CSS Project - Option 1: IMDB Data

Started: 3 Feb at 17:56

Quiz instructions

Answer the following quiz questions.

Only python files will be accepted when you submit your code, i.e .py files.

If you submit a jupyter notebook you will get zero.

As a researcher, you are tasked to do use ETL and EDA skills on a movie dataset to extract certain insights.

Using Pandas, use the "movie_dataset.csv" found in the "Week 1" section of Canvas.

Note, some column names have spaces which is not ideal. Some columns have missing values and it would be best to either fill or drop where appropriate those missing values to prevent a bias. Load and clean the data, make reasonable assumptions.

Once this has been done, answer the Quiz questions below.

Question 1   1 pts

What is the highest rated movie in the dataset?

○ The Dark Knight

○ Rogue One

○ Trolls

○ Jason Bourne

Question 2   1 pts

What is the average revenue of all movies in the dataset?

Note, since the answer will be effected by how you dealt with missing values a range has been provided.

○ Between 120 and 140 Million

○ Between 100 and 120 Million

○ Between 70 and 100 Million

○ More than 200 Million

Question 3   1 pts

What is the average revenue of movies from 2015 to 2017 in the dataset?

Note, since the answer will be effected by how you dealt with missing values a range has been provided.

○ More than 120 Million

○ Between 50 and 80 Million

○ Less than 40 Million

○ Between 100 and 120 Million

Question 4   1 pts

How many movies were released in the year 2016?

○ 297

○ 301

○ 113

○ 287

Question 5   **1 pts**

How many movies were directed by Christopher Nolan?

○ 4

○ 7

○ 5

○ 6

Question 6   **1 pts**

How many movies in the dataset have a rating of at least 8.0?

○ 81

○ 98

○ 78

○ 45

Question 7   **1 pts**

What is the median rating of movies directed by Christopher Nolan?

○ 7.4

○ 6.9

○ 8.6

○ 9.1

Question 8

1 pts

Find the year with the highest average rating?

○ 2006

○ 2016

○ 2008

○ 2007

Question 9

1 pts

What is the percentage increase in number of movies made between 2006 and 2016?

○ 421

○ 210

○ 575

○ 57

Question 10

3 pts

Find the most common actor in all the movies?

Note, the "Actors" column has multiple actors names. You must find a way to search for the most common actor in all the movies.

○ Mark Wahlberg

○ Chris Pratt

○ Bradley Cooper

○ Matthew McConaughey

Question 11

3 pts

How many unique genres are there in the dataset?

Note, the "Genre" column has multiple genres per movie. You must find a way to identify them individually.

○ 207

○ 15

○ 20

○ 30

Question 12

10 pts

Do a correlation of the numerical features, what insights can you deduce? Mention at least 5 insights.

And what advice can you give directors to produce better movies?

Edit   View   Insert   Format   Tools   Table

12pt ∨   Paragraph ∨ | **B**   *I*   U̲   A ∨   ✎ ∨   T² ∨ | ⋮

p   0 words   </> ↗ ⋮

Question 13   **1 pts**

Once you have completed the Quiz questions, create a public GitHub repository and upload a single python file called "css4p01.py" to it. Share the url ink for the your python file below. The code must show all the code used to load, analyse and clean the data, as well how you answered the Quiz questions. Make reasonable assumptions.

Edit   View   Insert   Format   Tools   Table

12pt ∨   Paragraph ∨ | **B**   *I*   U   A ∨   ✎ ∨   T² ∨ | ⋮

p   0 words   </> ↗ ⋮

Saved at 18:14   Submit quiz