C. Granados

Presentation Final Project Data Science

Coursera - IBM Data Science

May the 13th, 2020

Overview

Overview

1. Introduction

- In the last years Leipzig (Germany) has become a quite attractive
- The city offers many new possibilities
- Regarding to the food options, they are quite similar to other cities in Germany and Europe
- we need to explore the possibility to offer new options in a young and diverse market
- We want to offer a new option: Amazonas food! The flavors from the Amazonas forest!

Overview

2.1 Data Requirements

We need to...

- Analyze different areas of Leipzig
- Identify several important points through all the city
- Include known landmarks, museums, churches, gartens...
- Identify the zones with a large density of restaurants
- Analyze the sectors more visited as a function of the days of the week
- Obtain geographical coordinates of each selected point/sector
- Use Foursquare location data to obtain the venues nearby each selected point in the city

2.2 Data Acquisition and Cleaning

Data Sources

- Wikipedia article about Leipzig
- From it, we extract the names of known buildings, parks, museums, churches and touristic landmarks
- For each cathegory only some are saved (the list can contain more information than needed)

Data Cleaning

- We drop duplicates in the data frame
- We exclude famous names associated with the city, as Johann Sebastian Bach or Johann Wolfgang von Göthe
- ... Also some historical names as "DDR" or "East Germany"
- Finally, we drop names that have not a geographical coordinate (obtained from Google Maps)

2.3 Additional Features

- For each valid selected point of the city, we obtain its geographical coordinates
- We use Foursquare to obtain different venues around each point
- Foursquare not offer the trending values for Leipzig
- We use the explore option to obtain a list of venues
- The top 10 of venues are stored and merged with the cleaned data point in the city

Overview

3. Exploratory Data Analysis

-----Bach Archive-----

	Venue	Freq.
0	Sushi Restaurant	0.10
1	Theater	0.10
2	Hotel	0.05
3	Lounge	0.05
4	History Museum	0.05
5	Gym / Fitness Center	0.05
6	German Restaurant	0.05
7	Gastropub	0.05
8	Plaza	0.05
9	Deli / Bodega	0.05

-----Battle of Leipzig-----

	Venue	Freq.
0	Tram Station	0.4
1	Historic Site	0.2
2	Business Service	0.2
3	Monument / Landmark	0.2
4	African Restaurant	0.0
5	Museum	0.0
6	Plaza	0.0
7	Playground	0.0
8	Pet Store	0.0
9	Park	0.0

-----Leipzig Hauptbahnhof-----

	Venue	Freq.
0	Hotel	0.25
1	Modern European Restaurant	0.05
2	Restaurant	0.05
3	Irish Pub	0.05
4	Coffee Shop	0.05
5	Cocktail Bar	0.05
6	Shopping Mall	0.05
7	Sports Bar	0.05
8	Steakhouse	0.05
9	Newsstand	0.05

-----Leipzig Zoological Garden-----

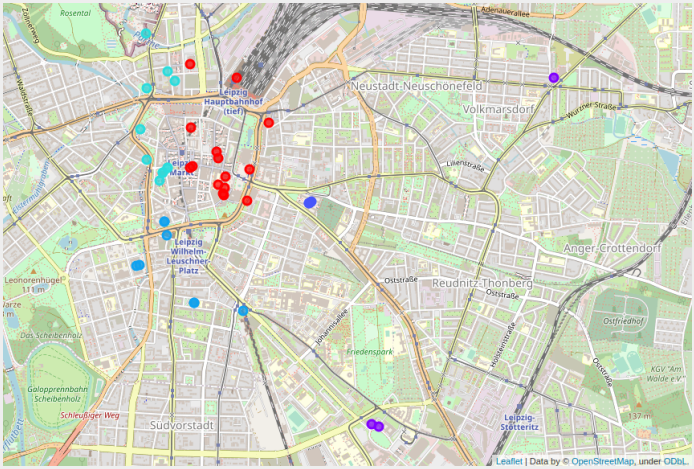
	Venue	Freq.
0	Zoo Exhibit	0.25
1	Hotel	0.10
2	Hotel Bar	0.05
3	Greek Restaurant	0.05
4	Zoo	0.05
5	Trattoria/Osteria	0.05
6	Theme Restaurant	0.05
7	Science Museum	0.05
8	Plaza	0.05
9	Playground	0.05

Obtained venues top ten and its frequency, using Foursquare, for four different points of interest in Leipzig.

Overview

4. Cluster Modelling

- We used the K-Means clustering method
- We used the `scikit learn` package
- We considered 10 different clusters



4. Cluster Modelling

Cluster 1:

```
L_merged.loc[L_merged['Cluster Labels'] == 0, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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	Auerbachs Keller	Hotel	Gym / Fitness Center	Ice Cream Shop	Boutique	Bistro	Spanish Restaurant	Café	Italian Restaurant	Sushi Restaurant	Plaza
1	Augusteum (Leipzig)	Hotel	Plaza	Church	Bed & Breakfast	Concert Hall	Museum	Café	Scenic Lookout	Deli / Bodega	Lounge
2	Augustusplatz	Hotel	Plaza	Bistro	Church	Coffee Shop	Nightclub	Concert Hall	Café	Museum	Scenic Lookout
8	City-Hochhaus Leipzig	Hotel	Plaza	Church	Bed & Breakfast	Concert Hall	Gym / Fitness Center	Café	Museum	Scenic Lookout	Deli / Bodega
9	Europahaus	Hotel	Museum	Plaza	Nightclub	Burger Joint	Drugstore	Bed & Breakfast	Bar	Café	Scenic Lookout
18	Hotel The Westin Leipzig	Hotel	Zoo Exhibit	Steakhouse	Art Museum	Bagel Shop	Coffee Shop	Hotel Bar	Modern European Restaurant	Restaurant	Shopping Mall
											Modern

4. Cluster Modelling

Cluster 2:

```
L_merged.loc[L_merged['Cluster Labels'] == 1, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
4	Battle of Leipzig	Tram Station	Historic Site	Business Service	Monument / Landmark	Event Space	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega
6	Center Torgauer Platz	Tram Station	Supermarket	Plaza	Pet Store	Grocery Store	Donut Shop	Clothing Store	Cocktail Bar	Coffee Shop	Concert Hall
14	German Museum of Books and Writing	Tram Station	Automotive Shop	Plaza	Event Space	Gym	Nightclub	Bar	Concert Hall	Construction & Landscaping	Cupcake Shop
15	German National Library	Tram Station	Plaza	Automotive Shop	Nightclub	Mediterranean Restaurant	Bar	Furniture / Home Store	Supermarket	Hockey Rink	Gym
30	Monument to the Battle of the Nations	Tram Station	Historic Site	Business Service	Monument / Landmark	Event Space	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega
47	Südfriedhof	Tram	Historic Site	Monument /	Event Space	Coffee Shop	Concert	Construction &	Cupcake	Currywurst	Deli /

4. Cluster Modelling

Cluster 3:

```
L_merged.loc[L_merged['Cluster Labels'] == 2, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
16	Grassi Museum	Hotel	Museum	Supermarket	Coffee Shop	Plaza	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Gym	Drugstore
22	Leipzig Museum of Applied Arts	Hotel	Museum	Supermarket	Plaza	Karaoke Bar	Coffee Shop	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Exhibit
23	Leipzig Museum of Ethnography	Hotel	Museum	Supermarket	Coffee Shop	Plaza	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Gym	Drugstore
33	Museum of Musical Instruments of the Universit...	Hotel	Museum	Supermarket	Plaza	Karaoke Bar	Coffee Shop	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Exhibit

4. Cluster Modelling

Cluster 4:

```
L_merged.loc[L_merged['Cluster Labels'] == 3, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
10	Federal Administrative Court of Germany	Café	Chinese Restaurant	Bar	Bank	Sushi Restaurant	Bakery	Beer Store	Bagel Shop	Tech Startup	Soup Place
19	Leipzig Bayerischer Bahnhof	Tram Station	Hookah Bar	Bike Shop	German Restaurant	Italian Restaurant	Drugstore	Dessert Shop	Nightclub	Chinese Restaurant	Café
37	New Town Hall (Leipzig)	Café	Bar	Nightclub	Clothing Store	Park	Chinese Restaurant	Gym / Fitness Center	Plaza	Cupcake Shop	Deli / Bodega
40	Propsteikirche, Leipzig	Café	Beer Store	Grocery Store	German Restaurant	Dessert Shop	Deli / Bodega	Cupcake Shop	Coffee Shop	Nightclub	Park
41	Reichsgericht	Café	Tech Startup	Soup Place	Beer Store	Bar	Sushi Restaurant	Bakery	Italian Restaurant	Bagel Shop	Park
43	St. Peter, Leipzig	Chinese Restaurant	Café	Sushi Restaurant	Hotel	German Restaurant	Drugstore	Dessert Shop	Cupcake Shop	Organic Grocery	Cocktail Bar
		Chinese		Sushi		German		Dessert	Cupcake	Organic	

4. Cluster Modelling

Cluster 5:

```
L_merged.loc[L_merged['Cluster Labels'] == 4, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
3	Bach Archive	Sushi Restaurant	Theater	Gym / Fitness Center	Coffee Shop	Plaza	Deli / Bodega	Burger Joint	Restaurant	Lounge	Shopping Mall
5	Cantor (church)	Sushi Restaurant	Theater	Clothing Store	Restaurant	Deli / Bodega	Shopping Mall	Plaza	Spanish Restaurant	Indie Movie Theater	Ice Cream Shop
12	Funkturm Leipzig	Exhibit	Supermarket	Restaurant	Auto Dealership	Coffee Shop	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega
13	G2 Kunsthalle	Sushi Restaurant	Spanish Restaurant	Tapas Restaurant	Lounge	Plaza	Burger Joint	Restaurant	Indie Movie Theater	Coffee Shop	Bar
17	Hochhaus Löhr's Carree	Hotel	Zoo Exhibit	Supermarket	Art Museum	Currywurst Joint	Hotel Bar	Indie Movie Theater	Italian Restaurant	Japanese Restaurant	Modern European Restaurant
20	Leipzig Debate	Hotel	Sushi Restaurant	Restaurant	Deli / Bodega	Shopping Mall	Plaza	Spanish Restaurant	Coffee Shop	Indie Movie Theater	Ice Cream Shop
25	Leipzig Gamescom	Hotel	Restaurant	Snack Place	Shopping Mall	Sushi Restaurant	Italian Restaurant	Bar	Currywurst Joint	Ice Cream Shop	

4. Cluster Modelling

Cluster 6:

```
L_merged.loc[L_merged['Cluster Labels'] == 5, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
36	New Leipzig School	Market	American Restaurant	Construction & Landscaping	Fast Food Restaurant	Concert Hall	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop	Donut Shop

Cluster 7:

```
L_merged.loc[L_merged['Cluster Labels'] == 6, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
7	Chimney of Stahl- und Hartgusswerk Bösdorf GmbH	Train Station	Zoo Exhibit	Exhibit	Coffee Shop	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop

Cluster 8:

```
L_merged.loc[L_merged['Cluster Labels'] == 7, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
24	Leipzig Panometer	Exhibit	Concert Hall	Cafeteria	Bus Stop	Cocktail Bar	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop

4. Cluster Modelling

Cluster 9:

```
L_merged.loc[L_merged['Cluster Labels'] == 8, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
26	Leipzig Trade Fair	Fast Food Restaurant	Hardware Store	Tram Station	Bus Stop	Fountain	Event Space	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint

Cluster 10:

```
L_merged.loc[L_merged['Cluster Labels'] == 9, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
28	Leipziger Baumwollspinnerei	Art Gallery	Bistro	Arts & Crafts Store	Pub	Rental Car Location	Café	Music Venue	Park	Zoo Exhibit	Concert Hall
35	Neo Rauch	Art Gallery	Park	Arts & Crafts Store	Café	Music Venue	Bistro	Event Space	Construction & Landscaping	Cupcake Shop	Currywurst Joint

Overview

5. Conclusions

- From the clustering results, we conclude that the third and fifth clusters are the most attractive
- The largest cluster is the first one, several restaurants in the sector
- The second cluster can be discarded, basically it is outside of the city
- The fourth cluster contain several coffee shops and bars, but another factors must be considered.
- The other clusters are “outliers” of the data
- For a second stage, we can include another restaurants with a similar profile
- The city regulations must be taken into account
- Include also the profiles on the availability of space and rent prices