

ITM 752: Data Visualization

Fall 2023

Group Project 2: Exploratory Data Analysis & Interactive Dashboard Design

Grade contribution: 15% of your overall grade

Brief Description

This assignment has two objectives:

- 1) To understand the process of using visualizations to perform **exploratory data analysis**. You will **formulate and answer a series of specific questions** about the data by visual exploration to **confirm or disconfirm your hypotheses** about the data.
- 2) To create an **interactive dashboard**. After answering the questions, you pose, you should create a **final visualization (an interactive dashboard)** designed to answer your most significant questions and allow the viewer to explore the dataset on their own.

You should maintain a word file that briefly **documents the questions you asked and the steps you performed from start to finish**. You will be tested utilizing the **advanced Tableau features** you learned, including **filters, calculated fields, sets, parameters, and LOD expressions**. Which of these features you will use depends on your research question but remember that your submission will be evaluated based on the way you utilize technical capabilities of Tableau.

Guideline on Exploratory Data Analysis

Step 1. Identify your ultimate goals and objectives. As a first step, put aside your data so that you do not let it limit how you think. Think broadly about some of the important questions that you may want to investigate for this project considering the overall project scope and the potential audience who might be interested in hearing the insights from this data. It's ok if you later find out that the data cannot sufficiently answer all your questions. You may discover ways to make your idea come through by introducing new data, or crunching the data you have, or simply have these high-level questions guide you throughout the project.

Step 2. Examine the data and data quality. Data cleaning, also called data cleansing or scrubbing, is the process of ensuring that your data is correct, consistent and usable by identifying any errors or corruptions in the data, correcting or deleting them, or manually processing them as needed. Before you conduct any analysis on the data, you need to detect and remove errors and inconsistencies from your data set in order to improve its quality.

Step 3. Pose an initial question that you would like to answer. As an example, here are a few testable questions from the other domains: Is there a relationship between melting point and atomic number? Are the brightness and color of stars correlated? Are there different patterns of nucleotides in different regions in human DNA?

Step 4. Assess the fitness of the data to answer your question. Inspect the data--it is invariably helpful to first look at the raw values. Does the data seem appropriate for answering your question? If not, you may need to start the process over. If so, does the data need to be reformatted or cleaned prior to analysis? Perform any steps necessary to get the data into shape prior to visual analysis.

After you have an initial question and a dataset, construct a visualization that provides an answer to your question. As you construct the visualization you will **find that your question evolves - often it will become more specific**. Keep track of this evolution and the other questions that occur to you along the way. Once you have answered all the questions to your satisfaction, think of a way to present the data and the answers as clearly as possible.

Before starting, write down the initial question clearly. And, as you go, maintain a document of what you had to do to construct the visualizations and how the questions evolved. Describe any transformations or rearrangements of the dataset that you needed to perform. In particular, describe how you got the data into the format needed by the visualization system. Keep copies of any intermediate visualizations that helped you refine your question.

After you have constructed the final visualization for presenting your answer, write a caption and a paragraph describing the visualization, and how it answers the question you posed.

Data Set: Movie & Netflix Data

The first dataset contains important statistics from a large sample of movies. The data includes the movie budget and revenue from different sources as well as ratings from [RottenTomatoes](#), [The Numbers](#) and [IMDB](#). The second data set includes titles from Netflix. You can select one of the datasets to create your dashboard or decide to use both of them.

The second dataset contains important recession historical data, you will aggregate the data into the current movie and Netflix data set and try to identify the correlation between the recession, movies US Gross revenue, Worldwide Gross revenue and the production budget variance over the years by plotting the data into the visualization using the type you see fit to complement the overall analysis. Using the recession dataset, you need to answer all the following three questions using the same visualization:

1. Is the movie's US revenue affected by the recession? If yes, show the correlation using the correct visualization.
2. Is the movie's worldwide revenue affected by the recession? If yes, show the correlation using the correct visualization.
3. Is the movie budget affected by the recession? If yes, show the correlation using the correct visualization.

Deliverables:

A copy of your final interactive dashboard with 4-to-6 visuals in Tableau (packaged workbook) via D2L.

In-class presentation: Use the interactive dashboard for your presentation (no slides). Each group will have 5 minutes for the presentation & 5 minutes for the Q/A. Include the following in your presentation:

1. Introduce your dashboard.
2. Briefly introduce your thought process, evolution of your research questions, challenges you encountered during the process, and the ways you solved them.
3. Introduce the technical features you used in Tableau.
4. A word file (750-1000 words in total) including:
 - a. Caption: This line is caption for your visualization
 - b. Screenshot of your dashboard
 - c. Paragraph 1: Brief explanation of your visualization and what it shows.
 - d. Paragraph 2: Briefly explain the data transformation and cleaning process. E.g., How you handled data diagnostics, missing data, transformation, and classification.
 - e. Paragraph 3: Brief description of the Tableau functionalities utilized and how/why you use them.
 - f. Paragraphs (4-6): To describe your data exploration process and what your final visualization shows. Describe the steps you have gone through, including the flow of your exploration, questions posed, and challenges. You may include some of the key visualizations that you have created during the process (e.g., the ones that have shaped

your thinking, helped you revise your research questions and/or pose follow-up questions, or provided you with a critical turning point in any other way)

Important Notes

This is a group assignment. No ZIP Files. No late submissions. No e-mail submissions. Each submission will be graded based on the analysis process, presentation, and the final visualization.

Grading

| <i>CATEGORY</i> | <i>EXPECTATIONS</i> | <i>GRADING</i> |
|---|--|----------------|
| Appropriate Data Diagnostics and Transformation | <ul style="list-style-type: none"> • Appropriate data diagnostics and transformation • Appropriate data classification (i.e., using optimum # of classes and method for breaking data into ranges) • Clear explanation on how you handled data diagnostics, transformation, and classification | 2 pts |
| Appropriate Data Exploration and Analysis Process | <ul style="list-style-type: none"> • Sufficient breadth and depth of data analysis, with appropriate research questions and follow up questions. • Brief explanation of the key steps in data exploration and visual creation process and how your visualization answers your questions | 4 pts |
| Design and Effective Communication | <ul style="list-style-type: none"> • The visualization is clear, clean, concise, and captivating. • The message of the caption is clear, effective, and captivating. • Chart types are suitable and best options for analysis. • The application of color is correct and clearly conveys meaning. • Fonts choices are conscious and consistent, proper grammar and spelling is used, and choice of position, size, and emphasis integrate elements into a visually appealing and engaging whole. All axes and text are treated appropriately. | 4 pts |
| Use of Technical Features & Interactivity | <ul style="list-style-type: none"> • The visual effectively leverages Tableau's functions including <i>calculated fields</i>, <i>parameters</i>, <i>filters</i>, <i>sets</i>, <i>LOD expressions</i> to present data effectively • Effectively uses marks shelf | 5 pts |