Amazon Prime TV Shows and Movies

Amazon Prime TV Shows and Movies

Exploratory Data Analysis

Project Description

Business Context

In today's competitive streaming industry, platforms like Amazon Prime Video are constantly expanding their content libraries to cater to diverse audiences. With a growing number of shows and movies available on the platform, data-driven insights play a crucial role in understanding trends, audience preferences, and content strategy.

This data set was created to list all shows available on Amazon Prime streaming, and analyze the data to find interesting facts. This dataset has data available in the United States.

This dataset has 2 csv files and it is a mix of categorical and numeric values.

Dataset Description

This data set was created to list all shows available on Amazon Prime streaming, and analyze the data to find interesting facts. This dataset has data available in the United States.

This dataset has two files containing the titles (titles.csv) and the cast (credits.csv) for the title.

This dataset contains +9k unique titles on Amazon Prime with 15 columns containing their information, including:

- id: The title ID on JustWatch.
- title: The name of the title.
- show_type: TV show or movie.
- description: A brief description.
- release_year: The release year.
- age_certification: The age certification.
- runtime: The length of the episode (SHOW) or movie.
- genres: A list of genres.

- production_countries: A list of countries that produced the title.
- seasons: Number of seasons if it's a SHOW.
- imdb_id: The title ID on IMDB.
- imdb_score: Score on IMDB.
- imdb_votes: Votes on IMDB.
- tmdb_popularity: Popularity on TMDB.
- tmdb_score: Score on TMDB.

And over +124k credits of actors and directors on Amazon Prime titles with 5 columns containing their information:

- person_ID: The person ID on JustWatch.
- id: The title ID on JustWatch.
- name: The actor or director's name.
- character_name: The character name.
- role: ACTOR or DIRECTOR.

Problem Statement

This dataset was created to analyze all shows available on Amazon Prime Video, allowing us to extract valuable insights such as:

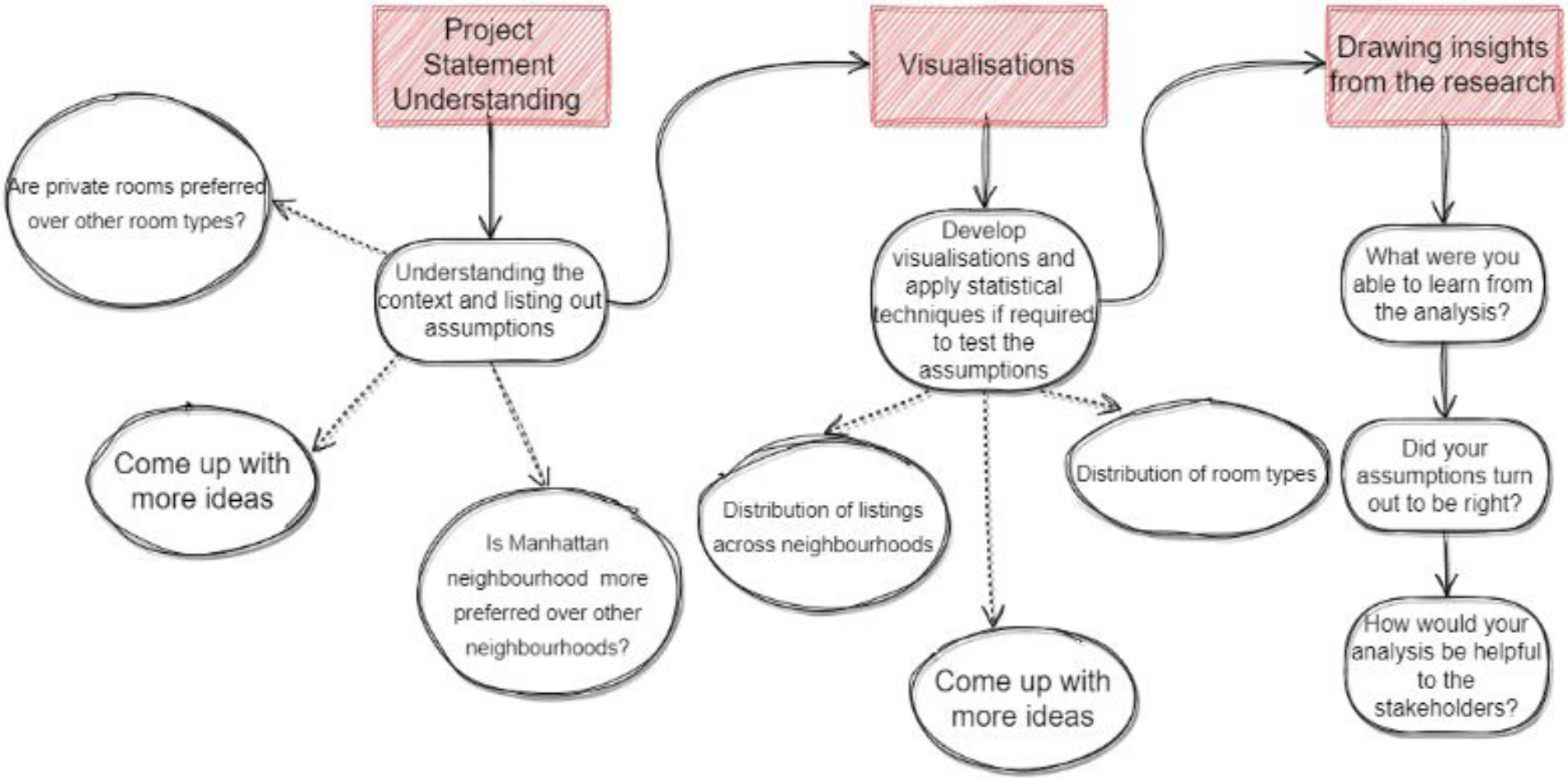
- Content Diversity: What genres and categories dominate the platform?
- Regional Availability: How does content distribution vary across different regions?
- Trends Over Time: How has Amazon Prime's content library evolved?
- IMDb Ratings & Popularity: What are the highest-rated or most popular shows on the platform?

By analyzing this dataset, businesses, content creators, and data analysts can uncover key trends that influence subscription growth, user engagement, and content investment strategies in the streaming industry.

Main Libraries to be used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualization and behavior with respect to the target variable. Use at least 5 different visualizations.
- NumPy for computationally efficient operations

Project Architecture:



Project Evaluation Criteria

- Efficient approach to the problem statement
- Data Exploration techniques and logic.
- Logic of handling missing values, and outliers.
- Visualization logic.
- Forming insights and explaining your understandings
- Understanding of how your project is useful to stakeholders.

Rubrics

Rubrics	Weightage
Summary and Technical Documentation in Collab Notebook	10
Exploration, Namely: Head, Tail, Summary, Data Dictionary	5
Looking for and Handling NaN/ Null/ Missing Values	5
Trying to Get Some Conclusion From Data, Correlation, Trends	10
Accomplish Various Milestones Given in the Problem Statement	10
Using Visualization (At least Five Different Types) for Presenting the EDA	15
Final Summary of Conclusion	5
Commented Code	5
Proper Output Formatting	5
Modularity of Code	5
Video Presentation	20
Fluency and Grammatical Accuracy in Video	5