Name – Suraj Vishwakarma,

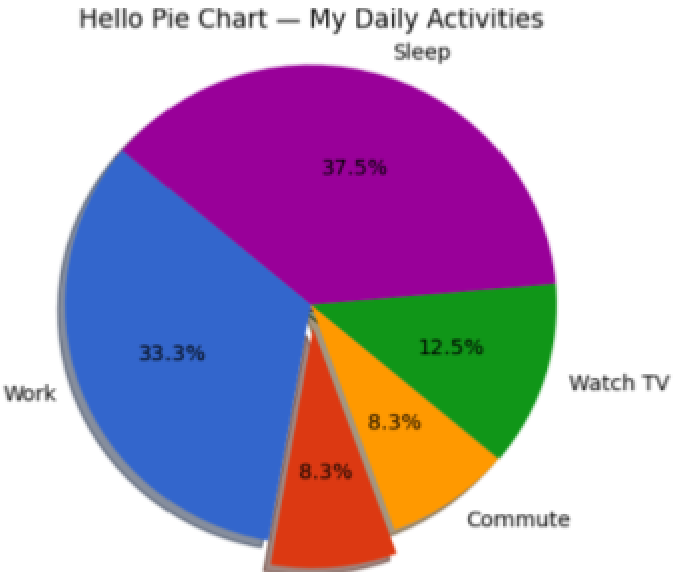
QID – 24510031

MCA 2nd Year

Assignment 1:

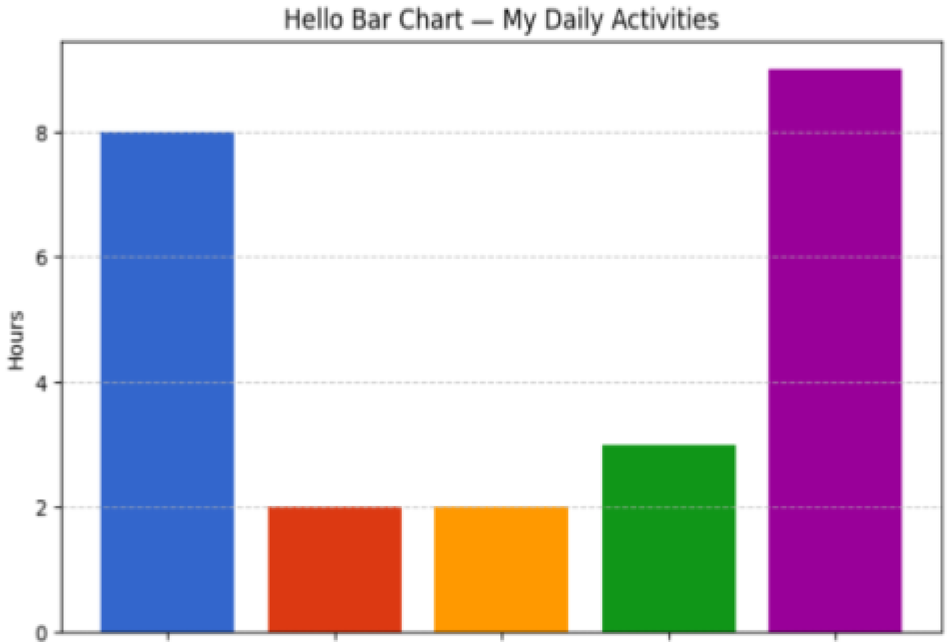
Hello Pie Chart

```
import matplotlib.pyplot as plt

activities = ['Work', 'Eat', 'Commute', 'Watch TV', 'Sleep']

hours = [8, 2, 2, 3, 9]

colors = ['#3366cc', '#dc3912', '#ff9900', '#109618', '#990099']

explode = (0, 0.1, 0, 0, 0)   # second slice highlight (Eat)

plt.pie(

    hours,

    labels=activities,

    colors=colors,

    autopct='%1.1f%%',

    shadow=True,

    startangle=140,

    explode=explode

)

plt.title("Hello Pie Chart — My Daily Activities")

plt.axis('equal')  # circle shape

plt.show()
```



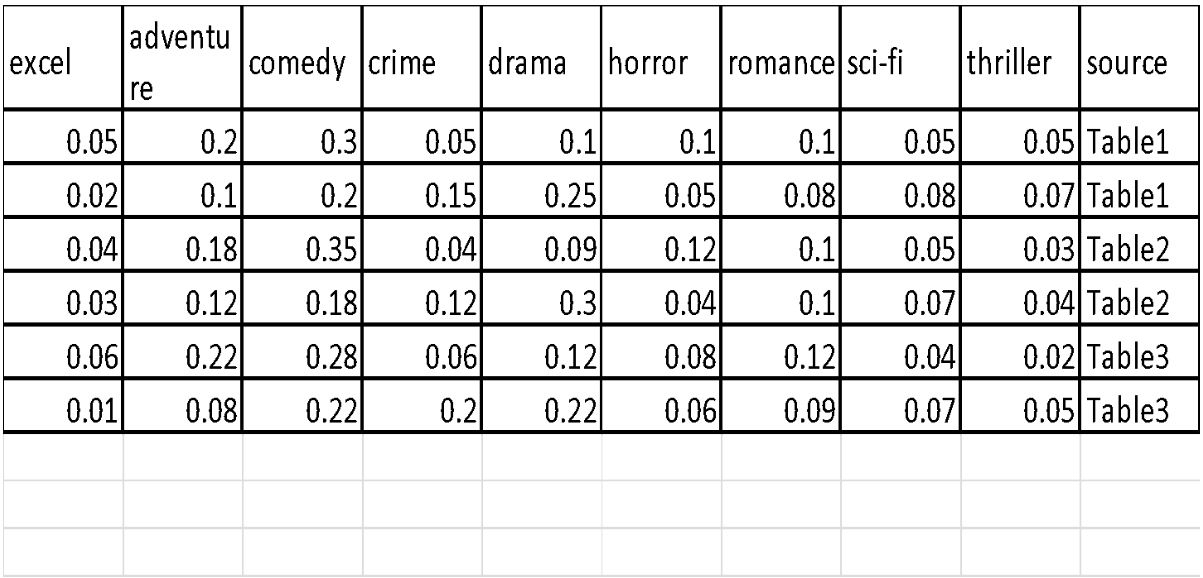
Hello Bar Chart:

```
import matplotlib.pyplot as plt

activities = ['Work', 'Eat', 'Commute', 'Watch TV', 'Sleep']

hours = [8, 2, 2, 3, 9]

plt.figure(figsize=(8,5))

colors = ['#3366cc', '#dc3912', '#ff9900', '#109618', '#990099']

plt.bar(activities, hours, color=colors)


plt.title("Hello Bar Chart — My Daily Activities")

plt.xlabel("Activity")

plt.ylabel("Hours")

plt.grid(axis='y', linestyle='--', alpha=0.7)


plt.show()
```



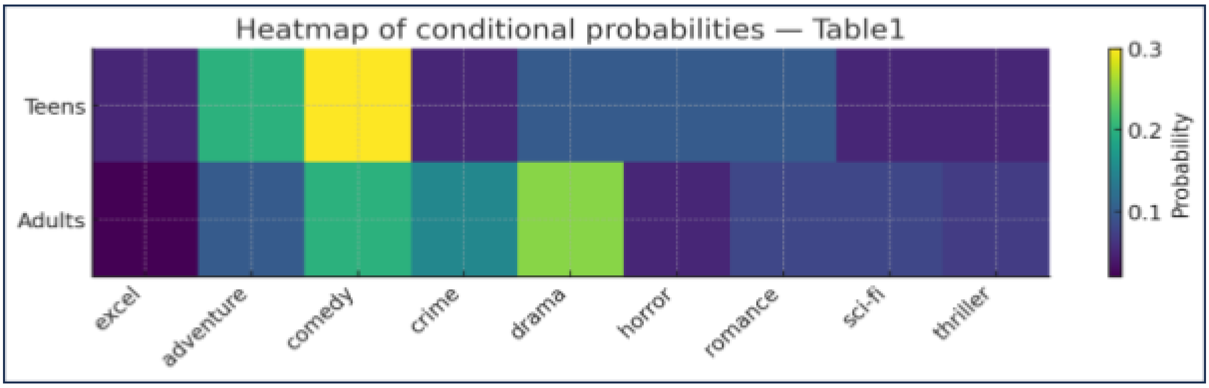
Assignment 2:

Conditional Probabilities for Movie Genres by Viewer Group

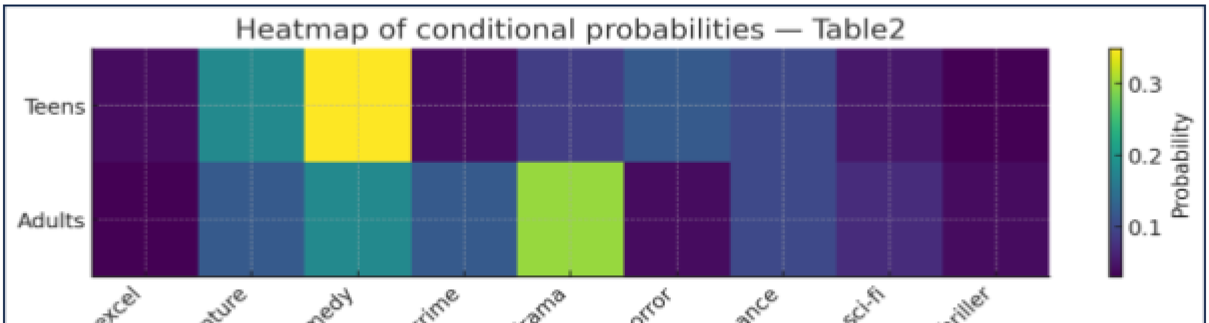


excel	adventure	comedy	crime	drama	horror	romance	sci-fi	thriller	source
0.05	0.2	0.3	0.05	0.1	0.1	0.1	0.05	0.05	Table1
0.02	0.1	0.2	0.15	0.25	0.05	0.08	0.08	0.07	Table1
0.04	0.18	0.35	0.04	0.09	0.12	0.1	0.05	0.03	Table2
0.03	0.12	0.18	0.12	0.3	0.04	0.1	0.07	0.04	Table2
0.06	0.22	0.28	0.06	0.12	0.08	0.12	0.04	0.02	Table3
0.01	0.08	0.22	0.2	0.22	0.06	0.09	0.07	0.05	Table3

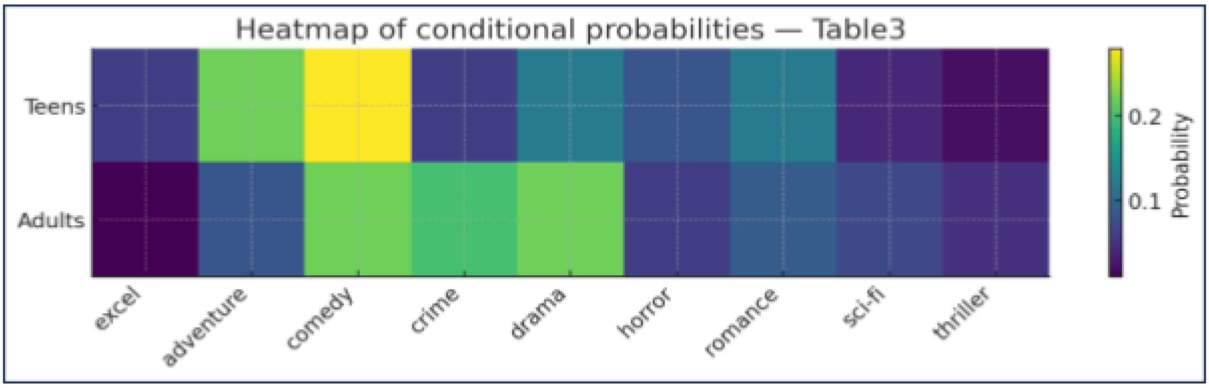
Heatmap of conditional probabilities — Table 1



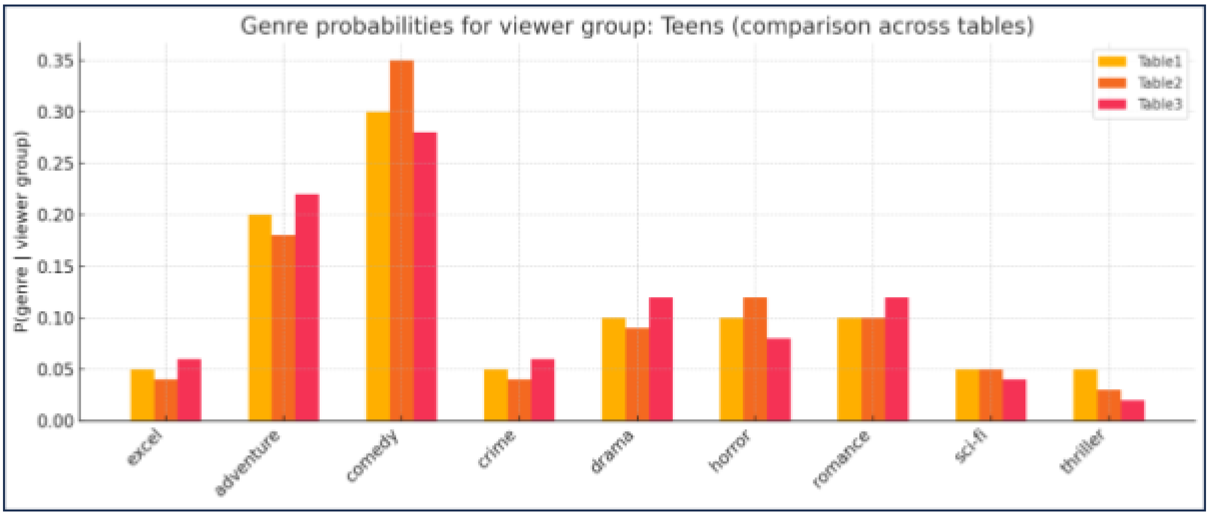
Heatmap of conditional probabilities — Table 2



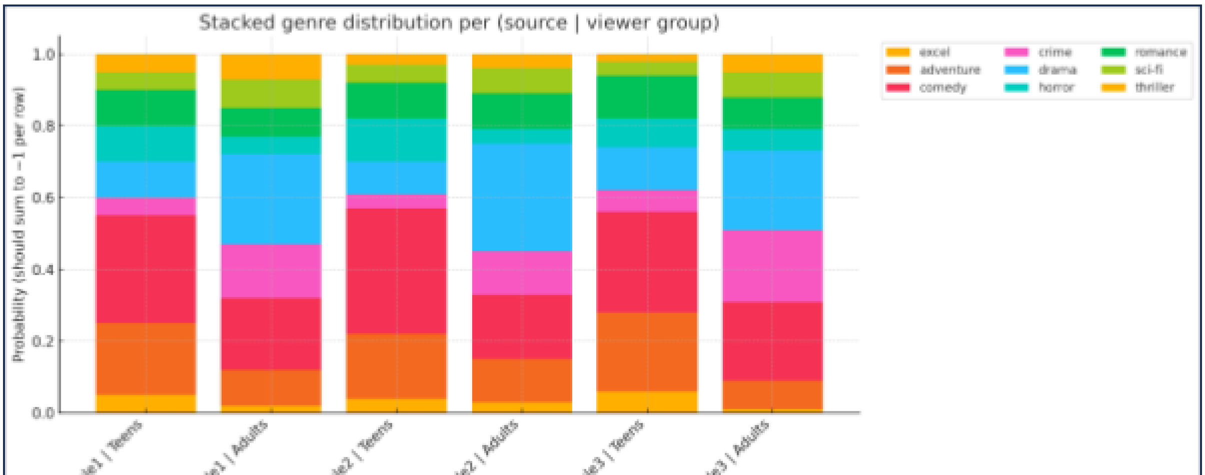
Heatmap of conditional probabilities — Table 3



Genre probabilities for viewer group: Teens (comparison across table)



Stacked genre distribution per (source | viewer group)



Assignment 3:

Regression Assignment - Exercise & Instructions

Objective

1. Learn the basics of regression (a supervised technique).

2. Compare several regression-based algorithms.

3. Understand how different metrics are used to evaluate regression.

4. Learn how to use Python libraries/packages.

5. Familiarize yourself with data manipulation.

Resources

- Regression Analysis

- Linear Regression Model

- Simple Linear Regression

- Regression Analysis - Step by step

Data Sources

- Kaggle

- UCI / Irvine ML Repository

- KDD

- Delve

- Best Free Dataset

Exercise Intro

This exercise will test your ability to train, develop and deploy a machine learning regression model. The dataset consists of 27 features and a label. Notice that several values are either missing or duplicated.

Exercise Steps

1. Download the dataset from Kaggle.

2. Perform data exploration:
  a. Display the first few records in the dataset.
  b. Display the number of rows and columns of the dataset.
  c. Display the dataset statistics (min, max, mean, etc.).
  d. Display the null value counts for each feature.
  e. Plot graphs to assist in data exploration (histograms, scatter plots, box plots, etc.).

3. Initial cleaning:
  a. Delete columns which mainly contain null values.
  b. Remove duplicate columns.
  c. Fill missing values in numeric columns if necessary (mean/median or other strategies).
  d. Display the number of rows and columns after cleaning.
  e. Display the features left after cleaning.
  f. Plot distributions (histograms) for selected features.

4. Analyze pairwise relationships between features (scatter matrix / pairplot).

5. Plot correlation heatmap from the pairwise plots.

6. Perform necessary data preprocessing (transformations) for final preparation:
  a. Scale numeric columns to [0,1] range if needed.
  b. Encode categorical data into one-hot vectors.
  c. Split the dataset into training, validation, and testing sets.

7. Perform hyperparameter tuning on models and compare different regression models (e.g., Linear Regression, Ridge/Lasso, Decision Tree, Random Forest, Gradient Boosting, SVR).

Notes

- Dataset has 27 features and may contain missing or duplicated values — inspect carefully.

- Keep copies of intermediate cleaned datasets so you can reproduce results.

- Record model performance using regression metrics such as MAE, MSE, RMSE, and R^2.