HW 4

your_name

date

Homework 4: That One That Brings It All Together

Before submission remove from HERE...

For this assignment include all your R code in this R Markdown file to support your answer. Your code goes in the code "chunks" below. If you open this file in RStudio, run `install.packages('rmarkdown')` (if you have not already) then you can run all the R code by clicking on the `Knit PDF` button. You can experiment with code in the R console and then paste your code into this document, or write it directly in this document and send it over to the console with the keyboard shortcut `command+enter` or `control+enter`. Please use the pipe operator to represent composite tasks.

Scenario

You have been hired as a data scientist consultant for a boutique coffee shop (Central Perk) in New York City. The coffee shop "believes" (whenever a client has a belief, you should aim to confirm or reject this belief with data) that their current customer base is fairly loyal and business is consistent year over year. Therefore, they are not **necessarily** interested in acquiring new customers–they feel like FRIENDS with the ones they have–rather they are interested in generating more revenue from their existing customer base.

Although the coffee shop believes they have a sense of their customer base, they realize they have never explored their data and do not have a good sense of their demand. Therefore, they ask you to analyze historical sales data to understand the general volume distribution (e.g., in a typical week, by hour, by day etc). In addition, they ask you to provide a narrative on the types of items being ordered. Their objective here is to understand the current customers' buying patterns (for example, but not limited to, average order size and time of day). Essentially, they want you to explore the data and provide them a general sense of behavior, in addition to identifying interesting/systematic patterns of volume and items, **that they can act upon**.

Moreover, using this data, the coffee shop wants you to use the information from your exploration to formulate a proposal of incentives/pricing adjustments that can be offered to customers to normalize demand and generate additional revenue. The believe demand is very high at certain times (on certain days) and very low other times, and they would like to smooth this demand out over time. You are allowed to make any recommendations you deem valuable (including increasing or decreasing the price)–for example, recommending a 20% discount (on specific products) at 3:30p on Tuesday. However, take care to ensure that these recommendations, if implemented, will not alienate customers or lead to an overall decrease in profits. The coffee shop not only wants these recommendations, but also clear, data-backed justifications for 1) why they should implement your proposed strategy and 2) why it will not alienate customers or decrease profits. For the purpose of profitability scenarios, you can assume a gross margin of 20%. Conclusions from data analysis should be demonstrated to be **robust** and well thought out.

If your proposal is effective, the data and learnings can be applied to additional locations and other restaurants. Therefore, be sure to consider this ultimate objective when making your recommendations. Please remember, that your audience will be the owners of the coffee shop as well as its investors, so your analysis (and the narrative surrounding it) needs to be clear, convincing, and suitable for the audience.

ANALYSES

You are welcome to use any methods, techniques, concepts, or analysis you have learned in this course or others. Although there is no requirement that a particular analysis is used, we will be evaluated on how well you use the various (types of) techniques in concert to solve the problem at hand. So be very thoughtful about how to use the various concepts and tools in novel ways and in conjunction with each other. However, in search for novelty, do not abandon interoperability and clarity. Whatever you choose, think carefully about how your work addresses the problem at hand. You will likely conduct multiple analyses to address the business question. For each analysis you decide to conduct, you must include the following in your R markdown document to get credit for the assignment:

1. Description and Rationale for the Chosen Analysis
2. Execution and Results (including code)
3. Interpretation
4. Conclusions

Description and Rationale for the Chosen Analysis

Name the analysis/technique/approach and write a short explanation that addresses:

- Why you chose this particular technique/approach. Given the question/problem, why use this approach vs. any others?
- How observations you've made or other analyses you've conducted led you to want to conduct this analysis.
- Your assumptions and justification (if any) for those assumptions.
- Follow the framework outlined in Visualizations That Really Work
  - Explain if your analysis is exploratory or declarative and why.
  - If the analysis is exploratory describe if (and how) it is either confirmation (i.e., testing a hypothesis) or exploration (i.e., mining for patterns).

Execution and Output (including code)

Show how you conducted the analysis, including the code, and the results/outputs.

Interpretation

How do you interpret the results? Reference specific output/results and relate the interpretation to the business problem.

Note: any output/result not specifically referenced in your interpretation is extraneous and therefore should not be included.

Conclusions

Based on your interpretation, what do you conclude? Why/how can you draw this conclusion? State the conclusion(s) as they relate to the business problem. A conclusion might answer an important question, provide a piece of the puzzle, and/or lead you to additional analyses.

Notes, Comments, Tips

- Your first set of analyses should be initial steps of EDA: inspect, clean, and prepare the data. Check and resolve (if necessary) data entry errors, missing values, outliers, etc.

- Following data cleaning and preparation, move on to data munging, data visualization, and data analysis.
- Remember that this is not a linear process, but cyclical. You should have multiple iterations of parts 1-4 above in your final R documentation.
- It's likely that you will conduct some analyses that don't yield useful results. Do not include these in your R documentation. I only want to see analyses that support your ultimate conclusions.
- As you move towards completing your analyses, you will need to consider all the conclusions you reached along the way, and how they fit together to solve the problem. This thinking and the conclusions should inform the final deliverables you create for the client.

. . . to HERE

Explanation of Approach and Goals 1 (Think and Describe It):

```
#Analysis (Do It)
```

Interperation from Approach:

Conclusions from Approach:

Based on conclusions from Approach 1, explanation of Approach and Goals 2:

```
#Analysis (Do It)
```

Interperation from Approach:

Conclusions from Approach: