

Final Project Information

DSCI 5360 – Data Visualization

Final Project

Objectives

1. The primary objective of this project is to demonstrate an understanding of the fundamental concepts related to visualization discussed throughout the course
2. This project should also demonstrate a sound working knowledge of Tableau
3. This project provides an opportunity to practice the art of storytelling with visualizations

Instructions

1. This project will consist of a report and a presentation which detail your visual analysis of a dataset of your choosing (sources for data discussed below)
2. There are no restrictions on the dataset you use, other than it cannot have been used in class
3. Your visualizations should
 1. Support a coherent story (no visualizations just for the sake of visualization)
 2. Adhere to the principles of visualization we discussed in class (i.e., pie charts may not be a good visualization choice, rainbow color schemes may not make sense, etc.)
- Be relevant to managerial decision making
4. The body of your report should be at most 5 pages and consist of
 1. An **Introduction** that describes the case or problem. This should be precise and clear enough that a layperson should be able to easily comprehend the nature of your case.
 2. An **Analysis** section which describes the specific analyses performed and the various Tableau functionalities utilized. **The actual visualizations can occupy an appendix if needed.**
- A **Conclusion** section, which summarizes findings, discusses managerial implications, provides recommendations, and discusses extensions (if necessary). An appendix may also include any formulae, calculations, relevant figures, etc. However, the main report should stand on its own.
5. Your Tableau file should
 1. Contain multiple, well-formulated, visualizations
 2. Contain at least one **Dashboard** and **Story**

- Be saved as a TWBX file.
6. This project must be original analysis completed independently by each student. Any deviation from these directives will be viewed as academic dishonesty, and a penalty will be imposed as per the syllabus.

Sources of Data

This project allows your group to walk through a visual analysis using real data that is of some interest to you. Data can come from two possible sources:

1. **First-hand data** directly from your job or your connections. If you choose this option, make sure you have approval from the data owner. This option is strongly recommended.
2. **Second-hand data** may be obtained from a variety of online sources. Some possible sources of data are listed below:
 - Kaggle (<https://www.kaggle.com>) – This site hosts data mining competitions. You do not have to compete in order to gain access to the data sets.
 - UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/>) – Free data sets hosted by the University of California Irvine.
 - Datahub (<https://datahub.io/dataset>) – An online repository of user-submitted data sets.
 - SQLBelle Blog (<https://sqlbelle.com/2015/01/16/data-sets-for-bianalyticsvisualization-projects/>) – A large list of data sets from a variety of sources on the web.

You are not limited to the sources described above. There are many sources of available data and as long as the dataset you select supports the types of visual analysis described in class, it should be fine. Please be sure that you pick a topic/dataset that is not only interesting but also understandable to you.