Python Report

Assignment 1 - Exploratory Data Analysis

Python is a high-level, interpreted, and versatile programming language that stands out for its simplicity and readability. Created by Guido van Rossum, Python has become immensely popular, serving as a favorite choice among developers worldwide. Its clean syntax and indentation-based code structure make it easy to learn and write, promoting better collaboration among programmers. Python's dynamism allows for faster development, as it does not require explicit variable type declarations.

**Libraries used in this assignment are Pandas, matplotlib.pyplot.**

**Task 1A: Dataset Details**

1. The dataset used for this project consists of a list of every Netflix movie and TV program and information on the cast, directors, ratings, year of release, length, and other factors. The "movies\_data" dataset is loaded using Python's Pandas package.

**Output**:

A screenshot of a computer

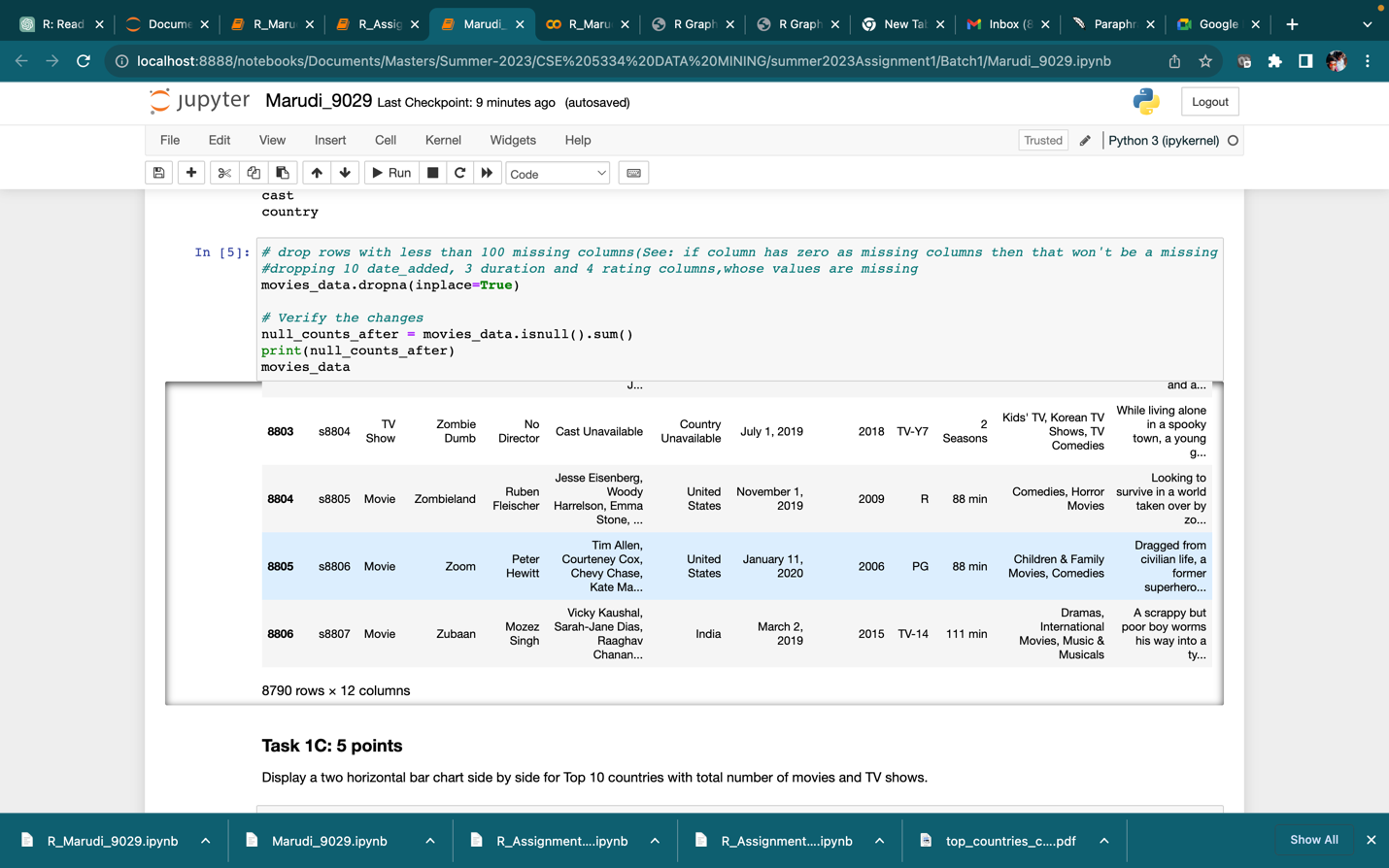
Description automatically generated

**Task 1B: Handling Missing Values**

1. We used Pandas' "isnull()" and "sum()" methods to check for null values in each column of the dataset. There are blank values in certain columns.

2. For certain columns with more than 100 missing data, we have filled in the blanks. We filled up the blanks for columns like "director," "country," and "cast" with titles like "No Director," "Country Unavailable," and, accordingly, "Cast Unavailable."

3. All of the remaining missing fields, such as "date\_added," "duration," and "rating," have had their corresponding rows deleted.



**Task 1C: Top 10 Countries with Movies and TV Shows**

1. The top 10 nations with the most films and television series are included. The number of films and TV series produced in each nation was represented visually using horizontal bar charts.

A screenshot of a computer

Description automatically generated

**Task 1D: Movies with Longest Duration from Each Country**

1. The director, date\_added, release\_year, length, and description of the movie from each nation with the longest runtime are shown

A screenshot of a computer

Description automatically generated

**Task 1E: Movies with Same Release and Added Year**

1. The names of the films with the same year of release and date of platform addition are

displayed.

A screenshot of a computer

Description automatically generated

**Task 1F: Directors and Their Movies/TV Shows Count**

1. The names of the directors, the year of their release, and the quantity of motion pictures and television episodes they were responsible for are all displayed. According to the overall count, the results are arranged in descending order.

A screenshot of a computer

Description automatically generated

**Task 1G: Documentary/Docuseries Movies and TV Shows**

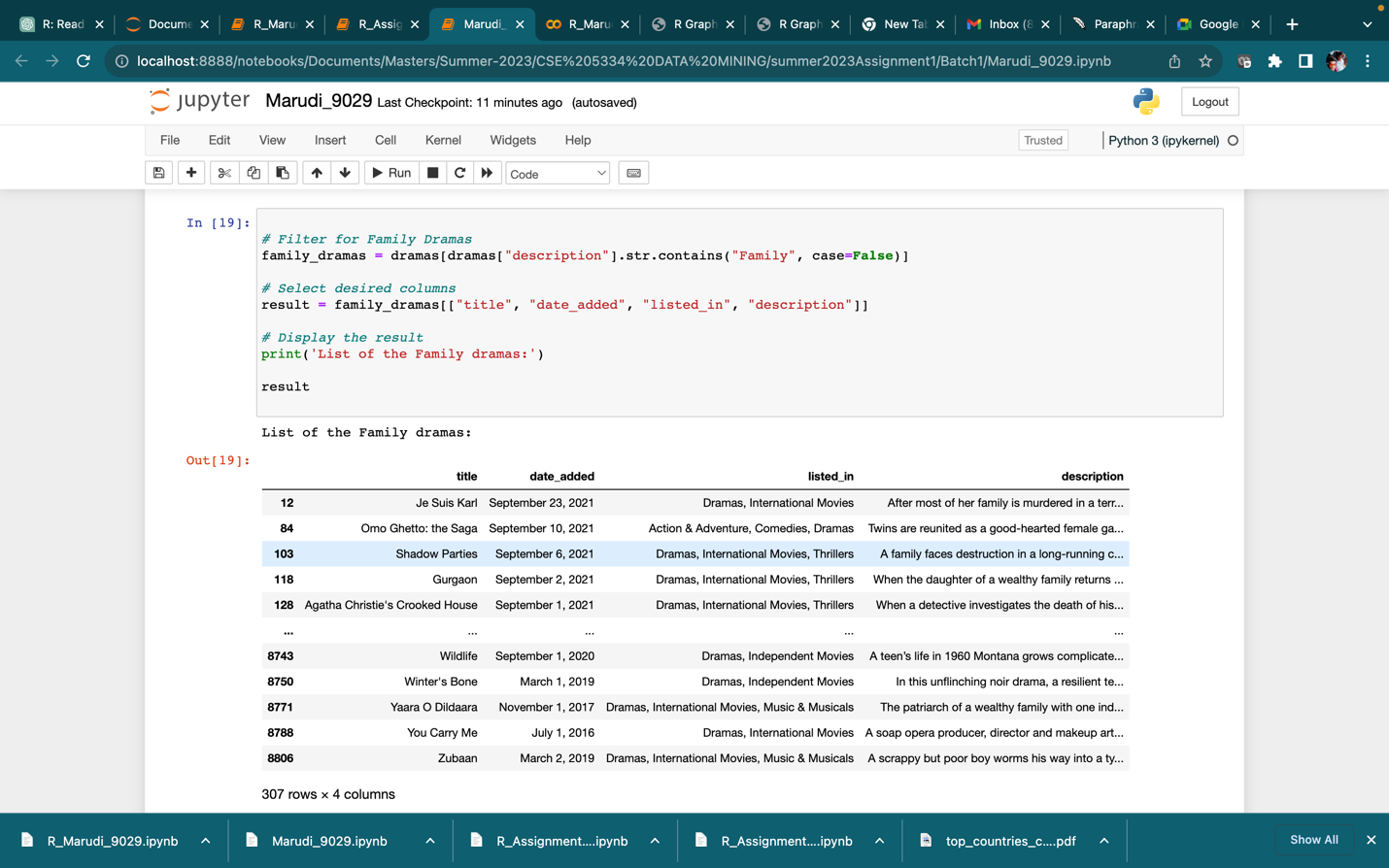
1. We have displayed the titles of the documentary and docuseries movies and TV shows, along with their directors, the date they were added to Netflix, and their category.

A screenshot of a computer

Description automatically generated

**Task 1H: Family Dramas**

1. We have displayed the titles of the Family dramas, along with their directors, the date they were added to the platform, and the category they are listed in.



**Task 1I: Distribution of TV Shows Based on Number of Seasons**

1. We have plotted a horizontal bar chart to show the distribution of TV shows based on the number of seasons. The TV shows are categorized into groups based on the number of seasons they have.

A screenshot of a computer

Description automatically generated

**Task 1J: Pie Chart for Movie and TV Show Ratings**

1. We have displayed a side by side pie chart showing different ratings for movies and TV shows.

