LAB 4 Assignments

Peform the given tasks in a Jupyter Notebook.
Also include Introduction, Theory and Conclusion in the same notebook.

Problem 1

On the same data that you performed wrangling on in the lab:-

- Clean the STARS column to extract new Directors and Actors columns.
- Then obtain a rating of each actor, by averaging the rating across all the movies that they've played in. And then plot the ratings of thus obtained top 10 actors.

Problem 2

Download the data for this assignment at:

- https://www.kaggle.com/datasets/tonygordonjr/spotify-dataset-2023/data

Tasks:

1. Import all 5 CSV files as Pandas DataFrames, print their schemas,
   then perform data cleaning if found to be necessary.
2. In `spotify-artist-data_2023.csv`:
   - Count the number of distinct artists.
   - Count the number of distinct artist genres.
   - Display a bar graph showing follower count of top 10 artists with the most followers.
3. In the `spotify_features_data_2023.csv`:
   - Calculate basic statistics of each feature (quartiles, min, max, variance etc.)
   - Plot the correlation between each feature using a heatmap.
   - Can any conclusion can be drawn from the heatmap?
4. In the `spotify_tracks_data_2023.csv`:
   - Find the top 10 tracks with highest popularity.
5. In `spotify-albums-data_2023.csv`:
   - Count total number of distinct albums.
   - Draw a pie chart showing distribution of album types (single, compilation or album).
   - Find the album with the most tracks.
   - Find the top 10 popular albums.

- Plot the average track duration of the top 10 albums in a bar graph.
- Plot number of tracks per album for the top 10 albums in a bar graph.

6. In the joined data `spotify_data_12_20_2023.csv`:
    - Find top 10 popular tracks.
    - Count the number of tracks of each artist.
    - Identify the artists with the highest average danceability.
    - Plot the average popularity of each genre.
    - Display line plot for the total number of album releases each year after 1920.
    - Display line plot for average album duration each year after 1920.