CAPSTONE PROJECT – PART 1: Report

Introduction

Beyond401K Inc (Fictitious Company) is a Houston,TX based Financial Services start up company which has developed an AI/ML driven Retirement Portfolio Management Solution targeting millennials. Now they would like to do a beta launch to attract early adopters for their solution. The obvious choice for them is to launch it in New York City because it's the financial capital of United States, and also the average income of the workforce is much higher. Their plan is to set up a self serviced kiosks in prominent locations, so that they can attract professionals during their break time. Along with the kiosks they also want to have their employees to answer questions and to help the customers sign up.

Having said that, as a start up, the company is constrained on resources, so they have to be careful in spending, and at the same time get a maximum return on this initiative.Since New York City is a big city in terms of its size, population, and business activities it becomes difficult to identify the appropriate locations that are most frequented by the population. Also, the city's boroughs compounds the selection because of the cultural diversity each one brings.

The key objective of this project is to explore and analyze the neighborhoods of the city, and come up with a suggested list of venues on ther respective neighborhoods.
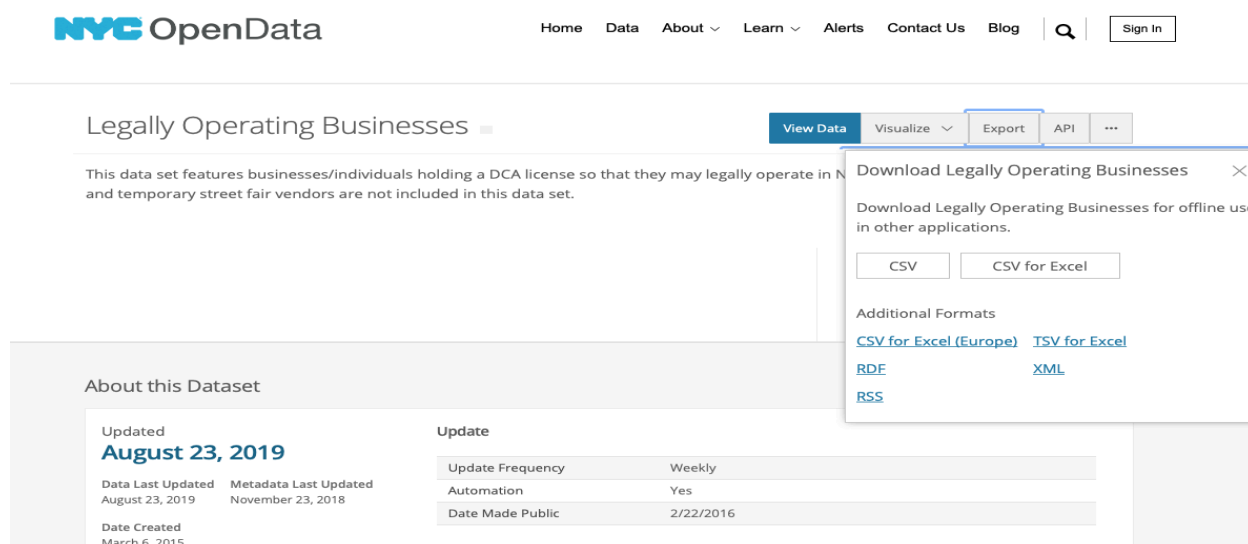
Datasets

To accompolish the stated objective, the following data requistes, and their sources are identified.

1) *Active Businesses in New York City*

The target customers for the yet to be launched Retirement Portfolio solution are the working professionals. For that data we need to get the database of the active businesses in the city, and it's available in NYC Open data portal,
url : (https://data.cityofnewyork.us/Business/Legally-Operating-Businesses/w7w3-xahh)

2) *GeoJSON of New York City*

To visualize the distribution of the businesses across the NYC boroughs, Choropleth maps could be utilized, for that GeoJSON file is required, and this available in NYC Open data portal in the below link,

url: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

3) *Demography Info New York City*

Apart from the businesses, to analyze the boroughs and it's population, demography data like borough's business volume, residents age, education, and income are required, which are available in the US Census portal.

url:https://www.census.gov/quickfacts/fact/table/newyorkcountymanhattanboroughnewyork,bronxcountybronxboroughnewyork,queenscountyqueensboroughnewyork,kingscountybrooklynboroughnewyork,richmondcountystatenislandboroughnewyork,newyorkcitynewyork/HSG010218

	New York County (Manhattan Borough), New York	Bronx County (Bronx Borough), New York	Queens County (Queens Borough), New York	Kings County (Brooklyn Borough), New York	Richmond County (Staten Island Borough), New York	New York city, New York
Housing units, July 1, 2018, (V2018)	886,249	532,487	865,878	1,053,767	181,199	
PEOPLE						
Population						
Population estimates, July 1, 2018, (V2018)	1,628,701	1,432,132	2,278,906	2,582,830	476,179	8,398,748
Population estimates base, April 1, 2010, (V2018)	1,586,360	1,384,603	2,230,578	2,504,717	468,730	8,174,988

4) *Borough Neighborhoods*

To perform k-means clustering, we need the details of the boroughs' neighborhoods, their geo coordinates. Those details can be accessed from New York universities open database.
url : https://geo.nyu.edu/catalog/nyu_2451_34572

5) *Foursquare API*

The key datapoint of this project is to identify the locations where the people frequently visits during anytime of the day. To find the places we need the details about the venues of the boroughs, and its available in Foursquare's developers portal. With its API, the below link could be accessed.
url:https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}