**NYC’s next coffee shop: Exploring the city’s caffeine-friendly neighborhoods**

1. **Introduction:**
   1. **Background**

New York City, one of the most vibrant cities in the world, has no shortage of coffee shops. The city is a caffeine lover’s dream and is almost brimming with baristas, cafes and coffee sellers and New Yorkers are spoilt for choices.

Setting up a coffee shop in the city is therefore no mean feat. You can be lest assured that there will be competition – a lot! The city has also one of the highest property rates in the country making it a cost intensive affair for someone who wants to rent or buy a place for business or otherwise. Added to that, if you are venturing into something quite as common as a coffee shop – it can be challenging!

* 1. **Problem:**

Location plays a key factor in the success of an enterprise – more so while catering to a refined consumer base in a city as NYC. A wide range of attributes affect the location felicity – proximity to probable customers, concentration of competitors, operational costs, and connectivity.

This project aims to advocate the neighborhood/s for opening a coffee shop in the Manhattan borough of New York City that are most optimal for business profitability and operational effectiveness.

* 1. **Stakeholders:**

The entities who will profit from the solution implemented from the project are: -

* **Entrepreneurs:** The solution is intended to recommend the most business conducive location to someone who intends to set up a coffee shop in New York City. It will enable the prospective coffee shop owner to narrow down the neighborhood choices.
* **Venture Capitalists:** Financers, banks, private equities who are financing the venture, can leverage the solution to suggest the most profitable locations and ensure profit maximization of their investment aided by data driven decision making.
* **Caffeine lovers:** Coffee enthusiasts can research about the next hip coffee destination in the city by exploring the suggestions recommended by the tool.

1. **Data sets and Data Acquisition**

The data of New York City’s neighborhoods were collected from NYC Open Data (Link: <https://data.cityofnewyork.us/City-Government/Neighborhood-Names-GIS/99bc-9p23>)

The records in the data set are in the following format: -

A screenshot of a cell phone

Description automatically generated

The data was loaded in a data-frame called data.   
The column ***the\_geom*** was processed top extract the latitude and longitude into two different columns - ***latitude*** and ***longitude***.

The data was loaded into a different data fame called manhattan\_data.

The columns ***the\_geom, OBJECTID, Stacked, AnnoLine1, AnnoLine2, AnnoLine3*** and ***AnnoAngle*** were dropped.   
The resulting data frame is in the following format.

A screenshot of a social media post

Description automatically generated

Using the **geopy** library, the latitude and longitude of New York City were extracted.

**A screenshot of a cell phone

Description automatically generated**

The latitude and longitude of New York City was found to be **40.7896239** and **-73.9598939.**

To determine which neighborhoods in Manhattan would be most conducive for setting up a coffee shop, a comparative analysis needs to be done for each neighborhood.

The two most critical factors affecting this decision would be demand and availability.

Most coffee drinkers tend to be students who attend schools, colleges and universities and working professionals. As a result, information about the number of schools, colleges, universities, and offices in each neighborhood would provide an insight into the potential customers in each neighborhood.

The number of coffee shops in each neighborhood would help to understand the distribution of competitors in each neighborhood.

For collecting the data about the coffee shops, cafes, offices, schools and universities in each neighborhood, **FourSquare** venue API was used. **FourSquare** is an independent location data and technology platform that provides location related data to various businesses.

Using the FourSquare venues API provided by FourSquare (<https://api.foursquare.com/v2/venues/explore>) we retrieved the venues within a radius of 500 meters of each neighborhood in a dataframe called ***manhattan\_venues*** in the following format.

A screenshot of a cell phone

Description automatically generated

The dataframe has an exhaustive list of venues for each neighborhood. However, the venues that ae relevant to the business problem are the ones that correspond to **coffee shops, cafes**, and places where the potential customers – **students, professionals**, can be found.

By analyzing the **Venue Category** column, the relevant categories were extracted programmatically. This was done as follows: -

1. A list of sample categories was formulated that contains relevant words like ***‘school’, ‘college’, ‘coffee’*** etc.
2. A list was created from the **Venue Category** column of the data frame.
3. The venue categories which contained one or more words in the sample categories were filtered out.
4. From this list of filtered categories, the list of ***relevant\_categor*ies** was formulated.

The list of ***relevant\_categories*** is as follows: -

*['Office', 'College Bookstore', 'Cafeteria', 'School', 'Coffee Shop', 'Café', 'College Arts Building', 'College Theater', 'Music School', 'Tea Room', 'Cooking School','College Academic Building', 'College Cafeteria', 'College Cafeteria', 'High School', 'Bubble Tea Shop']*

A close up of a newspaper

Description automatically generatedA screenshot of a cell phone

Description automatically generated

The rows from the data-frame ***manhattan\_venues*** that had a Venue Category not relevant to the business problem were filtered out.

The resultant data-frame is as follows: -

A screenshot of a cell phone

Description automatically generated

This data-frame is to be used for further analysis and decision making.