Landis: Data Science Exercise (Mecklenburg County)

**Pallavi Varandani**

To complete this exercise, I followed the following steps:

1. Created a scraper using Python and Selenium Web Crawler
   1. This scraper stores the extracted data into 2 locations:
      1. Local disk – CSV format
      2. SQL Database – localhost
         1. Pre-Step – Created the database and blank table before loading the data using python
         2. Further, in future we can split the table into various small tables and create a schema to have a better structure.
2. Create a R Shiny Dashboard which pull the data from the SQL database. The R Code does the following functions:
   1. Pulls the data from the database
   2. Cleans and Transforms the data
   3. Generates Sample Graphs
   4. Displays Sample Insights

Storage Choice:

For this exercise I choose localhost SQL database, because It is a small dataset to work with and can be easily handled on a localhost for an exercise. For the purpose of Schema, I just one table as a demonstration, further, we can split the table into multiple tables using Parcel ID as primary key.

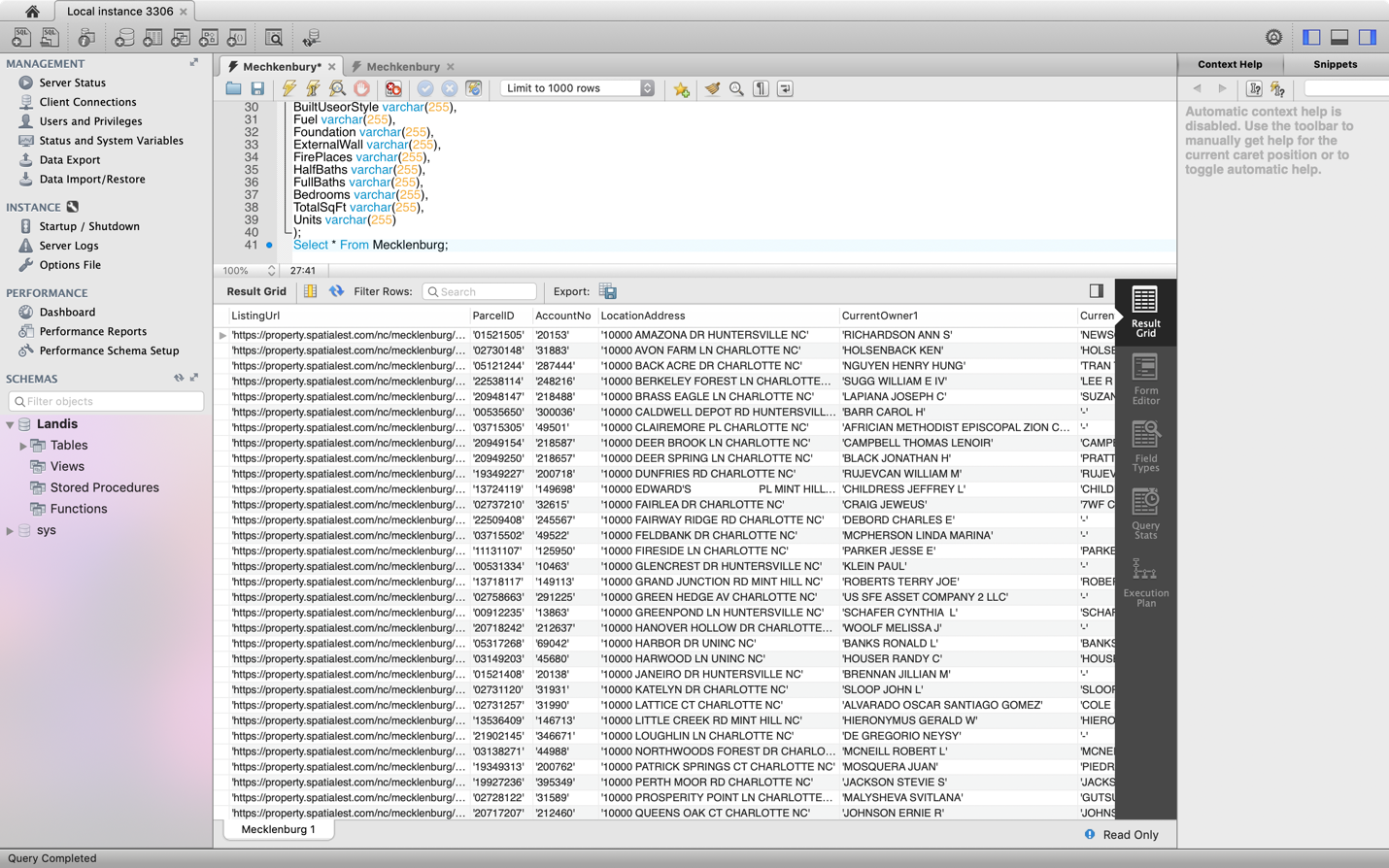
Insights:

From analyzing the data of Mecklenburg County, it can be said that sale price of the property depends a lot on the heat type and the fuel used. Further, it also depends on the property foundation and Bedrooms. There is a strong positive relation of Sale Price with Business Value and Total Appraisal Value, Following are some of the insights:

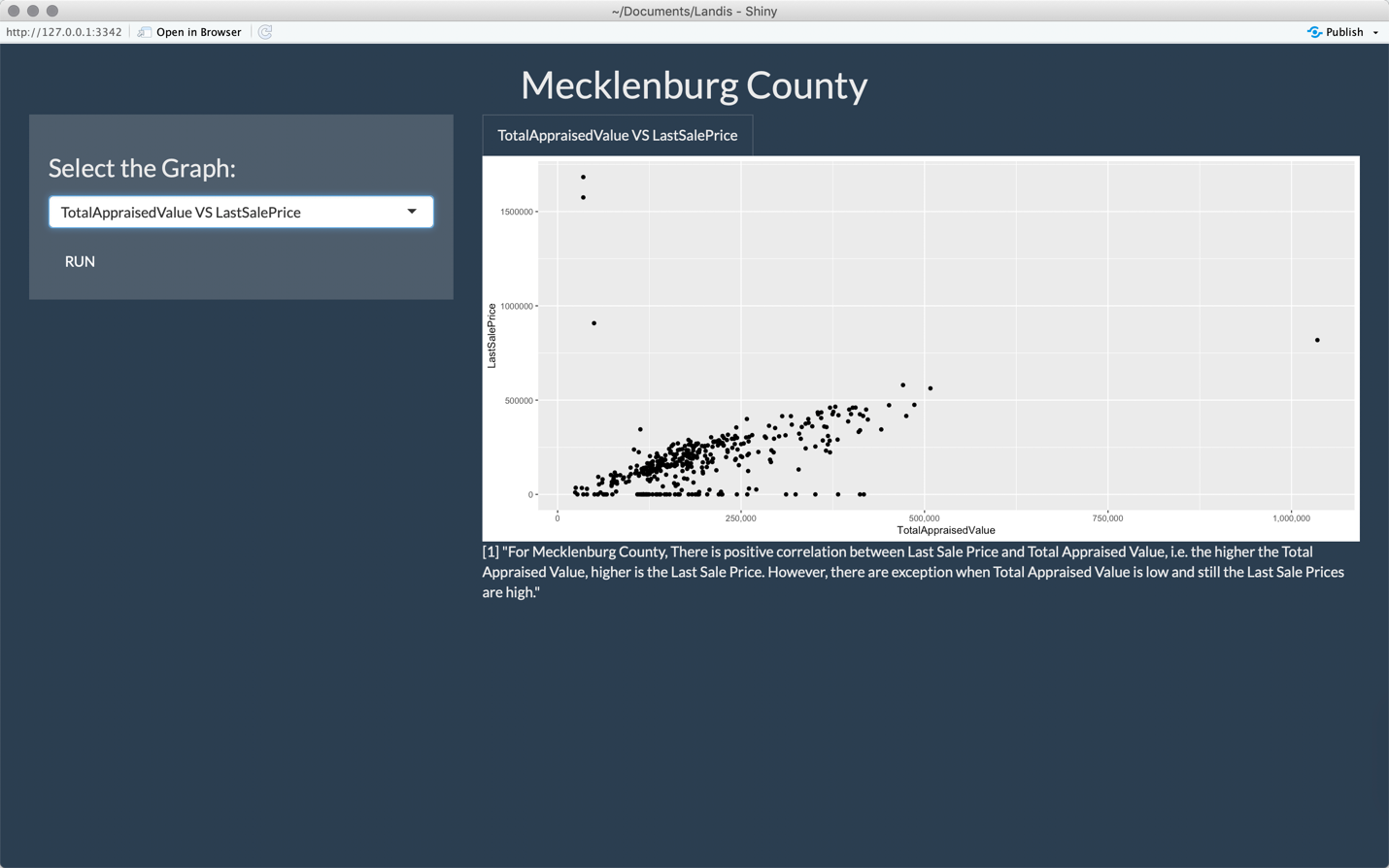
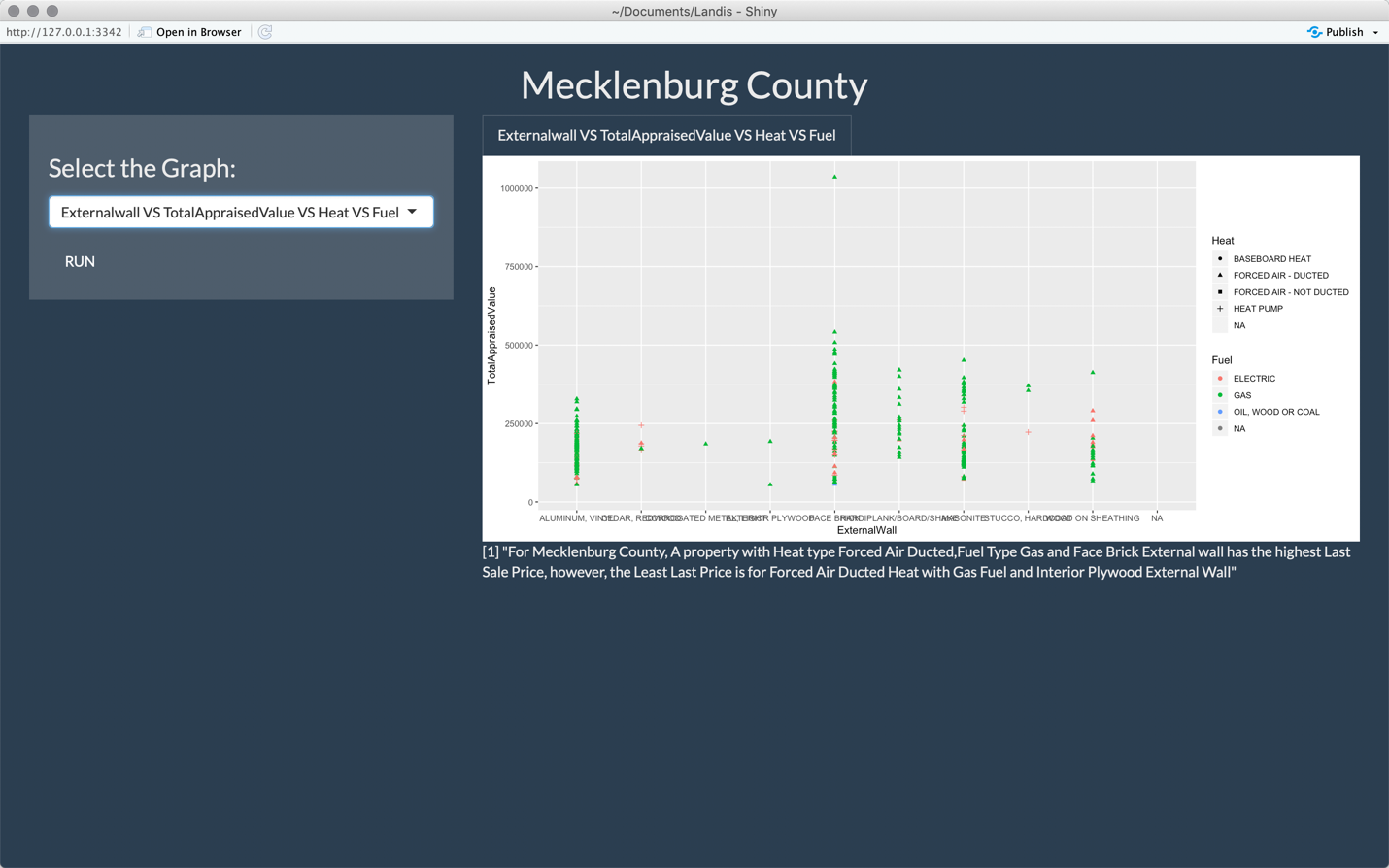
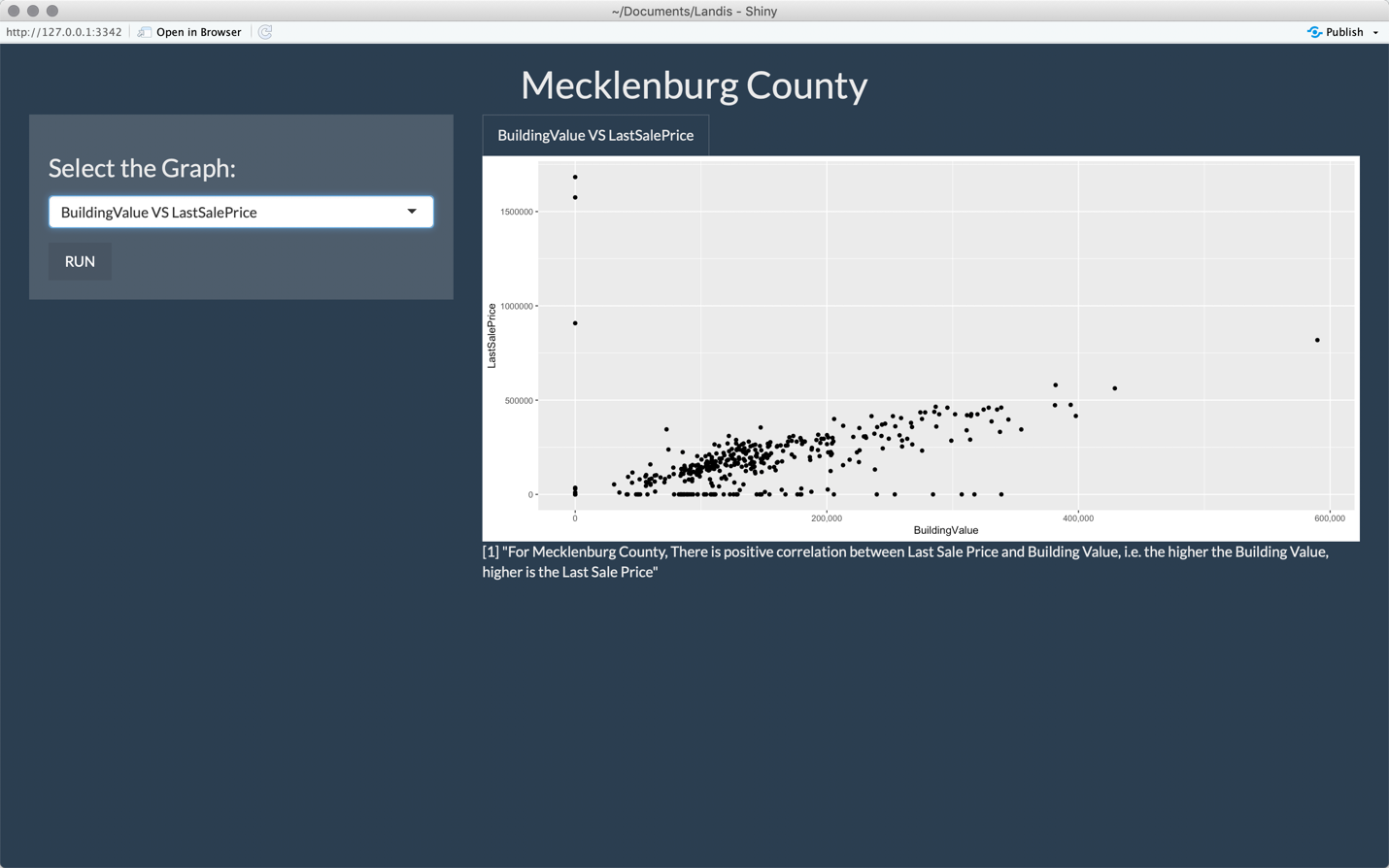
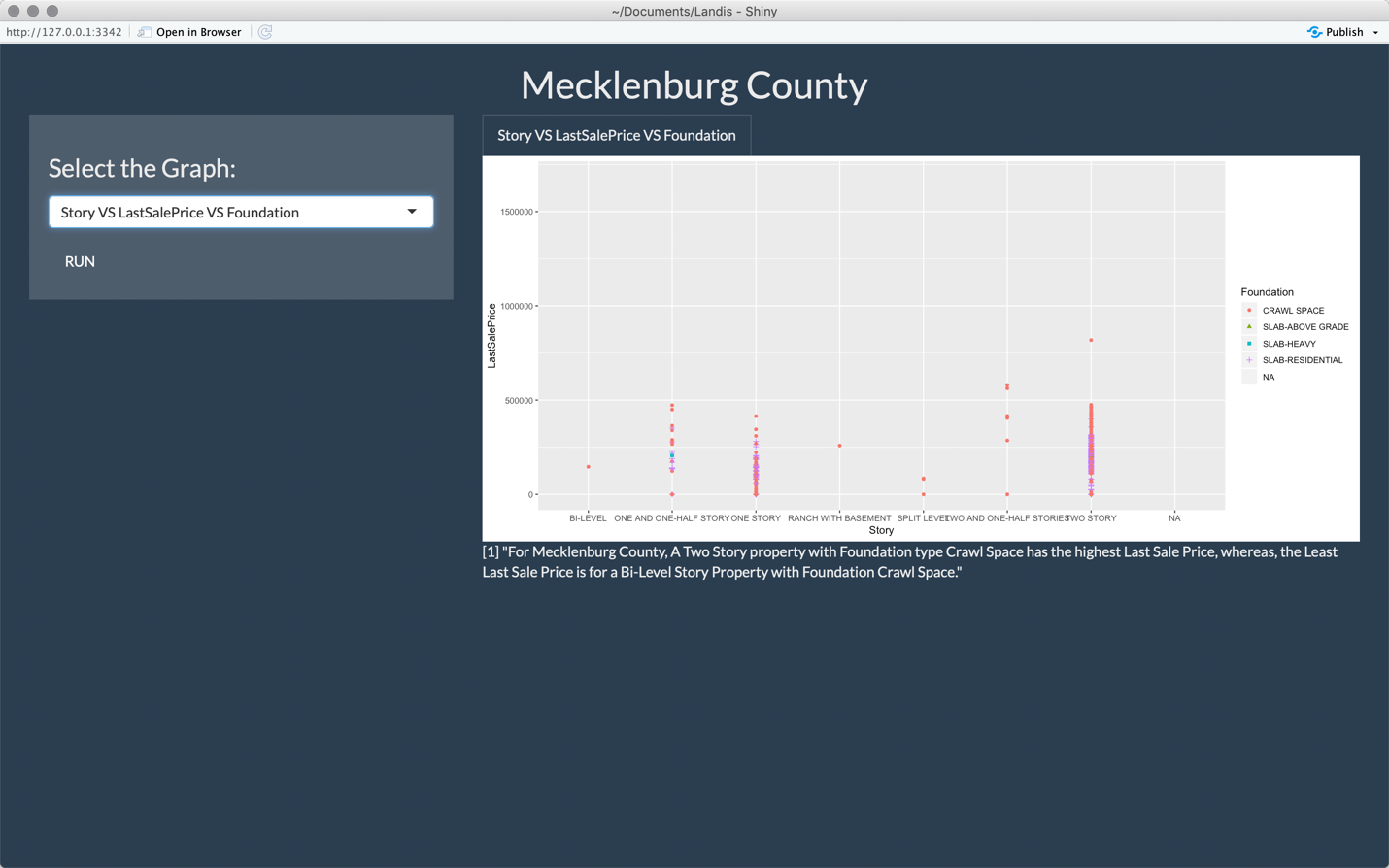
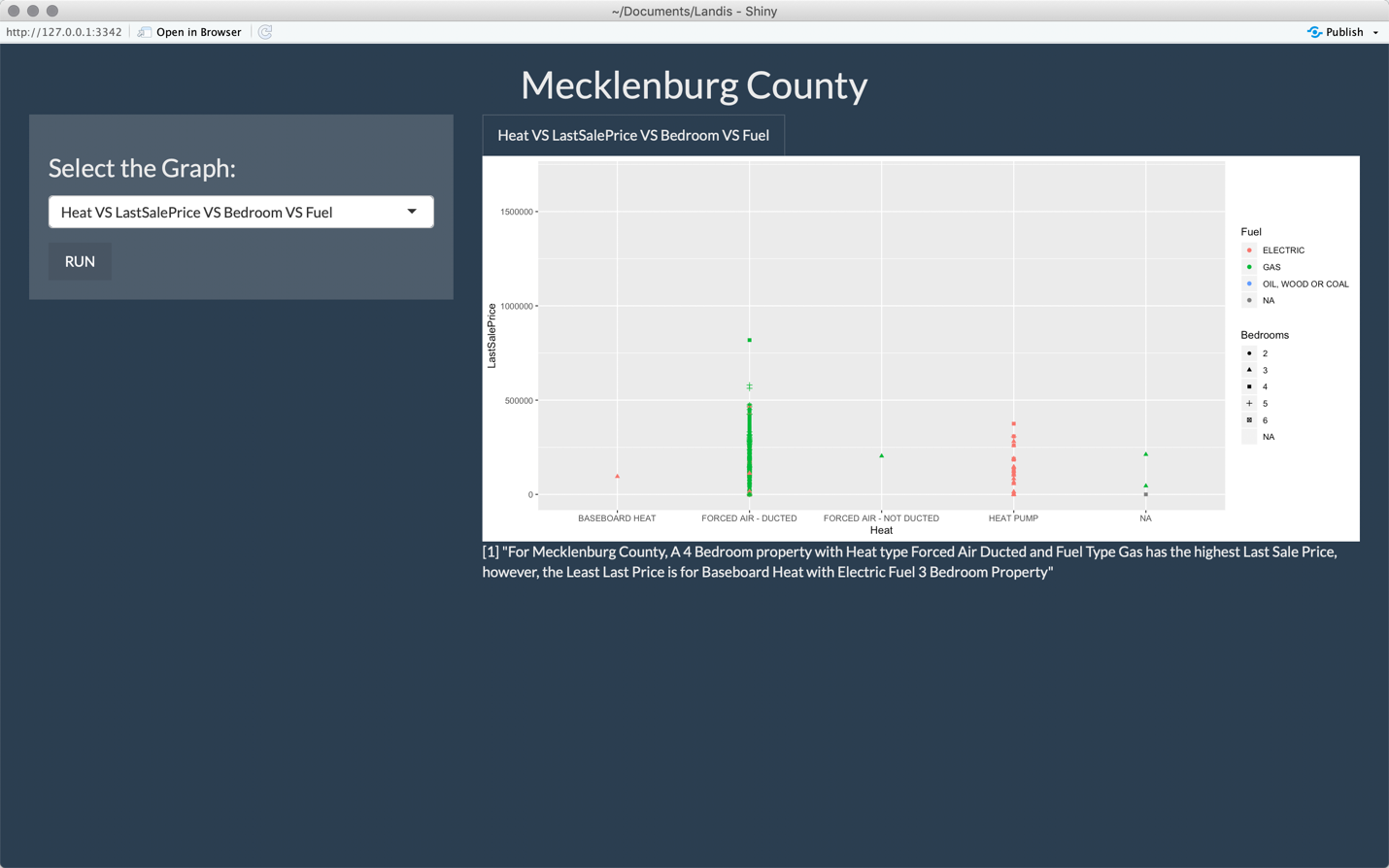
1. A 4 Bedroom property with Heat type Forced Air Ducted and Fuel Type Gas has the highest Last Sale Price, however, the Least Last Price is for Baseboard Heat with Electric Fuel 3 Bedroom Property.
2. A Two-Story property with Foundation type Crawl Space has the highest Last Sale Price, whereas, the Least Last Sale Price is for a Bi-Level Story Property with Foundation Crawl Space.
3. There is positive correlation between Last Sale Price and Building Value, i.e. the higher the Building Value, higher is the Last Sale Price.
4. A property with Heat type Forced Air Ducted, Fuel Type Gas and Face Brick External wall has the highest Last Sale Price, however, the Least Last Price is for Forced Air Ducted Heat with Gas Fuel and Interior Plywood External Wall.
5. There is positive correlation between Last Sale Price and Total Appraised Value, i.e. the higher the Total Appraised Value, higher is the Last Sale Price. However, there are exception when Total Appraised Value is low and still the Last Sale Prices are high.

These are just few insights, further analyzes can help to get more information about the property trends.

Screenshots of the outputs:



SQL Database



R-Shiny Dashboard