Capstone Project: Data Driven Methodology for Finding Suitable Places in Order to Open Restaurants in Blue Area, Islamabad, Capital of Pakistan

Applied Data Science Capstone
IBM Data Science Professional Certificate

Introduction

- Different flavours of food dishes are growing stronger in Islamabad, Pakistan like Asia, chinese etc.
- In Islamabad center (Blue area) and nearby places, Coffee shops, Pizza shops, fast food and tea points are opening everywhere, and are always full.
- This project aims to estimate the best localization to open such a business in Blue Area, Islamabad, Capital of Pakistan.

Introduction

- Prior launching any restaurant, it’s important to know if the business as a good opportunity.
- In order to do so, this report will try to gather data about other restaurant localization, competitors and best localization.
- These data could be use for a business plan afterward

Problem

- Which place should stakeholder select to open his new Restaurant in Blue Area, Islamabad?
- Restaurant needs to be strategically located inside the biggest concentration of workers in Blue area.
- Confirm any assumption by means of modeling and testing the data.
- Visually cluster common restaurants in Blue area.
- Place with high frequency of peoples.
    - Are neighborhood places populous?
    - Stack Holder wants to be able to judge which neighborhoods also may be poised to grow in restaurant numbers in coming years.

Problem

- Shortest travel time for his clients
- Overall lower run costs
- Increase in overall business
- Overall greater customer satisfaction

Data

- Following API's will be utilized to collect data
- Foursquare API:
  - For finding restaurant/venues
- Geopy API:
  - Reverse geolocalisation

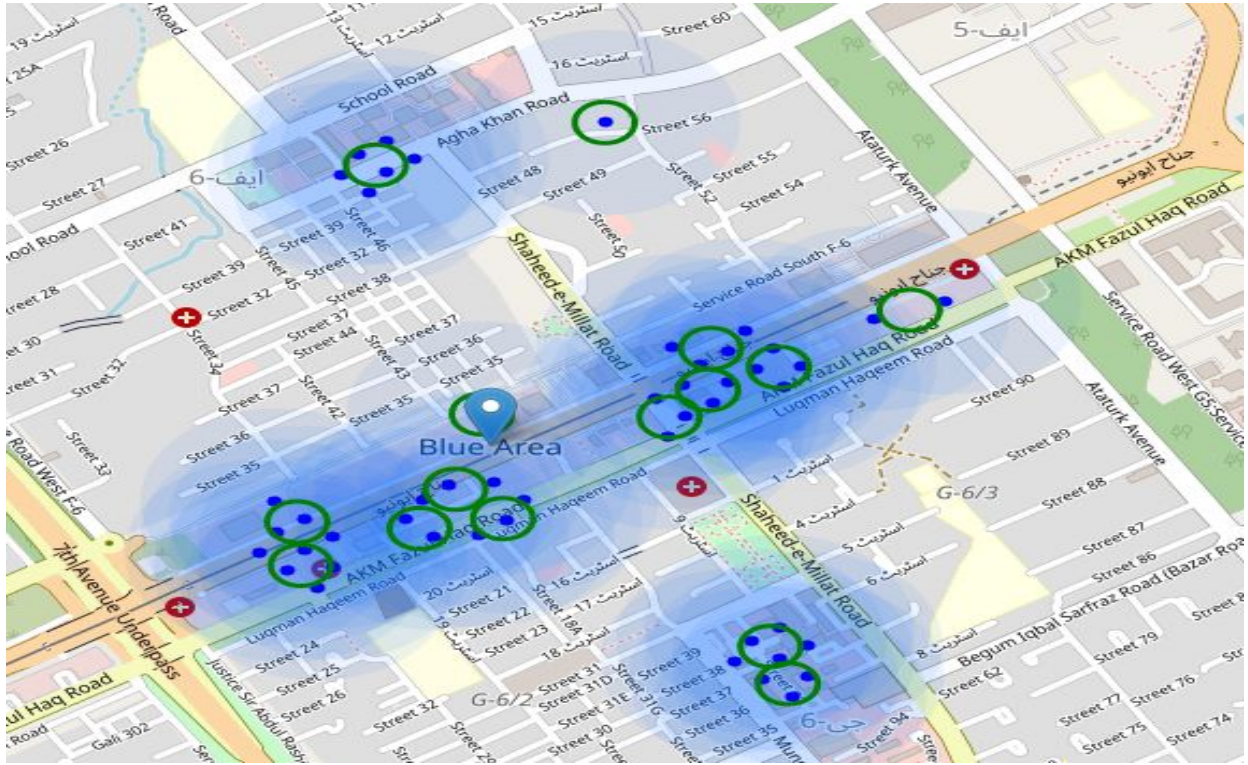
Methodology

- Firstly find the latitude & longitude of Blue Area, Islamabad
- Use Foursquare API to get info about neighborhood in Blue Area.
- We're interested in venues in 'food' category, but only the ones who can be competitors:
  - mean fast food, quick food, take away, healthy, not restaurant taking too long.

Methodology

- List all companies and universities to evaluate the customer pool.
- Spot out worker/people area more than competitors and try do determine where are clusters more of the worker.
- To do so the k-mean clustering method will be utilized.

Results


Blue Area
Agha Khan Road
School Road
Shaheed e-Millat Road
AKM Fazul Haq Road
Luqman Haqeem Road
Service Road South F-6
Service Road West G5-Service
Ataturk Avenue
7thAvenue Underpass
Shaheed e-Millat Road
Begum Iqbal Sarfraz Road (Bazar Roa
G-6/3
G-6/2
Street 56
Street 60
Street 90
Street 89
Street 88
Street 87
Street 86
Street 79
Street 77
Street 76
Street 75
Street 74
Street 62
Street 94
Street 24
Street 25
Street 26
Street 27
Street 28
Street 30
Street 31
Street 32
Street 34
Street 35
Street 36
Street 37
Street 38
Street 39
Street 41
Street 42
Street 43
Street 44
Street 45
Street 46
Street 48
Street 49
Street 50
Street 52
Street 54
Street 55
Gali 302

Results

```
================================================================
Addresses of centers of areas recommended for further analysis
================================================================

HSBC, Agha Khan Road, ایف-6, Blue Area, 1.1 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Melody Market, اسٹریٹ 100, جی-6, Blue Area, 1.1 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Service Road West F-6, Blue Area, جی-6, Blue Area, 0.2 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Milkshake Place, جناح ایونیو, ایف-5, Blue Area, 0.6 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
ایف-6, Blue Area, 0.6 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
PIA, جناح ایونیو, ایف-5, Blue Area, 1.2 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
31, Street 56, F-6/4, ایف-5, Blue Area, 1.3 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Melody Market, اسٹریٹ 100, جی-6, Blue Area, 1.2 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
ایف-5, Blue Area, 0.7 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Blue Area, جی-6, Blue Area, 0.4 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Oil & Gas Regulatory Authority, AKM Fazul Haq Road, جی-6, Blue Area, 0.7 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Goodwill Pharmacy, Shaheed-e-Millat Road, Blue Area, جی-6, Blue Area, 0.5 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
UBL Corporate Office, Service Road South F-6; Service Road West F-6, ایف-6, Blue Area, 0.1 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستان
AKM Fazul Haq Road, جی-6, Blue Area, 0.3 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
Shenwari Foods, AKM Fazul Haq Road, ایف-5, Blue Area, 0.8 <=    وفاقی دارالحکومت اسلام آباد, 4400, پاکستانkm from center
```

Discussion

- This analysis shows that we must consider other criterias than just number of restaurant.
- Blue area is a small part of Islamabad city, so the concentration of restaurant is quite high, in this analysis I tried to correlate the number of restaurant and quantity of potential customer.
- In opposite to what I was thinking, the center area is not very crowded, mainly because building are big.
- Also, I was able to discover that there are not a lot of competitor on this business area, which is very good.

Discussion

- 15 good potential places are found, I personally think that the one in the north is better.
- We must just take care of one thing, I think the API didn't return all data, we are missing a lot of companies, the map is still good, and the result can be trusted, but we should cross check data with other data source.
- In order to be more accurate, it could be possible to give a weight to customers for example, a university with 1000 student would then weight more than a haircut company with two employees.
-

Conclusion

- This project can be reused for other areas, just think about changing clustering size to adapt to your area.
- Also, I have created/modify a huge quantity of function in order to adapt.
- It's very far from being perfect, a lot of work can be done, other source of data can be found, but in the end the result seems to correlate with the real world, when we know the city, the area predicted seams correct.