Using Data Science to Indentify Business Opportunities in New York

- Introduction

Using Data Science techniques, I want to see if I can analyze New York City data to find out which type of businesses are most popular i.e have the most foot traffic. The metric I will use to gauge this is the number of check-ins each type of business gets.

Using this data I will explore areas where the most popular type of business is lacking, which would mean opening up such a business in these areas could prove to be profitable.

- Target Audience

The target audience is mainly small and medium business owners who would want to open new branches if their business is that which would fill the need gap of the locations, or people who have been wanting to start a business but haven't decided the type and location of the business yet.

- The Data

The data comes from Dingqi Yang from the following link:
https://sites.google.com/site/yangdingqi/home/foursquare-dataset.

It contains 227,428 check-ins in New York City. The data contains a file in tsv format. Each file contains 8 columns, which are:

- User ID (anonymized)

- Venue ID (Foursquare)

- Venue category ID (Foursquare)

- Venue category name (Foursquare)

- Latitude

- Longitude

- Time zone offset in minutes (The offset in minutes between when this check-in occurred and the same time in UTC)

- UTC time

After loading the data into python, I converted it to a pandas dataframe. The resulting dataset with the necessary columns looked like this:

	VenueID	CategoryName	Visitor Count	Latitude	Longitude
0	49bbd6c0f964a520f4531fe3	Arts & Crafts Store	7	40.719810375488535	-74.00258103213994
1	4a43c0aef964a520c6a61fe3	Bridge	37	40.60679958140643	-74.04416981025437
2	4c5cc7b485a1e21e00d35711	Home (private)	1	40.716161684843215	-73.88307005845945
3	4bc7086715a7ef3bef9878da	Medical Center	1	40.7451638	-73.982518775
4	4cf2c5321d18a143951b5cec	Food Truck	4	40.74010382743943	-73.98965835571289

The Analysis

I then created a dictionary to figure out which category is the most popular, and found that the category 'Bar' is the most visited.

After this, I needed to see which areas in New York City are lacking in bars. To do this, I checked the coordinates to count how many bars there are within a certain distance. I found that the neighborhoods Turtle Bay and Jamai had the least number of bars within 4 kilometers.

Results & Conclusion

Based on 2,000 venues and 227,428 check-ins for businesses in New York City, it was found that Bars are the most popular businesses in terms of foot-traffic. Further analysis revealed that the neighborhoods Turtle Bay and Jamai had the least bars in their vicinity.

For someone looking to start a new business or expand their current one, opening up a bar in Turtle Bay and/or Jamai could prove to be a profitable venture.