Data Science Capstone Project

Finding Optimal Location for Auto Mechanic Shop in State of Maryland

Introduction: Business Problem

In this project our goal is to find most favorable location for a car mechanic shop in Maryland, USA. This report will be compatible with shareholders' needs.

In accordance with shareholders request, we will find the best location which will be close to close to cities but not close range to any of competition. Using data science methodology and instruments such as data analysis and visualization, this project aims to provide optimal location.

Thus, we will compare Foursquare data to registered passenger vehicle data.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

Based on our business problem we will need:

- Passenger vehicle numbers in each city/county
- Geolocation coordinates of each city/county
- Number of existing mechanic shops in each city/county

Therefore, this project uses two main sources of data. Our crucial data is registered passenger vehicles in Maryland, USA. This data can be retrieved from State of Maryland's open data source - https://opendata.maryland.gov/w/kqkd-4fx8/gz96-f9ea?cur=7U4SpQ1_sHa

This data gives us important information on vehicle number from 2010 to 2020 in each county of Maryland. Even so, this data requires data cleaning, and wrangling.

To obtain geolocation, we will use Python's Geocoder package.

Finally, we will use Foursquare API to obtain number of existing mechanics in each county with their names as well. Then, this information will be visualized on Python's Folium package map.

Methodology

- Firstly, our data will be retrieved from State of Maryland's open data source - https://opendata.maryland.gov/w/kqkd-4fx8/gz96-f9ea?cur=7U4SpQ1_sHa For us to use this data, it will be cleaned and wrangled.
- Foursquare will be used to determine number of shops as well as their geolocations and names.
- Mechanic shops and vehicle numbers will be combined into one data frame.

Exploratory Data Analysis

After gathering all necessary data, and cleaning and wrangling, we made a new dataframe from combining existing data. A portion of our data shown below.

	COUNTY	COUNTY LATITUDE	COUNTY LONGITUDE	2020	NUMBER OF MECHANICS
1	ALLEGANY	39.657245	-78.655063	60551	45
2	ANNE ARUNDEL	39.049347	-76.512884	586982	50
3	BALTIMORE	39.290882	-76.610759	691845	50
4	BALTIMORE CITY	39.265954	-76.635278	292633	50
5	CALVERT	39.700388	-75.982445	98349	50
6	CAROLINE	34.791269	-82.282715	38335	50
7	CARROLL	39.635774	-75.806001	185749	50
8	CECIL	38.705311	-76.532315	98524	35
9	CHARLES	38.371881	-75.589176	156153	50
10	DORCHESTER	39.332102	-76.686443	33638	50

Number of registered cars on each county is labeled as '2020'. To see this data better, we have showed this on bar chart:
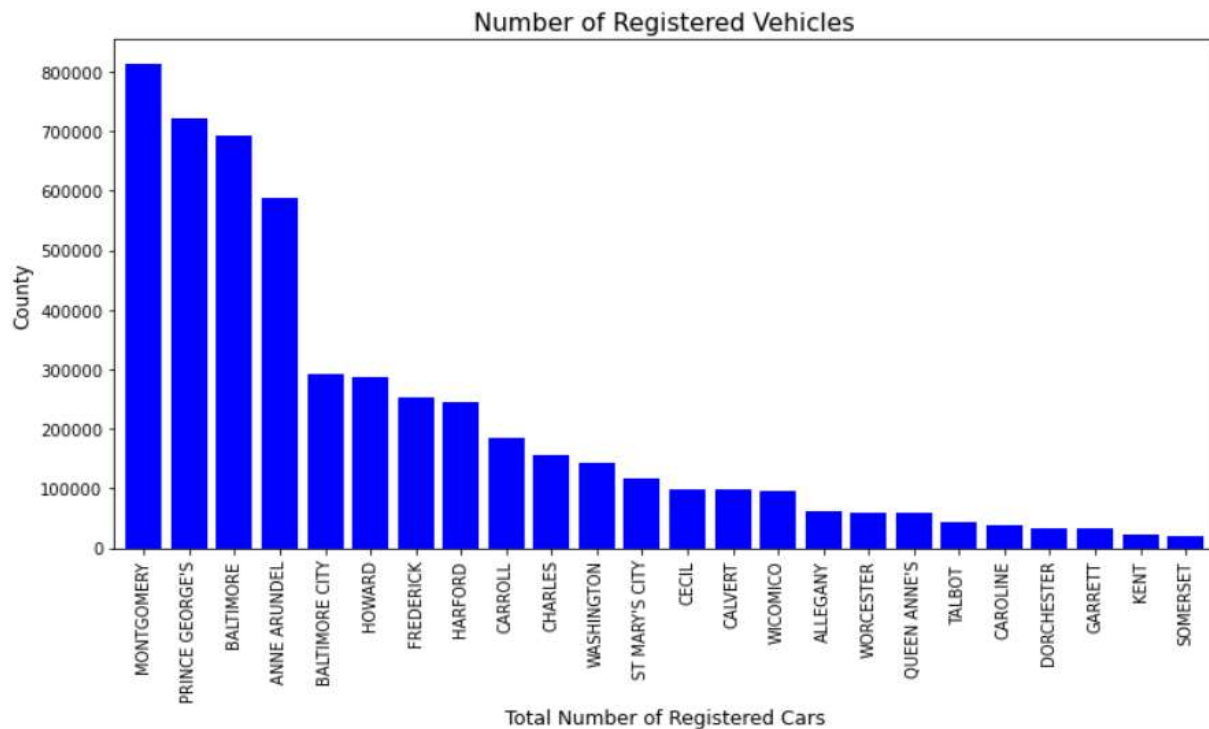
Ratio between number of cars and number of mechanics was not in our dataset so we must gather for our dataset. This will show how many cars for each mechanic in each county.

We added ratio column to our data as shown below.

|	COUNTY	COUNTY LATITUDE	COUNTY LONGITUDE	2020	NUMBER OF MECHANICS	RATIO
1	ALLEGANY	39.657245	-78.655063	60551	45	1345.58
2	ANNE ARUNDEL	39.049347	-76.512884	586982	50	11739.6
3	BALTIMORE	39.290882	-76.610759	691845	50	13836.9
4	BALTIMORE CITY	39.265954	-76.635278	292633	50	5852.66
5	CALVERT	39.700388	-75.982445	98349	50	1966.98
6	CAROLINE	34.791269	-82.282715	38335	50	766.7
7	CARROLL	39.635774	-75.806001	185749	50	3714.98
8	CECIL	38.705311	-76.532315	98524	35	2814.97
9	CHARLES	38.371881	-75.589176	156153	50	3123.06
10	DORCHESTER	39.332102	-76.686443	33638	50	672.76
11	FREDERICK	39.414219	-77.410927	253839	50	5076.78
12	GARRETT	39.544087	-79.249698	33105	14	2364.64
13	HARFORD	39.548545	-76.304899	245395	50	4907.9
14	HOWARD	39.318421	-76.615649	285699	50	5713.98
15	KENT	39.258755	-76.103845	22114	17	1300.82
16	MONTGOMERY	39.140627	-77.207561	813579	50	16271.6
17	PRINCE GEORGE'S	38.815114	-76.758858	722509	50	14450.2
18	QUEEN ANNE'S	38.996441	-77.018538	58257	50	1165.14
19	SOMERSET	38.967302	-77.095506	20557	50	411.14
20	ST MARY'S CITY	38.186492	-76.430120	116160	31	3747.1
21	TALBOT	38.754207	-76.191982	44642	29	1539.38
22	WASHINGTON	39.589350	-77.710254	142199	50	2843.98
23	WICOMICO	38.413458	-76.876637	95023	50	1900.46
24	WORCESTER	38.223118	-75.331947	59729	8	7466.12

Now, let us show this ratio in bar chart with number of cars.

There is a wide range values between these two values. So, we will use log scale bar chart to see the relation between these two values.

Registered Cars in 2020 and Ratio between cars and mechanic shops
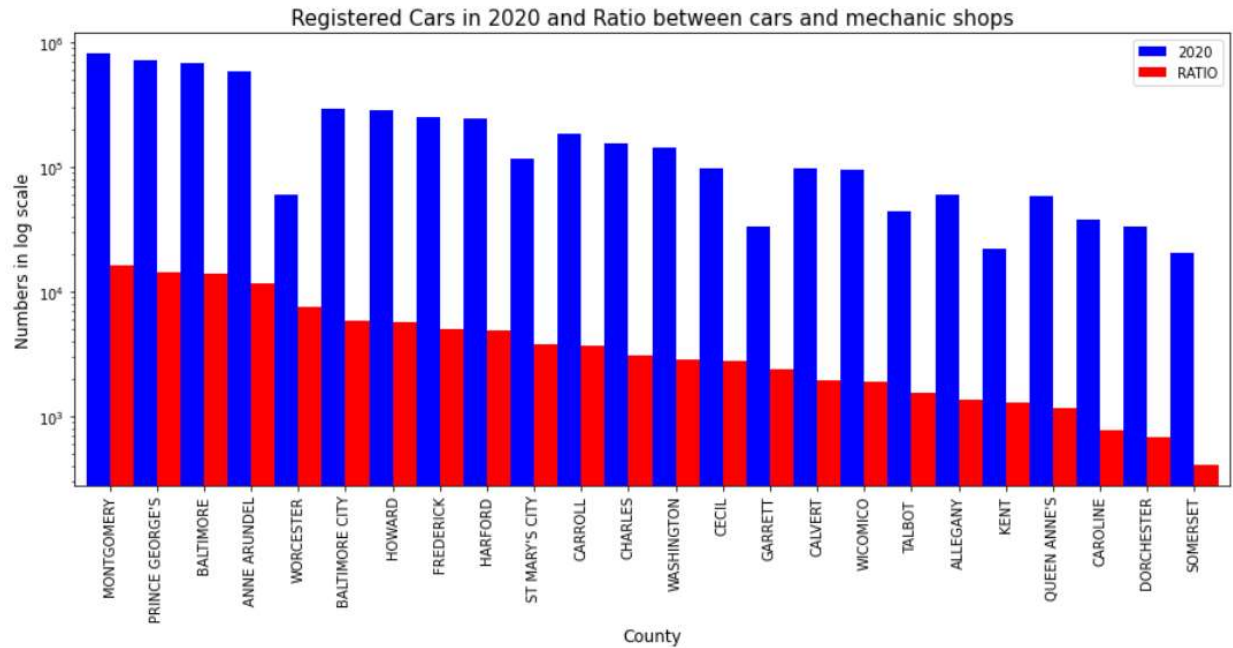
Ratio and number of cars in each county will determine where our mechanic shop would be. If you look at the chart, first six county looks very promising to open a shop. Also, these counties can be seen cluster map.

After this bar chart, we can clearly see that there are only four county's we can focus on. We will only focus on Worcester, Montgomery, Wicomico, and Anne Arundel counties. Therefore, we drop other countries on our dataframe.
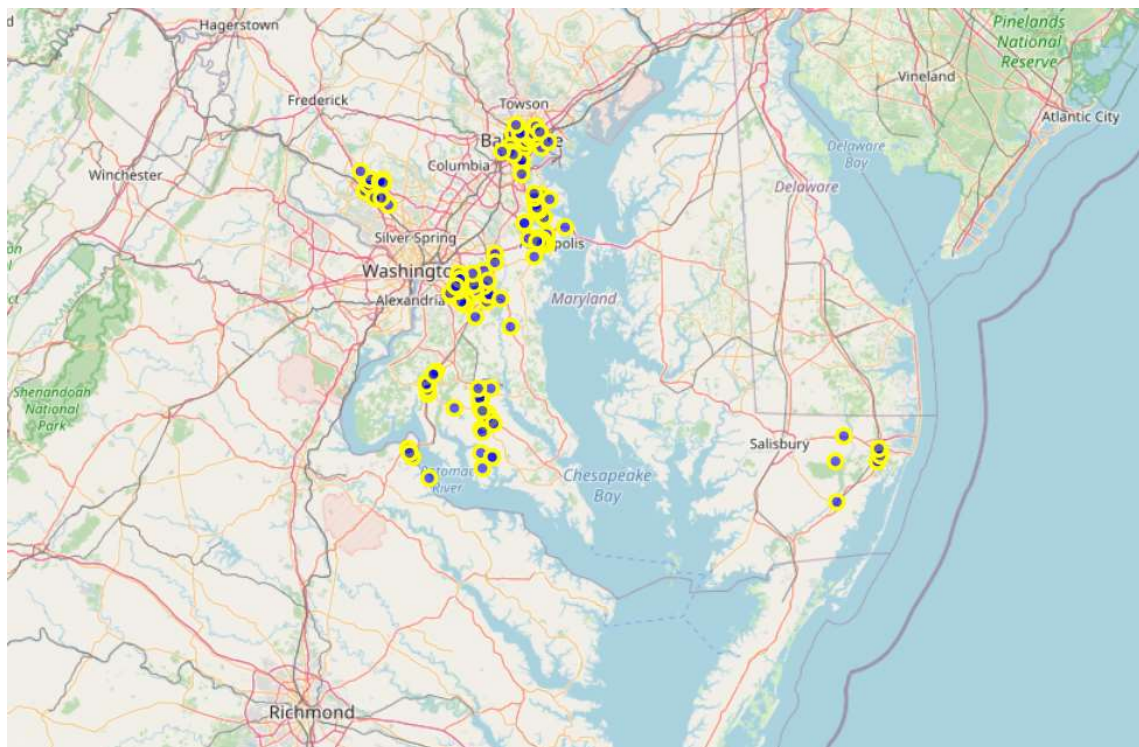
Map of clustered mechanic shops in count
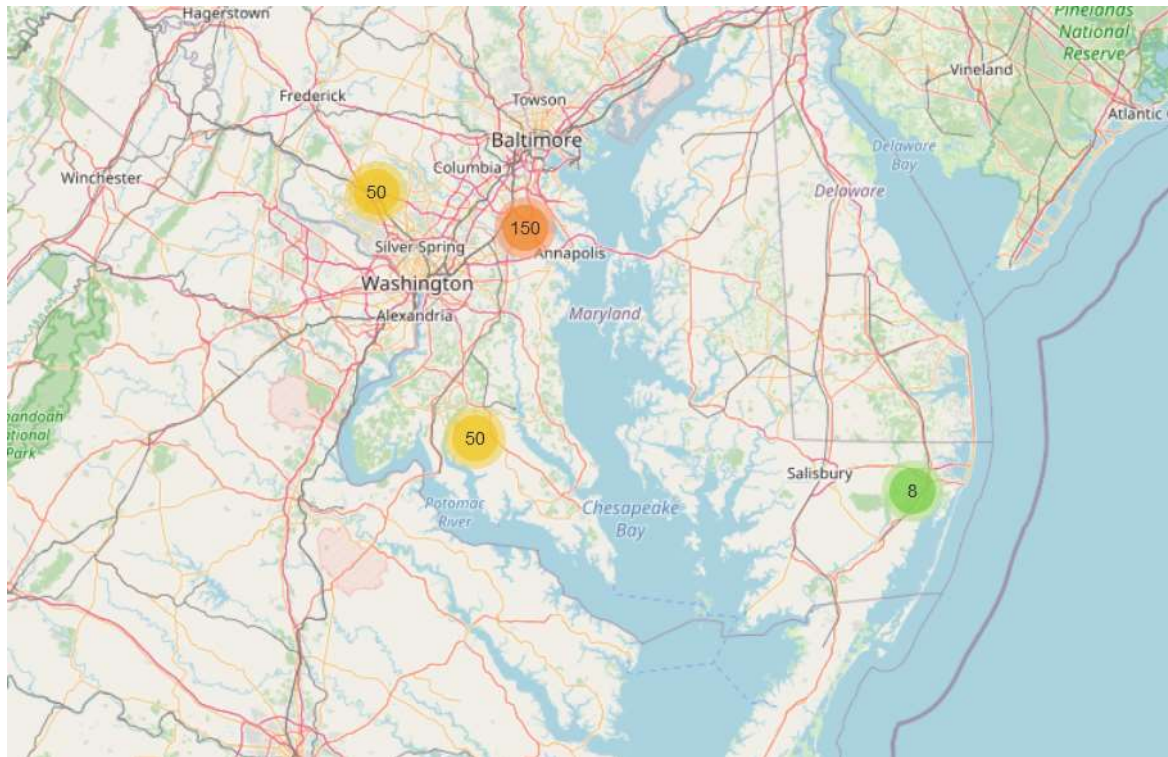
Results and Discussion

Our analysis shows that there is a great number of mechanic shops near capital city of Maryland. Highest concentration of mechanic shops was detected around city centers.

Last step of analysis, cluster map, shows that mechanic shops density getting higher around Baltimore. Thus, our focus should be on Worcester, Montgomery, Wicomico, and Anne Arundel counties. Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is remaining counties contain great number of potential new mechanic shop locations based on number of car and distance to existing mechanic shops. Purpose of this analysis was to only provide info on areas close to county centers but not crowded with existing mechanic shops.

Conclusion

Purpose of this project was to identify Maryland Counties to aid stakeholders in narrowing down the search for optimal location for a new car mechanic shop. By calculating mechanic shops density distribution from Foursquare data, we have first identified total number of shops in each county, and then we eliminate most of counties based on ratio analysis. We have showed remaining counties with

number of shops on map. Then we clustered those location to be used as starting points for final exploration by stakeholders.

There is a great number of mechanic shops near Baltimore area, which is 150 mechanic shops. It could be even greater number than 150, because we made our search limit to 150. Even so, there is a great number of registered cars in this area, Baltimore area shows no promising.

Final decision on optimal mechanic shops location will be made by stakeholders based on specific characteristics, which as mentioned on our analysis, number of car numbers for each mechanic shops, density of mechanic shops in each county, location distance from city centers etc.