2. Data

2.1 Python packages and Dependencies

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Matplotlib – Python Plotting Module
- Sklearn – Python machine learning Library
- Folium – Map rendering Library

2.2 Data sets

Data1: List of Montgomery County Zip codes

	ZIPCODE
0	20886
1	20895
2	20896
3	20899
4	20902

Source:https://www.zillow.com/browse/homes/md/montgomery-county/

This data set is used to determine the scope of our analysis.

Data 2: Maryland AGI by zip code

	ZIPCODE	Total Returns	Average_AGI
0	20854	22,390	252251.0
1	20815	14,380	239128.0
2	20817	16,420	229553.0
3	20816	7,470	227698.0
4	20818	890	179174.0

Source: https://datausa.io/profile/geo/maryland#economy

This data set reflects the power of consumption of each neighborhood.

Data 3: Maryland Census 2010 Data

	FID	ZCTA5CE10	FIRST_STAT	FIRST_GEOI	FIRST_CLAS	FIRST_MTFC	FIRST_FUNC	ZCTA5N	STATE	AREALAND	AREAWATR	POP100	HU100	NHW	N
0	1	20601	24	2420601	B5	G6350	S	20601	24	115635266	387684	24156	8722	9785	11
1	2	20602	24	2420602	B5	G6350	S	20602	24	35830723	352762	24955	9736	8466	13(
2	3	20603	24	2420603	B5	G6350	S	20603	24	44239637	219356	28967	10317	9625	15(
3	4	20606	24	2420606	B5	G6350	S	20606	24	7501011	1248760	431	230	377	
4	5	20607	24	2420607	B5	G6350	S	20607	24	54357590	448221	9802	3504	2165	6:

Source:

https://data.imap.maryland.gov/datasets/eb706b48117b43d482c63d02017fc3ff_1

This data set demonstrates the demographics of each neighborhood.

Data 4: Maryland Political Boundaries Json Data

	the_geom	OBJECTID_1	OBJECTID	AREA	PERIMETER	STATEZIP_2	STATEZIP_3	ZIPCODE1	ZIPCODE2	ZIPNAME	Shape_Leng	Shape
0	MULTIPOLYGON (((-79.3443788816394 39.657186418...	1	1	0	65617.799321	2	1	21531.0	2.402322e+09	Friendsville	65985.296521	1.71270:
1	MULTIPOLYGON (((-77.47019284905247 39.71844165...	24	24	0	3095.687185	25	24	21719.0	2.402122e+09	Cascade	3095.687185	3.67555(
2	MULTIPOLYGON (((-76.6490020078736 39.617757533...	25	25	0	41796.823787	26	25	21161.0	2.400521e+09	White Hall	41796.823787	5.88854:
3	MULTIPOLYGON (((-76.5221135638148 39.539045814...	73	73	0	18663.880331	74	73	21013.0	2.402521e+09	Baldwin	18663.880331	1.10893(
4	MULTIPOLYGON (((-76.48823506556019 39.42093878...	525	525	0	50123.651438	526	525	21236.0	2.400521e+09	Nottingham	50123.651438	2.28802

Source:

https://data.imap.maryland.gov/datasets/maryland-political-boundaries-zip-codes-11-digit/geoservice

This data set provides coordinates of each neighborhood for Foursquare API analysis.

2.3 Foursquare API

Foursquare API has a database of more than 105 million places. This project would use Foursquare API as its prime data gathering source. Due to request limitations, the number of venues per neighborhood parameter would reasonably be set to 1000, and the radius parameter would be set to 3000.

2.4 Preprocessing

2.3.1 Create a new empty data frame to combine the data derived from different sources
2.3.2 Fill out the new data frame one by one with the data sets mentioned above

2.3.3 Reset the index for a better view where is necessary

2.3.4 Convert data types to proper format where is necessary

2.3.5 Drop duplicate rows where is necessary

2.3.6 Drop rows with null values where is necessary

2.3.7 Data Normalization transforms values of several variables into a similar range

2.3.8 Get coordinates for each neighborhood

2.3.9 Visualize the data on map for a better understanding