The Battle of Neighborhoods: Buenos Aires data – Airbnb

Author: Máximo La Pietra – maximo.lapietra@ing.austral.edu.ar

Introduction/Business Understanding:

- How should a new businessperson decide where to open a particular accommodation?

€

- What factors should he look at before investing?

- Which neighborhood venues affect a user's rating for location of hostel?

Important Questions

- - 1) What is the determining factor / price of a publication?
- - 2) What is the percentage of occupation of the apartments?
- - 3) What is the average income of a host at AirBnB in Buenos Aires?
- - 4) Is there seasonality in reviews?
- - 5) Which neighborhood / neighborhoods dominate the AirBnB platform in BsAs?
- - 6) How does proximity to transportation affect hostel rating?

Target Audience

This project will be useful for two groups of audience:

1. Travelers: Help them make an informed decision while choosing a accommodation by providing an in-depth analysis of hostels and their neighborhood.

2. Businessperson: Provide useful information and models which can help them where to open their first/next business related to tourism and hosting travelers.

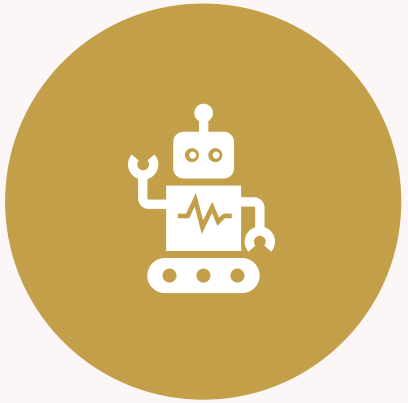
Data Requirements

- 1. [Airbnb Buenos Aires listings] (http://insideairbnb.com/get-the-data.html): This dataset was downloaded from the Airbnb webpage. This is the core dataset with which I'll work.
- 2. [Airbnb Buenos Aires reviews] (http://insideairbnb.com/get-the-data.html): Also, from the Airbnb webpage, in order to analyze the reviews in Buenos Aires.
- 3. [Foursquare API] (https://developer.foursquare.com/docs/api): This API will help me get the venues around the hostel which I will use for EDA and clustering.
- 4.[BuenosAiresNeighborhoods](http://cdn.buenosaires.gob.ar/datosabiertos/datasets/barrios/barrios.geojson): Thisgeojson file will help me get the location of the neighborhoods which I will use for EDA and clustering.

Methodology

FEATURE EXTRACTION → ONE HOT ENCODING

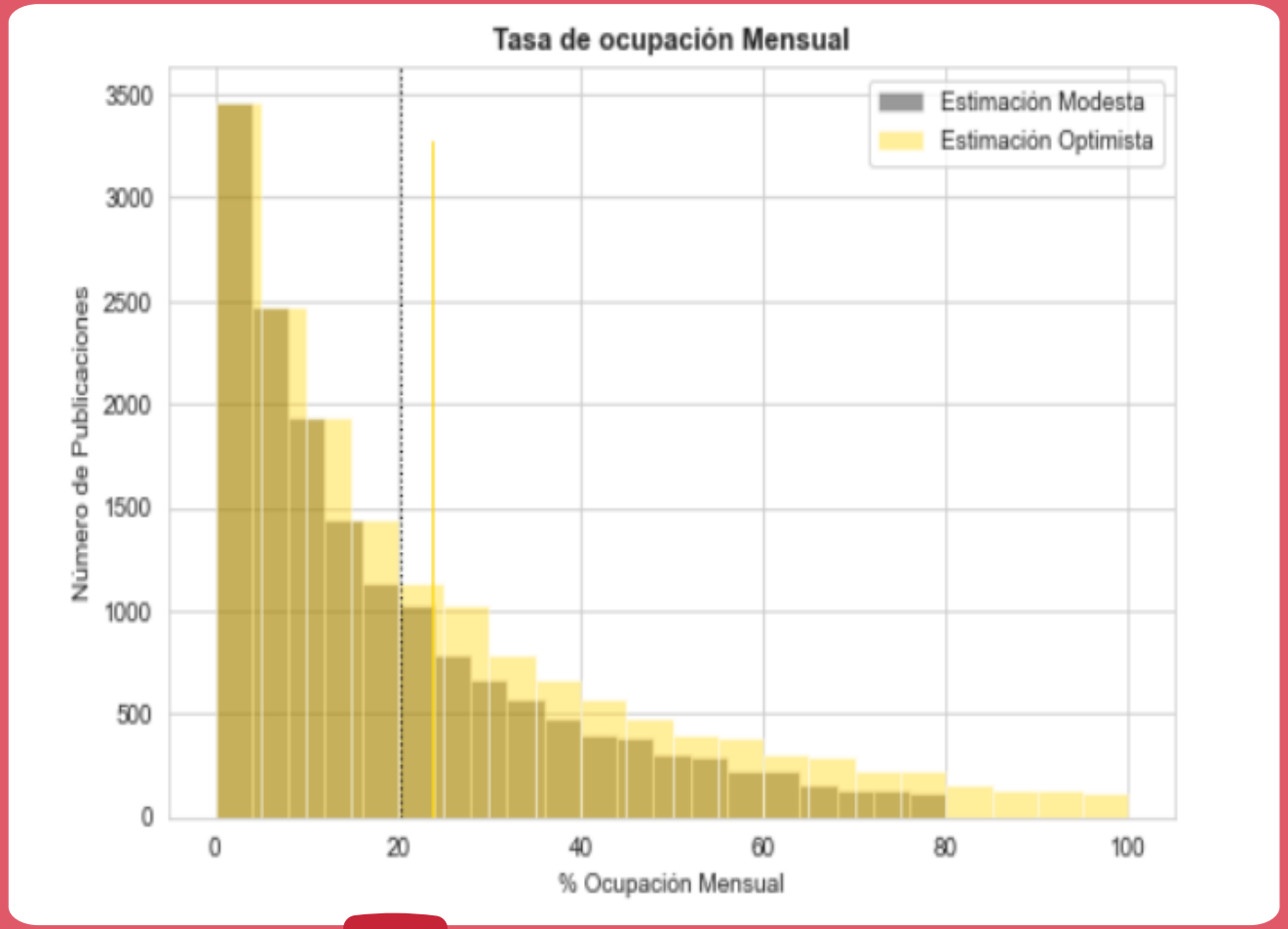
EXPLORATORY DATA ANALYSIS

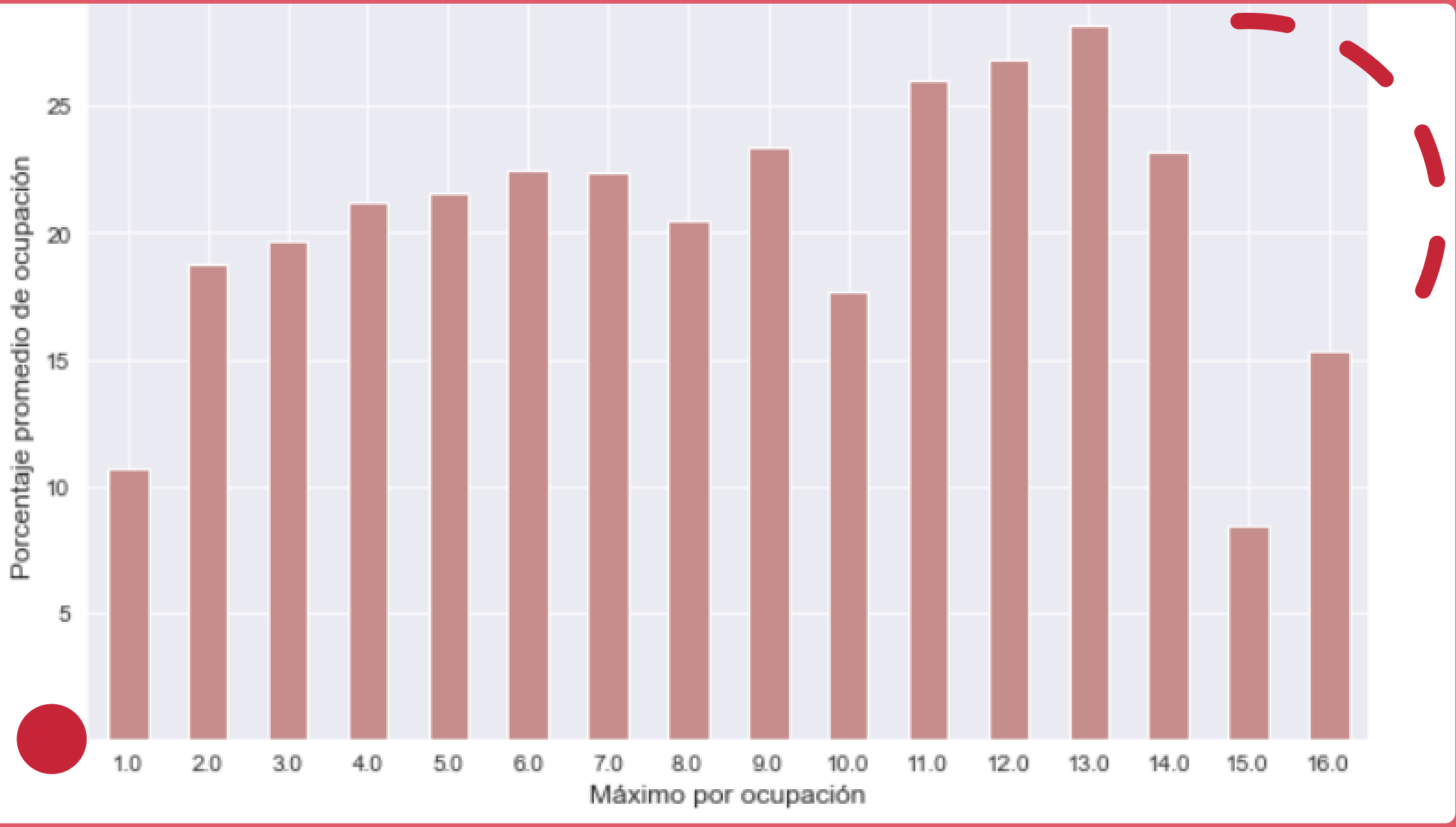


UNSUPERVISED MACHINE LEARNING → K-MEANS CLUSTERING

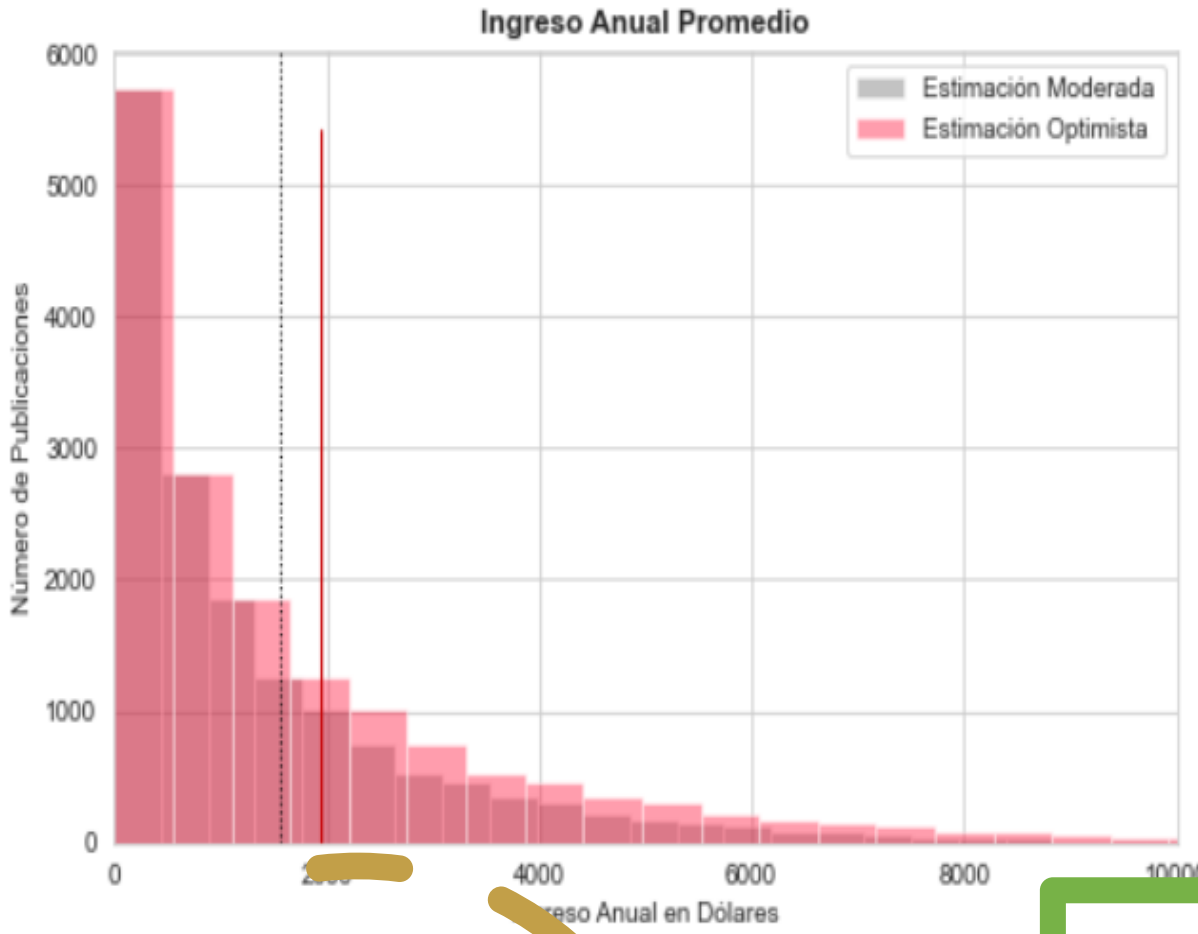
Analysis

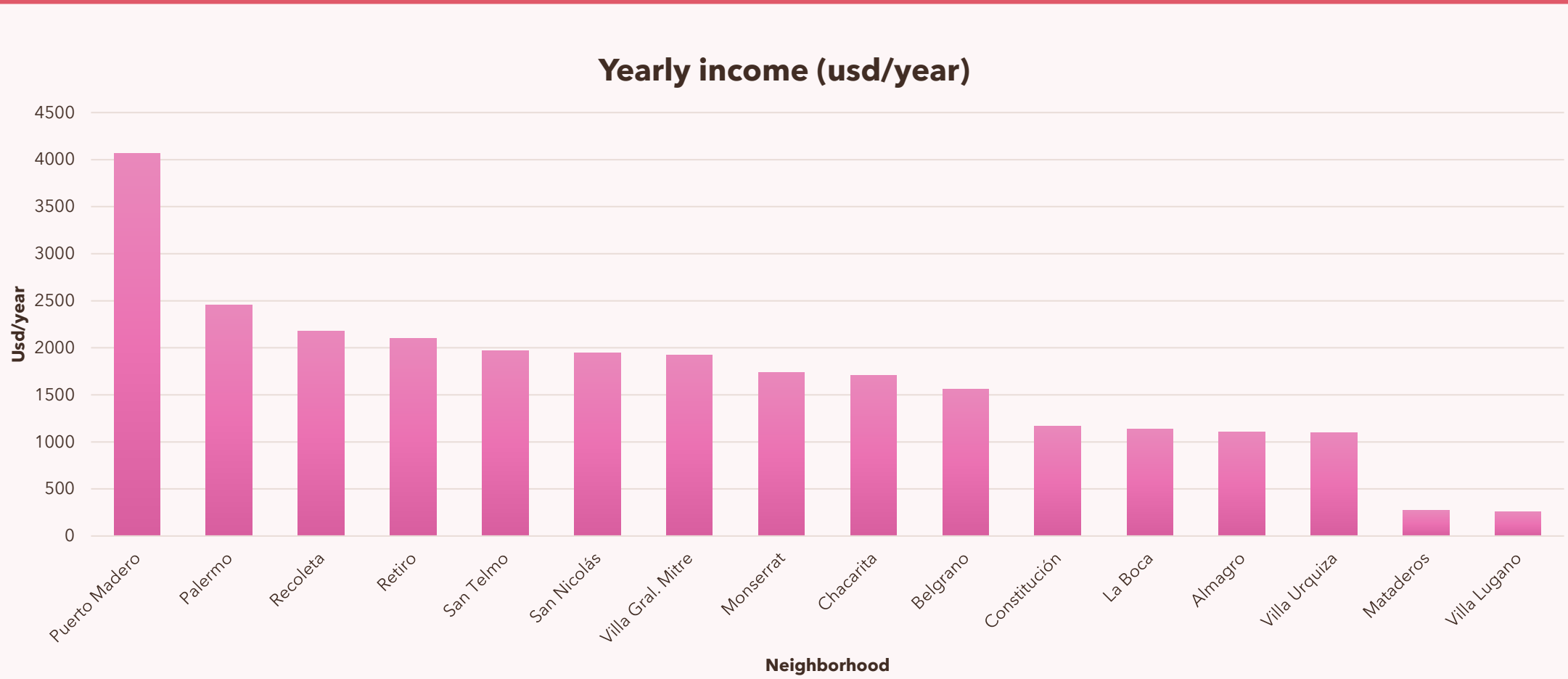
Monthly Occupancy rate

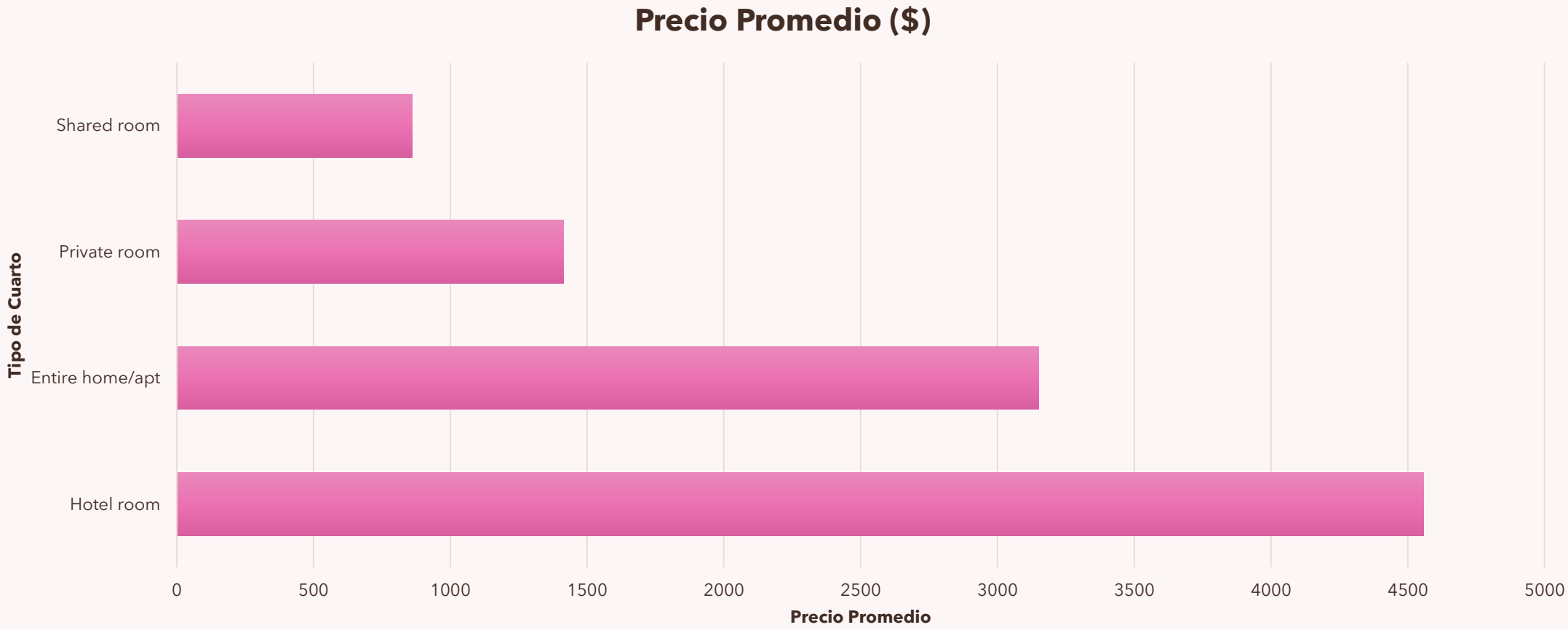


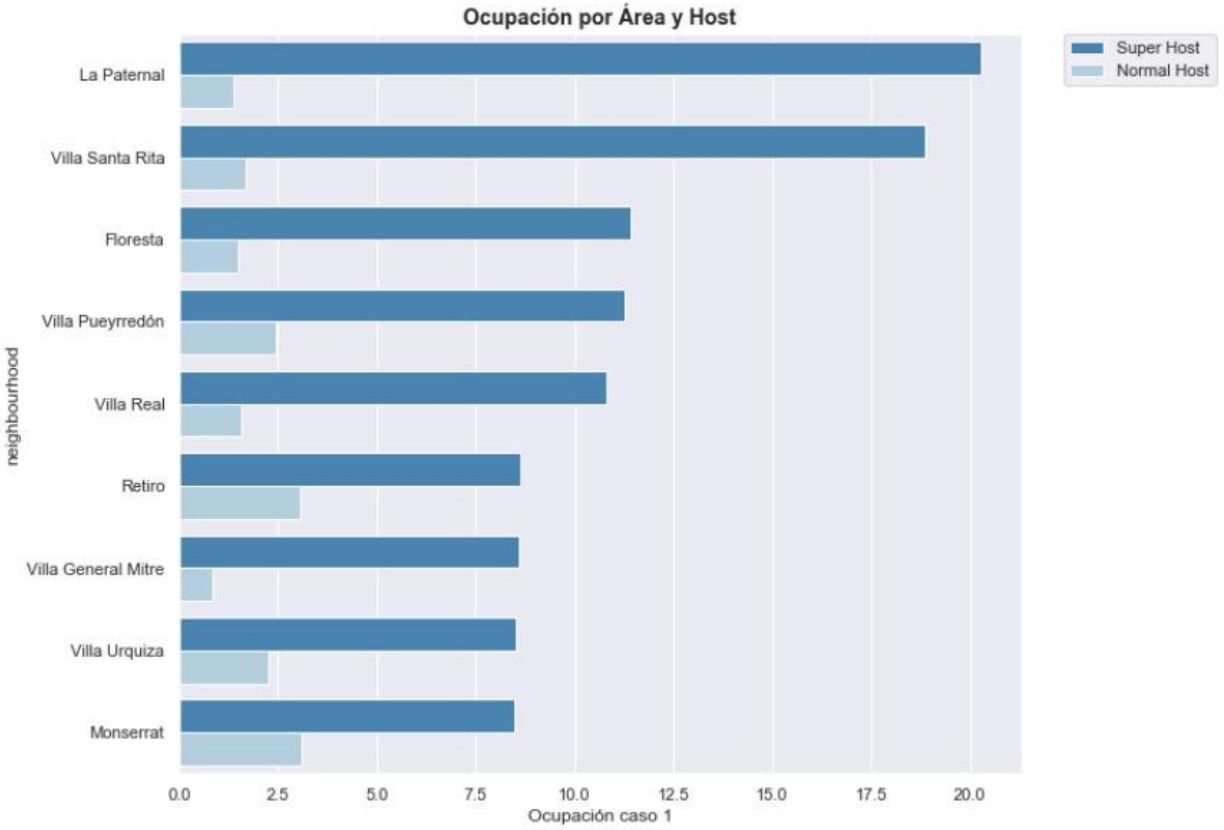

Porcentaje promedio de ocupación
25
20
15
10
5
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
Máximo por ocupación

Annual Income



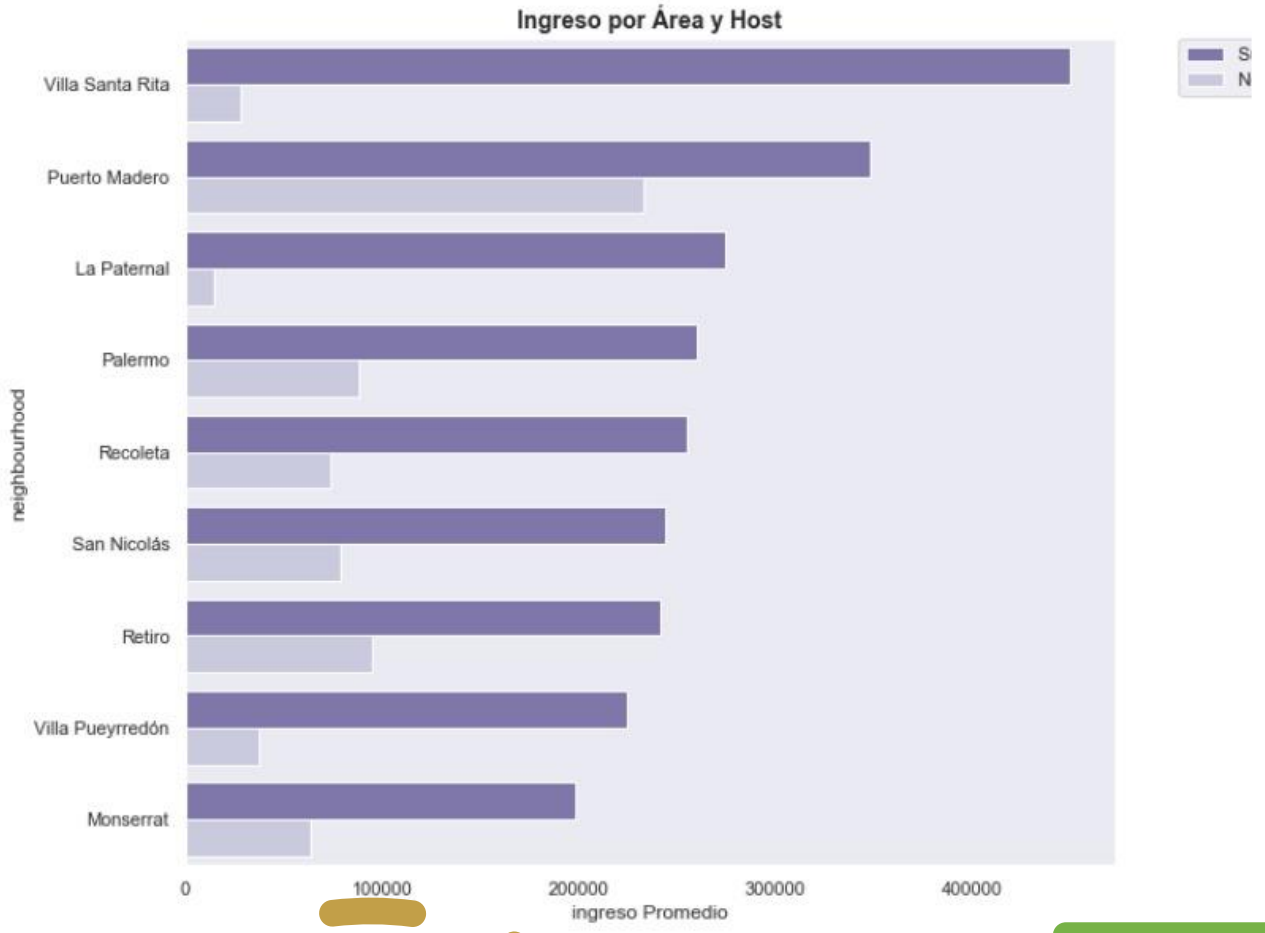

Yearly income (usd/year)
Usd/year
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Puerto Madero
Palermo
Recoleta
Retiro
San Telmo
San Nicolás
Villa Gral. Mitre
Monserrat
Chacarita
Belgrano
Constitución
La Boca
Almagro
Villa Urquiza
Mataderos
Villa Lugano
Neighborhood


Precio Promedio ($)
Tipo de Cuarto
Shared room
Private room
Entire home/apt
Hotel room
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Precio Promedio



Occupancy - SuperHosts

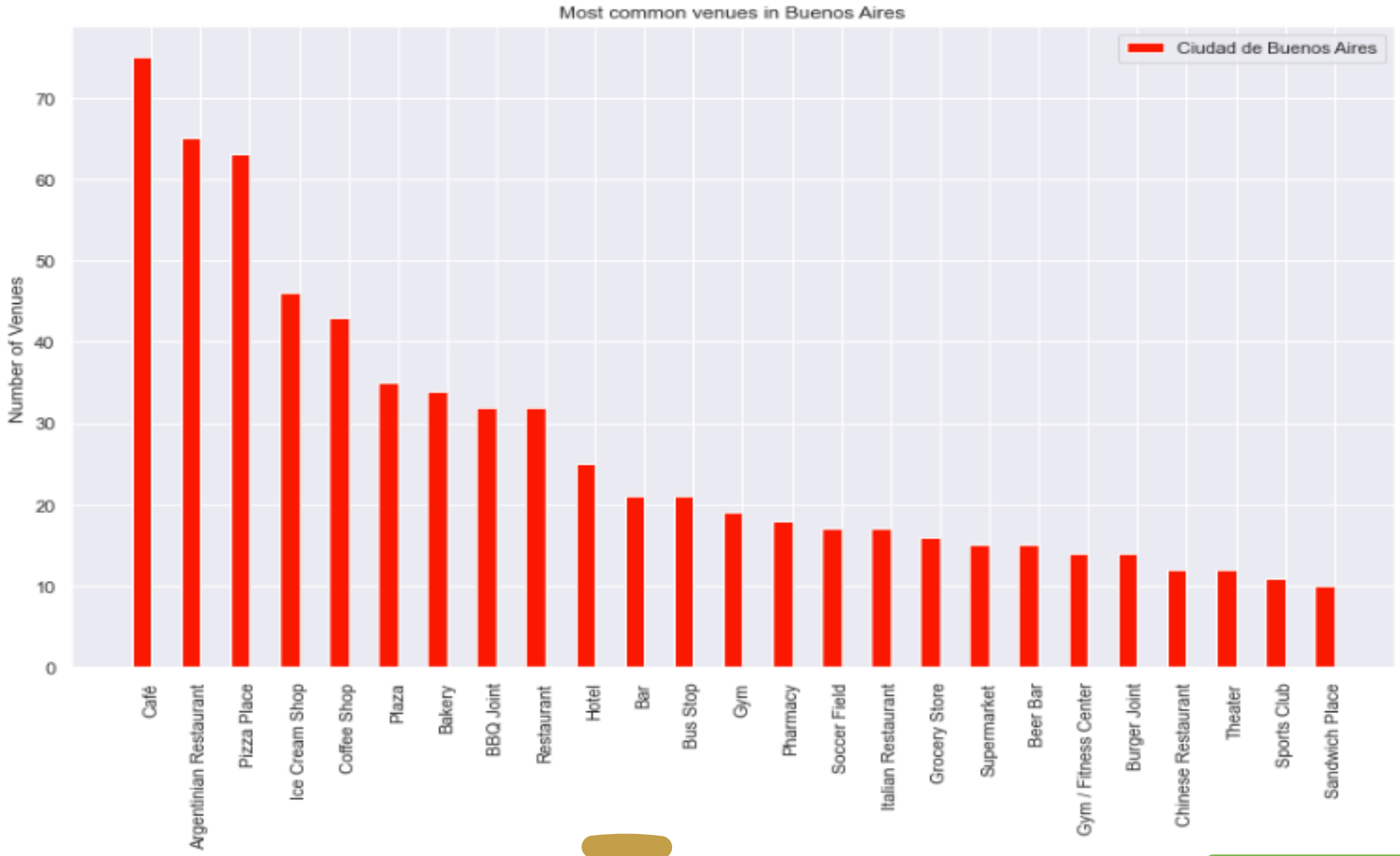
Income -- SuperHosts



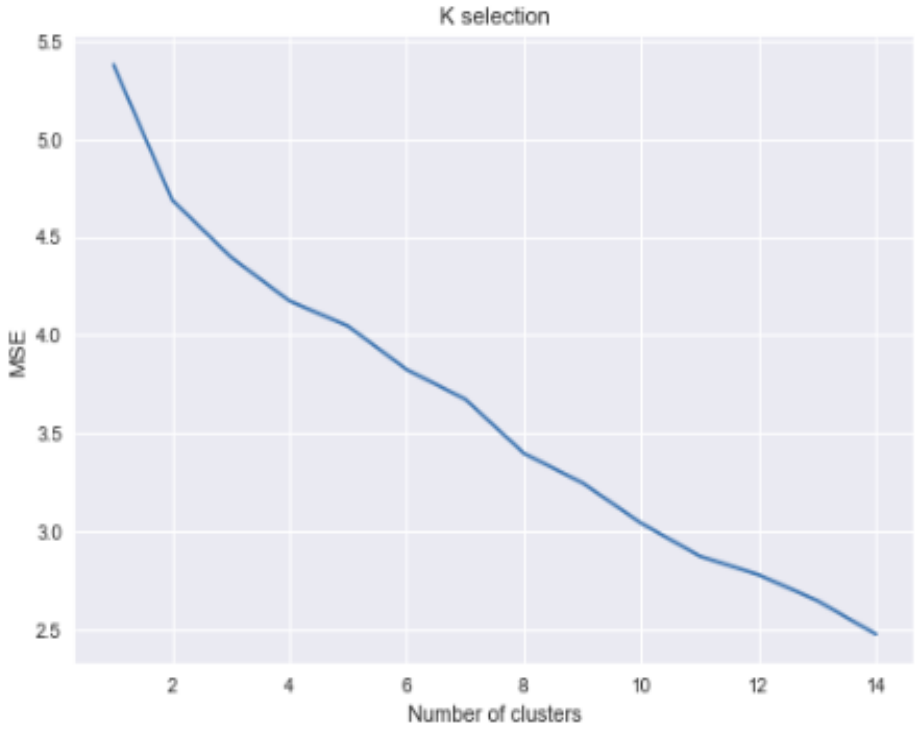
Seasonality -
- Reviews

Nordelta
Garín
Tigre
Del Viso
San Fernando
El Talar
San Isidro
Boulogne
Sur Mer
José C. Paz
Vicente López
San Miguel
Cuartel V
Villa Ballester
Santa María
General San
Martín
Trujui
Caseros
Villa Tesei
Moreno
Morón
Merlo
Avellaneda
San Justo
Parque San
Martín
Isidro Casanova
Lanús Oeste
Wilde
Quilmes
Gregorio de
Pontevedra
Banfield
9 de Abril

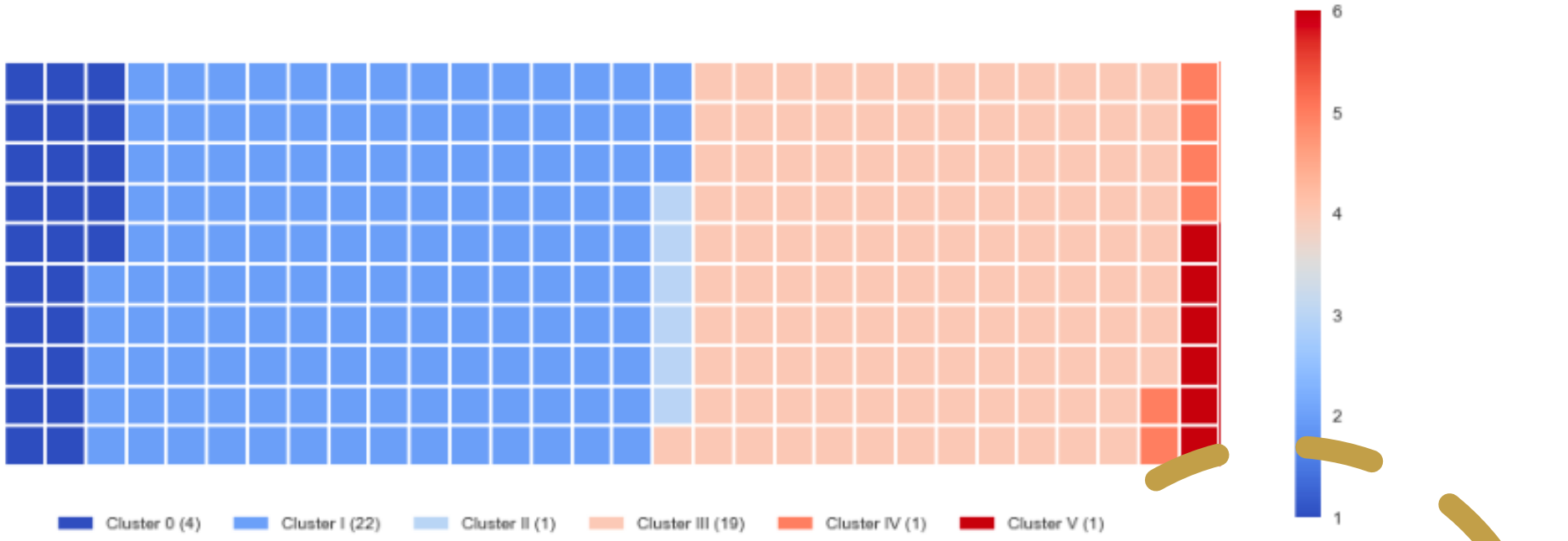
Venues


Most common venues in Buenos Aires
Ciudad de Buenos Aires
Number of Venues
0
10
20
30
40
50
60
70
Café
Argentinian Restaurant
Pizza Place
Ice Cream Shop
Coffee Shop
Plaza
Bakery
BBQ Joint
Restaurant
Hotel
Bar
Bus Stop
Gym
Pharmacy
Soccer Field
Italian Restaurant
Grocery Store
Supermarket
Beer Bar
Gym / Fitness Center
Burger Joint
Chinese Restaurant
Theater
Sports Club
Sandwich Place

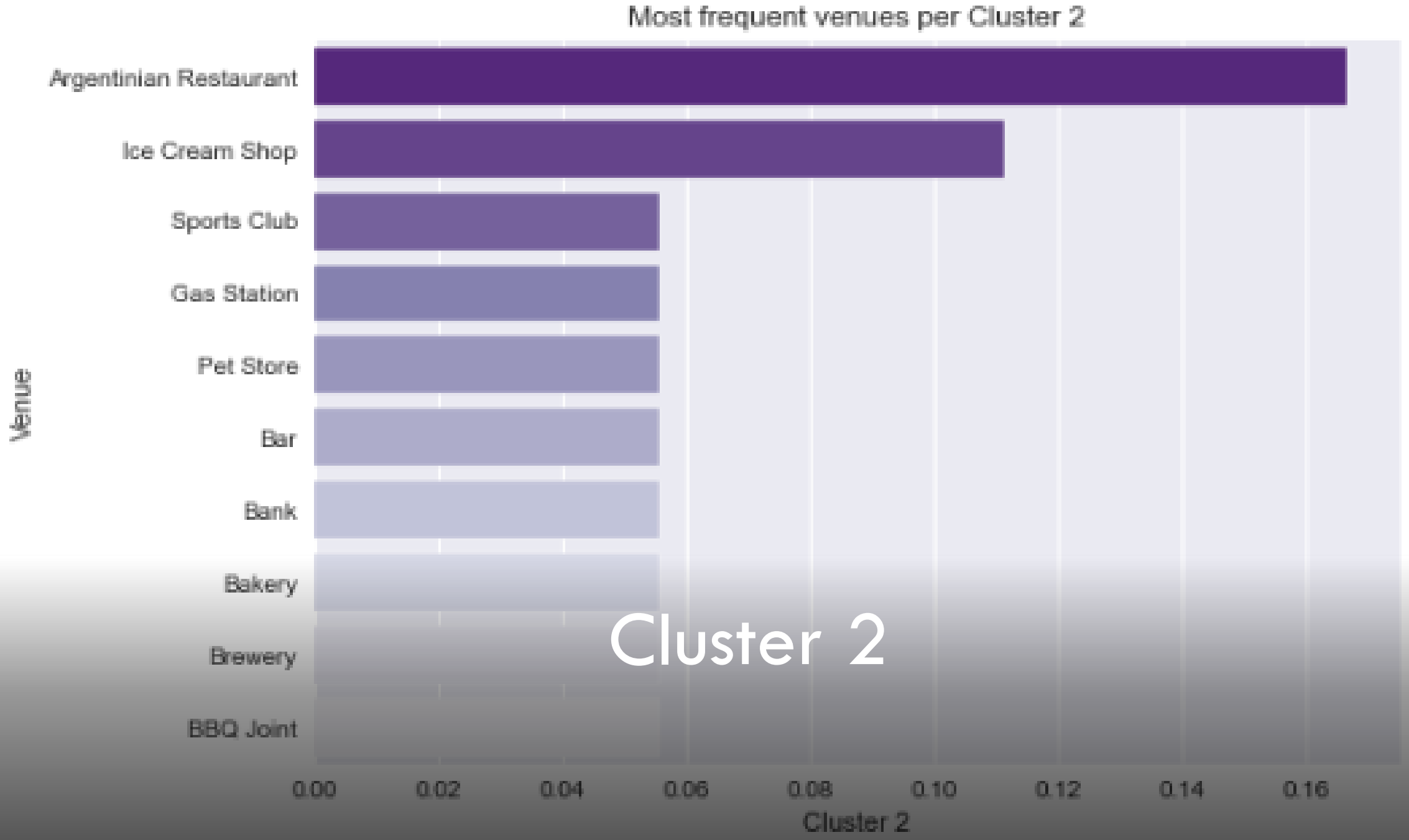
Clustering


K selection
MSE
Number of clusters
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2
4
6
8
10
12
14

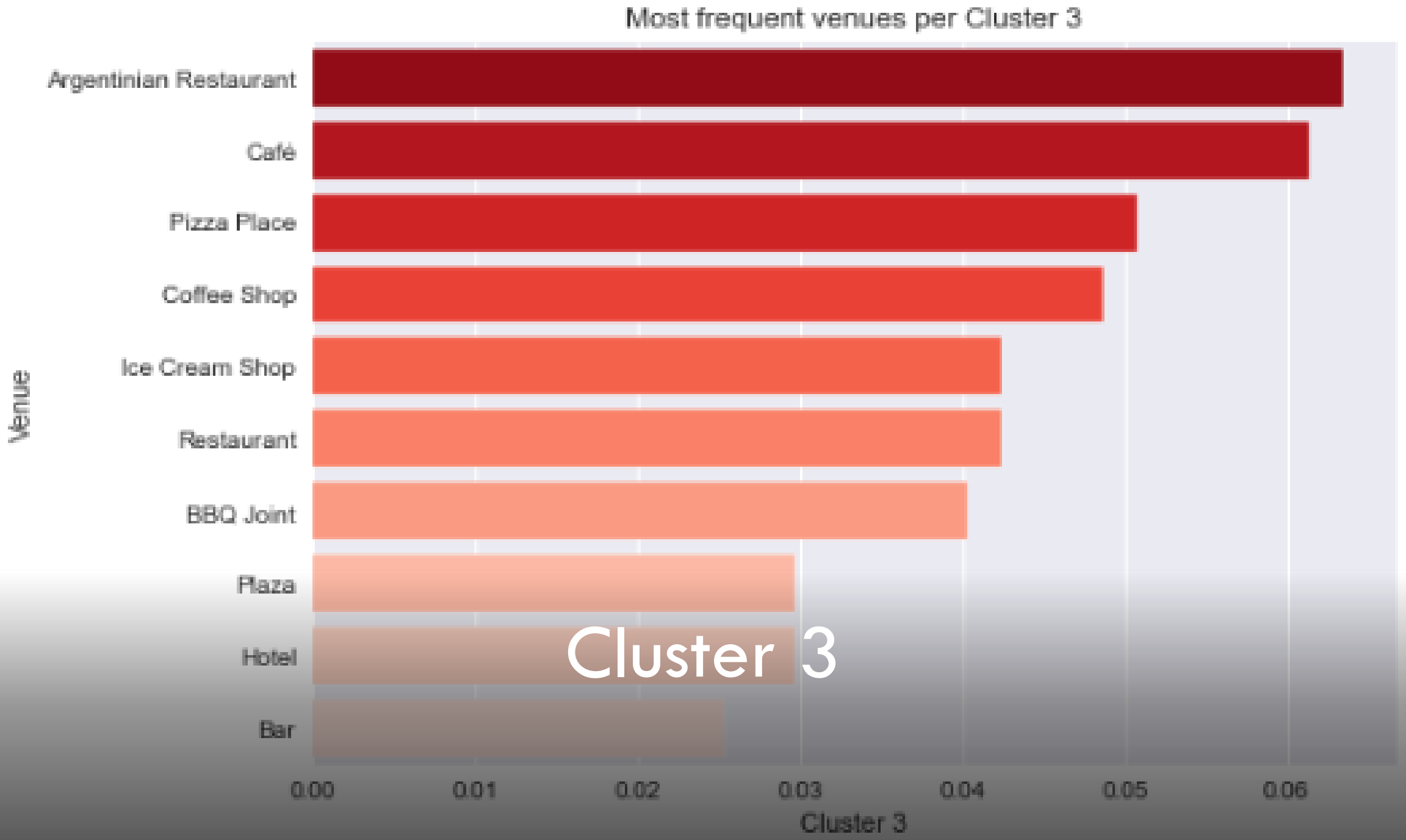
Villa Lynch
Buenos Aires
Ciudadela
AU1
Dock Sud
Avellaneda
San Justo
Villa Madero
Sarandí



Waffle Chart


Most frequent venues per Cluster 2
Argentinian Restaurant
Ice Cream Shop
Sports Club
Gas Station
Pet Store
Bar
Bank
Bakery
Brewery
BBQ Joint
Venue
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
Cluster 2

Cluster 2


Most frequent venues per Cluster 3
Argentinian Restaurant
Café
Pizza Place
Coffee Shop
Ice Cream Shop
Restaurant
BBQ Joint
Plaza
Hotel
Bar
Venue
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Cluster 3

Cluster 3

Conclusion

The price of a post is higher in Palermo, Puerto Madero and Recoleta.

Palermo has de widest offer of accomodations, and venues nearby.

Hotel rooms are the most expensive places to stay.

Being a SuperHost is a simple way of increasing profits.

- The starting price of hostels does not vary much depending on its distance from the city center.
- We can group Buenos Aires neighborhoods into 6 clusters.
- Cafes and Argentinian Restaurants are the most common venues (in clusters 2 and 3 especially).
- There is seasonality in Argentina: peaks of tourism in March, July and November.
- 1666 Usd is the approximate annual income for most of the posts in Airbnb.