CS145 - High-Level P3 Project Rubric

This document lays out some high-level expectations for what is a good open-ended portion of the third project. More specific details of grading (i.e. how many points or percentage points correspond to each bullet point) are omitted.

- **Analysis of your dataset (10%)**
  - +
    - It is clear that students understand the structure of their datasets -- data sizes, number of tuples they are focused on, and high level relationships between tables e.g., through simple diagrams, discussing objects, keys, and values if OKV, or by some other means that's clear and concise
    - If there are multiple tables, students should discuss how they are related.
  - -
    - Little to no effort in explaining the dataset. It is not clear to the grader that the student(s) behind the project understand the structure of the data they are working with.
    - Clearly misinterprets some aspects of the dataset.
    - Clearly provides some wrong or overly general (i.e. not specific enough) analysis.

- **Exploring your questions, with appropriate visualizations (60%)**
  - +
    - Considers a wide range of metrics that cover many facets of the data. The metrics chosen are not only what one would expect, but are insightful and well-tuned to the question asked. We expect your features (the choices of metrics that you put into your model) to be diverse.
    - It is clear that students understand a good set of patterns and trends in the data.
    - Charts are well-crafted. The choice of plot works well with the motivation behind it (the question it seeks to answer). Charts are neat and display the data clearly.
    - Students do not ignore things that they cannot explain in their visualizations; instead, they identify and give thought to why something unexpected or anomalous may be present
    - Provides hypotheses for some of the more interesting relationships observed.
  - -
    - No analysis/low effort
    - Only explores a few variables without reasoning about the choice for only using these few variables
    - Oversimplification of the dataset (attempts to make conclusions from a couple of simple queries)

- **Predictions based on your explorations (20%)**
  - +

        - We are not looking for any particular performance benchmark here, as long as the model is reasonable and the prediction problem is well-framed. That said, the model should not completely fail or be unusable for predictions.
        - Predictions are reasonably framed. A student does not seek to answer, e.g., a difficult NLP or computer vision using the models available to them on BigQuery.
        - Predictions use a reasonably sophisticated set of features
        - Feature engineering has clearly occurred: analysis/thought has been given for why a feature may or may not be a good
    - -
        - Does not have a holdout set for displaying final metrics
        - The features are haphazardly chosen and don't make too much sense for the prediction task at hand
        - Students only use only a small handful of features
        - The model fails heavily (e.g. achieves less than 1% accuracy, or less than .55 ROC_AUC), contains bugs, or is unable to generate predictions

- **Conclusion (10%)**
    - +
        - Gives good, insightful thought into analyzing what has been explored and visualized in the previous sections.
        - Has a keen eye for limitations in analysis, and is careful about not making overly definitive claims.
        - Identifies and attempts to offer explanations for aberrations in the data, visualizations, or predictions.
        - If something seems to be true, explains why with data-backed arguments. Conversely, if something seems to not hold, explains with the help of data or visualizations.
    - -
        - Makes highly simplifying assumptions about the data or visualizations, e.g., "this one SQL query shows this correlation, therefore x must be true"
        - Makes very strong "definitive" conclusions (e.g., "x happens exactly because of y")
        - Clearly does not demonstrate enough effort in reasoning about the data and plots obtained in the previous sections

FAQ

- How many models should I train?
    - It's fine for you to train a single model. We want you to spend time instead querying, visualizing, and reasoning about your data.
- How many plots should I have?
    - The real question is - when we read your exploration and analysis, do we see 6-10 clear features?
    - For a metric to be considered a feature that you train on, it should have a relationship to the dependent variable of your research question.

- How many features should I put into my model?
    - At least 6-10 diverse features is a good ballpark. By diverse, we mean that the features are not overly related with one another. We would also like at least a couple features to be things you compute or construct (more than selecting a column out of your dataset).
        - Example of not diverse: If predicting house prices, then sq ft of the house and sq ft of the land of the surrounding property are not diverse. It's effectively the same feature (you have explored the same dimension of information)
        - Something that would be diverse from sq. ft of a house is, say, the elevation of the house, or the demographics of the neighborhood, or the average income of the neighborhood.
        - Example of selecting a column: for predicting house prices, there's a column called "sq. ft".
        - Example of not selecting a column (something you have to compute or construct): You compute the distance to the nearest school.