

Capstone Project: The Battle of Neighbourhoods (Week 2)

Report

1) Business problem

Washington D.C is the capital of USA, which means that it is the capital of the World! Many people around the world visits this city yearly. This city contains many cultures around the world, it is known as the city of politics, business and deal making.

It is a strategic place to have a restaurant because they are need to do important businesses, restaurants gets a lot of visitors every two years because of the elections, it is a great opportunity during these elections to gain new customers. In addition, restaurants in D.C have less strict laws compared to other states in the US .

D.C contains a large metro area population, in 2021 there are around 5, 378, 00 people who live there, and it will not stop growing! In fact, every year there is an increase by around 1.1% to the population of D.C.

When someone owns a restaurant in D.C, there are some things to look for to have a better opportunity (*Problems to solve*):

- Which areas are not expensive for a restaurant owner: land area and renting cost for example?
- Which areas lack restaurants?
- Which area have the largest population?

2) Data

- **Foursquare API:**
Restaurants neighbourhoods in Washington D.C are explored using Foursquare API.
- **Washington D.C :**
Location of restaurants (latitude and longitude) are used to explore neighbourhoods of Washington D.C.

Source:

- a) <https://opendata.arcgis.com/datasets/071aa9b40a1a4b38a938f8f5058068a818.csv>.
- b) https://en.wikipedia.org/wiki/Neighborhoods_in_Washington,_D.C.

3) Methodology :

- 1) Data about Washington D.C is extracted from a csv file that contains the name of the neighbourhoods and their corresponding longitudes and latitudes.
- 2) From a Wikipedia webpage, data is extracted using web scraping which contains the list of Wards of different neighbourhoods.
- 3) Data extracted from the Wikipedia page is merged with the table that contains the data of the csv file

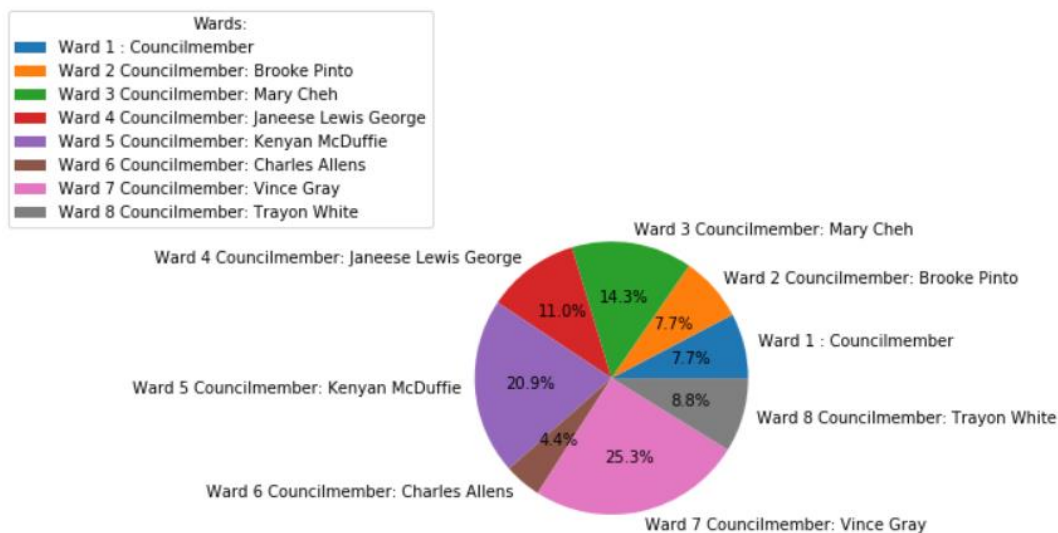
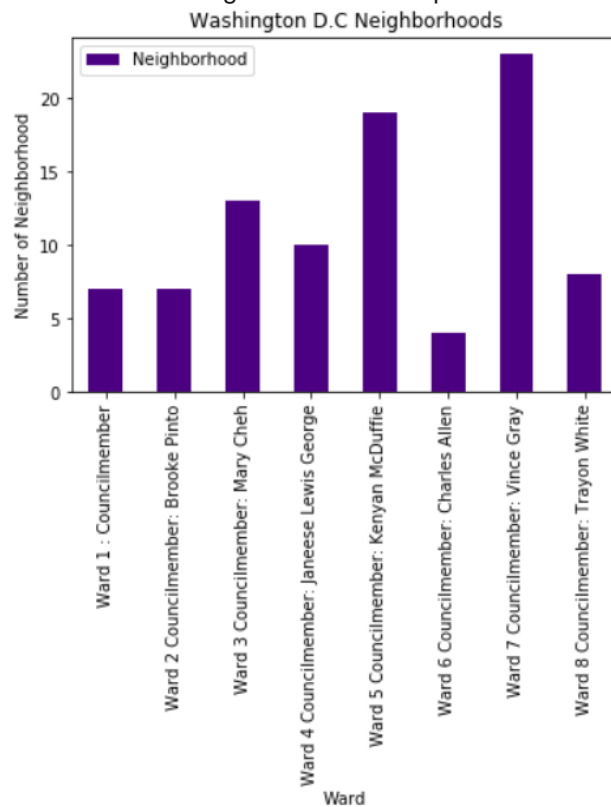
Output :

	Neighborhood	Latitude	Longitude	Ward
0	Fort Stanton	-76.980348	38.855658	Ward 7
1	Congress Heights	-76.997950	38.841077	Ward 8
2	Washington Highlands	-76.995636	38.830237	Ward 8
3	Bellevue	-77.009271	38.826952	Ward 8
4	Knox Hill/Buena Vista	-76.967660	38.853688	?
5	Shipley	-76.972902	38.846860	?
6	Douglass	-76.979645	38.847571	Ward 8
7	Woodland	-76.973926	38.856554	Ward 8
8	Garfield Heights	-76.972213	38.854085	Ward 8
9	Near Southeast	-77.000627	38.878384	?
10	Capitol Hill	-76.998477	38.883621	Ward 6
11	Dupont Park	-76.962818	38.872630	Ward 7
12	Twining	-76.960847	38.875588	Ward 7
13	Randle Highlands	-76.965804	38.869336	Ward 7
14	Fairlawn	-76.977452	38.869158	Ward 8
15	Penn Branch	-76.953424	38.869560	Ward 7
16	Barry Farm	-76.997281	38.859255	Ward 8
17	Historic Anacostia	-76.984678	38.863186	?
18	Columbia Heights	-77.030586	38.928008	Ward 1
19	Logan Circle/Shaw	-77.025340	38.909816	?

- 4) Many cells do not contain any data, so their rows have been deleted.
- 5) After the table have been cleaned, number of the total neighbourhoods have been displayed, by finding out the number of rows of the table.

```
In [8]: 1 WashingtonDC["Neighborhood"].count()
out[8]: 91
```

- 6) Data is visualized using a bar-chart and pie-chart



And this is a pie chart to show the information

So from these charts , we know that Ward 7 Councilmember: Vince Gray contains the most amount of Neighborhoods

And Ward 6 Councilmember: Charles Allen contains the least amount of Neighborhoods