Mattel Data Science and Analytics Technical Test

In this assignment, you will be analyzing a dataset of housing sales in Ames, Iowa spanning from January 2006 to mid-2010. This assignment should take ~4 hours to complete.

The dataset is called data.csv and the descriptions for each field are supplied in data_description.txt. Each row of data contains one home sale.

Please use one of the following tools to perform your analysis: • SQL (we recommend SQLite) o This tool may help: https://sqlitebrowser.org/dl/ • Python (we recommend using pandas, jupyter noteboks, and matplotlib) • R (we recommend using RStudio and ggplot)

Please provide (1) a write-up with answers and (2) your code for the following:

1) In each neighborhood of Ames, what is the median sale price for homes sold in 2006 which have an indoor square footage of greater than or equal to 2000 ft. (excluding porches, garages, decks, and veneers)

2) A client approaches you with a question about the local housing market. They're interested in whether more homes are sold at certain times of year than others? In other words, is there seasonality? Provide a visualization and a brief description of your findings. (For the visualization, you may use R plots, Python's matplotlib, Tableau, Excel charts, or whatever tool you like).

3) You're a contractor consulting for a client who wants to remodel their home and then sell it on the market. The home the client occupies is a 3 bedroom / 2 bathroom 1500 sq. ft. house. They're deciding between the following options for the remodel:

    a) Adding a new bedroom measuring 130 sq. ft.

    b) Adding a new half bathroom measuring 80 sq. ft.

    c) Expanding the living room by 400 sq. ft.

Assume that the cost of all three remodel options is equal. Based on the data provided, which option do you think will provide the greatest predicted increase in home value and why? What other information would you seek out that might help you make the decision? Please provide any visualizations, tables, etc. to support your findings. Note: we're not looking for one "right answer" here; it's more important to explain your reasoning and the limits of how this data can inform this decision.

4) You own a single-family home (i.e. BldgType = "1Fam") with 4 bedrooms that you are looking to rent out or sell. Assume you can generate a yearly rent that is 10% of the estimated sales price. Your options include:

    a) Convert the home into a duplex and rent both units.

    b) Rent the home as is.

    c) Sell the home for market value.

Assume that cost is negligible for our purposes. Which option maximizes revenue received in 5 years? 10 years? 15 years? List out all assumptions you are making in your calculations and outline your thought process. Note: as in question 3, we're not looking for one "right answer" here; it's more important to explain your reasoning and the limits of how this data can inform this decision.