**Group 4 Project**

**Jiaming Zhu(71345111);**

**Jingwen Ni(63745156);**

**Sara Shen(61832733);**

**Project Overview:**

The topic of this research paper is to argue the reason that housing is so unaffordable in New York. Many factors may influence the housing price in New York, but in this research, we only consider the two major factors, which are the household income and the population density. Thereby, the dependent variable is the Housing Price in New York and the two independent variables are Household Incomes and Population density in New York.

Due to the global pandemic, housing supplies nationwide were being influenced a lot. The pandemic-induced drop only exacerbated that. There are a lot of people who live in New York, and, as we are known, the housing supply in New York City is low, thereby the market tends to favor homesellers over homebuyers. New Yorkers have relatively fewer choices when supply is low, especially during this pandemic. This situation reveals that more people are looking for fewer houses, which drives prices higher. Based on the data from ***New York City Real Estate Market Overview 2020,*** “The median list price of homes in New York, NY was $799K in August 2020… The median listing price per square foot was $562. The median sale price was $665K. On average, homes in New York, NY sell after 182 days on the market.” The price in New York is sufficiently higher than other countries. Therefore, in this project, we want to discuss the relationship between housing price, average income, and population density. High housing prices may result in outflowing of talent. Also people who work and study in NewYork have to live in other cities like NJ or Long Island which will cost them a lot of time in commuting every day. What’s more, people will be evicted if they lose their job suddenly which will cause heavy mental pressure. People would be unsatisfied and thus decrease social stability.

**Project structure:**

The project will provide a data report first. This data report contains data sources and data access. The second part provides summary statistics about median household income, population density and housing price trends. We will plot this relation and visualize it in this project. The third part contains our hypothesis and a comparison between housing price-population density relations and housing price-household income relation. The final part of our project will consist of a linear regression and a discussion of project results: whether this result is consistent with our hypothesis or not.

**Data Report:**

The data of this project comes from various sources. The data source of independent variable Median Household Income in New York from 1990 to 2019 is from the US census Bureau. This can be accessed at [https://www.statista.com/](https://www.statista.com/statistics/205974/median-household-income-in-new-york/). This data is updated annually. The data source of independent variable population density in New York from 1960 to 2018 can be accessed from [https://www.statista.com/](https://www.statista.com/statistics/304695/new-york-population-density/). This data is updated annually. The data source of dependent variable New York city’s housing price is available at [https://fred.stlouisfed.org](https://fred.stlouisfed.org/). This data is updated annually.

**Key Variables:**

The first **independent data** behind this project is the **Median Household Income** in New York from 1990 to 2019, and it is from the US census Bureau. This research topic is to talk about why housing is unaffordable in the metropolis area, and we use New York as an example, thereby we focus on the data of New York. The Median Household Income in New York gives us specific information about the income for people living in New York. The regional data provides access to measures of economic activity in New York and a long time series from 1990 to 2019. This data was measured annually. In order to calculate the household income, we need to add each person’s income that lives in the home who is 15 years old or older, regardless of whether they are relatives or not. Plus, the household income is calculated before taxation. Therefore, we speculate that median household income may have a positive relation with our dependent variable, the New York housing price.

The second **independent variable** is the **population density** in New York from 1960 to 2018, and it is from the US census Bureau. The population density in New York can show us how crowded New York is, and we can speculate that New York has a large population base. Thereby, it is reasonable to assume that the housing market has a large amount of buyers and demand in New York. Population Density measures how many people live in a certain area. Population density is calculated by dividing the number of people by the area, in our data, the unit is the number of people per square kilometre.

Thirdly, the **dependent variable** I have is **New York city’s housing price**. The data I found from Fred is the Condo Price Index in New York City from 1996 to 2020 *(https://fred.stlouisfed.org/series/NYXRCSA)* which is representative of New York housing prices because most people in New York are living in condos. By comparing the Median Household Income in New York, we can figure out the reason why most people feel housing in New York City is too expensive to afford. Also from the second independent variable: The Population Density, we can know why New York Housing price is higher than other small cities. It is not hard to understand because the population density in metropolis especially in New York City is much higher than others. When the demand for housing is increasing, the price will go up. Hence the dependent variable: “Housing Price” has a positive relation with both Median Household Income and Population Density.