Introduction / Business Problem

Market research is one of the very important activities in the business planning of new and existing businesses. Whether a business seeks to launch a new product, expand on the existing range of products and services or to optimize current products and services, market research plays a crucial role in informing the decisions of management about the direction of a business.

In a study (*The Top 20 Reasons* 2019) published on the CBInsights website, 42% of business start-ups fail because there is simply no market need for their products or services.

An entrepreneur has recently moved to the city of New York and is looking to start a small business in an affluent neighbourhood. He does not have intimate knowledge of the city and has hired our data analytics company to analyse the city and make recommendations on the type and location of the start-up. He wants to start a business that is underrepresented in a specific neighbourhood in order to avoid market competition.

Methodology and Data Sources

This section outlines the method used and the sources of the data for the research.

Methodology

In order to recommend the type of business start-up in New York City; taking into consideration what is happening in all the neighbourhoods within all five boroughs in NYC, the following shall be done;

- A clustering, of all the neighbourhoods in New York City based on demographic data from the US Census conducted in 2010.
- An analysis of each cluster to identify the most popular types of businesses in each cluster using Foursquare data.
- Identify types of businesses that are not part of the cluster.
- Make recommendation on the type and location of the business start-up.

Data Sources

The following data is required for each one of the cities:

1. Geographical data – list of boroughs and neighbourhoods, neighbourhoods' latitude and longitude data.
    a. New York City - https://geo.nyu.edu/download/file/nyu-2451-34572-geojson.json

b. From the json file, a pandas data frame was created with the head of the data frame as shown in the figure below.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

2. 2010 New York City Census dataset
   a. The dataset will be mined to extract the following information for the year 2010;
      i. Persons per square mile
      ii. Proportion with less than high school diploma
      iii. Proportion with high school diploma and/or some college
      iv. Proportion with bachelor's degree or higher
      v. Proportion of households with income less than $15k
      vi. Proportion of households with income $15k - $30k
      vii. Proportion of households with income $30k - $50k
      viii. Proportion of households with income $50k - $100k
      ix. Proportion of households with income greater than $100k
      x. Proportion employed as management or professional
      xi. Neighbourhood average affluence
      xii. Neighbourhood average disadvantaged

```
nyc_census_df.head(32)
```

	index	STATEFP	COUNTYFP	TRACTCE	PUMA5CE	tract_fips10	totpop10	popden10	phispanic10	pnhwhite10	pnhblack10	p
88	43283	36	005	026701	03701	36005000100.0	9741.0	15968.850	0.339185	0.066831	0.562981	
89	43284	36	005	026702	03701	36005000200.0	4734.0	26897.730	0.658428	0.005281	0.261301	
90	43285	36	005	027300	03701	36005000400.0	5211.0	14803.980	0.664940	0.026290	0.256956	
91	43286	36	005	027700	03701	36005001600.0	5127.0	27417.110	0.652038	0.017944	0.324751	
92	43287	36	005	027900	03701	36005001900.0	2420.0	3811.024	0.591322	0.125620	0.257438	
93	43288	36	005	028100	03701	36005002000.0	8523.0	55705.880	0.655051	0.007040	0.306817	
94	43289	36	005	028300	03701	36005002300.0	4779.0	119475.000	0.725466	0.020297	0.254237	
95	43290	36	005	028500	03701	36005002500.0	4785.0	97653.060	0.788506	0.026959	0.158830	
96	43291	36	005	028700	03701	36005002701.0	3100.0	96874.990	0.601935	0.000000	0.398064	
97	43292	36	005	028900	03701	36005002702.0	4703.0	77098.360	0.755475	0.003827	0.228152	

3. Venues data – list of venues and business type for each borough

   Using Foursquare, New York City venues data was collected

```
# type your answer here

LIMIT = 100 # limit of number of venues returned by Foursquare API

radius = 500 # define radius

# create URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    neighborhood_latitude,
    neighborhood_longitude,
    radius,
    LIMIT)
url # display URL
```

References

1. *The Top 20 Reasons Startups Fail 2019*, 26 April 2020, https://www.cbinsights.com/research/startup-failure-reasons-top/