1. **Introduction**
   1. **Problem**

For this project, my business problem centered around the scenario where a contractor is looking to open a business in the state of New York. They want to ensure the success of their business by seeking out location where there is a high number of local residents with a fixed income under a popular venue in the area. With more residents working with a fixed income, the contractor hopes it will lower the financial risk in meeting product demand.

* 1. **Interest**

This type of information is beneficial to startups, entrepreneurs, and small businesses who aim to expand or promote their services to similar areas.

1. **Data**

The NYC Open Data offers public data provided by New York groups and agencies. This capstone project utilizes the ‘NYCHA Development Data Book’ dataset for several of its features. It contains information on New York’s different boroughs such as income and clustering in an area. While the dataset does have a ‘density’ feature which would help in understanding the former, its definition makes it irrelevant to the analysis of this data.

The contractor wants to know more about the citizens in the area rather than knowing more about the area itself. The features regarding fixed income for residents are important as this type of information can help guarantee the amount of profits gain in an area.

1. **Methodology**
   1. **Data Processing**

We begin the project by loading the data and selecting the relevant features for the business problem. Therefore, we collect features regarding fixed income, boroughs, and address. The data is then summarized and compared among the different boroughs of the state of New York.

* 1. **Obtain coordinates**

In order to visualize the locations from the dataset, we must convert the street addresses into coordinates. Using the geocode library, we can convert the street addresses from the dataset into latitude and longitude coordinates. This processing also needed additional modifications to the naming conventions of some streets as the method would return coordinates in a different location. For example, entering simply ‘CAULDWELL AVE’ returned coordinates for a location in Europe.

* 1. **Cluster location**

To help visualize the locations, we cluster the different locations based on the different boroughs. This makes it easier to visualize and compare the location data on a map.

* 1. **Venue data**

Using Foursquare, we can obtain venue data using the coordinate values obtained earlier. With the venue data, we can compare different venues in each borough to gain insight on which venues are popular in the area.

1. **Results**

With the cleaned dataset, we can then group all the rows based on borough. We can observe that the Manhattan and Brooklyn boroughs have a similar percentage and total number of individuals with a fixed income. Furthermore, a cluster map to display the different locations in relation to the borough shows Manhattan and Brooklyn having a densely packed number of locations. However, the Manhattan borough has a larger number of locations albeit in a larger radius centered on the borough.

1. **Discussion**

While this data was useful in this business problem, additional data features could have been a factor in this analysis. Information such as shipment availability for inventory or popular hours customers would arrive to a venue are useful to determine the best location to start a business. Another factor to consider is what kind of business the contractor plans to open. Depending on the type of business, targeted information related to that kind of venue could also provide some insight on which location to start in. Additional information regarding the income of residents would provide some insight of the area as well. While the features regarding residents’ fixed income guarantees the amount of profit gained to minimalize risk, more information to generate a risk report would also help the contractor in understanding the location.

1. **Conclusion**

Based on the map and grouped data, we can infer that either the Manhattan or Brooklyn borough would be a good location to start a business. Both boroughs have a similar number of residents with a fixed income. Also, locations obtain from the dataset show that the locations are very clustered near each other which implies that these residents are fairly clustered as well. The venue data implies that those centered around food, such as cafes, restaurants, and bars, are fairly popular in the Manhattan and Brooklyn boroughs.