Applied Data ScienceCapstone by IBM and COURSERA

Finding Out the Firm Opportunitites in Stuttgart (Baden-Württemberg)

HASAN EMRE TEKCE

8.11.2020

1. Introduction:

- **The Battle of Neighborhoods in Baden-Württemberg, Germany| Business Proposal | Introduction**
- **Introduction:**
- The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Baden-Württemberg, Germany.
- We are also particularly interested in rural areas. We would also prefer locations with middle level population.
- We will use our data science powers to generate more information about promising cities based on these criteria. Advantages of each area will then be clearly expressed.
- Lots of people are migrating to various states of Germany and needed lots of research for good housing prices and reputated schools for their children. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

1. Introduction(cont.)

- This Project aim to create an analysis of features for a people migrating to Baden-Württemberg to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both freash and waste water and excrement conveyed in sewers and recreational facilities.
- It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.
- **Problem Which Tried to Solve:**
- The major purpose of this project, is to suggest a better neighborhood in a new city for the person who are shiffting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.
- 1.Sorted list of house in terms of housing prices in a ascending or descending order

2. Data Section

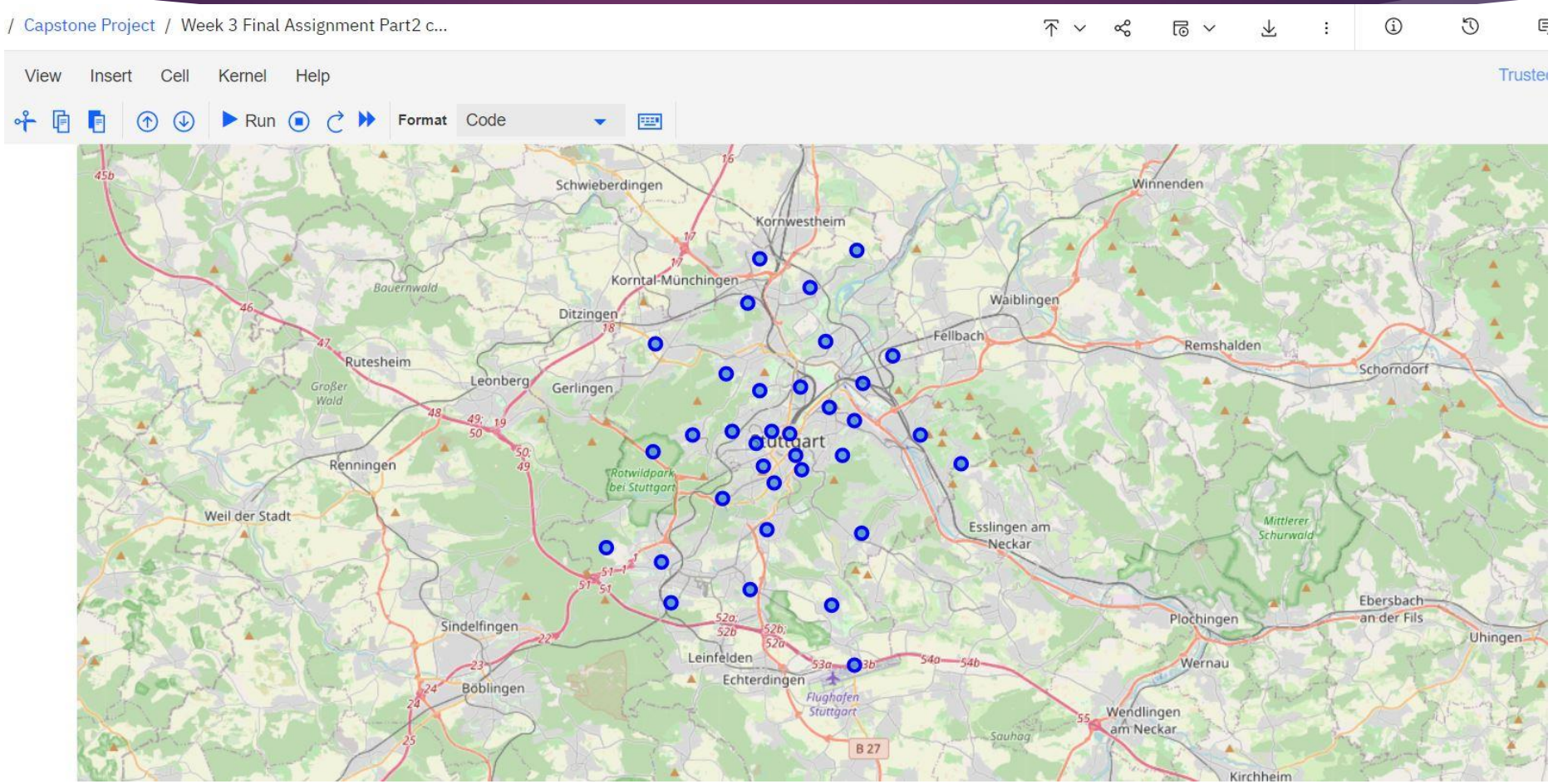
- **The Battle of Neighborhoods | Data Description**
- **Data Description:**
- Approximate addresses of centers of those areas will be obtained using geopy.geocoders
- □ Number of venues, their type and location in every city will be obtained using Foursquare API
- □ List of cities and their populations of Baden-Württemberg state will be obtained using https://www.suche-postleitzahl.org/downloads

2. Data Section(cont.)

Foursquare API Data:
We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:
`1. Neighborhood 2. Neighborhood Latitude 3. Neighborhood Longitude 4. Venue 5. Name of the venue e.g. the name of a store or restaurant 6. Venue Latitude 7. Venue Longitude 8. Venue Category`

	ort	plz	bundesland	neighbourhood
0	Stuttgart	70173	Baden-Württemberg	Akademiegarten
1	Stuttgart	70174	Baden-Württemberg	Goethestraße
2	Stuttgart	70176	Baden-Württemberg	Elisabethenstraße
3	Stuttgart	70178	Baden-Württemberg	Hasenbergstraße
4	Stuttgart	70180	Baden-Württemberg	Fangelsbachstraße
5	Stuttgart	70182	Baden-Württemberg	Heusteigstraße
6	Stuttgart	70184	Baden-Württemberg	Gröberstraße
7	Stuttgart	70186	Baden-Württemberg	Gröberstraße
8	Stuttgart	70188	Baden-Württemberg	Heubergstraße
9	Stuttgart	70190	Baden-Württemberg	Landhausstraße
10	Stuttgart	70191	Baden-Württemberg	Friedhofstraße
11	Stuttgart	70192	Baden-Württemberg	Schoderstraße
12	Stuttgart	70193	Baden-Württemberg	Falkertstaffel
13	Stuttgart	70195	Baden-Württemberg	Talwiesenweg
14	Stuttgart	70197	Baden-Württemberg	Herbsthalde
15	Stuttgart	70199	Baden-Württemberg	Binsenplattenweg
16	Stuttgart	70327	Baden-Württemberg	Abelsberg

2. Data Section(cont.)



3. Methodology Section

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Using K-Means Clustering Approach

K-Means Clustering Approach

In [46]:
```
# Using K-Means to cluster neighborhood into 3 clusters
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering)
kmeans.labels_
```

Out[46]:
```
array([0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 1, 2], dtype=int32)
```

In [55]:
```
neighborhoods_venues_sorted.insert(0, 'Cluster Labels 4', kmeans.labels_)

Scarborough_merged =df.iloc[:16,:]

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
Scarborough_merged = Scarborough_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='neighbourhood')

Scarborough_merged.head()# check the last columns!
```

Out[55]:

3. Methodology Section(cont.)

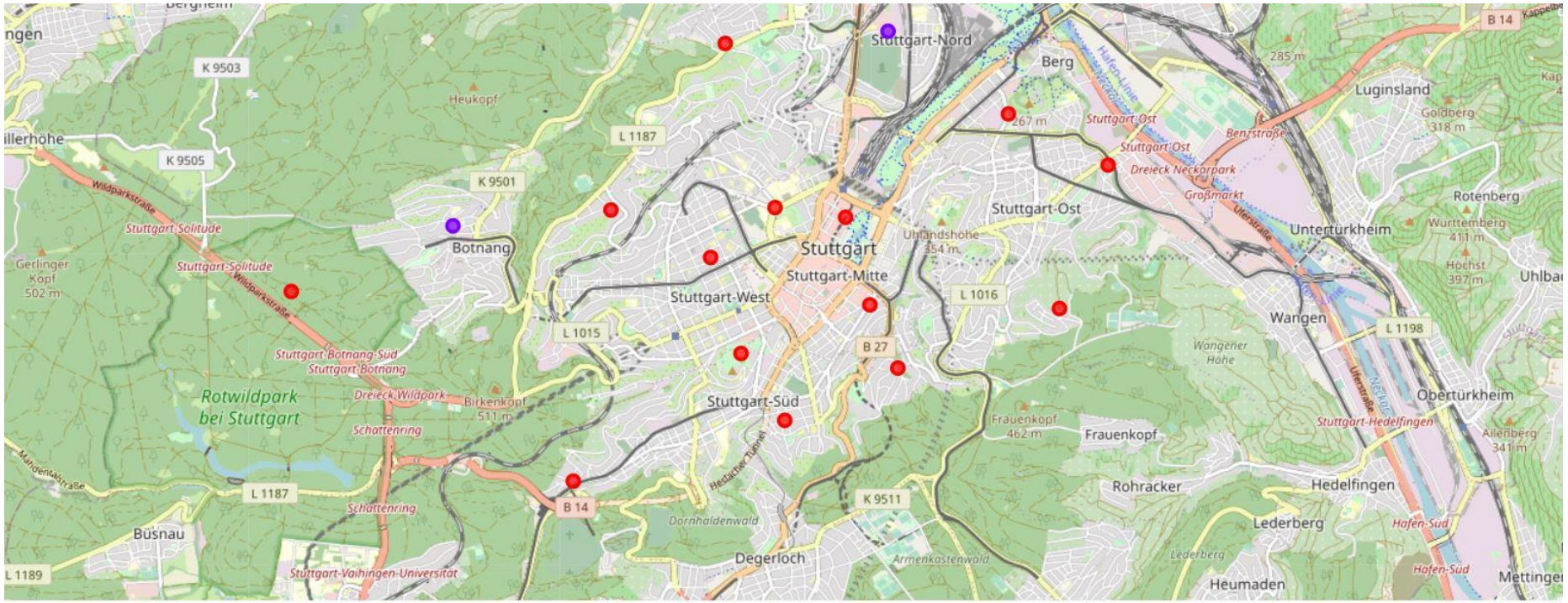
Most Common venues near Neighborhood

Out[45]:

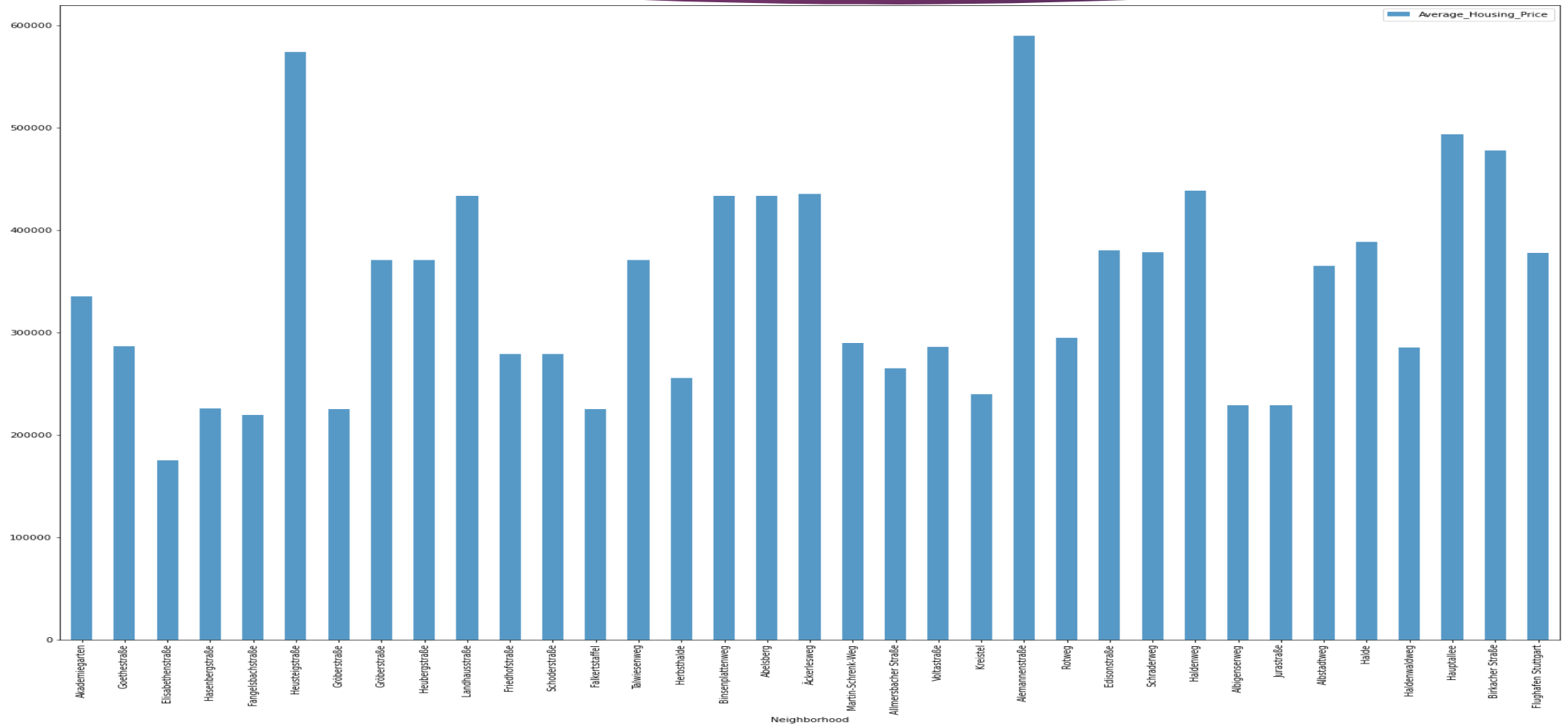
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Abelsberg	Bakery	Wine Shop	Gastropub	Doner Restaurant	Big Box Store	Plaza	Light Rail Station	German Restaurant	Flower Shop	French Restaurant
1	Akademiegarten	German Restaurant	Hotel	Plaza	Italian Restaurant	Bar	History Museum	Park	Café	Coffee Shop	Supermarket
2	Albigenserweg	Gastropub	German Restaurant	Italian Restaurant	Asian Restaurant	Ice Cream Shop	Sushi Restaurant	Mexican Restaurant	Miscellaneous Shop	Movie Theater	Chinese Restaurant
3	Albstadtweg	Hotel	Theater	Casino	Metro Station	Italian Restaurant	Gym / Fitness Center	Bus Stop	Spa	Chinese Restaurant	Multiplex
4	Alemannenstraße	Hotel	Bus Stop	Theme Park	Light Rail Station	Museum	Burger Joint	Supermarket	Modern European Restaurant	Climbing Gym	Food & Drink Shop

K-Means Clustering Approach

4. Results Section



4. Results Section(cont.)



5. Deduction

- **The Location:**
- Stuttgart is a popular destination for new immigrants in Germany to reside. As a result, it is one of the most diverse and multicultural areas in the Stuttgart, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into Germany has been one of on the rise.
- **Foursquare API:**
- This project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

6. Conclusion Section

- **School Rating by Clusters**
- **Conclusion:**
- In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 35 different lattitude and logitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices have been made.
- I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.