**STAT 515: Final Project - Progress Report**

Due before April 23rd , 2025, in-person during office hours. **Hard Copy** must be submitted.

|  |  |
| --- | --- |
| **Section Number:** | 005 |
| **Group Number:** | 5 |
| **Names of the Members:**  **(As stated on Canvas)** | Phong Danh Nguyen  Cheng-hsun Hsu  Abdullah Yusaf Khan |

|  |  |  |
| --- | --- | --- |
| **Title of the Project:** | Predicting Real Estate Sales Price in the state of Connecticut | |
| **Describe the data set(s) that were considered:**  **(Provide Background information, including the source and the variables)** | This data set contains all real estate sales in Connecticut from 2001 to 2022 with a value over $2000. The dataset was discovered on <https://catalog.data.gov/dataset/real-estate-sales-2001-2018>, in which the data set originated from the Connecticut state website at <https://data.ct.gov/Housing-and-Development/Real-Estate-Sales-2001-2022-GL/5mzw-sjtu/about_data>. The data was provided in csv format, with variables:   |  |  |  | | --- | --- | --- | | Variable name | Description | Data Type | | Serial Number | Serial number | Int | | List Year | Year the property was listed for sale | Int | | Date Recorded | Date the sale was recorded locally | Chr | | Town | Town name | Chr | | Address | Address | Chr | | Assessed Value | Value of the property used for local tax assessment | Num | | Sale Amount | Amount the property was sold for | Num | | Sales Ratio | Ratio of the sale price to the assessed value | Num | | Property Type | Type of property including: Residential, Commercial, Industrial, Apartments, Vacant, etc | Chr | | Residential Type | Indicates whether property is single or multifamily residential | Chr | | Non Use Code | Non usable sale code typically means the sale price is not reliable for use in the determination of a property value. See attachments in the dataset description page for a listing of codes | Int / Chr | | Assessor Remark | Remarks from the assessor | Chr | | OPM remarks | Remarks from OPM | Chr | | Location | Lat / lon coordinates | Lat,lon | | |
| **Explain the reasons why the above data set(s) were selected:** |  | |
| **Research Questions:** |  | |
| **Explain the rational behind each question and how questions relate to the data set:** |  | |
| **Provide an overview of the statistical methods that will be used to answer the research questions:** |  |
| **Explain the rationale behind the methods used:** |  |
| **Any Challenges or issues faced/facing:** |  |
| **Comments (If any):** |  |