Data Science Midterm Project / Exam
Due: Oct. 24th, 11pm

Attached is a dataset that shows list of Netflix movies, TV shows, etc.

I would like the class to conduct various types of analyses on this dataset. I will highlight what I expect, but I don't want this project to be based on only what I have requested.

Use Numpy, Pandas, stats, etc. Examples of libraries imported are in Jupyter Notebook files sent to you in previous sessions.

Use your creativity, experience, and understanding to be curious and analyze the dataset.

When you enter any organization, or start on any new data analysis project, you will need to gain understanding on what's there, and then bit by bit gain insights.

I want to see Python code, I want to see visuals (charts, plots, lines), and I want to see results, and your insights. Being able to translate your data into insights is gold. Practice this as much as possible.

So, some areas I want you to highlight:

Data cleanup and prep:
What does data look like?
What data is missing?
Why is data missing for some movie, shows, and not for others?
If data missing, what should I replace it with so that data makes sense?
Are dates in correct format?
Are there duplicate rows? Are you sure they are duplicates?
What else should I do to get good overview of data before I start digging in?
Show me details on what else you did to prep, and cleanup data.

Descriptive Statistics:
Find out various types of descriptive stats. You have seen examples of various types.
What do you find about the data when you run descriptive stats through it?

Distribution:
Tell me about data distribution. For example:

- Data has list of movies and TV shows.
- How many movies, how many TV shows? Which is more?
- Display data based on ratings, durations, by country, etc.
- Is there any one TV show that is popular? Is there any one movie that is popular?

Visualize:

- What is the most common genre?
- Look at length of movies.
- Are there more movies or shows produced in certain years?
- Display using pie chart, bar graph, any other chart you feel is useful to show

Relationships:

- Use correlation to show relationships for:
    - Type vs Rating
    - Genre vs Rating
    - Type vs Release Year
    - Which other correlation is useful?

- Are there positive correlations for which Director produces which type of Genre?
- What type of correlations show negative relationships?
- Show me using charts, graphs.

Your Insights:

- In addition to above guidelines, what additional analysis can you do to show me more insights about data?
- What additional correlations could show meaningful insights?
- What additional visualizations should show meaningful insights?
- There is a Description column that holds text. What analysis can we do with this text? Should we look for word frequency? Can we do relation testing with Description and Genre? What would make sense?