Search for accommodation in Madrid

Capstone Project (by César Portales)

Introduction

Madrid Familiar Tours (MFT) is a company dedicated to organizing guided tours of the main monuments of Madrid and its surroundings. Visits last (depending on the package you hire) of between three and five days. Request that we make it easier to find accommodation that meets the following requirements:

- ✓ Maximum distance of 1.5 kms. To the Prado Museum

- ✓ Accommodations must be of the whole apartment type

- ✓ Information on Price, Kitchen, Kids Friendly, Smoking, Pets, Internet Access, Welchair Accessibility,

- ✓ Information about points of interest near the accommodation

Data Sources

For this purpose, we will use:

✓ A public Airbnb file in csv format (airbnb-extract-Madrid.csv), which contains information about accommodation in Madrid.

   The Dataset contains the following fields: Property ID, Neighborhood, Property Type, Room Type, Amenities, Latitude, Longitude, Property URL, and Property Name.

✓ As far as points of interest near the accommodation are concerned, we will need to use the Foursquare location data.

	ID	Neighbourhood	Property Type	Room Type	Amenities	Price	Latitude	Longitude	URL	NAM
0	6478626	Retiro	Apartment	Private room	TV,Wireless Internet,Kitchen,Heating,Essential...	28.0	40.406953	-3.670894	https://www.airbnb.com/rooms/6478626	Habitaci¢ con ba¤c
1	4941335	Salamanca	Apartment	Entire home/apt	TV,Internet,Wireless Internet,Air conditioning...	58.0	40.425904	-3.681647	https://www.airbnb.com/rooms/4941335	C¢mod estudio e la cal Goy
2	12510355	Salamanca	Apartment	Entire home/apt	TV,Internet,Wireless Internet,Air conditioning...	70.0	40.428352	-3.687155	https://www.airbnb.com/rooms/12510355	Studio Terrace Barr Salamanc

Methodology

We'll use jupyter notebooks and the programming language will be Python

```python
# mathematical functions, management of arrays and datasets
import math
import numpy as np

from scipy import stats
from scipy.stats import norm
import pandas as pd

import warnings
warnings.filterwarnings('ignore')
```

```python
#graphics management functions

import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
%matplotlib inline

cm = plt.cm.RdBu
cm_bright = ListedColormap(['#FF0000', '#0000FF'])

import seaborn as sns
```
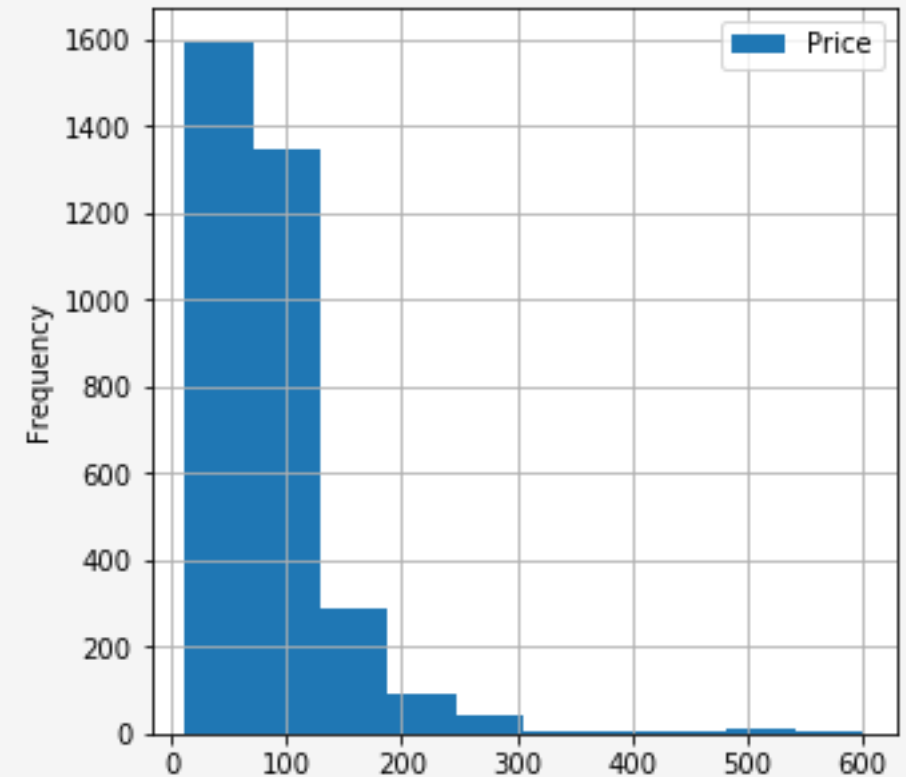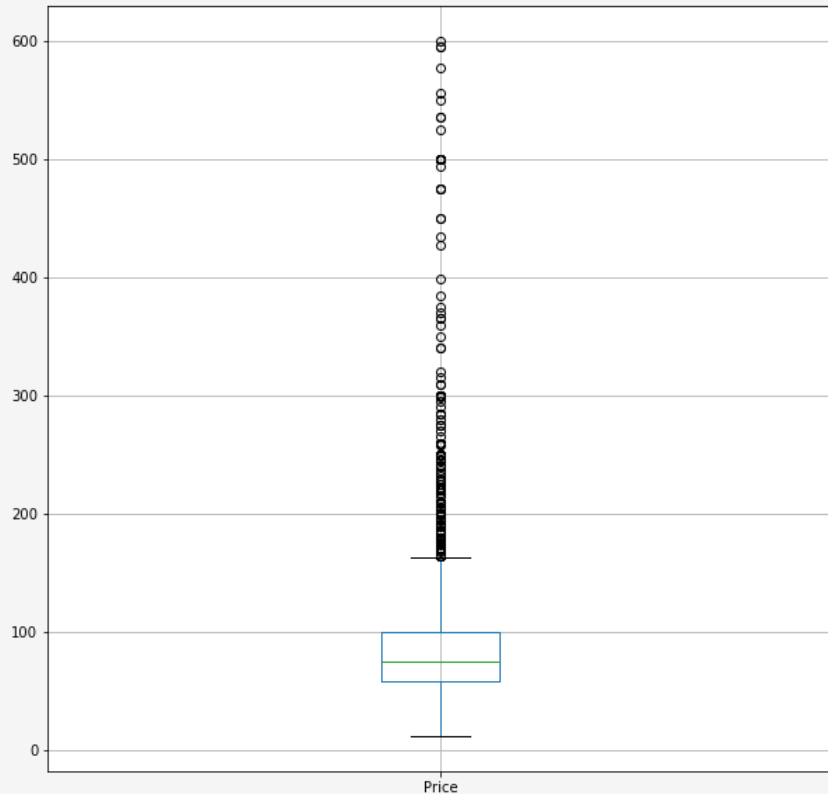
```python
# preprocessing and clustering functions
#from sklearn.preprocessing import LabelBinarizer
from sklearn.preprocessing import MinMaxScaler # to normailize data
from sklearn.cluster import KMeans # clustering
# NLP for extract Amenities
from sklearn.feature_extraction.text import CountVectorizer
import string
```

```python
# geolocation functions
!pip install geocoder
import geocoder # import geocoder
!pip install geopy
from geopy.geocoders import Nominatim
from geopy.distance import geodesic # To calculate distances to the Prado museum
from geopy.extra.rate_limiter import RateLimiter #necessary to pause between each search
import requests # library to handle Foursquare requests

# graph the maps
!pip install folium
import folium
geolocator = Nominatim(user_agent="mad_explorer")
```

Data cleaning I

The price data has too many outliers. It is decided to work only with accommodation priced less than 250 euros

Data Cleaning II

Using NLP techniques, you get the necessary information from the Amenities field

	ESSENTIALS	KITCHEN	SMOKING ALLOWED	WHEELCHAIR ACCESSIBLE	Pets	InternetAccess	TempControl	KidsFriendly
0	1	1	1	0	0	1	1	1
1	1	1	0	0	0	1	1	0
2	1	1	1	0	1	1	1	1
3	1	1	0	0	0	1	1	1
4	1	1	0	0	0	1	1	1

Neighborhood exploration

The Foursquare API is used to get information about the main points of interest within 500 meters of each neighborhood

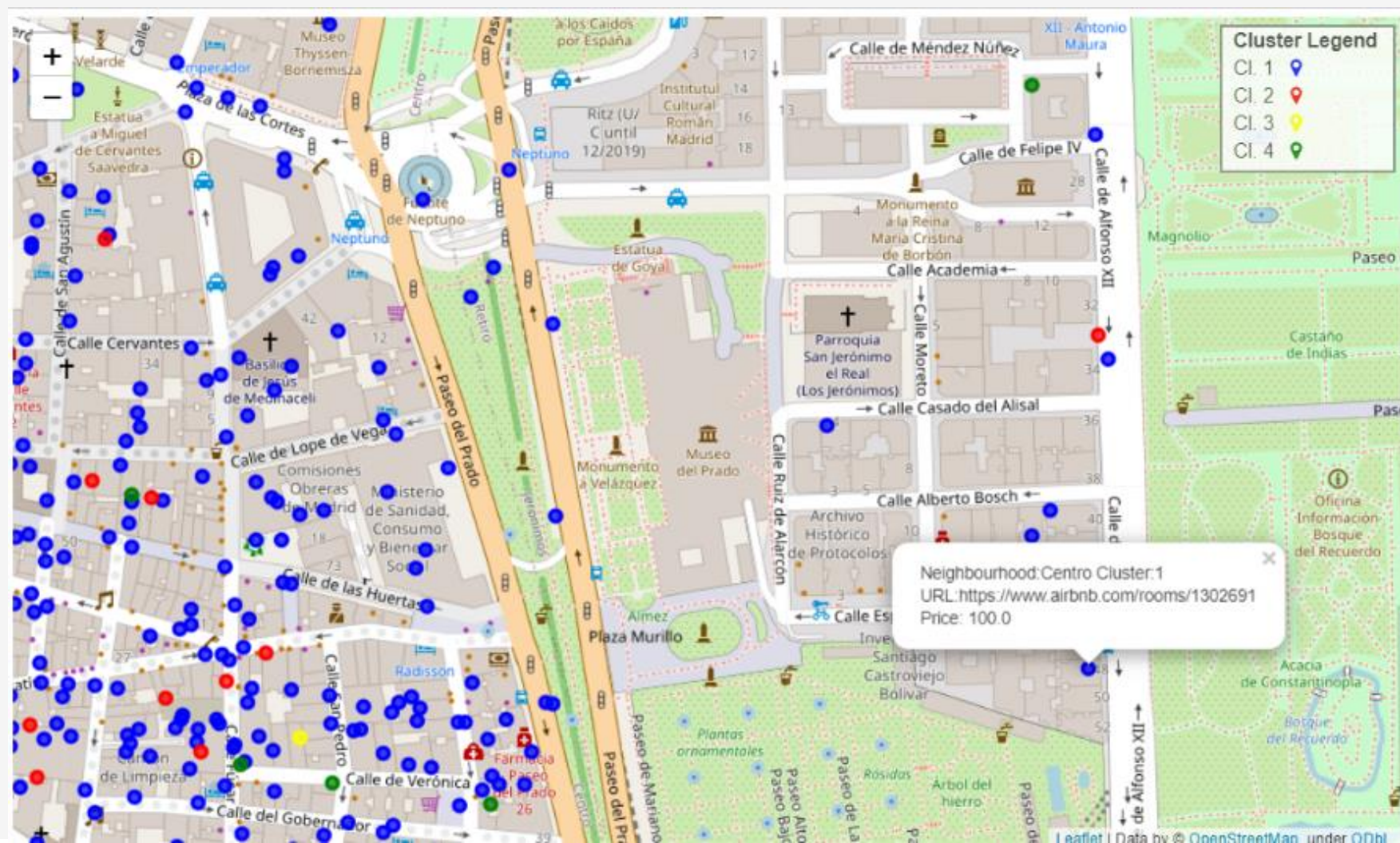|	neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arganzuela	Tapas Restaurant	Restaurant	Beer Garden	Bakery	Spanish Restaurant	Museum	Chinese Restaurant	Café	Mediterranean Restaurant	Market
1	Centro	Plaza	Hotel	Hostel	Tapas Restaurant	Gourmet Shop	Wine Bar	Ice Cream Shop	French Restaurant	Bistro	Restaurant
2	Chamberí	Tapas Restaurant	Spanish Restaurant	Bar	Restaurant	Café	Theater	Plaza	Bakery	Gastropub	Italian Restaurant
3	Retiro	Spanish Restaurant	Bar	Grocery Store	Art Gallery	Food & Drink Shop	Tapas Restaurant	Supermarket	Burger Joint	Italian Restaurant	Museum
4	Salamanca	Spanish Restaurant	Restaurant	Bar	Burger Joint	Seafood Restaurant	Hotel	Plaza	Gymnastics Gym	Coffee Shop	Tapas Restaurant

Clustering

The classification carried out would allow us to select by the following groups:

✓ **Cluster 1: In Centro Ward (or very close), pets are allowed, smoking is not allowed, and very high probability of being Kids Friendly**

✓ **Cluster 2: In Centro Ward, pets are not allowed, smoking is not allowed, and not kids friendly**

✓ **Cluster 3: In Centro Ward, pets are not allowed, smoking is not allowed, and kids friendly**

✓ **Cluster 4: Not In Centro Ward, pets are not allowed, smoking is not allowed, and very probably kids friendly**

Labels	Price	Latitude	Longitude	Dist_Museos	Arganzuela	Centro	Chamberí	Retiro	Salamanca	ESSENTIALS	KITCHEN	SMOKING ALLOWED	WHEELCHAIR ACCESSIBLE	Pets	InternetAccess	TempControl	KidsFriendly
0	0.304548	0.545882	0.269896	0.668313	0.000000	1.000000	0.000000	0.000000	0.000000	0.926065	0.988095	0.000000	0.085213	0.000000	0.981203	0.979950	1.000000
1	0.293149	0.537250	0.260128	0.655259	0.000000	0.998239	0.001761	0.000000	0.000000	0.889085	0.955986	0.123239	0.026408	0.000000	0.957746	0.940141	0.000000
2	0.295151	0.560264	0.290779	0.705152	0.400631	0.000000	0.009464	0.356467	0.233438	0.917981	0.981073	0.138801	0.151420	0.151420	0.940063	0.984227	0.810726
3	0.289366	0.534205	0.263884	0.655839	0.000000	1.000000	0.000000	0.000000	0.000000	0.945455	0.985455	0.000000	0.076364	1.000000	0.996364	0.978182	0.865455
4	0.306681	0.525110	0.271621	0.678320	0.000000	0.994118	0.001961	0.000000	0.003922	0.917647	0.992157	1.000000	0.090196	0.521569	0.968627	0.949020	0.939216

Accommodation distribution map

Conclusions and future directions

✓ The offer of tourist accommodation provided is of great interest, having the vast majority of essentials accommodations, kitchen, internet access at a reasonable price and with a lot of close supply when it comes to food.

✓ Iit would be recommended to improve the wheelchair access part. It is clear that in the downtown area (with the oldest houses in Madrid) it was going to be difficult to meet this requirement (it is not therefore a problem of the dataset that has been used).

If you want to satisfy this type of tourism, it would be advisable to contact companies transporting specialized travelers, since accommodations with wheelchair access are far from the points where you are going to work.