Finding the Best Neighborhood in Berlin for Opening a Café cum Restaurant | Data Science Capstone

Since the last few months, I have been working to achieve an IBM Data Science Professional certification. Through this course, I have learned about data science and also acquired various skills and in-depth understanding about different tools that the data scientists need to solve the given problems. As a Capstone for the certification, I am required to define busines problem in the city of my choice, solve it by scraping data from the web and location data from Foursquare, and produce full report. This report contains only the work of first week – Background Discussion, Problem Description, and Data Description - for finding the best spot in Berlin to open a Café cum Restaurant. Detailed python code for this project is located on this GitHub Link

https://github.com/alpeshlaxmanvora/Coursera_Capstone_Project/blob/master/Week4/Capstone_Project-Battle_of_Neighborhoods_Part1.ipynb

1 Introduction

1.1 Background Discussion

Berlin is the largest city of Germany by both population and area. Its 3.77 million registered inhabitants also make it the most populous city of the European Union. Since 1990 after the fall of the Wall, Berlin is the capital of Germany and it established itself as a city of unlimited possibilities for travelers. The mixture of its historical significance, various festivals, diverse architecture, spectacular nightlife, and contemporary arts makes this city a magnet for tourists. Berlin's 3.77 Million inhabitants are hosts to nearly 34 million overnight stays and 14 million visitors in year 2019 [1], which makes Berlin a top European destination – ranked third after London and Paris.

In addition, Berlin is also Europe's leading economic force and its economy is growing rapidly for several years. In 2019, the nominal gross domestic product in Berlin was € 153.3 billion [2], which is 3% higher compared to the previous year and the number of people employed increased by 2.4% to more than 2 million [3].

All in all, the city's rapidly growing economy, office businesses and tourisms make Berlin one of the best places to start up a new business, especially in Restaurants as the office businesses and the tourism are an important driver for restaurant industry.

1.2 Business Opportunity (Problem Description)

Now a days, most people, especially business employees, are eating more meals out of the home, because the work and lifestyle pressure often do not allow sufficient time for most of employees to prepare their own cuisine. Therefore, they rely on the Cafés and restaurants for their dietary needs, especially for breakfast and lunch. Similarly, tourist or business

visitors also rely on the cafés and restaurants for their every meal. Berlin, where both the office businesses and tourism are growing rapidly since several years, provides huge opportunities for restaurant business. Reasonably priced restaurant in popular tourist and highly dense office places can attract both audiences – office employees and tourists. With given this scenario, we will analyze various factors to determine the best place in Berlin to open a café cum restaurant, where breakfast cum lunch or even brunch can be served. This report outlines some basic assumptions, data sets and data processing, and detailed analysis which can help us to select the best spot in Berlin for opening a café cum restaurant. In this process, we have assumed that money is not an issue for stating the business.

1.3 Target Audience

The key audience of this report would be anyone who wants to invest or open a restaurant in Berlin, or anyone in Berlin looking for a delicious breakfast, brunch, or lunch. The analysis will also help to office employees to find reasonable breakfast/lunch/brunch place close to office area.

2 Data

In order to find best spot in Berlin, various set of data like boroughs, neighborhoods, area, population, price of rent or buying property, number of registered companies, and number of visitors, etc. are required. Unfortunately, all sets of data are not available from any single source. Hence, the various sets of data are retrieved from different sources and Foursquare API is used to retrieve the venue data.

Altogether, below 5 sets of data are used for our analysis:

1. Berlin Boroughs, Neighborhoods, Area and Population Data
2. Berlin Registered Business Data
3. Berlin Registered Tourist Data
4. Rent or buying price of property in Berlin
5. Foursquare API

2.1 Berlin Boroughs, Neighborhoods, Area and Population Data

This set of data is retrieved by scrapping the Wikipedia Page - "Verwaltungsgliederung Berlins" [4] and placed it into a pandas data frame using python. This data frame contains lists of every neighborhoods of Berlin with their boroughs, size of area, and population in each neighborhood. This set of data will help us to figure out the densely populated neighborhoods or boroughs in Berlin. In the scrapped raw data, many columns contained the trailing newlines (\n), and also commas as decimal separators and dots as thousand separators due to German conventions. After cleaning and processing the raw data frame by coding in python, we get resulting data frame with standard conventions as follow:

```
Data types of each columns:
 Neighborhoods          object
Boroughs               object
Area_sqkm             float64
Population              int32
Population Density    float64
dtype: object

The shape of dataframe is: (96, 5)
Total numbers of Boroughs in Berlin are: 12
Total numbers of Neighborhoods in Berlin are: 96
```

Out[3]:

	Neighborhoods	Boroughs	Area_sqkm	Population	Population Density
1	Mitte	Mitte	10.70	101932	9526.0
2	Moabit	Mitte	7.72	79512	10299.0
3	Hansaviertel	Mitte	0.53	5894	11121.0
4	Tiergarten	Mitte	5.17	14753	2854.0
5	Wedding	Mitte	9.23	86688	9392.0
...	...	...	...	...	...
92	Waidmannslust	Reinickendorf	2.30	10958	4764.0
93	Lübars	Reinickendorf	5.00	5174	1035.0
94	Wittenau	Reinickendorf	5.90	24306	4120.0
95	Märkisches Viertel	Reinickendorf	3.20	40258	12581.0
96	Borsigwalde	Reinickendorf	2.00	6826	3413.0

96 rows × 5 columns

As we can see there is a total 12 boroughs and 96 neighborhoods in Berlin. The detailed python code for scrapping, wrangling, cleaning, and processing the data is given in the *section 3.1.1* of *Jupyter Notebook in GitHub*.

2.2 Berlin Registered Business Data

This data is downloaded from *"Amt für Statistik Berlin-Brandenburg"* website [5] in csv format. This downloaded csv file is placed into a pandas data frame using python. This data frame contains the number of registered businesses in each borough of Berlin. For each borough, the numbers of registered businesses are also divided into four groups according to number of employees in each business. But, we would like to have the total number of registered businesses in each borough because it can give us a rough indication of most busy areas of berlin by business employees, at least on the workdays. After

processing the dataset by coding in python, we get the resulting data frame with total number of registered businesses in each borough as follow:

```
Data types of each columns:
 Boroughs            object
Total Businesses     int64
dtype: object

The shape of dataframe is: (12, 2)
Total numbers of registered businesses in Berlin are: 192199
```

Out[6]:

	Boroughs	Total Businesses
1	Charlottenburg-Wilmersdorf	29324
2	Mitte	28553
3	Pankow	22628
4	Friedrichshain-Kreuzberg	20978
5	Tempelhof-Schöneberg	18926
6	Steglitz-Zehlendorf	14400
7	Neukölln	12511
8	Treptow-Köpenick	11316
9	Reinickendorf	9966
10	Lichtenberg	8233
11	Spandau	7692
12	Marzahn-Hellersdorf	7672

We can see that there are total 192,199 businesses in Berlin. The Charlottenburg-Wilmersdorf and Mitte boroughs have significantly more business. It is also seen that there are four boroughs in which the total number of businesses are more than 20,000.

2.3 Berlin Registered Tourist Data

This data is downloaded from *"Amt für Statistik Berlin-Brandenburg"* website [5] in csv format. This downloaded csv file is placed into a pandas data frame using python. This data frame contains the number of visitors and overnight stays visitors in each borough of Berlin for year 2019. This data can give us a rough indication of most busy areas of berlin by tourists. After processing the dataset by coding in python, we get the resulting data frame with total number of tourists in each borough as follow:

```
Data types of each columns:
 Boroughs                 object
Total Visitors in 2019    int64
dtype: object

The shape of dataframe is: (12, 2)
Total numbers of tourists visited Berlin in year 2019 are: 48087709
```

Out[9]:

	Boroughs	Total Visitors in 2019
1	Mitte	20809270
2	Charlottenburg-Wilmersdorf	9157433
3	Friedrichshain-Kreuzberg	6495909
4	Tempelhof-Schöneberg	2948106
5	Pankow	1924337
6	Lichtenberg	1791645
7	Neukölln	1331199
8	Treptow-Köpenick	951862
9	Spandau	877379
10	Reinickendorf	732466
11	Steglitz-Zehlendorf	711672
12	Marzahn-Hellersdorf	356431

We can see that Berliner inhabitants hosted nearly 48 million visitors in year 2019. Out of 48 Million tourists, nearly 21 Million tourists have visited or stayed overnight only in Mitte borough.

2.4 Rent or buying price of property in Berlin

Even though we have assumed that money is not an issue, the information about the rent or buying price of property can help us to give rough idea of how expensive it will be to maintain a café cum restaurant business in each neighborhoods or borough. Fortunately, we found the average rent price in € per month-m^2 in year 2019 for each neighborhoods of Berlin from "Homeday – mein Immobilienmakler" website [6]. From there, the data is retrieved by scrapping the web page [6] and placed it into a pandas data frame using python. The scrapped web-data contained commas as decimal operators due to German conventions. After cleaning and processing the scrapped web-data by coding in python, we get resulting data frame, with standard conventions, contained the rent price in each neighborhood of berlin:

```
Data types of each columns:
 Neighborhoods                        object
Rent Price in 2019 [€/m²-month]      float64
dtype: object

The shape of df_berlin dataframe is: (96, 2)
```

Out[12]:

	Neighborhoods	Rent Price in 2019 [€/m²-month]
1	Tiergarten	14.0
2	Friedrichshain	13.6
3	Grunewald	13.0
4	Rummelsburg	13.0
5	Moabit	12.6
...	...	...
92	Neu-Hohenschönhausen	8.0
93	Hellersdorf	7.9
94	Falkenberg	7.5
95	Marzahn	7.3
96	Wartenberg	7.2

96 rows × 2 columns

3 Conclusion for 1st Week Work

All required data are retrieved from various sources by scrapping and wrangling; and then stored them into different pandas data frames. Required data processing and cleaning also have been done. As results, we have four data sets:

1. Lists of berlin neighborhoods with their boroughs, areas, and population
2. Number of registered businesses in each borough
3. Total number of tourists in year 2019 in each borough
4. Rent price in each neighborhood of berlin

So as next step we can merge the data frames and import the Geospatial data; and then use the Foursquare data to obtain the venue data. Altogether, we can start our battle of neighborhoods for opening a café cum restaurant in Berlin.

Note: As some data are available only for boroughs, we must group all data on boroughs and start battle of boroughs.

4 References

[1]. Berlin.de, The Official Website of Berlin, Germany, accessed 21 July 2020, https://www.berlin.de/sen/wirtschaft/wirtschaft/branchen/tourismus/tourismus-in-zahlen/

[2]. Berliner Wirtschaft 2019 stark gewachsen, Wirtschaft aktuell, Senatsverwaltung für Wirtschaft, Energie und Betriebe, Germany, April 2020.

[3]. Economic Development, The Official Website of Berlin, Germany, accessed 21 July 2020, https://www.berlin.de/en/business-and-economy/economic-center/5611367-4011028-economic-development.en.html

[4]. Verwaltungsgliederung Berlins, accessed 21 July 2020. In Wikipedia. Retrieved from https://de.wikipedia.org/wiki/Verwaltungsgliederung_Berlins

[5]. Amt für Statistik Berlin-Brandenburg; Germany, accessed 21 July 2020, https://www.statistik-berlin-brandenburg.de/datenbank/inhalt-datenbank.asp

[6]. Homeday.de, Homeday mein Immobilienmakler, accessed 25 July 2020, https://www.homeday.de/de/blog/mietpreise-berlin-2019/