Semester Long Project

Deliverable #2 – The Check-In

Overview

Your pitch was successful, your client decided to engage with you for a project, but most of all they have decided to invest time and money into you. It has been weeks since the client accepted your proposal, and now they want a check in to see all of the progress you have made. It is time to come through on some of the promises you have made.

Core components of Deliverable #2

1. Identify valuable components of the dataset

The consultant should have a solid understanding of what is in the data, and be able to communicate to their stakeholder some of the key data points.

2. Define and communicate limitations of the data

The consultant should communicate issues and limitations of the data. Is there missing information that will cause issues? If so, how do you plan on overcoming these challenges? Are there any indications of poor data quality?

3. Demonstrate preliminary insights

What meaningful insights can you derive at this early stage of the data analysis? These insights should be presented in a professional format and leverage visualizations.

4. Clean and professional code

Alongside the visual deliverable, the client would like to see your entire code base. Make sure it is clean, uses Markdown, it is commented and easy to navigate.

Format of Delivery

Format of delivery is open to the students preference though I have a slight bias to slides. For this delivery you will not need to present it.

Rubric

- **Data analysis and demonstration of analytic curiosity: 25%**
    - Demonstrates the consultant spent a significant amount of time (at least ~15+ hours) exploring the dataset
    - Defines limitations of the data

    - Derives valuable and useful insights for their stakeholder
- **Presentation of the information – 25%**
    - Information is presented at the right level for their stakeholder, always know your audience
    - Business product looks professional
- **Demonstrate practical use of skills learned in class – 25%**
    - Leverages Pandas or Dplyr functions well
    - Creates useful and professional visuals
    - Use of Matplotlib, ggplot2 etc
- **Quality and extensiveness of code – 25%**
    - Code is clean, commented, makes use of documented, and housed in separate ipynb files where necessary