PHASE 1 PROJECT

Introduction

- Microsoft wants to venture into the film creation industry and we have access to data in different databases and dataframes to help Microsoft come up with a strategy to us in the new venture.
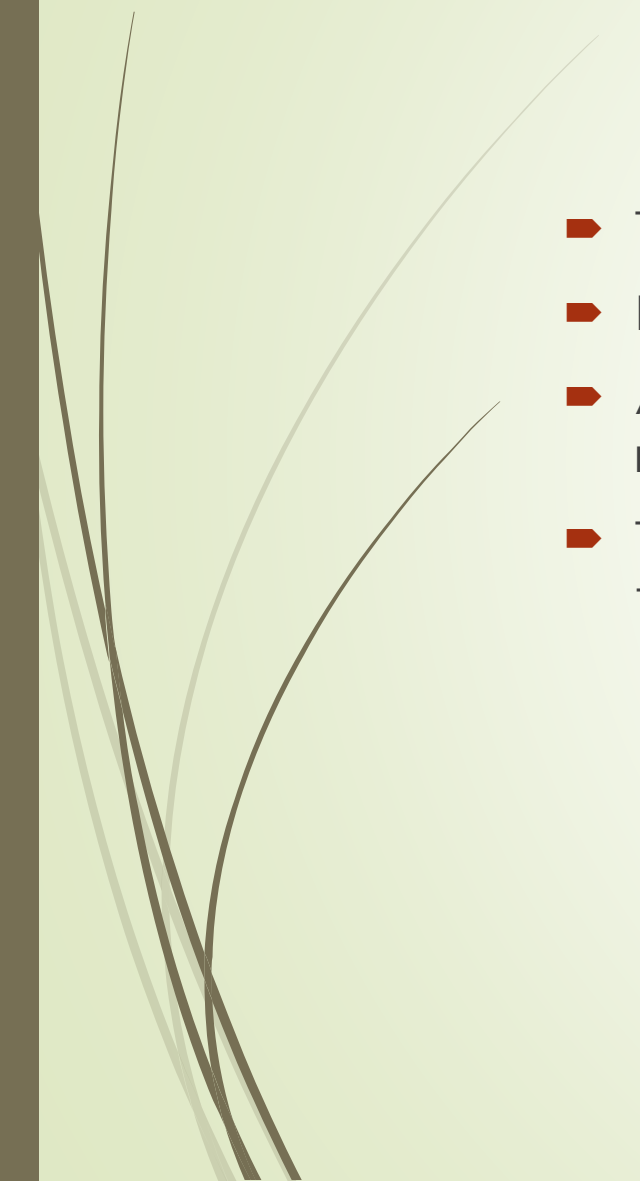
Objectives

- To identify the most popular journals to put Microsoft on the movie creation radar.

- To make projections on what genres can generate the most revenue for Microsoft

- To evaluate what studios Microsoft could use for a higher revenue.

- To establish the studios that Microsoft can collaborate with for high quality films, or see who their greatest competitors would be.
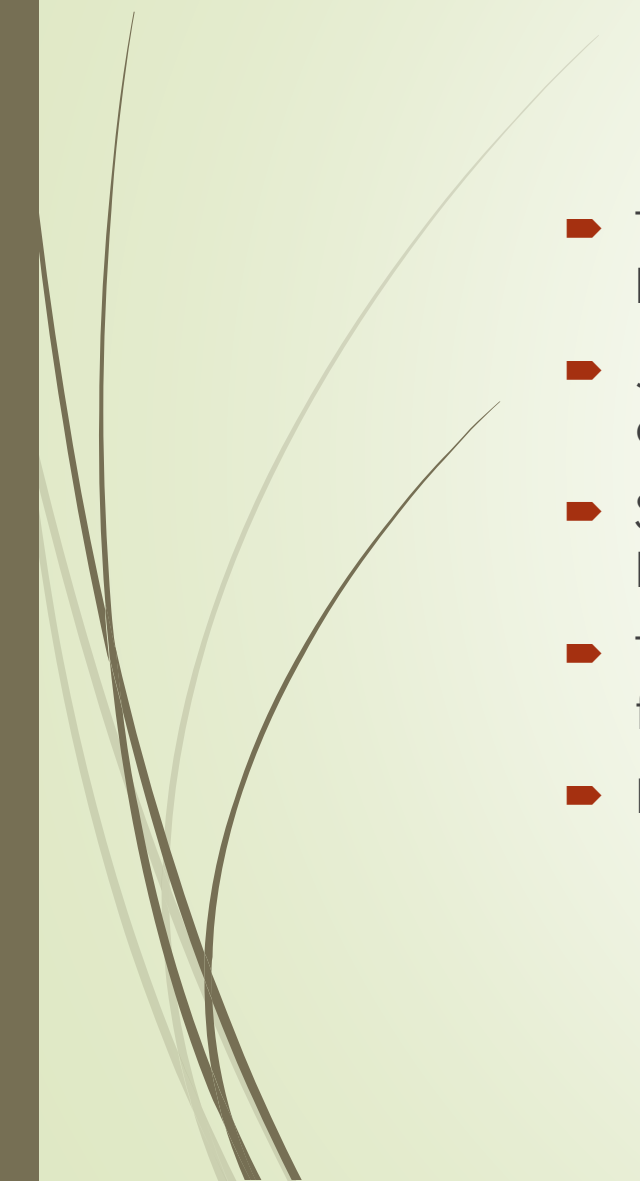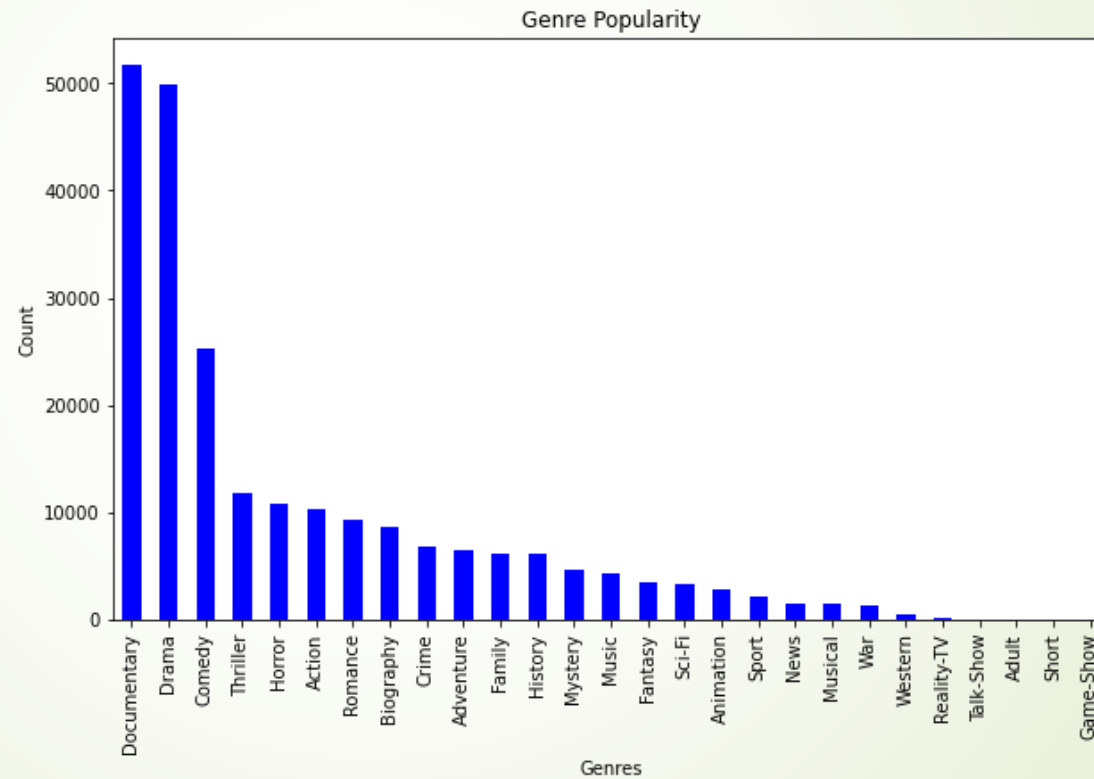
Data cleaning

- The first step is checking and cleaning the data.

- Data is cleaned by checking for any missing values and duplicates.

- After finding the missing values, the next decision to make is whether to remove, replace and keep the null values.

- The decision is made based on what data is required, and how the action taken impact all the data.
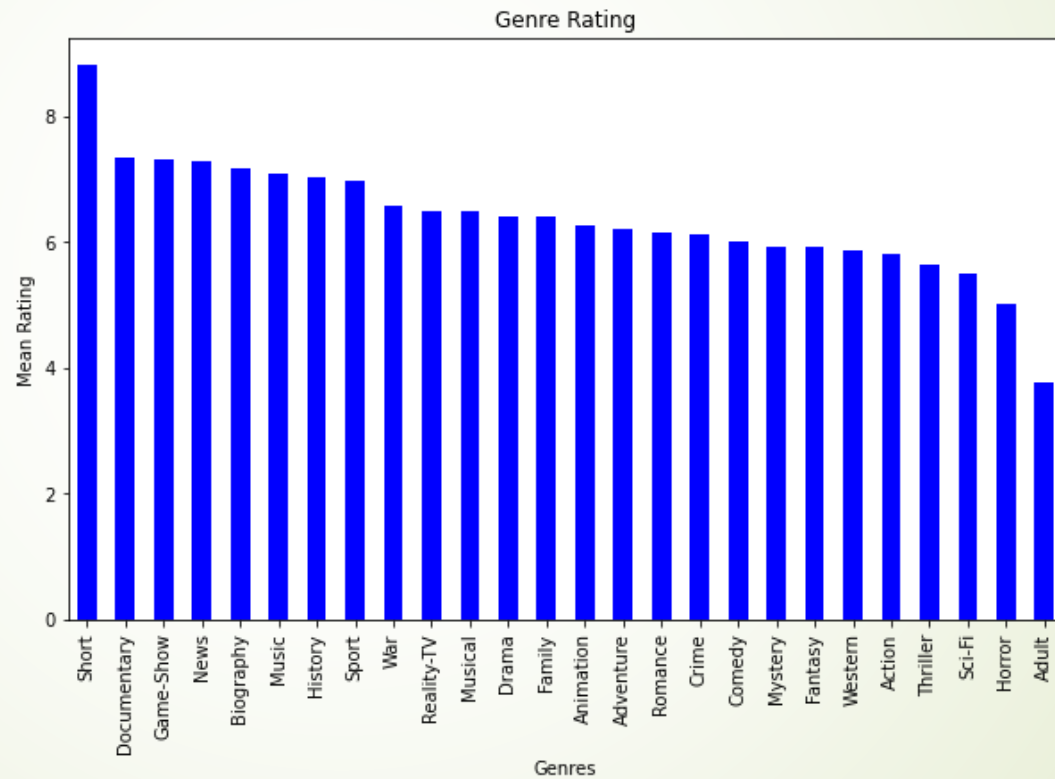
Data Analysis and Visualization

- The next step is for analyzing the data to get outcomes and results, and to help plot graphs.

- Joining tables helps in creating connections and comparisons between different datasets and dataframes.

- Sometimes comparisons are made by picking out columns or rows for a better analysis.

- There are several comparisons and connections used to make bar graphs for better visualization of the data.

- Below are some of the plots derived from the analysis:

A bar graph showing the cumulative genre popularity

A bar graph showing the cumulative mean genre rating

A bar Showing Studios with the Highest Gross Revenue

Top Total Revenue Generating Studios