Homework 1

team_member_names

date

Homework 1: Mobile Application Market

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 already have not) 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 are a data scientist for a consulting firm. Your team is tasked with helping your client better understand the mobile application market as they consider entering this potentially lucrative space.

Your job is to investigate and draw insights from data on the market, and present your client with a well-structured narrative (supported by analysis) in order to assist them in developing a plan to enter this market. Your client audience includes product managers, marketers, and business strategists. They appreciate statistical evidence and technical results; however, they want to understand how those things relate to managerial decision making about the project.

ANALYSES

You are welcome to use any analysis techniques you have learned, including your statistics training to test hypotheses. Whatever you choose, think carefully about how your work addresses the problem at hand. For example, you might consider ideas such as the following:

- Understanding which kinds of apps are "successful" (however defined), and how that differs across region (US vs. China) or platform (Google Play vs. Apple App Store).
- How different factors, such as file size, relate to success.
- Understanding app portfolio composition for different developers, including the distribution of types, price points, region, stores, etc..

These are merely potential ideas; there are many ways to analyze data and answer a question, so be creative as you develop your approaches. Simply doing the analyses describe above will result in little credit. Do not just jump into analysis; think critically about the business problem and the actual question(s) you want to answer. Then plan how to accomplish it, and execute.

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.

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, by 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: