IS 6850 - Final Project

For the final project, we're going to treat this like a real world data project. You have the liberty of choosing the data you want to work on, and your approach to analyzing this data. This is both to give you the opportunity to work on something interesting to you, as well as accommodate varying levels of experience in the group.

The big thing we ask of you - you need to provide useful and actionable analysis, using the skills you learned in this class.

Grading

The project proposal will account for 10% of your final grade and the final project report will account for a remaining 20%.

A Proposal

The proposal must address each of the sections below.

- Data you plan to use
- Techniques you intend to use
- Questions you intend to ask
- Visualizations you intend to create

Please submit this proposal to Canvas in PDF format.

You need to get your proposal to us by July 16th. We will review your proposal by July 19th and approve it or request changes. The project will be due on July 26th.

We understand that projects often change inflight. If you discover that you need to make an adjustment as you are working through your project, communicate with us immediately so we can provide assistance and approve your changes.

Data

A great data project starts with interesting data. Choose public data that you are interested in working with. You can choose one of the Google Cloud Public Datasets or choose a dataset from some other source.

Use Interesting Techniques

We won't specify exactly which analytical techniques you should use, but your project should demonstrate new things you've learned in advanced SQL. This could include window functions, statistical functions, multistage transformations of data, pivoting or unpivoting, extraction of nested data, joining with a dim dates table, etc. If you have multiple years of experience in SQL analytics, consider going beyond the core material of the class by doing something with BQML (BigQuery machine learning) or GIS (geographic) functions.

Ask Interesting and Actionable Questions

Think about the practical impact of the dataset you've chosen. For example, COVID19 data is important in determining public policy in 2020 and 2021. Business datasets inform forecasting and strategizing. Datasets on patent filings might reveal trends in technology and business. Ask a question that speaks to the practical impact of the dataset. Ask questions that can lead to an action or a result.

Create A Visualization

Your final report should include some kind of visualization. This could be in Google Data Studio (free BI tool that integrates with BigQuery), Google Sheets, Excel, plot.ly, or any graph or visualization tool that you prefer. (Note: https://coronavirus.utah.gov/case-counts/ is a great example of insightful visualizations. Many of the graphs are rendered in plot.ly.) Note that if you use Excel or Google Sheets for visualization, your analysis should still be done in SQL given that this is an advanced SQL class.

Your visualization could be a basic graph of your results, data points on a map if you use GIS functions, a visualization of clusters if you utilize clustering algorithms in BQML, etc. Your visualization should provide some insight into the question that you are answering in your project.

The Final Report

Assemble a final report. In this report, you should do the following.

1. Explain the technical details of the dataset that you chose to investigate.
2. Explain the importance of the dataset from a policy or business perspective.
3. Explain the question that you asked, and the importance of this question.
4. Explain your conclusions
5. Provide your visualizations

6. Methodologies
    a. How did you analyze the data?
    b. Provide all queries and code.
    c. Explain your visualization methodology.
7. Deliverables
    a. Your final report should be in PDF format, submitted through Canvas.
    b. Code
        i. Code and README should be in a publicly accessible git repository like Github, Gitlab, Bitbucket

            OR

        ii. Any supporting documentation or code should be submitted in a zip file.