2020

Arora, Ankit

Capgemini

10/24/2020

Battle of Neighborhood



**Contents**

[**NEW FAST FOOD STORE PREFERENCES** 2](#_Toc54554382)

[**Introduction** 2](#_Toc54554383)

[**Data Collection** 2](#_Toc54554384)

[**Methodology** 3](#_Toc54554385)

[**Results** 4](#_Toc54554386)

[**Discussion** 4](#_Toc54554387)

[**Conclusion** 4](#_Toc54554388)

[**References** 5](#_Toc54554389)

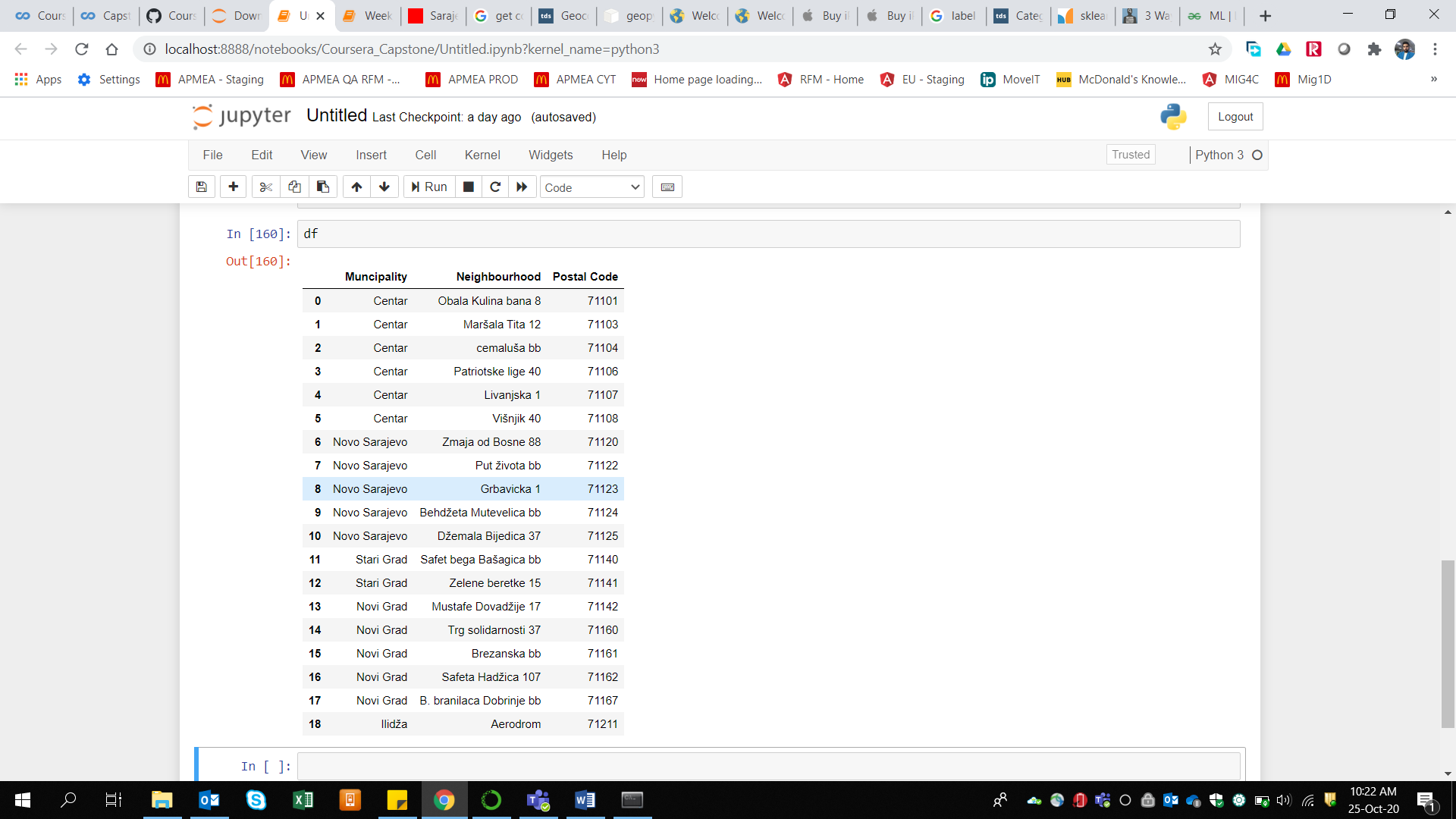
# **NEW FAST FOOD STORE PREFERENCES**

## **Introduction**

Whenever one have to start a business we need to first do market research in order to know the current trends of market as well as to find the premium location to set up an office where we can grab the attention of customers the most. Thus, we have to keep check on demographical, geospatial and environmental factors. Other major factor is to find potential competitors and the surroundings. In this proposal we are determined to evaluate above said features in order to find the best location in a city for opening of fast-food store. The proposal is intended to be for business management and thus the intention from the plan is to establish a new store and to get the business out of it. Initially, this setup will be helpful to determine the best location for a fast food store, but we can also build similar kind of models for other businesses and commodities.

## **Data Collection**

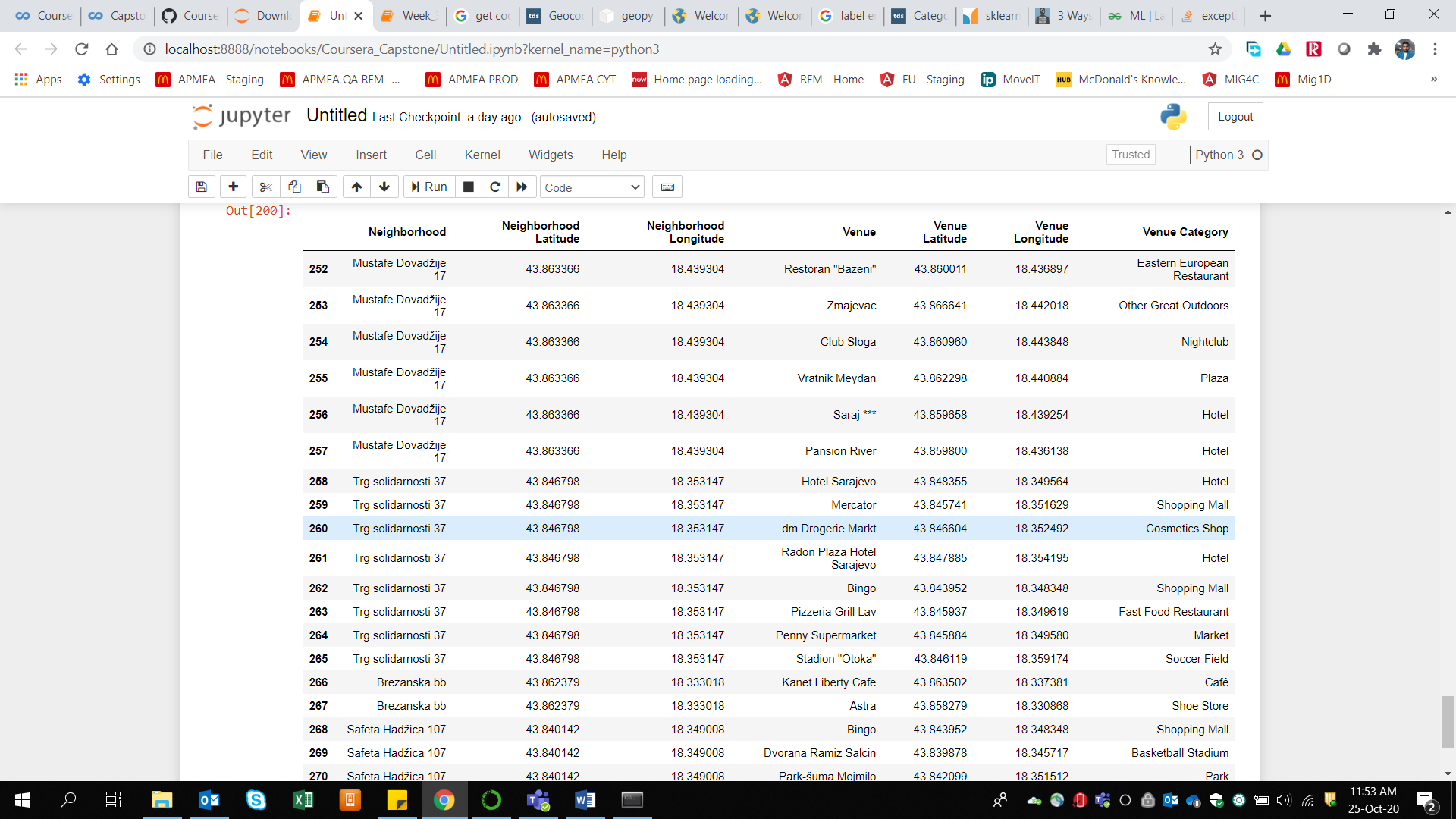
In order to accomplish the analytics task, we need to gather and process huge amount of data related to city locality, census, demographics, meteorological data and current businesses in the area. We have targeted Sarajevo for our analysis which is capital city of Bosnia in European continent and have collected the data for Municipalities and their neighborhoods from PostalCodeCountry.com. The data consist of Postal Codes, Municipality, and Neighborhood. Similarly, we have used geopy.geocoder module of Python to get longitude and latitude data of the localities.



*Fig. 1: Postal addresses of Sarajevo*

## **Methodology**

Methodology is the process that shows the approach followed to define and solve a business problem. In the introduction section, we have defined the business problem, then in data collection we have provided the reference of data we have taken for Sarajivo, Bosnia. We will be creating Jupyter notebook for interactive data analysis using Python programming language. We will be using KMeans clustering algorithms for explanatory data analysis and to find relation between independent and dependent variable. Venue density in the neighborhood will help determine the accuracy of our model. But before that we need to focus on Data Pre-processing and Feature Engineering which includes finding and removing outliers, filling up with NULL or missing values, Feature generation and data merging from various sources and datasets.



*Fig. 2: Venues in various Neighborhoods of Sarajevo*

## **Results**

As stated earlier, our business problem says itself what we are expecting from the project. We want to know the perfect location to setup a store for our business. We have taken Postal data for Sarajevo city of Bosnia and then used Foursquare API to find the venues within the regions. Then we have calculated the percentage of Venue Categories available in each Neighborhood which provided us with the most common venues per area when sorted. This was the first step for data pre-processing and Feature Engineering. Then we have applied the KMeans Clustering Algorithm to find the best cluster. This provide us with the category of venues as most common to uncommon and we can easily start a store in the location which have less common fast-food restaurant in the area.

## **Discussion**

This forum is to discuss the potential risks related to the detection of store location and also the key points which can improve our overall project. The list of few of the points is given below:

1. Population density and other demographical features to be considered too
2. Market divisions and residential blocks to be checked for the location detection.
3. From the above report we can only conclude about the likelihood of stores and the most common venues in the localities, whereas having a venue in locality doesn’t assure the sales and thus our store can fail if we neglect the sales trends of other stores.
4. Attraction points of the area/locality to be added in the model.

## **Conclusion**

This report concludes that we have a long path to go in order to process our plan and to deploy it in real-time. Our initial approach is very nice where we are checking the density of the various categories of venues in a neighborhood and also putting it as the most common venues in the location. Thus there are chances of profit if we start a store which is less common in the area as the number of competitors will be less or none. But that is just a part of it, we have to check the other points mentioned in discussion section before putting this idea in live environment.

## **References**

Postal Codes: http://www.postalcodecountry.com/zipcode/Sarajevo-1-Bosnia%20Herzegovina

Venues: (Foursquare API) <https://api.foursquare.com/v2/venues/explore>

Geocoder reference: <https://pypi.org/project/geopy/>