Term Project 1 - Milestone 1:

Proposal

New York Home Price Analysis

DSC 680

Prof. Amirfarrokh Iranitalab

Carlos Cano

Milestone 1 – Proposal

- Topic – Describe and name your project in 1-2 sentences max.

This project will focus on the relationship between prices and square footage within New York Housing Market as well as the relationship therein among other variables as a price indicator/ predictor.

- Business Problem – Describe the business problem your project is trying to solve and/or the research questions you will explore.

This research will be able to be used to predict the cost of a potential home sale based on the provided data. This could aid in assisting sellers understand the market and assist in facilitating an opening asking price. This in turn will be used to give the seller a realistic expectation for their real estate. Furthermore, this could help market price stabilization as sellers wouldn't keep their homes on the market and continuously decrease the price and inflate the market expected purchase price.

- Datasets – Where are you getting your data? Describe the data that you will use to solve the problem.

The data is provided by Kaggle, which provides free datasets for Data Science uses. The data contained in the link below consists of 17 variables related to each property and data point. From this a broad understanding of data is captured which allows for an apples-to-apples approach. Price, Number of baths, Number of beds, as well as Square Footage are the more commonly understood variables and will be showcased primarily. Subsequent variables will aid in the model create, while other may be removed for redundancy or posterity but justified with explanation as such.

https://www.kaggle.com/datasets/nelgiriyewithana/new-york-housing-market

- Methods – What analysis method will you use to complete this project? Note: this is just a proposal, your project can adapt as you work on it.

This analysis will implore regression analysis to calculate the relationship causality, a brief exploration maybe expanded to showcase the linear relation as well as the logistic approach among different variables.

- Ethical Considerations – What are some potential ethical concerns of this topic or analyzing the data?

Data manipulation is always of ethical concerns in Data Science applications: over processing, over deletion, and creation of unneeded dummy variables. Also of special note is data origin collection, as this can introduce initial biases in the data. Project bias is also an ethical concern as it can create tunnel vision in applying various Data Science techniques and processes.

- Challenges/Issues – What are some issues and challenges do you think you might face?

The initial challenge will be sourcing a dataset that will be of EDA merit that can be used for analysis purposes that uses the culmination of knowledge garnered through this degree and previous experience. Additionally, interpretation and analysis of data implementation for published work is of concern.

- References – What sources will you use to validate your results and support your project topic?

The below links will be used to verify the efficacy of this model analysis. By testing the relationship among the variables, it can be used to test the price relationship among the price and square footage.

1. https://www.elikarealestate.com/square-foot-calculator/

2. https://www.zillow.com/home-values/6181/new-york-ny/