Predicting A New Location For a Mechanic Shop in the DC Area

Connor Ruch

June 24, 2021

1. Introduction

1.2 Background/Problem

Could you use location data to predict where to open a new mechanic shop? I will try to find out if you can in this report. I have chosen DC for the availability of the data. This kind of analysis could be done elsewhere with the right access to datasets. I want to see if I can use vehicle crash data combined with location data for the mechanics shops already open in DC to see if there is an area of the city where there is crashes but not many mechanics. My assumption for this report is that in the event of a vehicle crash the car(s) are most likely taken to the nearest mechanic afterwards and this would lead to business. I am also assuming that crashes will tend to happen in similar areas over time due to faulty roads, high traffic areas, or unsafe areas. Many more factors should be consdered when opening a shop but I will focus on these small few.

2. Data

2.1 Data Sources

My data sources for this report will be the foursquare location data providing the location and other info on all the mechanics in the DC area, and the dataset of Washington, DC vehicle crashes from Kaggle at this url "https://www.kaggle.com/gauravduttakiit/accidents-in-washington-dc"

2.2 Data Cleaning

The foursquare data was gotten by using requests package and cleaned. A radius was set as to get mechanics that were in a similar radius to the vehicle crash data I had available. The categories were cleaned as well, and any unneccessary columns were dropped from the table (Table 1)

]:

Table 1

	name	categories	location.lat	location.lng	location.postalCode
0	Auto Alliance	None	38.902200	-77.033240	20005
1	Auto Alliance	Office	38.900070	-77.021512	NaN
2	Automated Graphic Imaging	None	38.905868	-77.032239	20005
3	Auto Ching Ching Mercedes	Road	38.905593	-77.043451	20001
4	Exotic Auto Detail	Car Wash	38.904973	-77.048837	20037

The vehicle crash data downloaded from Kaggle and then imported to the report from the local file. If was filtered by the latitude and longitude of the outer most mechanic shops in order to limit the data within the same radius of the shops that were picked up from the Foursquare API data. The data was then scanned for the ID, latitude, and longitude of each incident and reformed into a dataframe. (table-2)

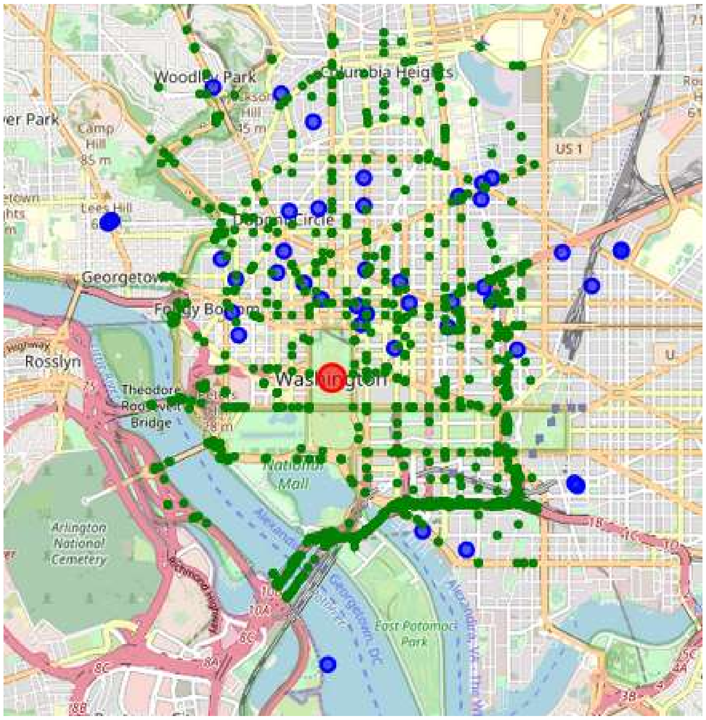
Table-2

	ID	lat	long
805	A-2575025	38.872988	-77.042354
1189	A-3309587	38.873060	-77.042273
991	A-2884633	38.873160	-77.042170
975	A-2871809	38.873160	-77.042170
974	A-2871799	38.873160	-77.042170

3. Methodology

3.1 Folium

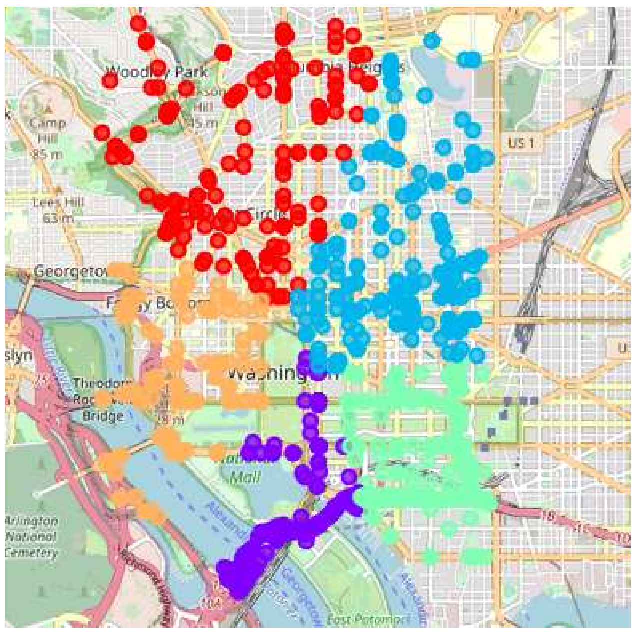
I used folium to map out the location of the mechanics and the crashes in the area for a visualization.



3.2 Clustering

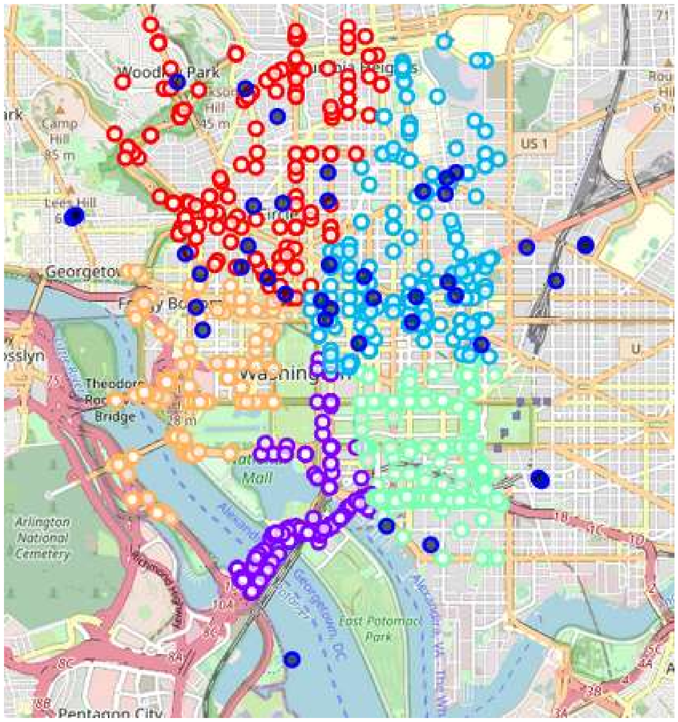
I clustered the crashes together using the Kmeans algorithm package from sci-kit learn in order to cluster the crashes in even groupings by location. I first assigned each data point to a cluster label then plotted the data.

	Cluster Labels	lat	long
0	4	38.888493	-77.052330
1	1	38.882065	-77.027855
2	3	38.894791	-77.014091
3	1	38.879242	-77.035889
4	1	38.882214	-77.028603



3.3 Adding in shops

after I clustered the crashes together I superimposed the location of the mechanics shops onto the map in order to see the distribution. The crashes are the different colors but with a white fill and the mechanics area all dark blue outline with a black fill to distinguish them a bit more.



3.4 Grouping shops

I then grouped the shops into one of the clusters of the crashes. I did this in order to see if there was a cluster with fewer shops than the rest, hopefully to show that there would not be much competition in the area given an even number of crashes. I normalized the distances beteween the shops and each clusters center in order to get the cluster closest to each shop and assign it to that cluster. Then I got a count of the shops in each cluster after the assignment

	name	categories	location.lat	location.lng	location.postalCode	cluster
0	Auto Alliance	None	38.902200	-77.033240	20005	1
2	Automated Graphic Imaging	None	38.905868	-77.032239	20005	1
3	Auto Ching Ching Mercedes	Road	38.905593	-77.043451	20001	4
4	Exotic Auto Detail	Car Wash	38.904973	-77.048837	20037	0
6	Auto Insurance Washington	Insurance Office	38.914553	-77.017018	20001	1

- The code below will show the mechanics frame after they have been assigned to clusters and a count by cl
- This should show whether any cluster has too few mechanics.

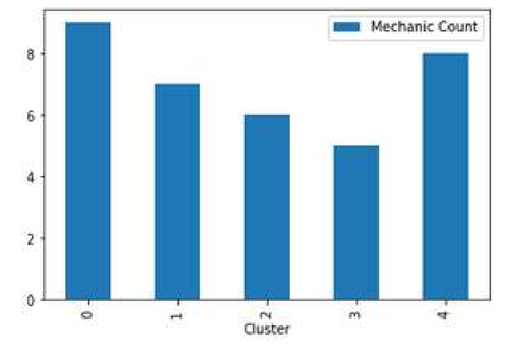
```
count_frame = df_filtered.groupby('cluster').count()
count_frame
```

cluster	name	categories	location.lat	location.lng	location.postalCode
0	9	8	9	9	9
1	7	5	7	7	7
2	6	6	6	6	6
3	5	5	5	5	5
4	8	8	8	8	8

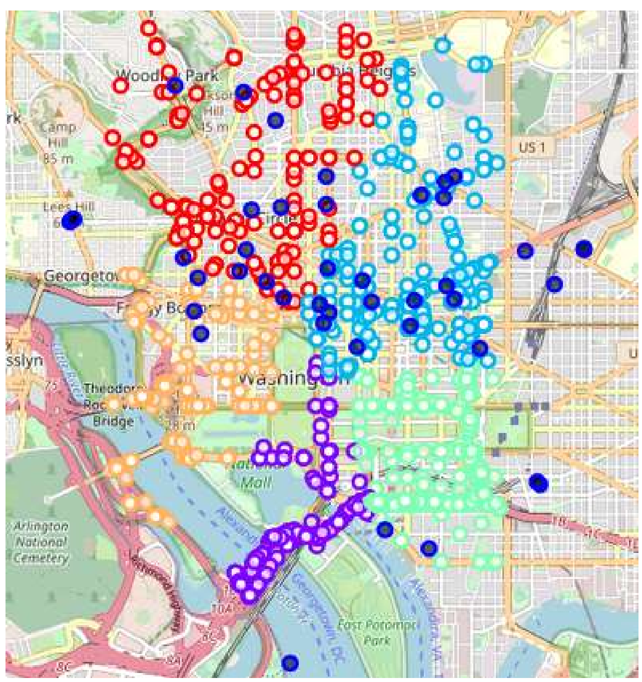
4. Results

We can see from the results that the area of cluster 3 had the fewest number of mechanics. If the assumptions in this report are correct it could mean that this would be a good area in which to open a new shop.

	Cluster	Mechanic Count
0	0	9
1	1	7
2	2	6
3	3	5
4	4	8



Looking again at the map, it shows cluster 3 (appearing in light green) would seem to be the best place for a new shop, down in southeastern DC. and cluster 0 (appearing in red) had the most shops, this could mean stiff competion for the new owner.



5. Discussion

Some observations:

- Cluster 3 seemed to be the best place for a new shop given that there were few other mechanics in the area despite a similar cluster of crashes.
- Cluster 0 had the most mechanics in the area followed by cluster 4. It would seem these areas would have the most competition from competing shops.

  - Given that it probably takes a lot more analysis to probably determine the best place for a new shop, some more data and analysis could be done in order to figure out more about the market for auto maintenance
- Recomendations:
  - to a new business owner that he consider the analysis and given the other shops in the area that he open a shop in the geographic zone covered by cluster 0 (in light green) if possible

6. Conclusion

In conclusion the data seems to show some area for a new shop to be opened in DC.

- It is very difficult to determine as a certainty what the market demand is for a new shop just based on the crash data.
- Certainly more analysis could be done, especially to analyze some of the assumptions in the report
- Some assumptions made include:
  - That in the event of a crash the car(s) would be most likely taken to the nearest mechanic in the area and,
  - That the prime business for a mechanic is derived from motor vehicle accidents.
- There may be far more variables that determine the success of a mechanic shop but it is out of the scope of this report.
- Some other factors for shop success may include:
  - Number of residents in the area who own a vehicle,
  - Regulations by the state required auto inspection which may lead to more business for a shop,
  - Average amount of vehicle care in the area compared to other areas, etc.
- Again more analysis should be done when deciding to open a new business but this showed some promise in analyzing the potential for demand and competition!