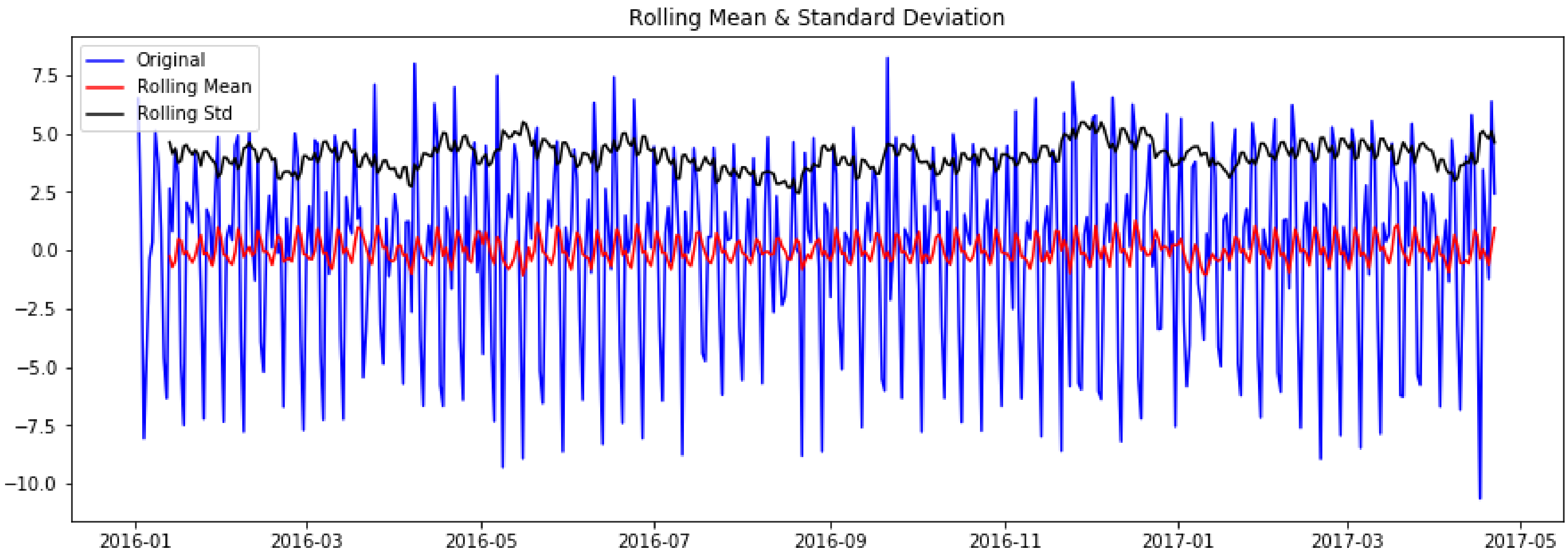
Estacionariedad

test_stationarity(air.visit_mean)



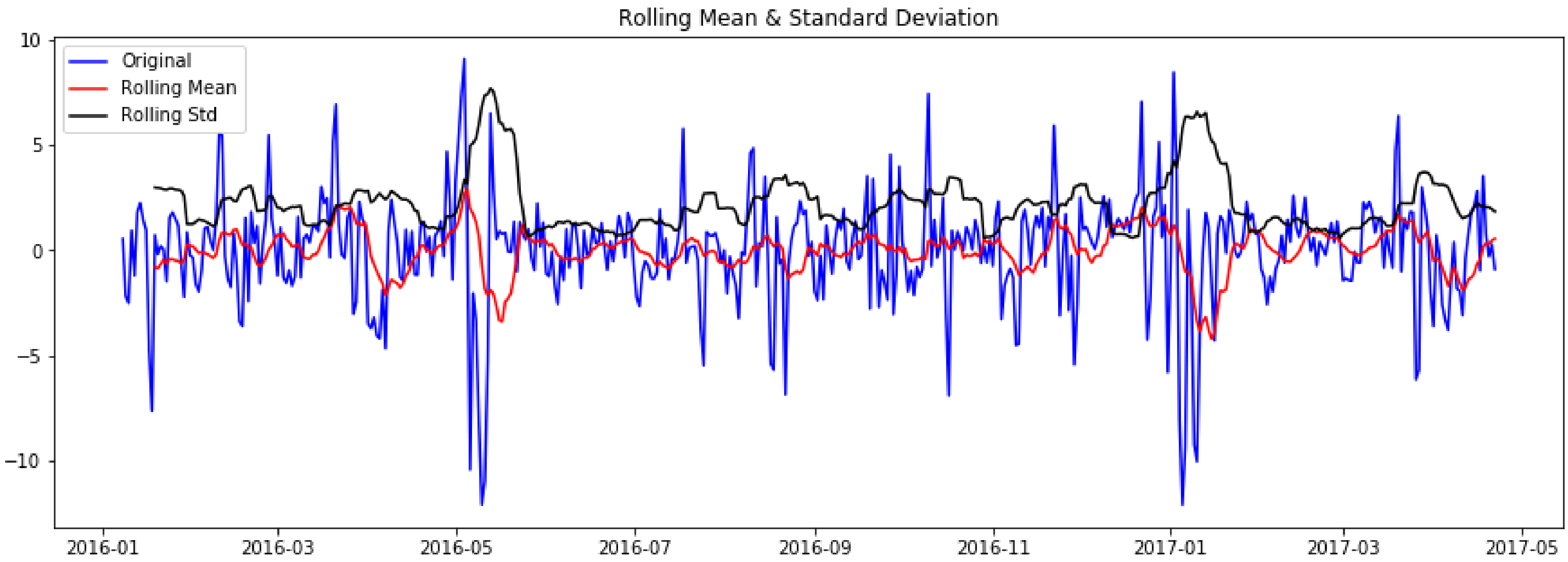
```
air['visit_mean_diff'] = air.visit_mean - air.visit_mean.shift(1)
test_stationarity(air.visit_mean_diff.dropna(inplace=False))
```

Estacionariedad



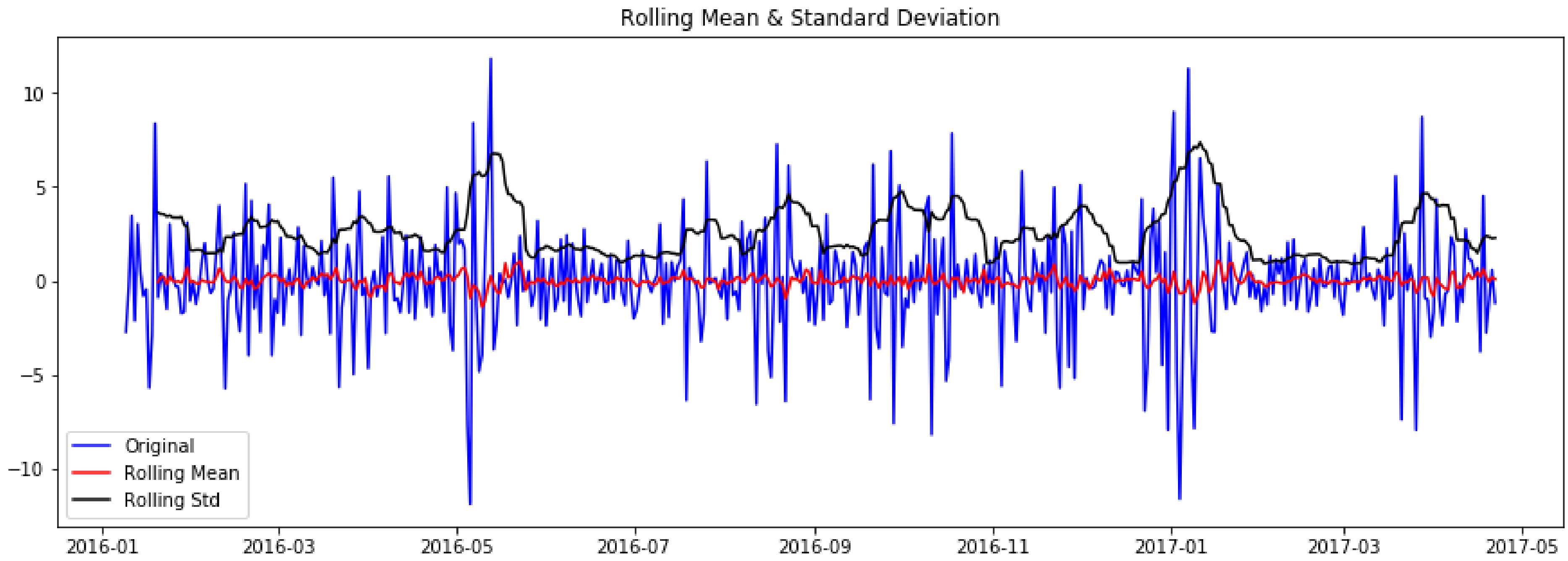
Estacionariedad

```
air['visit_mean_seasonal'] = air.visit_mean - air.visit_mean.shift(7)
test_stationarity(air.visit_mean_seasonal.dropna(inplace=False))
```



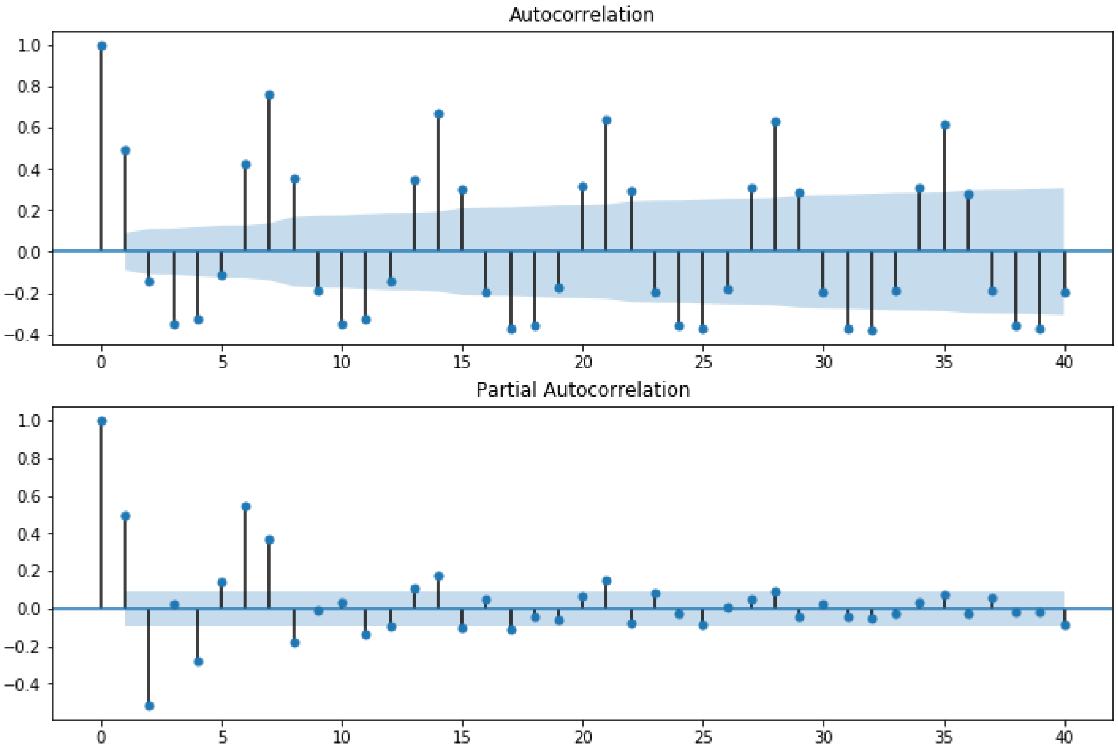
Estacionariedad

```
air['visit_mean_seasonal_diff'] = air.visit_mean_diff - air.visit_mean_diff.shift(7)
test_stationarity(air.visit_mean_seasonal_diff.dropna(inplace=False))
```



Estacionariedad

```
fig1 = sm.graphics.tsa.plot_acf(air.visit_mean, lags=40, alpha=.05, ax=ax1)
fig2 = sm.graphics.tsa.plot_pacf(air.visit_mean, lags=40, alpha=.05, ax=ax2)
```



Features adicionales

GENERALES

- Día / Mes / Año
- Feriados
- Cant. Restaurants (x zona)
- Total de locales
- Total reservas (HPG / AIR)
- Clima

PARTICULARES

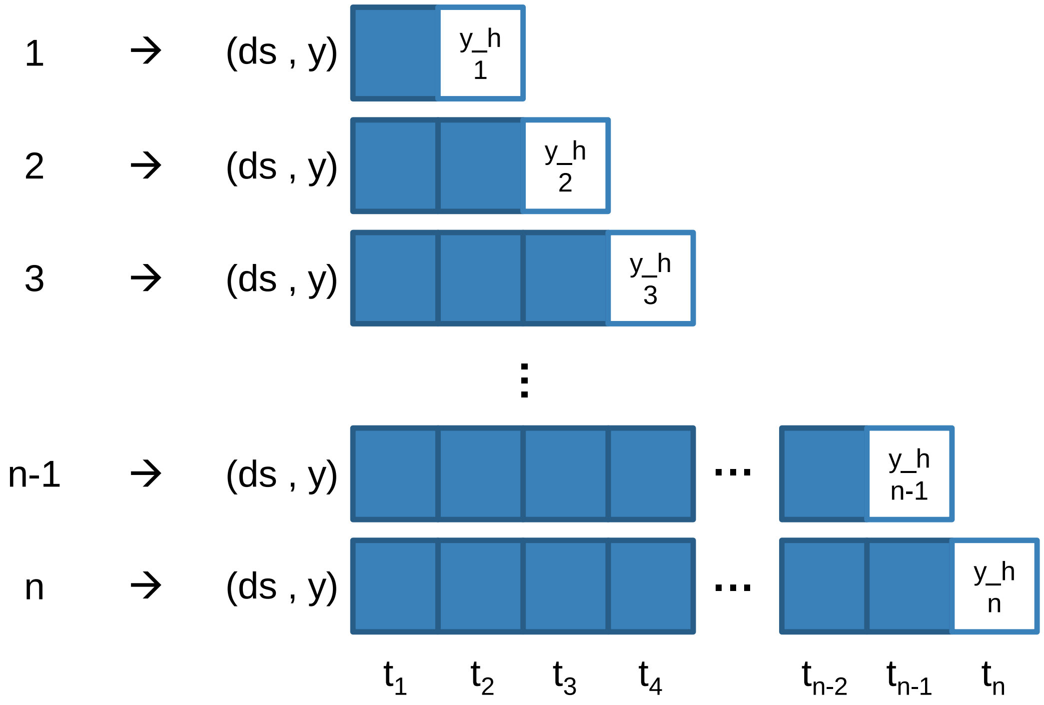
- Tipo de comida
- Ciudad / Área
- Latitud / Longitud
- Anticipación reservas
  1, 2, 7 o más días

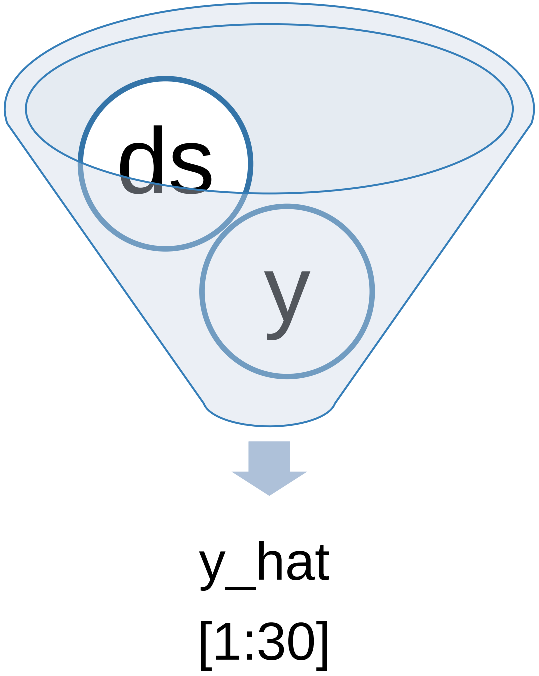
AUTO Regresor

- Predicción de visitas en el corto plazo (1~30 días)

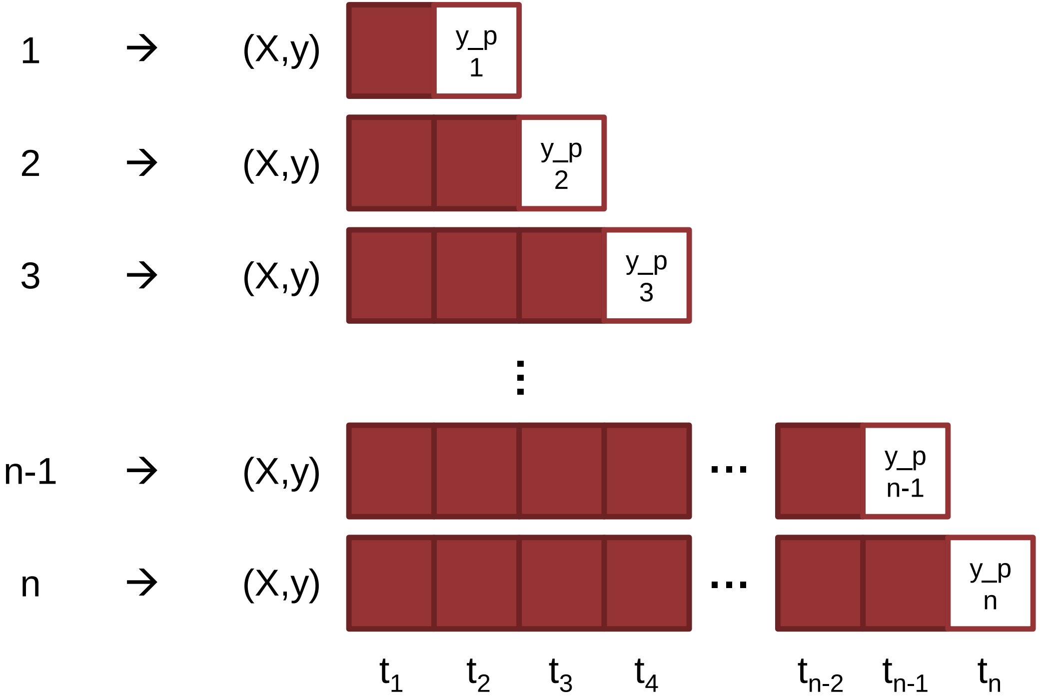
METODOLOGÍA

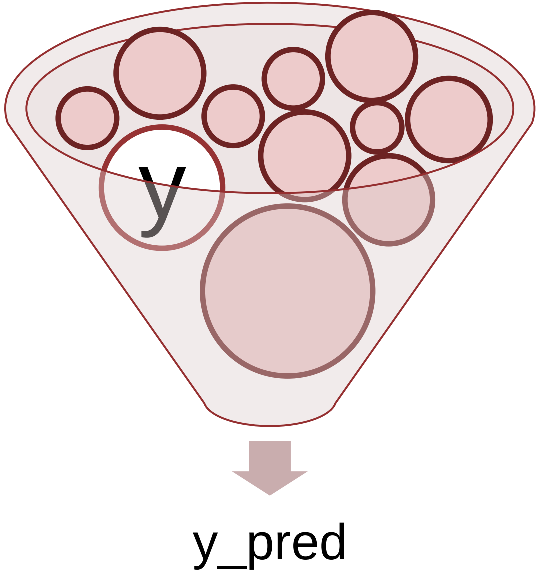
Prophet (facebook)



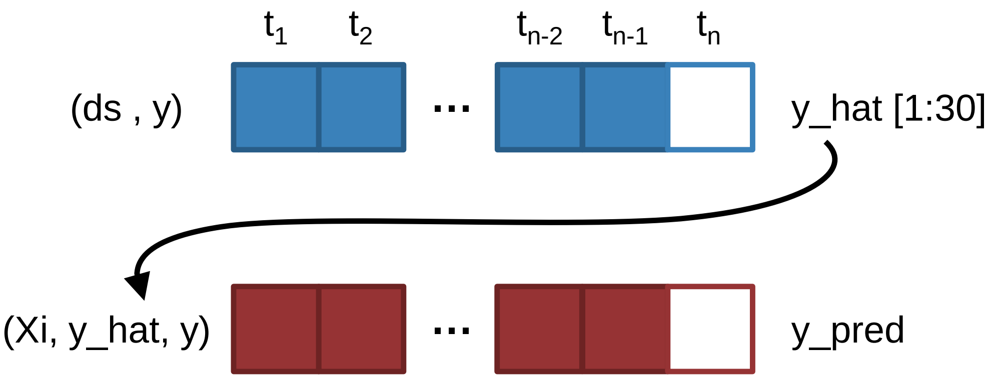


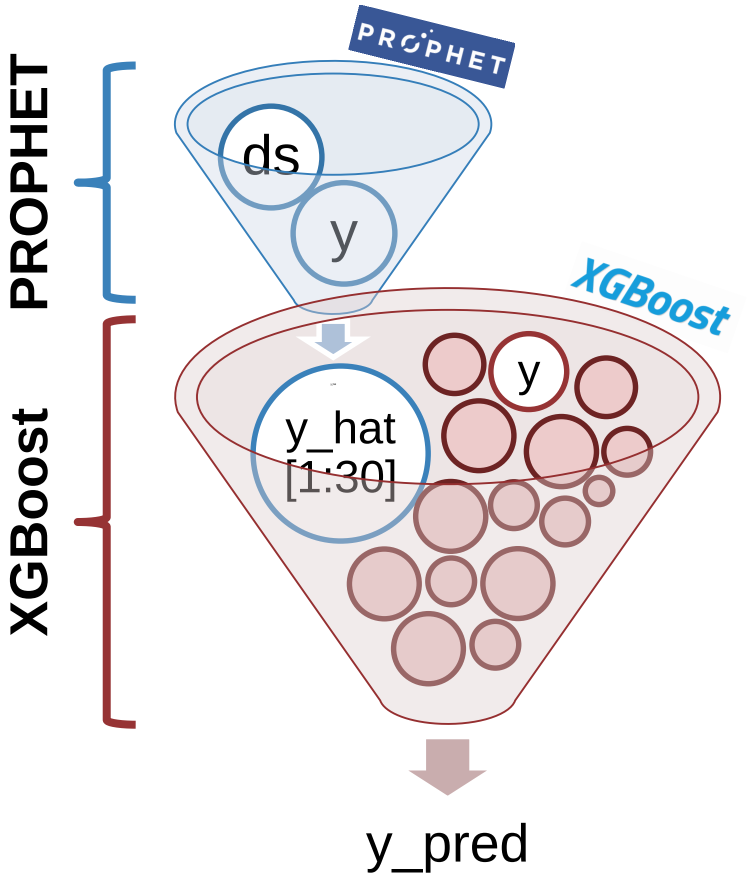
XGBoost (eXtreme Gradient Boosting)



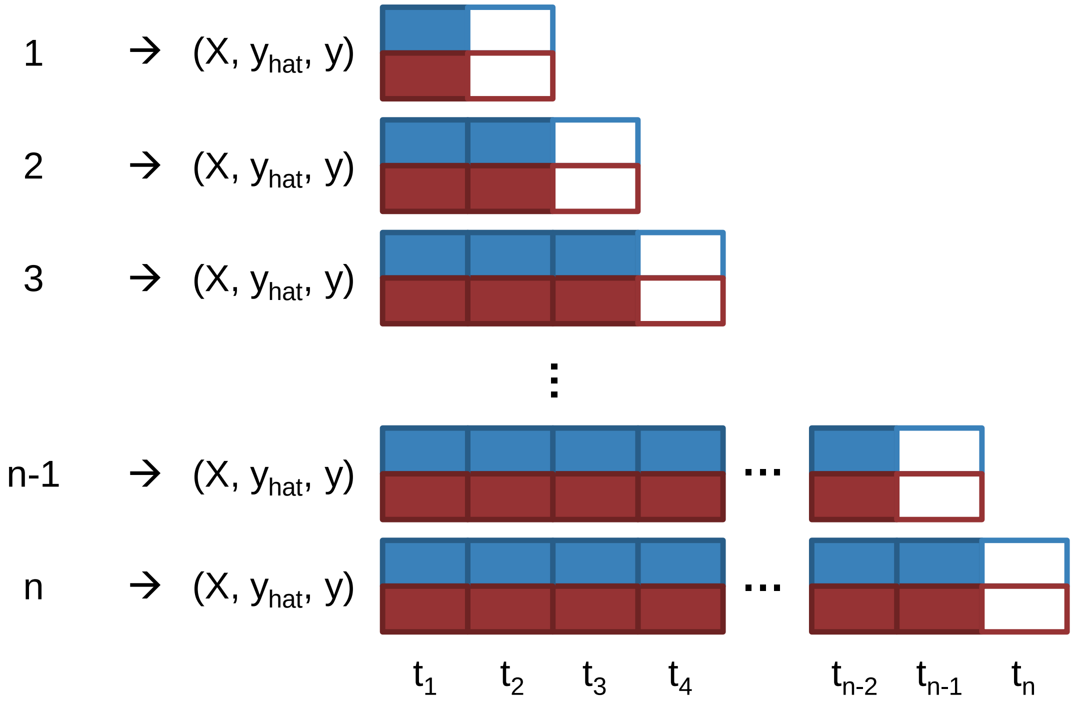


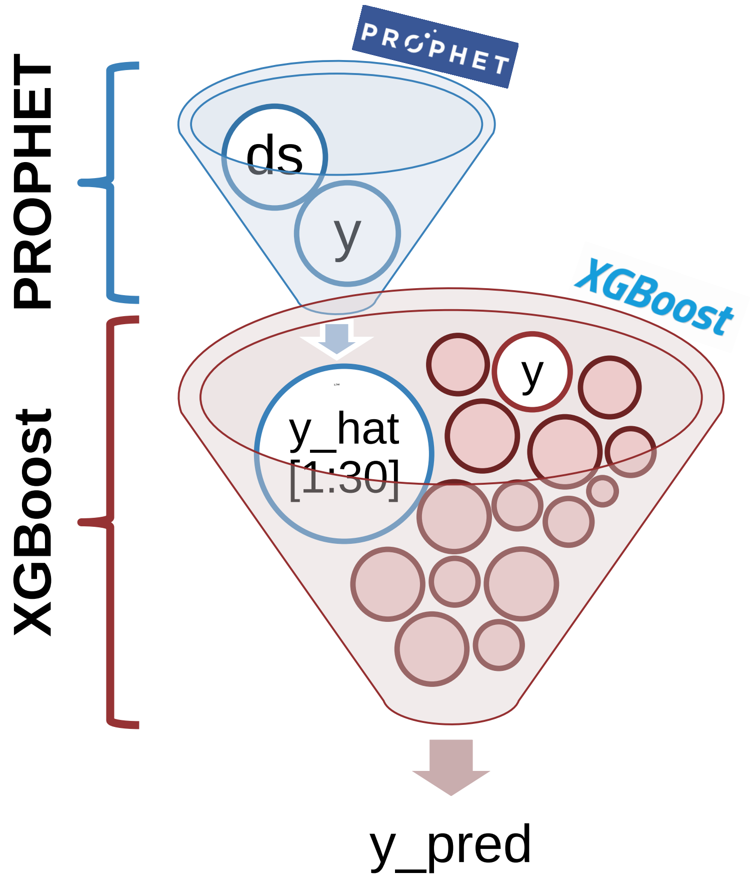
Prophet + XGBoost



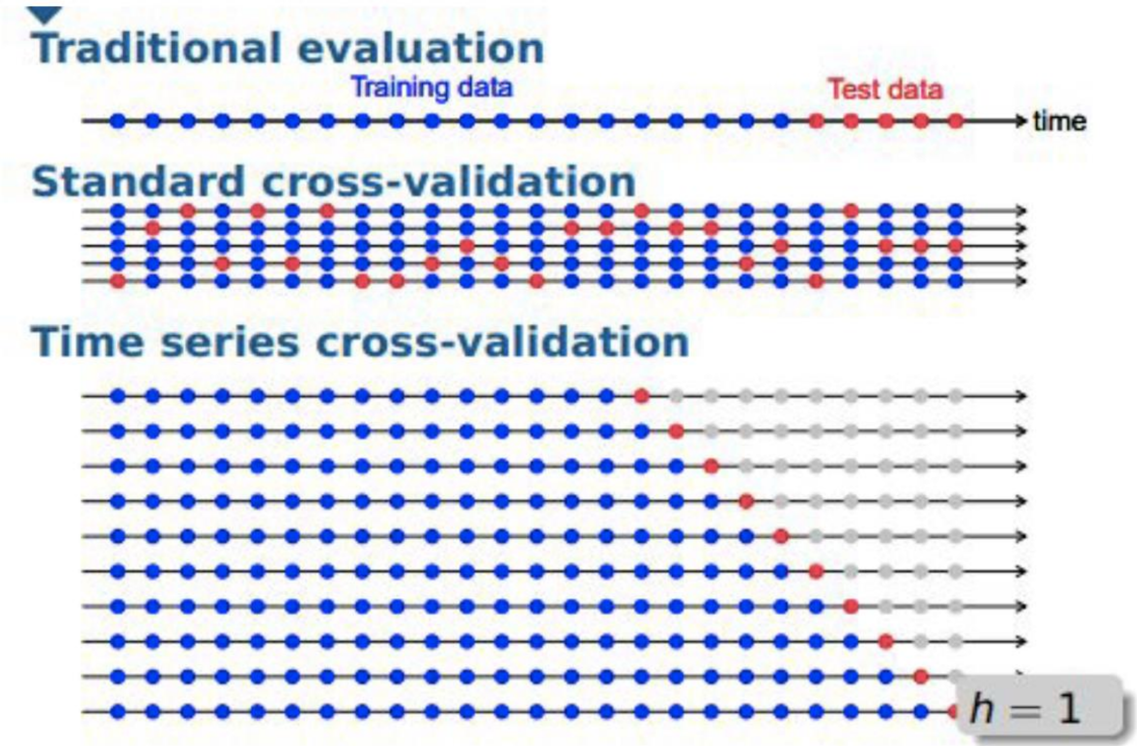


Prophet + XGBoost

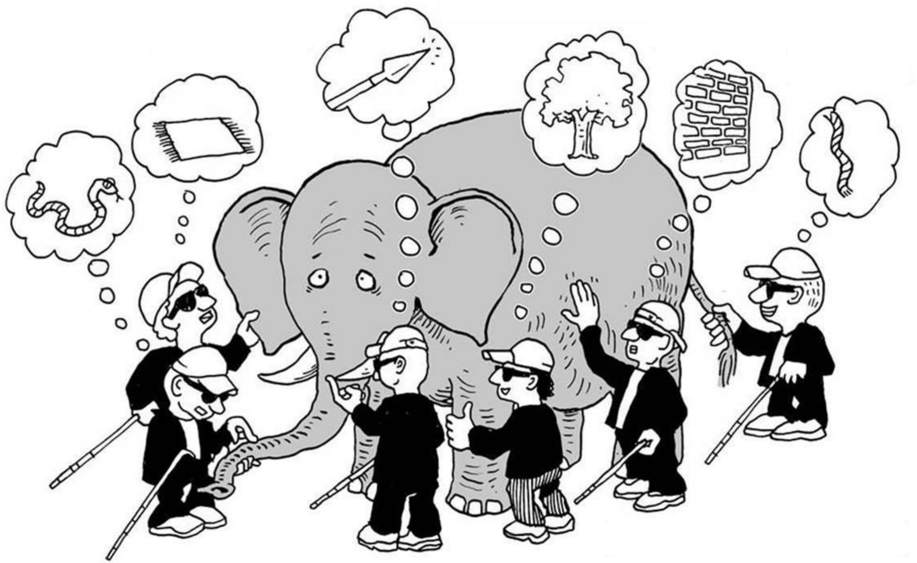


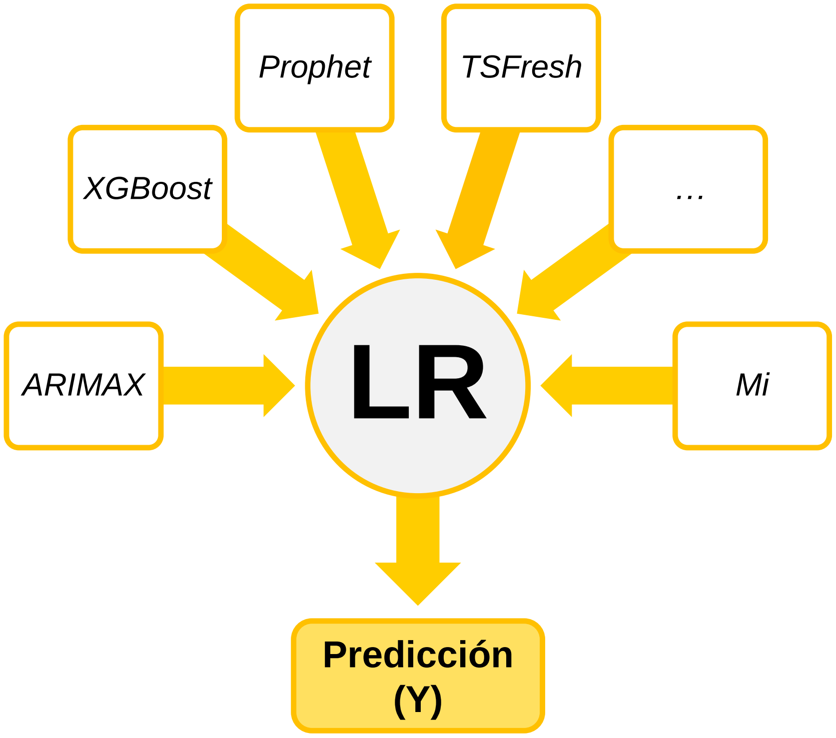


Validación Cruzada



Otro enfoque





RESULTADOS

Resultados

CONCLUSIONES

Conclusiones

- Múltiples enfoques posibles (modelos)
- Importancia del análisis temporal de los datos (TS)
- Features adicionales propios de la Serie de Tiempo
- Información útil a incluir:
  - "capacidad máxima del local"
  - "reserva concretada o no" (Flag_reserva_caida)
- A partir de eso, se podría estimar:
  - Qué reservas se podrían caer
  - Determinar hasta cuánto hacer overbooking

Aplicaciones múltiples

- Reservas turísticas
  - Hotelería
  - airbnb
  - Excursiones
- Pasajes aéreos
- Turnos médicos
- Espectáculos
- Otros

Muchas Gracias!

Preguntas ?