Forecasting Tourism in Barcelona City for the Summer 2021

Capstone Project

Carlos López Martínez
Ignacio Matías Montorfano
Elena Picazo Gurina
Adrià Valls Sánchez

UNIVERSITAT DE BARCELONA
Data Science and Big Data

OUTLINE

¿ What Will be the Number of Visitors for Summer 2021 in Barcelona City?

1. Data Gathering
2. Explanatory Analysis
3. Modelling
4. Results

DATA GATHERING

DATA BASE

TRAVELERS

Number of travelers
Number of foreign travelers
Number of spanish travelers

Source: INE

TRANSPORTATION

Number of flights
Number of ships

Source: European Comission, Ministerio de Transportes, Movilidad y Agenda Urbana, Port de Barcelona

HOTELS

Average price for a night
Average of Overnight stays
Number of open hotels
Estimation of available rooms
Percentatge of ocupation
Number of workers

Source: INE

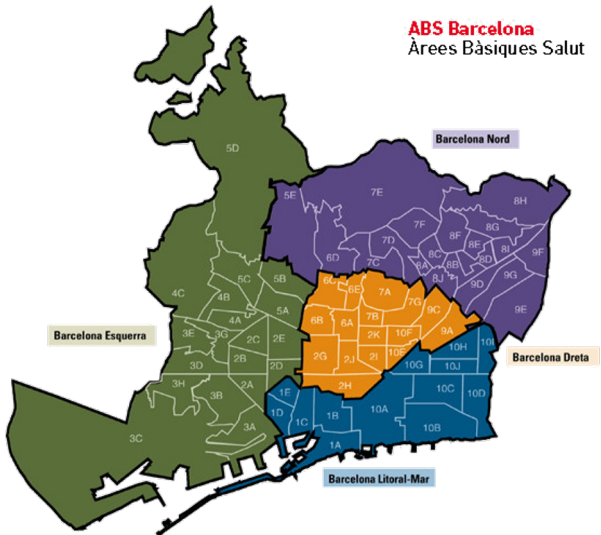
COVID

COVID Cases
Number of people with first vaccination dose
Number of people with second vaccination dose
Number of PCR tests
Number COVID tests different from PCR

Source: Gencat

INPUTS

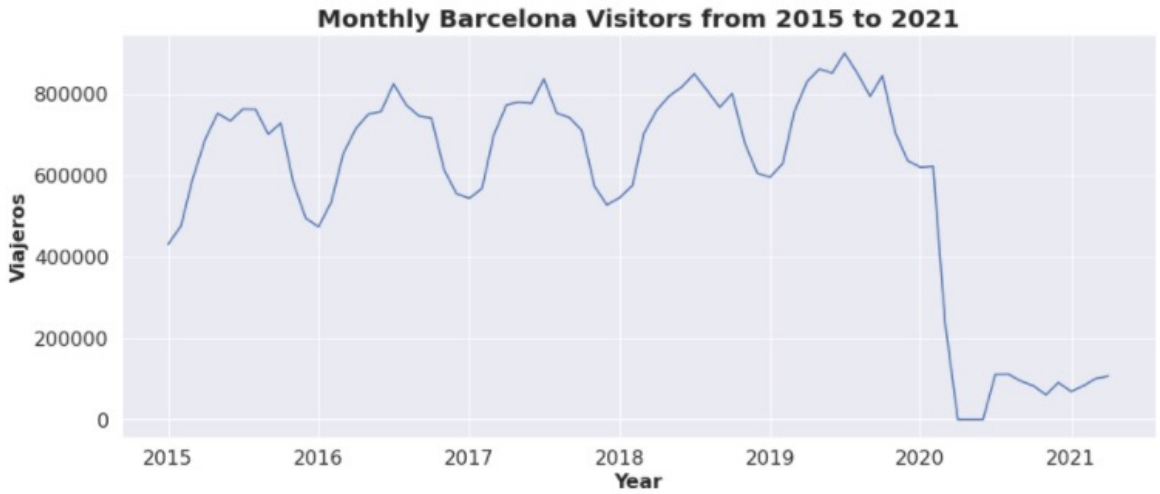
Month
Temperature
Number of Google searches
Unemployment
Others



EXPLANATORY ANALYSIS (i)

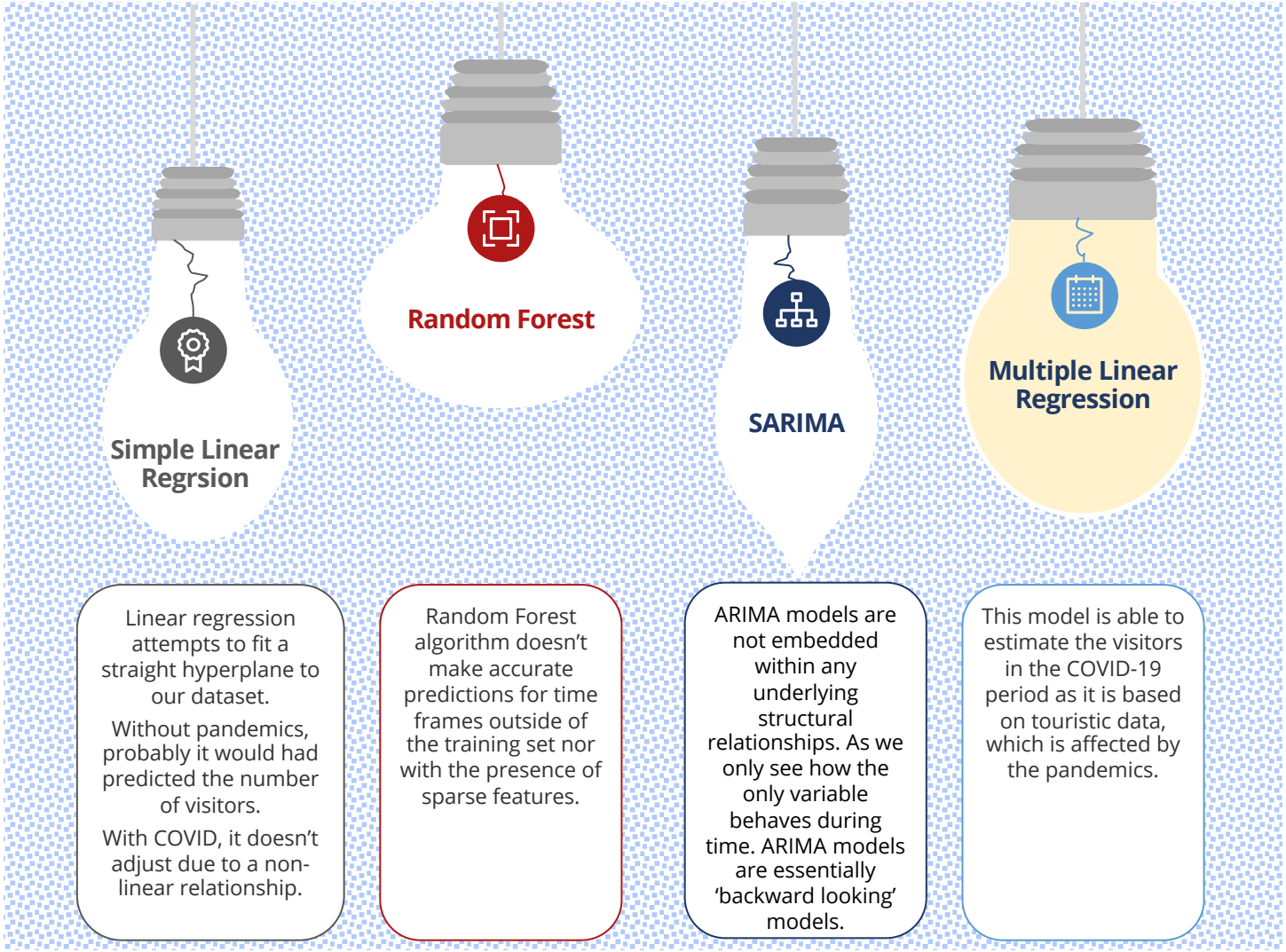
Once the database has been merged, we analyze the behavior of our **dependent variable 'Viajeros' that refers to number of visitors in the city of Barcelona.** This analysis shows:

- The number of visitors has a seasonality pattern, with the spring and summer periods having the highest volume of tourists
- Likewise, a positive trend can observe between years, from 2015 to 2019 the number of visitors has increased

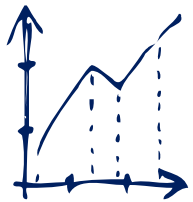


MODELLING

Models applied



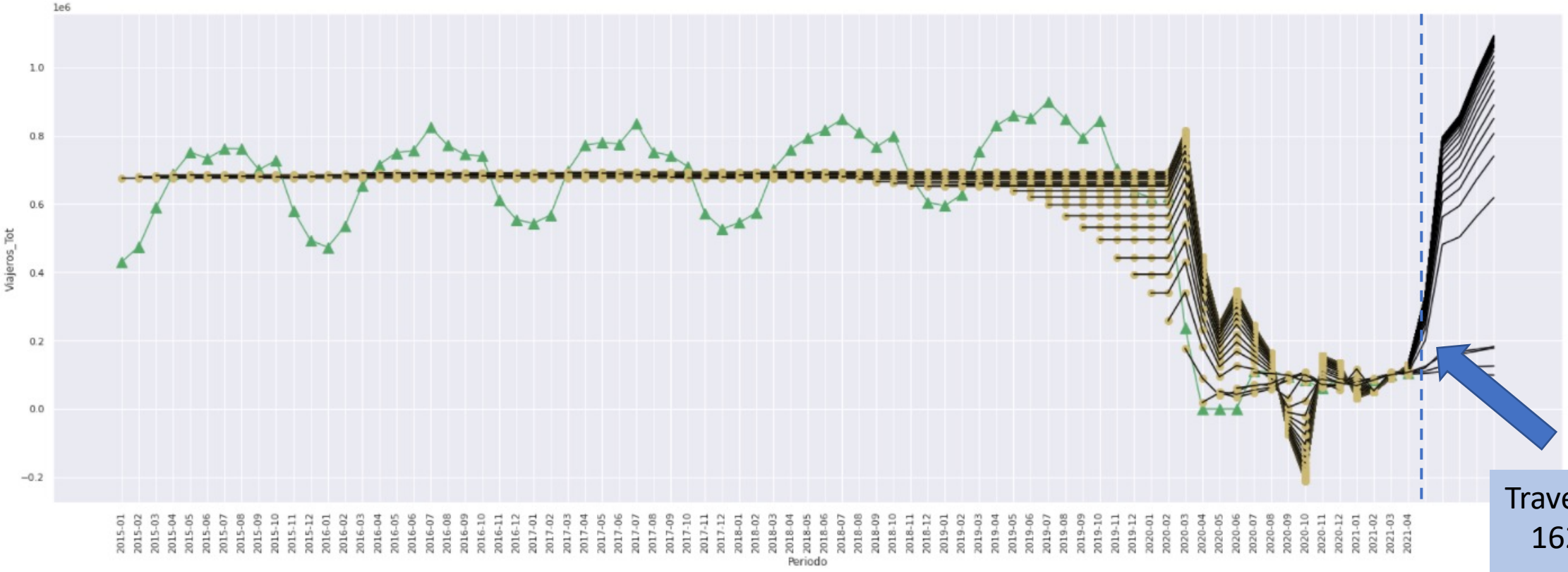
RESULTS: Multiple Linear Regression



We have observed that the level of estimation of the tourists for the summer 2021 depends on the training period considering samples from the non-COVID period.

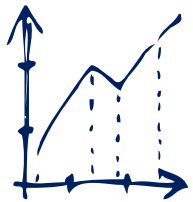
In the following, we make all the trainings considering scenarios that range from the beginning of the database until the drop of tourist due to the lock-down in Spring 2020.

We consider only COVID features. **No touristic features to avoid correlations**



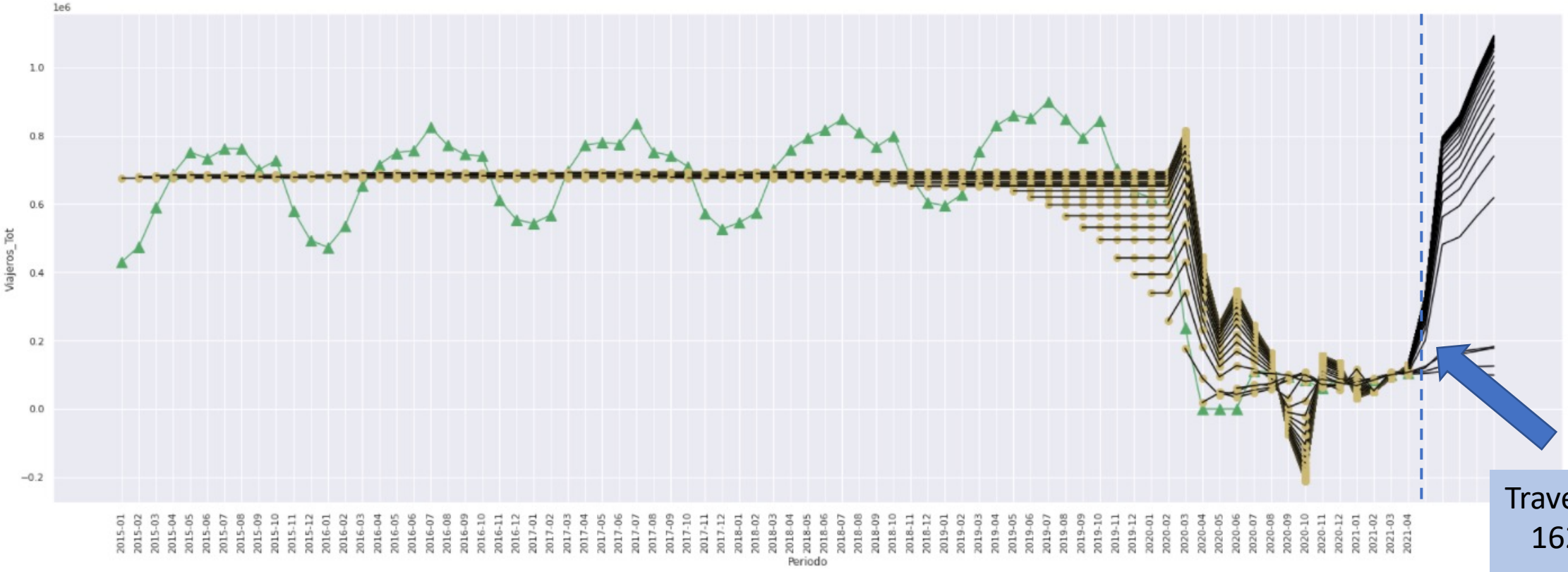
Travelers May 2021
162000 approx.
INE

CONCLUSIONS: Multiple Linear Regression



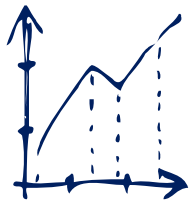
The capability to forecast a **growth in the number of tourists for summer 2021** depends on both:

1. **To observe de decrease in the number of tourists in March 2019 due to the close-down**. If not observed, no growth is predicted
2. The growth in the number of tourists in summer 2021, if the close-down period in May 2019 is observed, is linked to the **vaccination process**

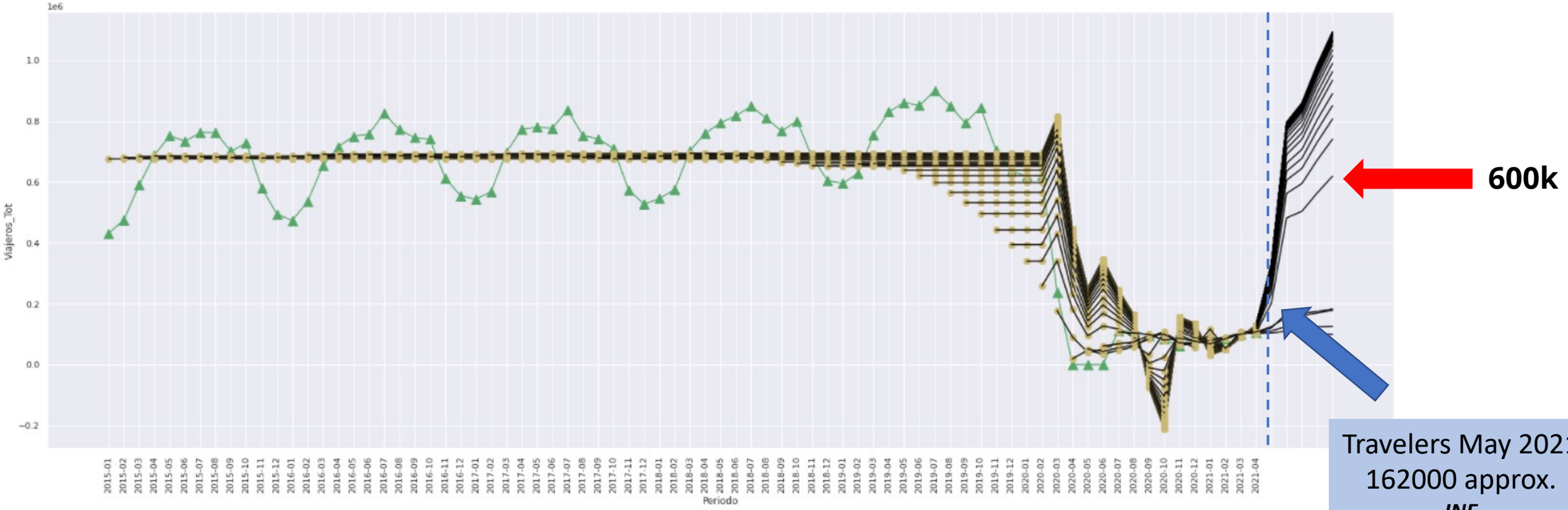


Travelers May 2021
162000 approx.
INE

FORESCAST: Multiple Linear Regression

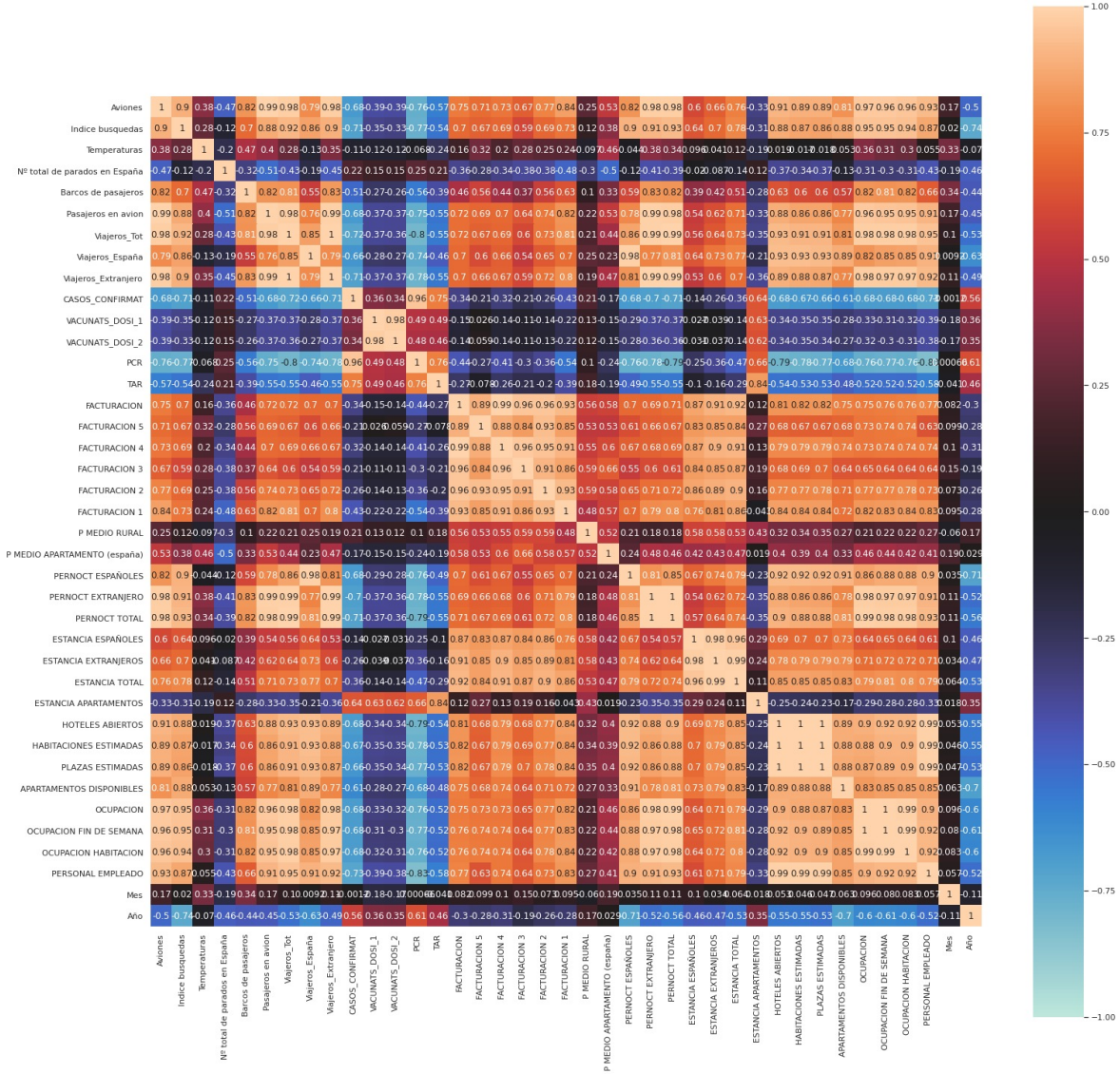


Using the data observed in May 2021 (not used in the overall process) to calibrate the curves, our forecast for summer 2021 is to have a **visitors peak of 600k in July 2021**

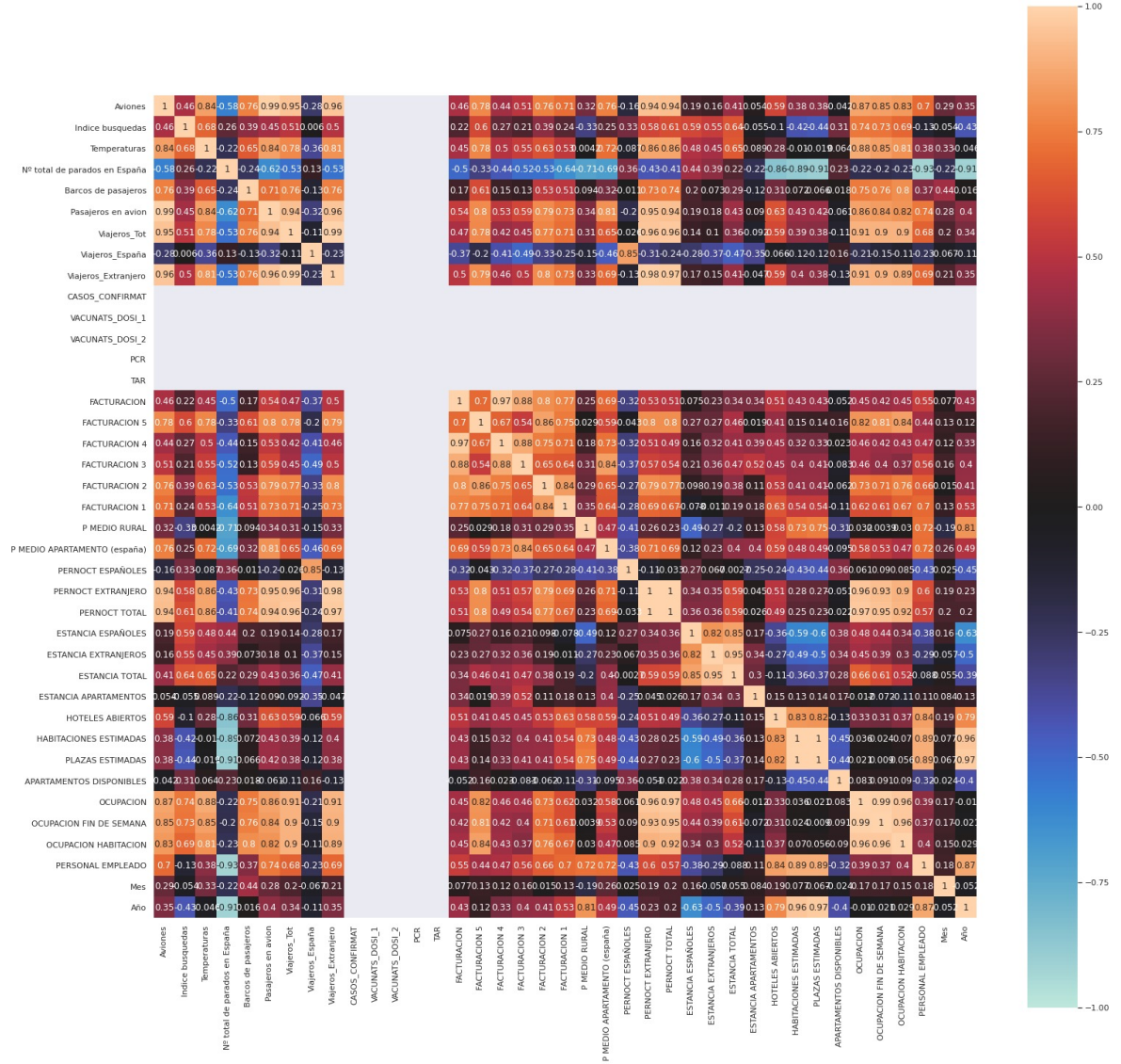


Travelers May 2021
162000 approx.
INE

EXPLANATORY ANALYSIS - CORRELATIONS

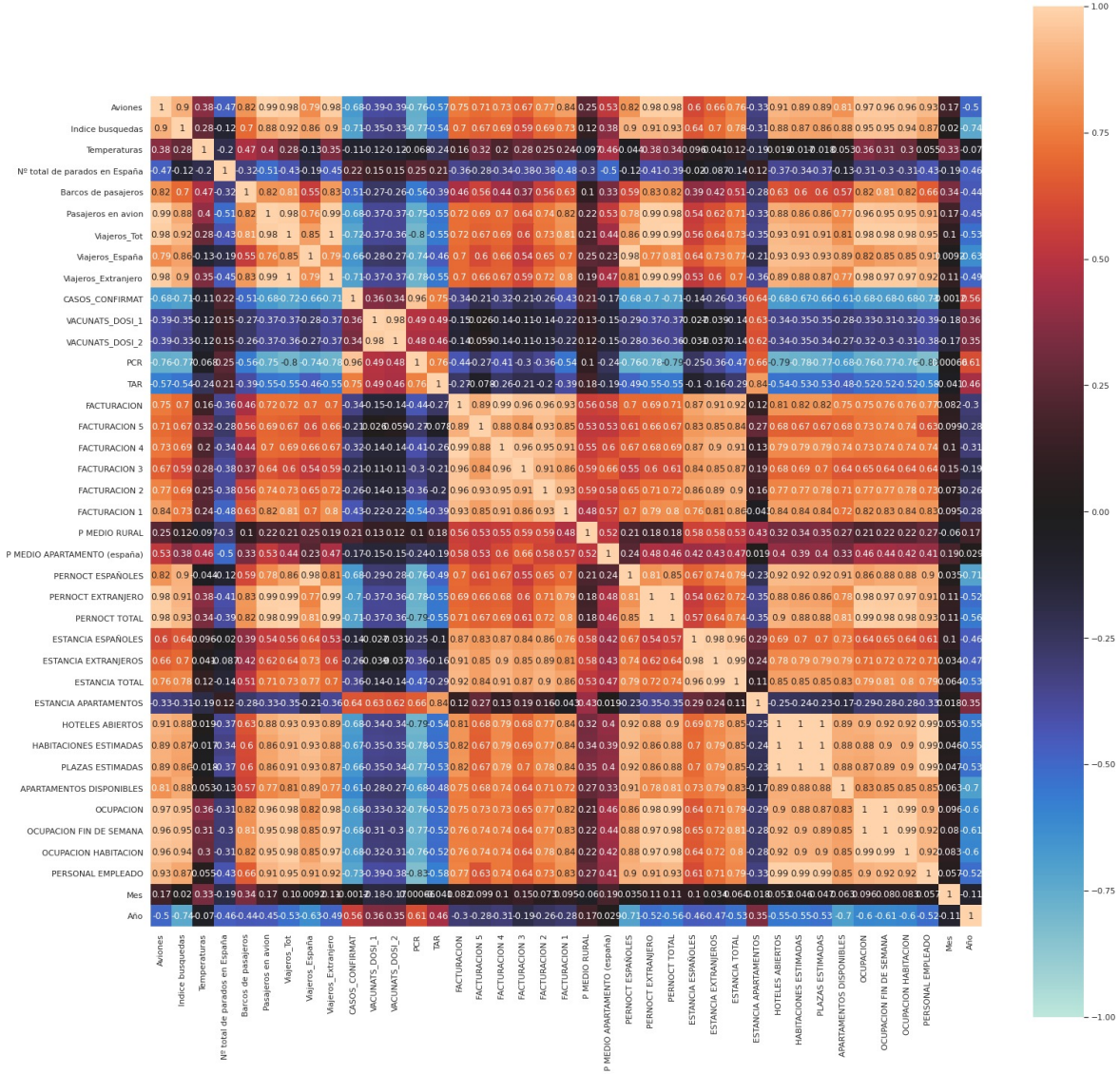


Blind correlation

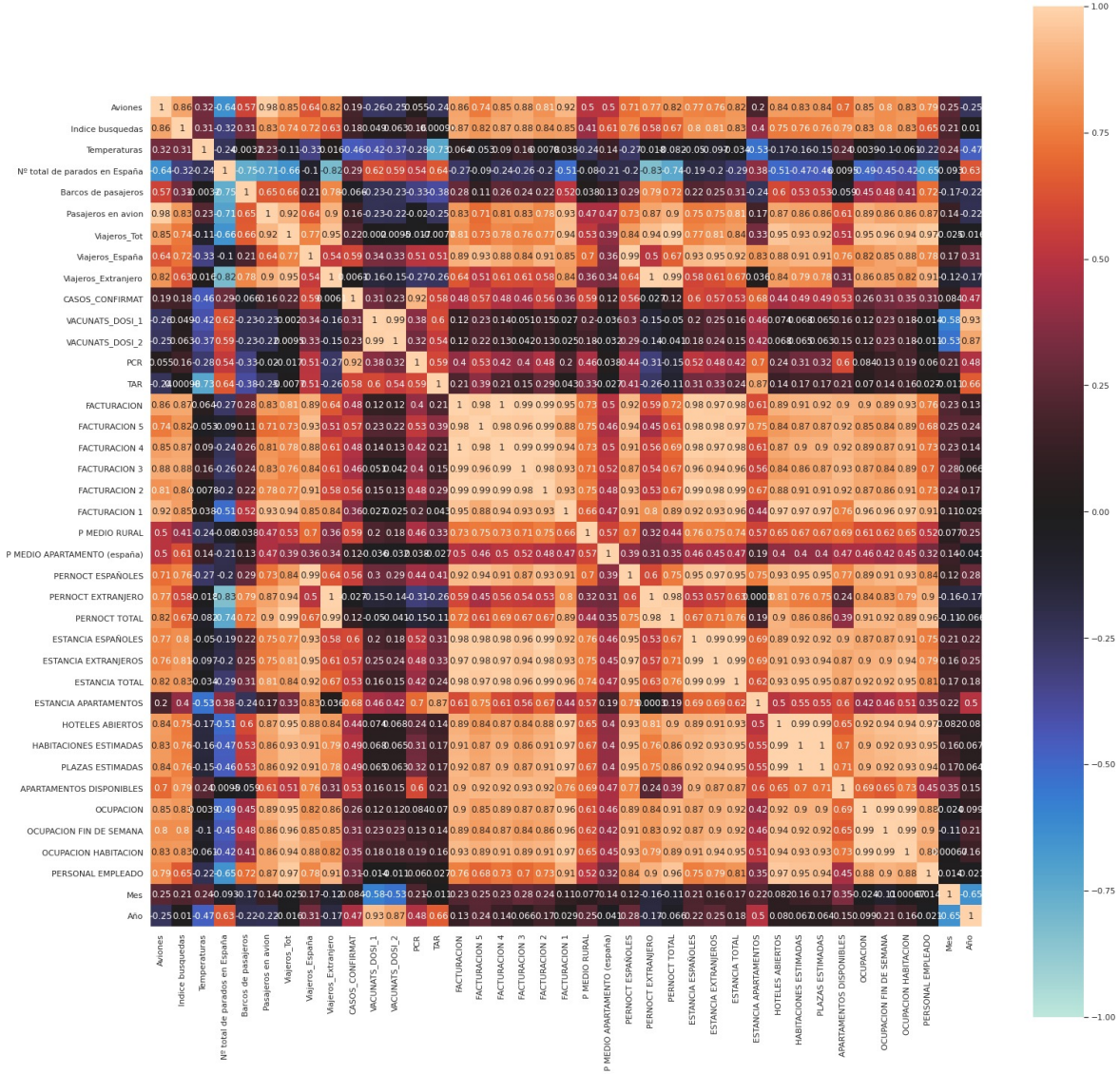


Correlation **before** pandemics

EXPLANATORY ANALYSIS - CORRELATIONS



Blind correlation



Correlation **in** pandemics