



Open a new restaurant

Where to open a new  
restaurant in Lisbon

João Amador  
2/2/2021

# OUTLINE

---



- Executive Summary
- Introduction
- Methodology
- Results
  - Data Loading
  - Data Cleansing
  - Visualization
  - Clustering
- Conclusion

# EXECUTIVE SUMMARY

---



In this report we will identify the best place to open a new restaurant

We will use public data and data science techniques to analyze and give results that will help a future restaurant owner take a decision

# INTRODUCTION

---



This report will be replying to the following questions :

- What is the concentration of restaurants in Lisbon districts
- What is the concentration of restaurants per postal code
- Classify each district by income revenue
- Classify each district as residential or work

With this indicators we will give a score for each district that will allow a future restaurant owner decide where he/she should open is new restaurant

# METHODOLOGY

---

The origin of our data

- FourSquare data to collect restaurants per district
  - The data from annual census population in Lisbon [INE DATA](#)
- will provide the population concentration, their academic level and with their academic level we will deduct income level
- The list of postal codes in Lisbon

Our methodology

- Analyze foursquare data to gather restaurants concentration
- Analyze the restaurants distribution
- Clsuter the restaurants per postal code

# Results – data loading

Using the FourSquare API we have searched all the restaurants in lisbon within a 5 kilometer radius, the results contain venue name, address and also geolocation data that we will use later to cluster and group the restaurants by districts

[102]:

	venue.name	venue.location.address	venue.location.lat	venue.location.lng	venue.location.postalCode	venue.location.cc	venue.location.city	venue.location.state
0	Miradouro Sophia de Mello Breyner Andresen	Largo da Graça	38.716454	-9.131557	1170-165	PT	Lisboa	Lisboa
1	Café da Garagem	Rua da Costa do Castelo, 75	38.714806	-9.132889	1100	PT	Lisboa	Lisboa
2	Jardim da Cerca da Graça	Nan	38.717911	-9.132196	Nan	PT	Lisboa	Lisboa
3	Miradouro da Senhora do Monte	R. Senhora do Monte	38.719072	-9.132793	1170-254	PT	Lisboa	Lisboa
4	The Food Temple	Beco do Jasmin, 18	38.715881	-9.134385	1100-289	PT	Lisboa	Lisboa

# Results – data cleansing

The Statistics in our dat shows that there are lines with empty Postal Codes, City and , State, Neighborhood

```
restaurants.count()

venue.name          100
venue.location.address   89
venue.location.lat      100
venue.location.lng      100
venue.location.postalCode  62
venue.location.cc       100
venue.location.city     95
venue.location.state    95
venue.location.country   100
venue.location.formattedAddress 100
venue.photos.count     100
venue.location.neighborhood 14
cluster_label          100
dtype: int64
```

We use the method dropna to delete all rows that contain empty Postal Codes

```
restaurants.dropna(subset=['venue.location.postalCode'], inplace=True)
restaurants.count()
```

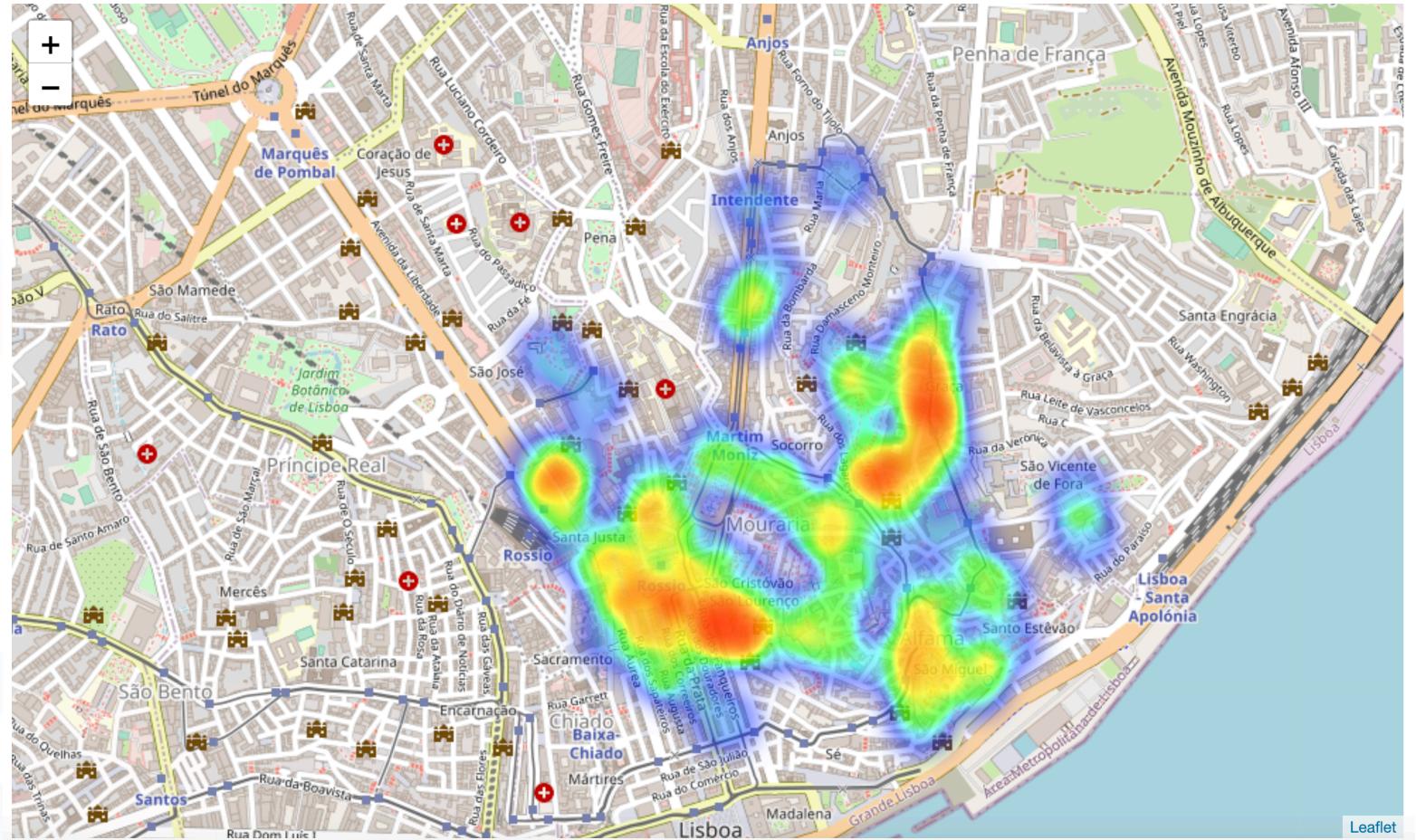


# Results - Visualization

Let's visualize the distribution of our restaurants in a Lisbon map to understand their concentration.

We see there is a higher concentration in the center of Lisbon, and along the main avenues

And almost none in the east and west side of town

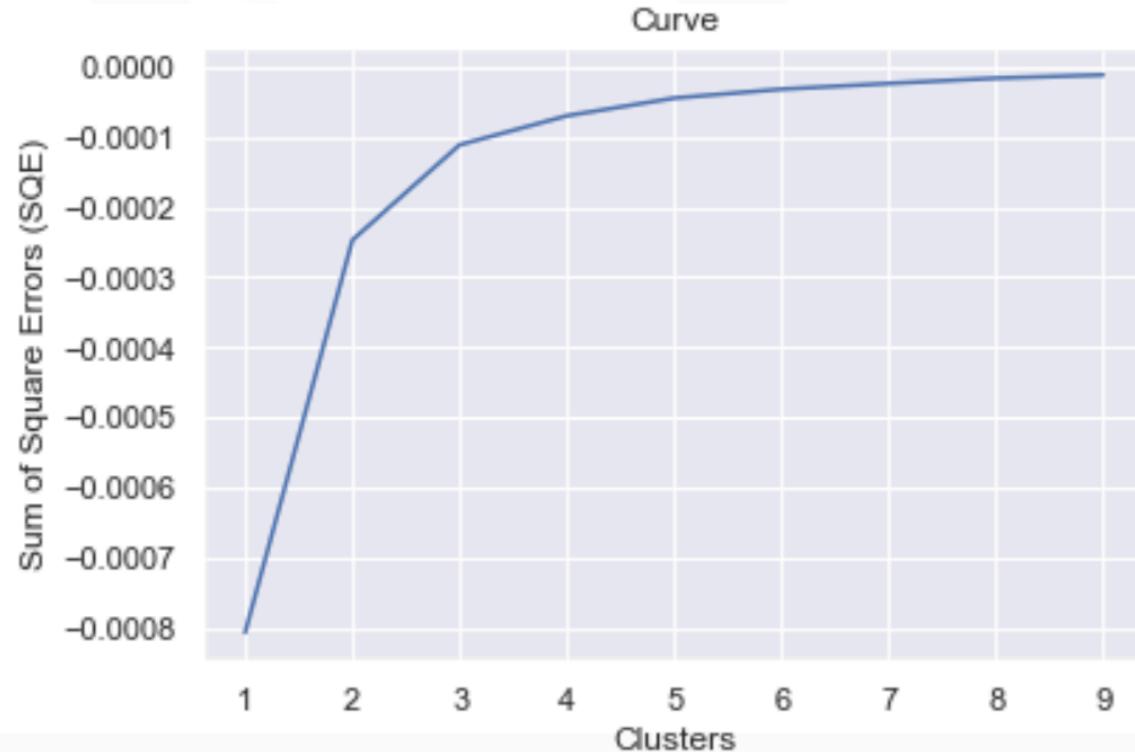


# Results - Clustering

Now we need to cluster all the restaurants into different clusters.

To use KMeans we have to find correct k ( number of clusters).

It's not easy to find the correct number of clusters, for this we will use the elbow method, the idea is to run kmeans thru the dataset n times, and for each value calculate the Sum of Square Errors, by plotting this in a graph we can see that at the optimum K the graph stabilizes and there is no improvement.

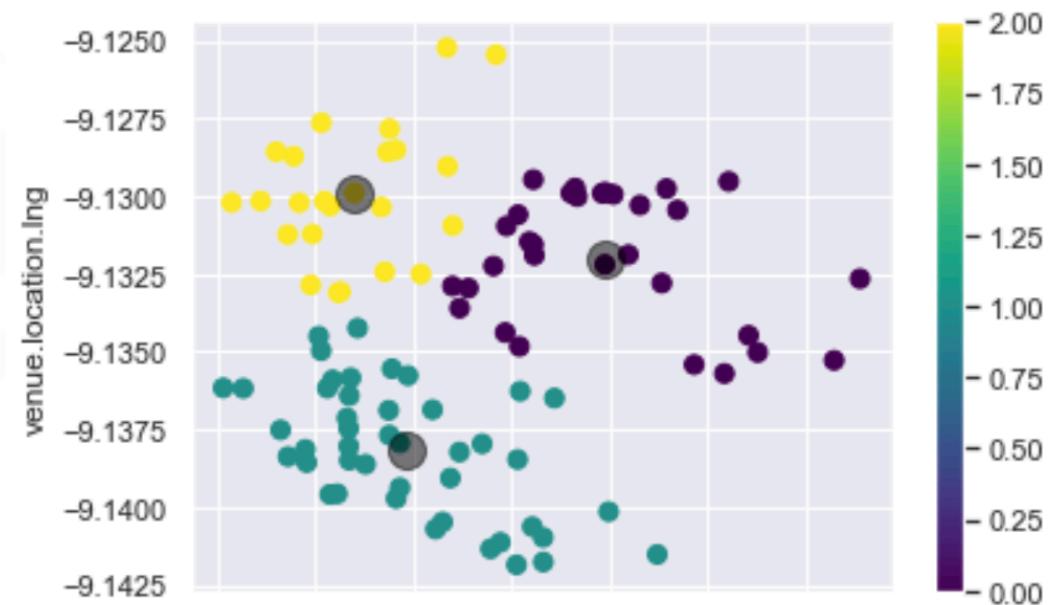


# Results - Clustering

The results of the kmean clustering group the restaurants by 3 clusters.

This gives three areas of the city where the concentration of the restaurants is higher.

And allows to group the restaurants by postal codes / areas



# Conclusion

---

- The data allows us to help a future restaurant owner to chose the location of a new restaurant.
- It's clear that the city center is too saturated and new venues should prefer to the east and west side of town
- The FourSquare Data is an excellent source for this studies but at the moment there is a big gap in the total existent restaurants and the FourSquare data , there should be an improvement of the quality of the data as more and more turists come to Lisbon and update their preferences thru FourSquare