Programming for Analytics | Module 1 Exam
Please read these instructions carefully

Unless otherwise specified, you may use any function, package, or method to complete the requirements of this exam. I expect you to adhere to the best practices we have discussed in class including using intuitive, consistent variable names, using comments in your code, piping commands together instead of making multiple objects, etc. You are welcome to ask me questions, but you must work on this assignment ***individually***.

The deliverables are as follows:

- An HTML document submitted via Canvas but **NOT** published via RPubs.
- A self-contained RMarkdown file submitted via Canvas that creates the HTML document.

The RMarkdown file and its associated HTML output document should include everything needed to complete the tasks of this exam. As would be expected from a professionally published document, all your code, R messages and warnings should be hidden from the end viewer unless otherwise stated.

Xavier Property Sales

The leadership of Xavier University are interested in how the COVID-19 pandemic related economic policies have influenced the prices of residential properties in the Cincinnati neighborhoods adjacent to the university campus. Additionally, the city of Cincinnati has expressed concern surrounding housing affordability in the city at large. To assess these interests and concerns, you have been hired to investigate recent property sales in the neighborhoods in and around the Xavier University campus. To support your work, you have been provided data from the Hamilton County Auditor recording every residential property sale in the area from January 1, 2018, to December 31, 2021.

The data you will use for this project is stored in a table where each row represents the transference of a residential property in one of nine neighborhoods adjacent to the university campus. Each row records the details of the transaction including the following list of variables. A complete data dictionary is available at: http://asayanalytics.com/xu_prop-xlsx

- Parcel ID
- Year built
- Bedrooms
- Purchaser name
- Date
- Bathrooms
- Street address
- Sale amount
- Total rooms
- Property use code
- Neighborhood
- Finished SQFT

You can learn more about the Hamilton County Auditor and view every property sale in the county by visiting their government page here: https://www.hamiltoncountyauditor.org/. As a part of your analysis, you are expected to make the data more accessible to lay persons and illustrate trends in residential property transactions visually using a variety of visualizations, tables, and a compelling narrative presented in a professional blog-style webpage. All work contributing to the final document is expected to be done in R. More specific instructions and requirements follow.

Instructions:

Use the following Xavier property sales data to create a professional webpage from an R Markdown document. The expectations and scoring rubric are attached below. 200 total points are possible. The dataset is available as a CSV from the links below. You are welcome to knit from the URL directly. Only use the mirror link if the original stops working.

Data:	http://asayanalytics.com/xu_prop-csv
Mirror:	http://asayanalytics.com/xu_prop-mirror
Dictionary:	http://asayanalytics.com/xu_prop-xlsx

You are writing this document for a lay person reading a blog or similar style website. You may occasionally add comments intended for a data scientist in a code chunk. Your final .HTML document should not include any R code or RMarkdown syntax.

Please read through the **entire** exam before beginning!

Section	**Expectation**	**Points**
Introduction	**1.1** Provide a brief introduction that explains the purpose of the document and how your analysis will help an individual better understand property sales trends around the Xavier University campus. **1.2** Load all package libraries in a code chunk at the beginning of the document. Include explanations as necessary. Do NOT include commands to install packages.	5
Data Preparation: (Cleaning & Wrangling)	**2.1** Download and load the data set directly from http://asayanalytics.com/xu_prop-csv **2.2** Perform the following data cleaning tasks. Explain your process in the markdown document as necessary. Address the following data preparation tasks: A. **Data Errors:** Inspect each column for blatant data entry and transcription errors and address any obvious anomalies for which there is no explanation. Where your data cannot be properly validated, do ***NOT*** omit entire rows. Instead, change invalid data to be missing. The following list of possible corrections is provided *only* as an example, and it is not exhaustive: o Sales cannot occur on impossible dates such as February 30. o Homes cannot be built in a future date. o Sales values cannot be negative, but they **can** be missing. o Data types for each vector should be properly defined in the original data. B. **Variable Creation:** Create variables in the original dataset for each of the following. This is not an exhaustive list of variables you may need to calculate or mutate, but these should be the **only** ones saved in the original data frame: o A fully functional date vector capable of accepting requests using functions such as wday(), year(), month() and ***delete*** the now redundant month, day and year vectors in the original data. o A dummy variable indicating whether the property is a multifamily dwelling o A discrete variable indicating the following using a non-hardcoded formula: ▪ The property value is within 1 standard deviation of the mean value. ▪ The property value is more than 1 standard deviation above the mean value. ▪ The property value is less than 1 standard deviation below the mean value. ▪ The property value is missing.	40
Introduction to sections 3-5: For sections 3 through 5 below, you may need to mutate a new variable to accomplish these tasks. Do **not** modify your original data object in any way for these activities or make new objects unnecessarily. Instead, use your dplyr family of functions and pipe the output directly into your visual. For any visualization, you are welcome to adjust axis labels, trim values for outliers, etc., but make sure you comment on whatever adjustments you make. You should also confirm the adjustments are correct for the relationship, data types and the visualization you are using.		

Section	**Expectation**	**Points**
Simple Trends & Analysis	Create a single **visualization** (not a table) illustrating each of the following comparisons. Facets still count as only one visualization because there is only one ggplot canvas. Briefly comment as requested. **3.1** The distribution of single-family dwelling home sizes in square feet. Judging strictly by the visualization, would you say SQFT is normally distributed? **3.2** The ratio of full bathrooms to bedrooms for each neighborhood. What phenomena do you suspect might explain this result? **3.3** The total value of home transactions processed in each month for each neighborhood. Judging strictly by this visualization, does the housing market appear to exhibit seasonality?	45
Directed Analysis	Use any number of your own visuals, tables, or other calculations to answer the following question. Your response should include an explanation of the approach used and an interpretation of the result. **4.1** If you were gifted a residential property in this area and you intended to sell it for the highest price: ○ In what neighborhood would you want it to be located? ○ What features (size, rooms, bedrooms, etc.) would you want it to have? ○ How old would you want it to be? ○ What time of year or day of the week would you want to sell (or does it matter?)	30
Self-Directed Analysis	Use your own visualizations, tables, or other calculations to evaluate the following prompts. You should be comprehensive but not unnecessarily verbose with your introductions, justifications, and interpretations. Even if you do not find support for the stated claim, you should defend your position with empirical evidence. **5.1** In recent years, government officials for the city of Cincinnati have expressed concern at the growing number of investment firms purchasing residential homes in Cincinnati as investments with the intent of converting what would otherwise be owner occupied housing into rental properties. In fact, a Cincinnati government entity even outbid more than a dozen such investment firms to buy nearly 200 homes in January, 2022 alone and the city intends to buy even more properties in the future. Using any methodology you prefer, attempt to show support for or against the claims made by the city of Cincinnati that over the last four years, residential properties are increasingly becoming owned by corporations rather than by individuals and that this phenomena is contributing to the increased price of housing in the area.	50
Formatting & Miscellaneous Requirements	**6.1** Proper coding style is followed. Code is well commented where necessary. Proper naming conventions are followed, complicated commands are systematically piped from one operation to the next, and the global environment is free of extraneous objects and subsets. **6.2** .RMD file fully executes without errors and produces the HTML report submitted by the student. **6.3** Summary statistics and visualizations used to evidence a point of view are not egregiously misleading and are appropriately explained.	30