

Part 2: Practical Application

4. Data Cleaning and Preparation

Problem Statement: Given a dataset with missing values and inconsistencies, clean and prepare the data for analysis.

Data Cleaning and Preparation Documentation

The following steps outline the process taken to clean and prepare the IMDB_Movies.csv dataset for analysis using Python. The steps address missing values and inconsistencies in the dataset.

Dataset link:

1. Imported Libraries and Load Data

2. Inspected the Data

- Viewed the first few rows
- Checked dataset shape
- Checked for missing values
- Got a summary of numerical statistics
- Checked data types

3. Handled Missing Values

- **'duration'**: Normally distributed, imputed with mean.
- **'num_critics_for_reviews'**: Not normally distributed, imputed with median.
- **'director_facebook_likes'**: Imputed with median.
- **Categorical columns**:
 - **'director_name', 'actor_2_name', 'actor_1_name', 'actor_3_name'**: Imputed with "Not mentioned".
 - **'color', 'plot_keywords', 'language', 'country', 'content_rating'**: Imputed with mode or "Not specified".
- **'actor_2_facebook_likes', 'actor_3_facebook_likes', 'actor_1_facebook_likes'**: Imputed with median.
- **'facenumber_in_poster'**: Imputed with median.
- **'Gross'**: Imputed with median.
- **'num_user_for_reviews'**: Converted to numeric and imputed with median.
- **'budget'**: Imputed with median.
- **'title_year'**: Interpolated linearly to fill missing values.
- **'aspect_ratio'**: Imputed with median.

4. Verified Missing Values

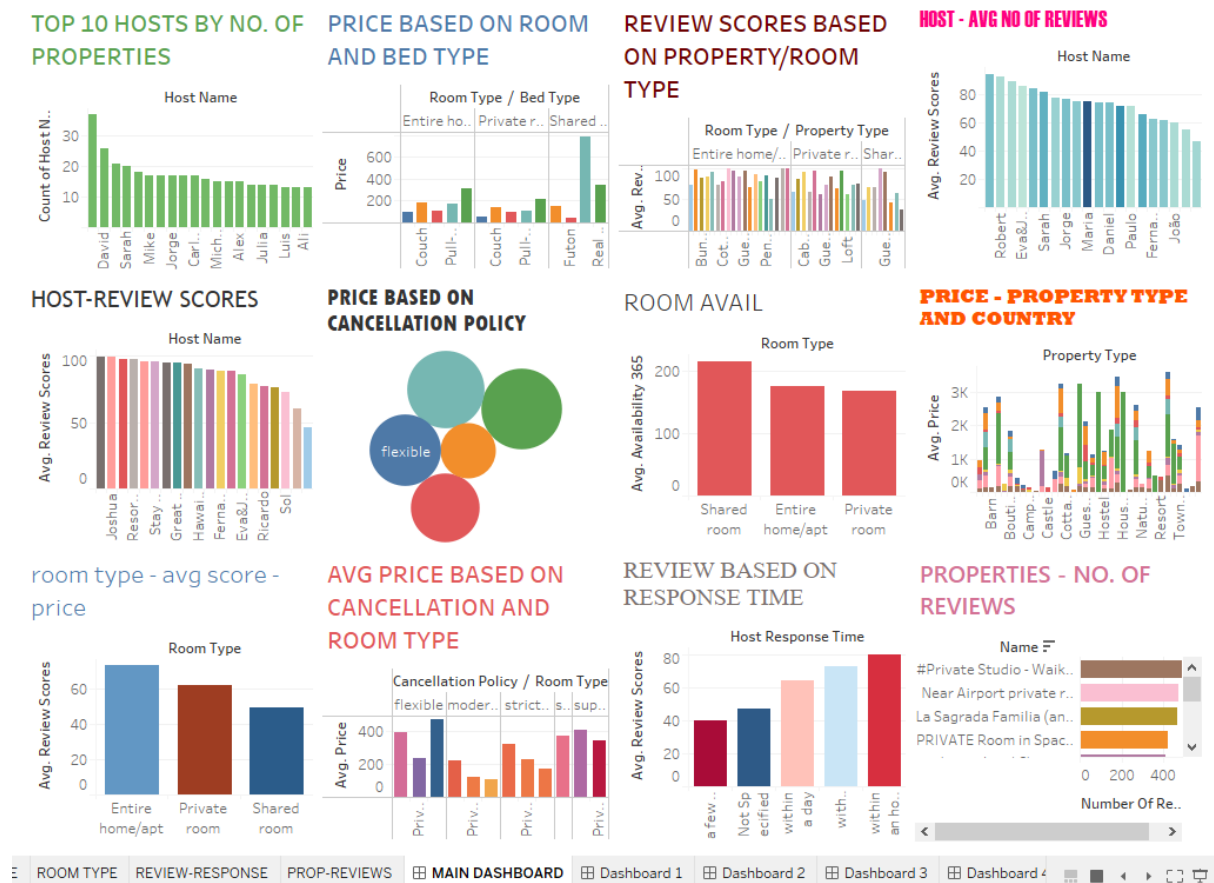
Summary of Actions:

- Imputed missing values in numerical columns using mean or median based on data distribution.
- Filled missing values in categorical columns with mode or a default value.
- Handled inconsistencies in data types and standardized values where necessary.
- Used interpolation for linear values and conversion to numeric types for specific columns.

5. Tableau Visualization

In this task, I have created an interactive Tableau dashboard to analyse Airbnb data, aimed at providing actionable insights into various aspects of the Airbnb rental market. The dataset includes information such as property types, locations, prices, and guest reviews, which I used to build visualizations that highlight key metrics and trends.

The final dashboard is shown below:



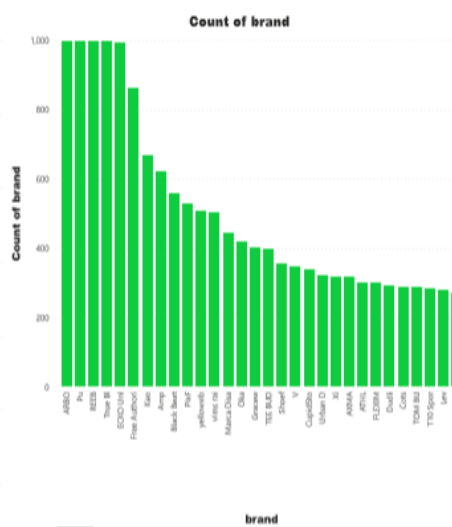
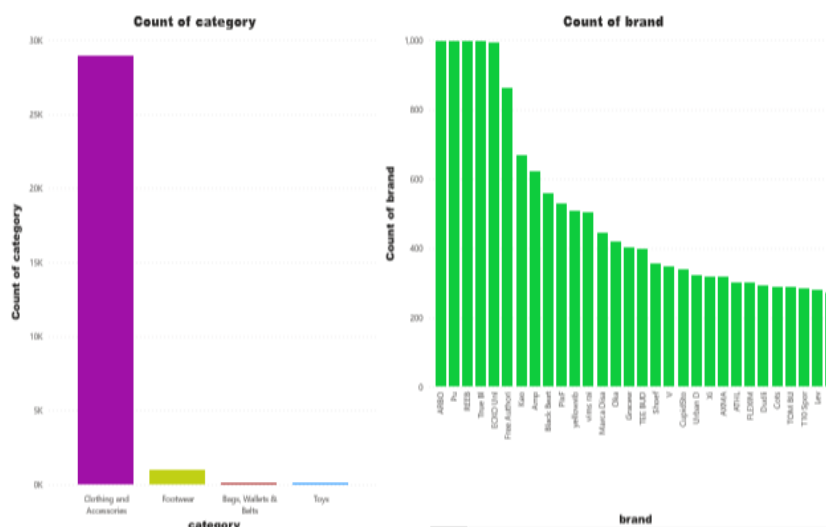
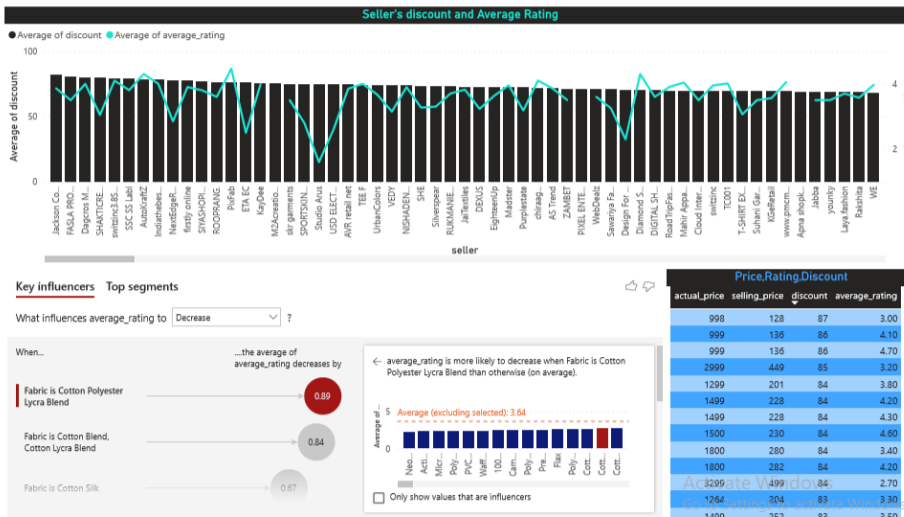
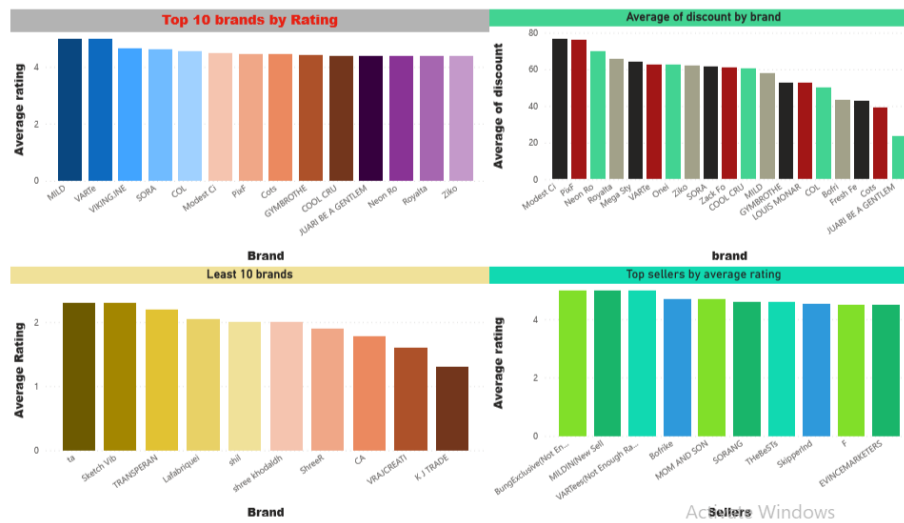
6. Power BI Report

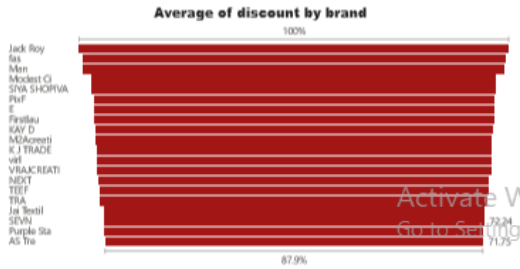
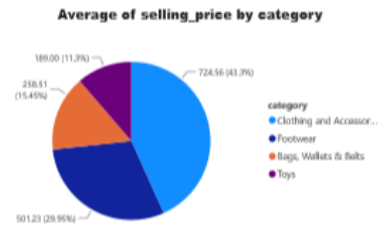
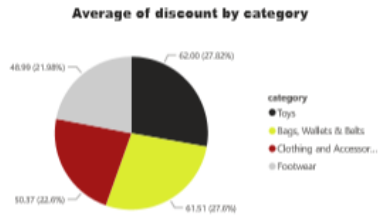
In this task, I have developed an interactive Power BI dashboard for analysing e-commerce data.

The objective is to provide a detailed overview of key performance metrics, trends, and insights that can drive strategic decisions in the e-commerce domain.

It provides insights into various aspects of product data collected from a major Indian e-commerce platform.

Here are the PowerBI visuals:





Activate V
Go to Settings