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

1. Introduction:

1.1. 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!

1.2. 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.

1.3. 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.

2. 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: -

```
In [1]: import pandas as pd

In [2]: data = pd.read_csv("NHoodNameCentroids.csv")

In [3]: data.head()
```

Out[3]:

	the_geom	OBJECTID	Name	Stacked	AnnoLine1	AnnoLine2	AnnoLine3	AnnoAngle	Borough
0	POINT (-73.8472005205491 40.89470517661004)	1	Wakefield	1	Wakefield	NaN	NaN	0	Bronx
1	POINT (-73.82993910812405 40.87429419303015)	2	Co-op City	2	Co-op	City	NaN	0	Bronx
2	POINT (-73.82780644716419 40.88755567735082)	3	Eastchester	1	Eastchester	NaN	NaN	0	Bronx
3	POINT (-73.90564259591689 40.895437426903875)	4	Fieldston	1	Fieldston	NaN	NaN	0	Bronx
4	POINT (-73.91258546108577 40.89083449389134)	5	Riverdale	1	Riverdale	NaN	NaN	0	Bronx

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.

```
In [62]: manhattan_data = data[data['Borough'] == 'Manhattan'].reset_index(drop=True)
         drop_columns = ['the_geom', 'OBJECTID', 'Stacked', 'AnnoLine1', 'AnnoLine2', 'AnnoLine3', 'AnnoAngle']
         for col in drop_columns:
             del manhattan_data[col]

In [63]: manhattan_data.head()
```

Out[63]:

	Name	Borough	latitude	longitude
0	Marble Hill	Manhattan	40.87655077879968	-73.91065965862988
1	Chinatown	Manhattan	40.715618422314364	-73.99427936255985
2	Washington Heights	Manhattan	40.85190252555309	-73.93690027985242
3	Inwood	Manhattan	40.86768396449919	-73.92121042203905
4	Hamilton Heights	Manhattan	40.82360428481197	-73.94968791883373

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

```
In [55]: from geopy.geocoders import Nominatim

In [56]: address = 'Manhattan, NY'
         geolocator = Nominatim(user_agent="ny_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         print('The geograpical coordinate of Manhattan are {}, {}.'.format(latitude, longitude))

         The geograpical coordinate of Manhattan are 40.7896239, -73.9598939.
```

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.

```
In [28]: manhattan_venues.head()
```

Out[28]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.87655077879968	-73.91065965862988	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.87655077879968	-73.91065965862988	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.87655077879968	-73.91065965862988	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.87655077879968	-73.91065965862988	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.87655077879968	-73.91065965862988	Dunkin'	40.877136	-73.906666	Donut Shop

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_categories*** 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']
```

```
In [38]: # Examiing the venue category
venue_categories = list(set(manhattan_venues['Venue Category']))
print(venue_categories)
```

```
['Resort', 'Rest Area', 'Soup Place', 'Shipping Store', 'English Restaurant', 'Fast Food Restaurant', 'Moving Targe
t', 'Israeli Restaurant', 'Vegetarian / Vegan Restaurant', 'Bus Line', 'Food Truck', 'Bike Shop', 'Trail', 'Ski Sho
p', 'Bike Trail', 'Medical Center', 'Nightclub', 'Music Venue', 'Skating Rink', 'Fountain', 'Wine Bar', 'Boat or Ferr
y', 'African Restaurant', 'Used Bookstore', 'Dumpling Restaurant', 'Garden Center', 'Bar', 'Non-Profit', 'Health Food
Store', 'Rock Club', 'Health & Beauty Service', 'Coworking Space', 'Outdoors & Recreation', 'Bank', 'Burger Joint',
'Arcade', 'Jewish Restaurant', 'Vietnamese Restaurant', 'Breakfast Spot', 'Food & Drink Shop', 'Beer Store', 'Golf Co
urse', 'Yoga Studio', 'Italian Restaurant', 'Playground', 'Miscellaneous Shop', 'Antique Shop', 'Gymnastics Gym', 'Bi
ke Rental / Bike Share', 'Grocery Store', 'Movie Theater', 'Hobby Shop', 'Dry Cleaner', 'Sculpture Garden', 'Office',
'Ramen Restaurant', 'Rock Climbing Spot', 'Soccer Field', 'Southern / Soul Food Restaurant', 'Japanese Restaurant',
'Tourist Information Center', 'Big Box Store', 'Gastropub', 'BBQ Joint', 'Karaoke Bar', 'Shoe Store', 'Brazilian Rest
aurant', 'Video Store', 'Theater', 'Farmers Market', 'Indie Movie Theater', 'German Restaurant', 'Asian Restaurant',
'Park', 'Himalayan Restaurant', 'American Restaurant', 'Creperie', 'Caribbean Restaurant', 'Supplement Shop', 'Pet St
ore', 'Kosher Restaurant', 'Sushi Restaurant', 'Ethiopian Restaurant', 'Persian Restaurant', 'College Bookstore', 'La
tin American Restaurant', 'Roof Deck', 'Food Court', 'Boutique', 'Donut Shop', 'Department Store', 'Jazz Club', 'Dim
Sum Restaurant', 'Sporting Goods Shop', 'Hotel', "Women's Store", 'Cafeteria', 'Liquor Store', 'Hot Dog Joint', 'Cult
ural Center', 'Clothing Store', 'Flower Shop', 'Cambodian Restaurant', 'Tree', 'Opera House', 'Christmas Market', 'Ba
by Store', 'Pool', 'Mexican Restaurant', 'School', 'Memorial Site', 'Public Art', 'Gift Shop', 'Scandinavian Restaura
nt', 'Spa', 'Kids Store', 'Weight Loss Center', 'Seafood Restaurant', 'Sports Club', 'Australian Restaurant', 'Soba R
estaurant', 'Pizza Place', 'Discount Store', 'Optical Shop', 'Hostel', 'Garden', 'Venezuelan Restaurant', 'Veterinari
an', 'Laundry Service', 'Czech Restaurant', 'Afghan Restaurant', 'Pie Shop', "Doctor's Office", 'Gaming Cafe', 'India
n Restaurant', 'Bistro', 'Athletics & Sports', 'Chocolate Shop', 'Leather Goods Store', 'Toy / Game Store', 'Lingerie
Store', 'Fried Chicken Joint', 'Volleyball Court', 'Plaza', 'Fish Market', 'Tailor Shop', 'Tiki Bar', 'Cosmetics Sho
p', 'Library', 'Moroccan Restaurant', 'Sake Bar', 'Spiritual Center', 'Juice Bar', 'Beer Garden', 'Flea Market', 'Dan
ce Studio', 'Lounge', 'Arts & Crafts Store', 'Hotel Bar', 'Design Studio', 'Gym / Fitness Center', 'Street Art', 'Con
cert Hall', 'Tapas Restaurant', 'Social Club', 'Bridal Shop', 'Kebab Restaurant', 'Convenience Store', 'Pub', 'Duty-f
ree Shop', 'Salad Place', 'Hardware Store', 'Strip Club', 'Photography Studio', 'Theme Park Ride / Attraction', 'Dine
r', 'Thrift / Vintage Store', 'Gay Bar', 'Basketball Court', 'Szechuan Restaurant', 'Board Shop', 'Bed & Breakfast',
'Financial or Legal Service', 'Tennis Court', 'Market', 'Mobile Phone Shop', 'Theme Restaurant', 'Shanghai Restauran
t', 'Speakeasy', 'Food Stand', 'Irish Pub', 'Taiwanese Restaurant', 'Metro Station', 'Deli / Bodega', 'Heliport', 'Do
g Run', 'Mediterranean Restaurant', 'Club House', 'Paella Restaurant', 'Cheese Shop', 'Skate Park', 'Bagel Shop', 'Ba
seball Field', 'Record Shop', 'Monument / Landmark', 'Sandwich Place', 'Wings Joint', 'Harbor / Marina', 'Bus Statio
n', 'Hotpot Restaurant', 'Udon Restaurant', 'Spanish Restaurant', 'Auditorium', 'Noodle House', 'Salon / Barbershop',
'Historic Site', 'Austrian Restaurant', 'Accessories Store', 'Dessert Shop', 'Whisky Bar', 'Filipino Restaurant', 'Gr
eek Restaurant', 'Wine Shop', 'History Museum', 'Nail Salon', 'Gas Station', 'New American Restaurant', 'Climbing Gy
m', 'Beer Bar', 'Snack Place', 'Smoothie Shop', 'Steakhouse', 'Paper / Office Supplies Store', 'Coffee Shop', 'South
Indian Restaurant', 'Chinese Restaurant', 'Museum', 'Caucasian Restaurant', 'Russian Restaurant', 'Butcher', 'Shoppin
g Mall', 'Furniture / Home Store', 'Café', 'Modern European Restaurant', 'Restaurant', 'Pet Service', 'College Arts B
uilding', 'College Theater', 'Burrito Place', 'Bookstore', 'Korean Restaurant', 'Music School', 'Kitchen Supply Stor
e', 'Poke Place', 'Malay Restaurant', 'Cuban Restaurant', 'Massage Studio', "Men's Store", 'Falafel Restaurant', 'Art
Museum', 'Physical Therapist', 'Outdoor Supply Store', 'General Entertainment', 'French Restaurant', 'Cupcake Shop',
'Rental Car Location', 'Scenic Lookout', 'Comedy Club', 'Pharmacy', 'Waterfront', 'Smoke Shop', 'Bakery', 'Japanese C
urry Restaurant', 'Pet Café', 'Tea Room', 'Gym Pool', 'Candy Store', 'Dive Bar', 'Video Game Store', 'Hawaiian Restau
rant', 'Jewelry Store', 'Circus', 'Community Center', 'Cooking School', 'Empanada Restaurant', 'Indie Theater', 'Leba
nese Restaurant', 'Outdoor Sculpture', 'Frozen Yogurt Shop', 'Performing Arts Venue', 'Exhibit', 'College Academic Bu
ilding', 'Argentinian Restaurant', 'Daycare', 'Sports Bar', 'College Cafeteria', 'Residential Building (Apartment / C
ondo)', 'Organic Grocery', 'Supermarket', 'Boxing Gym', 'Adult Boutique', 'Drugstore', 'Middle Eastern Restaurant',
'Pilates Studio', 'Art Gallery', 'Gym', 'Ice Cream Shop', 'Peruvian Restaurant', 'Cocktail Bar', 'Gourmet Shop', 'Are
pa Restaurant', 'High School', 'Cycle Studio', 'Taco Place', 'Swiss Restaurant', 'Molecular Gastronomy Restaurant',
'Tennis Stadium', 'Recreation Center', 'Turkish Restaurant', 'Martial Arts Dojo', 'Building', 'Piano Bar', 'Thai Rest
aurant', 'Tattoo Parlor', 'Newsstand', 'Eastern European Restaurant', 'Hookah Bar', 'Event Space', 'Electronics Stor
e', 'Bubble Tea Shop', 'Bridge']
```

```
sample_categories = ['school', 'college', 'university', 'tea', 'coffee', 'office', 'center', 'cafeteria', 'café']
filtered_categories = []
for venue in venue_categories:
    venue_lower = venue.lower()
    for sample_venue in sample_categories:
        if sample_venue in venue_lower:
            filtered_categories.append(venue)
print(filtered_categories)
```

```
In [42]: #Filerring out the relevant categories
relevant_categories = ['Office', 'College Bookstore', 'Cafeteria', 'School', 'Coffee Shop', 'Café', 'College Arts Build
```

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: -

```
In [18]: manhattan_venues = manhattan_venues[manhattan_venues['Venue Category'].isin(relevant_categories)]
```

```
In [19]: manhattan_venues.head()
```

Out[19]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	Marble Hill	40.87655077879968	-73.91065965862988	Starbucks	40.877531	-73.905582	Coffee Shop
9	Marble Hill	40.87655077879968	-73.91065965862988	Starbucks	40.873755	-73.908613	Coffee Shop
19	Marble Hill	40.87655077879968	-73.91065965862988	Starbucks	40.873234	-73.908730	Coffee Shop
44	Chinatown	40.715618422314364	-73.99427936255985	Yaya Tea Garden	40.716177	-73.994902	Tea Room
49	Chinatown	40.715618422314364	-73.99427936255985	Teado	40.717184	-73.994883	Bubble Tea Shop

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