Coursera Capstone

Applied Data Science Capstone

Opening Juice Bars in Dhaka, Bangladesh

Amdadul Haque

April 2020

Introduction

- Fresh fruits and vegetables can change everything
- People are aware of healthy lifestyle and food habits and rapidly changing preferences of consumers
- Eventually, Juices and smoothies are becoming a vital part of consumer health and convenience
- Juice bars can help people in this regard
- Business prospects makes the idea sustainable

Business Plan

- Dense population has more prospect in Dhaka city
- Availability of current businesses in this regard
- Suitable location of Juice bar will add competitive advantage over other areas
- So, In the city Dhaka how can we recommend investors to setup Juice bars?
- How data science methodology and machine learning techniques can help?

Data

- Performance of machine learning algorithms highly depend of proper datasets. It requires:
  - List of neighborhood of Dhaka city
  - Location information of the neighborhoods for better visualization
  - Venue data specially related to Juice Bar Industries in Dhaka.

Data Acquisition

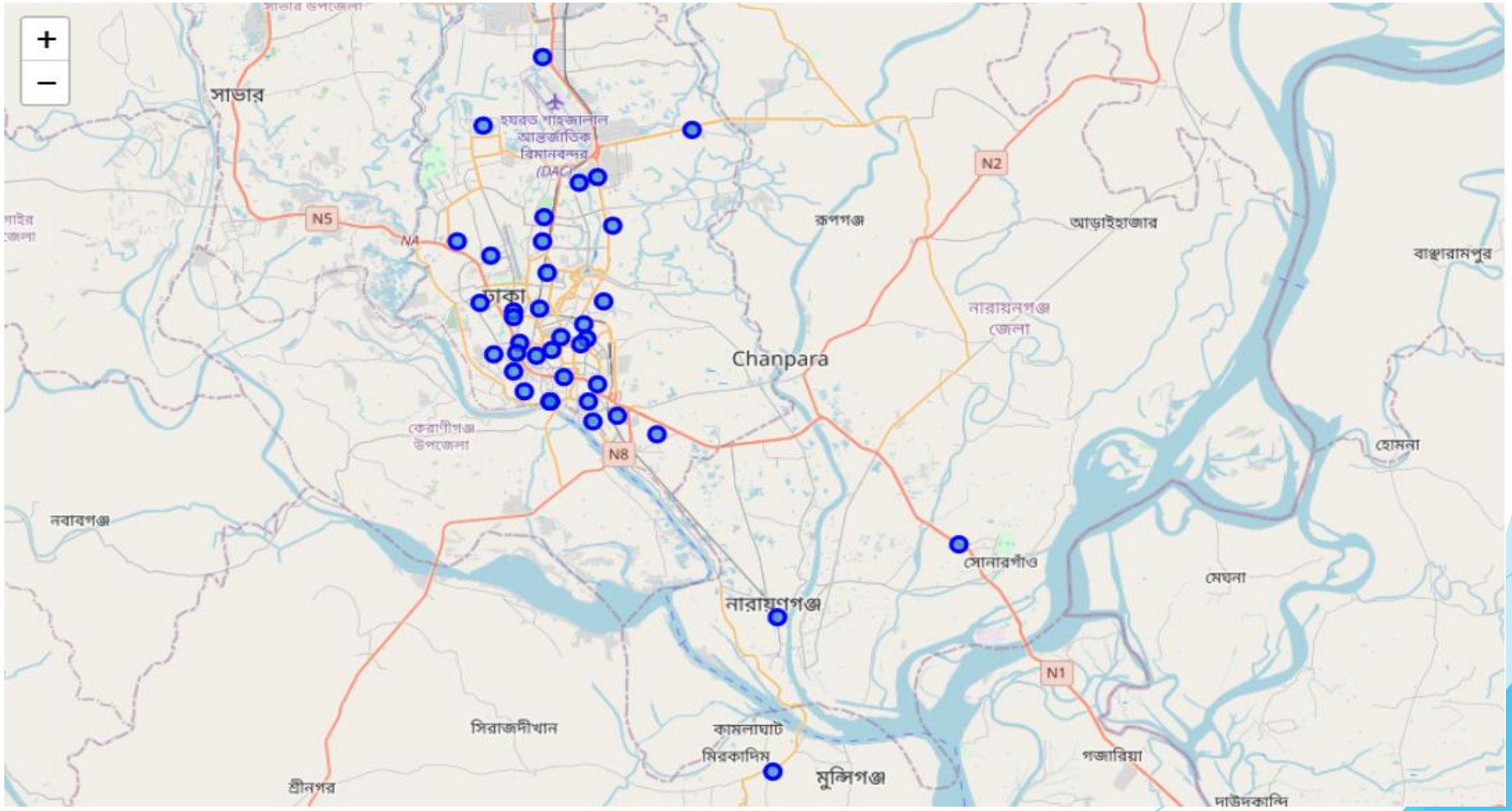
- Neighborhood data is not so much publicly available for Dhaka city. The Wikipedia page *(https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Dhaka)* of Dhaka, Bangladesh contains 42 neighborhoods
- The geographical coordinates for the neighborhoods will be extracted using Python Geocoder Packages which will give latitude and longitude information.
- And finally venues information will be collected using Foursqure API.

Methodology

- Dataset Preparation
  - Download and install required Python packages i.e. Numpy, Pandas, geopy, geocoder, matplotlib, Scikit-learn, folium and BeautifulSoup.
  - Scrap Wikipedia page of Dhaka City, parse using BeautifulSoup package.
  - Create pandas dataframe from scrapped web data. It come up with a dataset of 42 neighborhood of Dhaka city.
  - Getting the geographical data i.e. latitude, longitude of neighborhoods using geocoder package and update dataframe.

Methodology

- Create map of Dhaka City superimposed on top



Methodology

- Explore the neighborhoods using Foursquare API
  - We used Foursquare API to get the top 100 venues that are within a radius of 2000 meters and found 1598 venues information.
  - Finally updated the dataframe from applying machine learning algorithms.

Methodology

- Explore the neighborhoods Data, dataset come up with 115 unique venue categories.

```
print('There are {} uniques categories.'.format(len(df_venues['VenueCategory'].unique())))
# print out the list of categories
df_venues['VenueCategory'].unique()[:100]
```

```
There are 115 uniques categories.

array(['Electronics Store', 'Asian Restaurant', 'Theme Park',
       'Fast Food Restaurant', 'Shopping Mall', 'Steakhouse',
       'Indian Restaurant', 'Market', 'Department Store', 'Bus Line',
       'Bookstore', 'Shawarma Place', 'Bakery', 'Restaurant', 'Nightclub',
       'Art Gallery', 'Street Food Gathering', 'Bike Shop', 'Plaza',
       'Multiplex', "Dentist's Office", 'Convenience Store', 'BBQ Joint',
       'Café', 'History Museum', 'Fried Chicken Joint', 'Clothing Store',
       'Park', 'Ice Cream Shop', 'Italian Restaurant', 'Pharmacy',
       'Pizza Place', 'Food', 'Burger Joint', 'Thai Restaurant',
       'Scenic Lookout', 'Chinese Restaurant', 'North Indian Restaurant',
       'Bistro', 'Hotel', 'Diner', 'Boutique', 'Recreation Center',
       'Food Court', 'Soccer Field', 'Arts & Crafts Store', 'Shoe Store',
       'Historic Site', 'Coffee Shop', 'Bus Station',
       'Other Great Outdoors', 'Performing Arts Venue',
       'American Restaurant', 'Flea Market', 'Fish Market', 'Juice Bar',
       'Golf Course', 'Seafood Restaurant', 'Gym / Fitness Center',
       'Japanese Restaurant', 'Supermarket', 'Hobby Shop', 'Lounge',
       'Turkish Restaurant', 'Mexican Restaurant', 'Sushi Restaurant',
       'Donut Shop', 'Grocery Store', 'Portuguese Restaurant',
```

Methodology

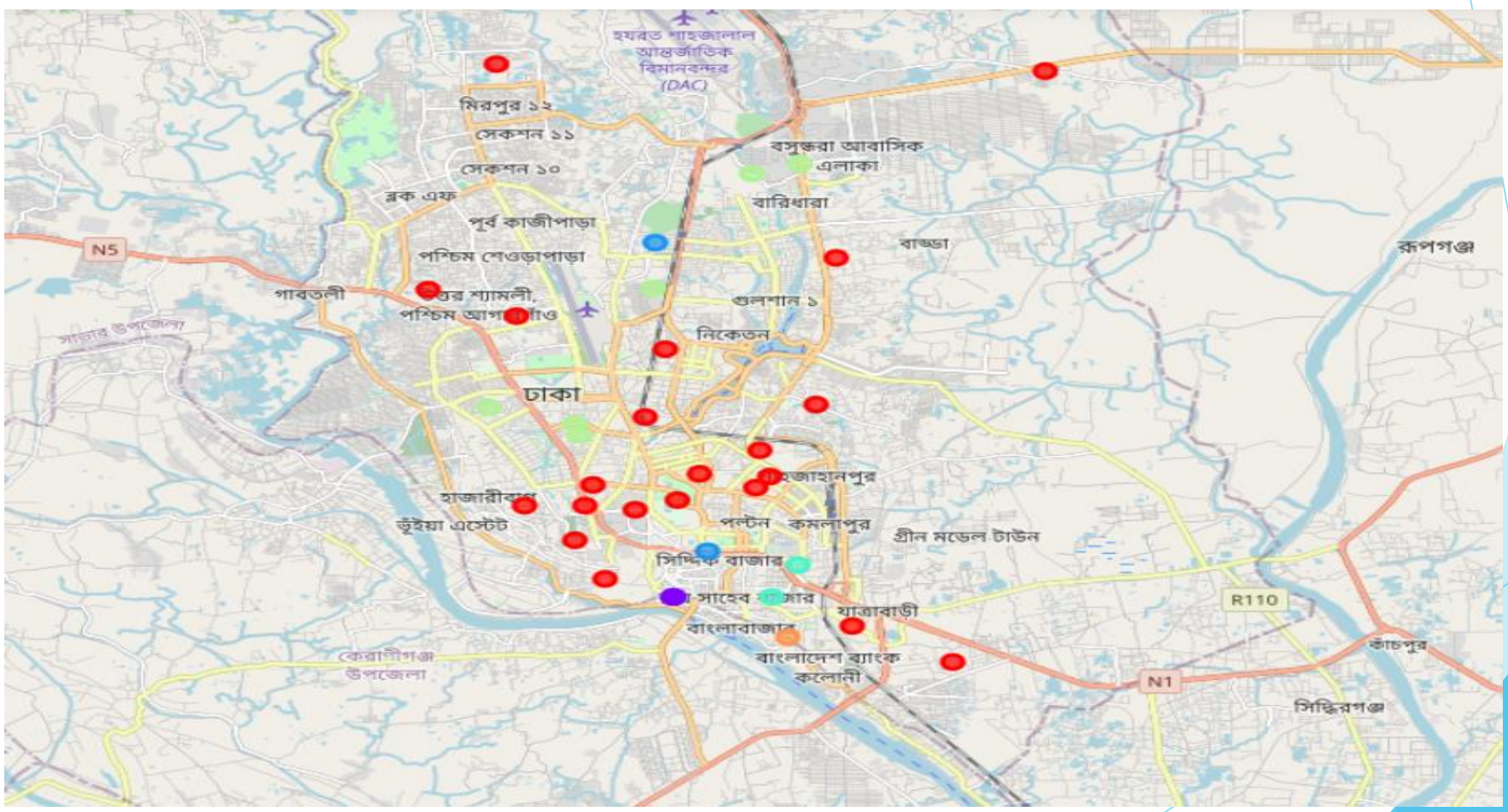
- Clustering using KMeans Algorithm
  - Perform data preprocessing of categorical data like venues using one hot encoder algorithm.
  - Analyze each neighborhood data and group venue categories by neighborhood and by taking the mean of the frequency of occurrence of each category.
  - Create a new DataFrame for Juice Bar data only
  - We run k-means to cluster the neighborhoods in Dhaka into 6 clusters on newly created grouped dataset of Juice Bar.
  - Fit the algorithm on dataset
  - Find the clusters with clusters levels

Results and Findings

- Results of 6 clusters of k-means.
  - Cluster 0: It is the biggest cluster but has **ZERO** concentration of Juice Bars in the neighborhoods.
  - Cluster 1: This cluster has high concentration number of Juice Bars in the neighborhoods
  - Cluster 2 and Cluster 3: Moderate Concentration
  - Cluster 4 and Cluster 5: These clusters are small and has low concentration of Juice Bars in the neighborhoods.

Results and Findings

- Clusters are represented in different colors superimposed in the map.



Recommendation and Future Aspects

- There are no of Juice Bar in the city of cluster 0
- It depicts that cluster 1 to cluster 5 has very few number of Juice bar with good concentration. That is, Juice bar is quite popular some areas neighborhoods in Dhaka city
- So opening up multiple Juice bar in the areas of cluster 0 with have chance of flourishing business.
- Better dataset and enterprise Foursquare API may provide better results

Conclusion

- Prospect of opening Juice bar in the Dhaka city with most popular clustering algorithm and tools.
- From identifying business area, extract data from available sources, analyze dataset, explore location API and finally found cluster that has zero concentration of Juice bar in cluster 0 and It also depicts that Juice bar is quite popular in the city.
- Thus, opening multiple Juice bar in has a good prospect of business expansion

Thank You