Homework 1

team_member_names

date

Homework 1: Association Rules and Sports Analytics

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

As data continue to be a ubiquitous component of competition across firms and industries, data analytics has found a home in more and more organizations. Potentially sparked by money ball, sports analytics is on the rise to the extent that many major sports teams have a data scientist on staff to help give the team an edge over the competition. To that end, you have been hired by AS Roma to be their resident data scientist. The coach of the team wants you to find patterns that he can exploit to increase success on the field and decrease failure; specifically, he wants to find non-obvious patterns because experience has taught him many of the obvious ones. He suggest that you use association rules because he once heard how Walmart used them to find a non-obvious pattern and sell more pop-tarts during hurricanes. At first you think his request does not make sense because association rules are traditionally used to analyze purchases, but then you realize the coach may be on to something, and so you set out to see what patterns you can uncover! You have been provided data (`euro_soccer.sqlite`) corresponding to matches played between European soccer teams from 2008 to 2016, as well as information concerning the teams and players. More information concerning the data can be found here. Additional information on how to use dplyr to munge data in databases can be found here; I provide a simple examples in file `example_munging.R`.

ANALYSES

You are welcome to use any methods, techniques, or concepts for analysis you have learned in this course or others. However, association rules must be a central piece of your approach. Recall that in order to use association rules you need to clearly define your "items", your "transactions", and a measure of being "selected" (into the basket). Whatever you choose, think carefully about how your work addresses the problem at hand. For example, you might consider ideas such as the following:

- Your transaction can be matches
- Items can be
  - Players (where players are "selected" if they play in the match)
  - Match outcomes (where matches are "selected" if the outcome is a win)
  - Betting odds of winning (where odds are "selected"" if the odds where not in your team's favor)

Such an analysis could help you understand which combinations of players on the field tend to correspond to your team winning games that it was supposed to lose. Furthermore, instead of considering players specifically, you could consider items to be particular player characteristics (e.g., speed).

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 described 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. Be sure to provide a general (but concrete) description of the set of problems/questions you are addressing and how they form a logical chain to solve the overall business problem provided in the prompt. Check here for more information about the data.

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 of the specific problem/question you are addressing.
2. Description and Rationale for the Chosen Analysis
3. Execution and Results (including code)
4. Interpretation
5. 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. However, please do not include extraneous outputs or table/graph dumps. Once you generate outputs that help you understand, please create simple and compelling results that precisely demonstrate the understanding that you want the client to understand. *Only* include these results; others are too difficult to comprehend and are extraneous. Additionally, all results should be properly labeled with headings/variables that are clear and professional.

Interpretation

How do you interpret the results? Reference specific output/results, explain what the result is saying, and what should be taken away from it.

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 1-5 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

Team Solution

Overview

Strucutre the client's initial question into a set of problem statements. Explain generally how these problems logically connect, what approach you will take, what is the goal of this approach, and how its answer will address the client's initial problem.

Details

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 next question, its approach, and goal:

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

Interperation from Approach:

Conclusions from Approach:

Conclusion

Given the logical chain of problems, approaches, and conclusions, draw a general set of conclusions to the clients original question.

Recommendations

Given the conclusions, what recommendations would you make to the client. Recommendations are have to take into consideration contexts, constraints, cost/benefits, business challenges, etc. Recommendations lay out plans of how to solve the clients problem(s), and support for why each of the steps in the plan are critical and effective.