Project Proposal

Project Title: Netflix Binge

Team Members: Bao Huynh, Chadwick Nguyen, Mary Colasanto, and Terril Vallikalam

Project Description/Outline: Analyze a 2019 dataset of Netflix movie and tv show titles

Research Questions to Answer:

1. Distribution of Movies and TV Shows Over Time
    a. Number of movies vs. tv shows
    b. Bin movies/show by year
    c. Length of movies/shows vs. release date
    d. Ratings/genres of films made over time
2. Country Collaboration
    a. How many movies and tv shows have multiple countries involved?
    b. Which country produced the most movies and tv shows?
    c. Comparison of US and rest of the world movie and tv show production
    d. Country of origin vs. when it was added to Netflix
    e. See which genre is popular in each country

Datasets to be Used:

- https://www.kaggle.com/shivamb/netflix-shows?select=netflix_titles.csv

Rough Breakdown of Tasks:

- Everyone: Clean data!!!!!!
- Bao: Country Collaboration
- Chadwick: Distribution of Movies and TV Shows Over Time

- Mary: Distribution of Movies and TV Shows Over Time
- Terril: Country Collaboration

Schedule:

Thursday 01/28: Clean Data.

Saturday 01/30: Everyone works on their parts.

Tuesday 02/02: Analysis should be finished.

Thursday 02/04: Have Powerpoint finished.

Saturday 02/06: Project 1 Due and Presentation

Technical Requirements

The technical requirements for Project 1 are as follows.

* [ ] Use Pandas to clean and format your data set(s)

* [ ] Create a Jupyter Notebook describing the **data exploration and cleanup** process

* [ ] Create a Jupyter Notebook illustrating the **final data analysis**

* [ ] Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)

* [ ] Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation

* [ ] Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions

* [ ] Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots.

Presentation Guidelines

You are free to structure your presentations to your liking, but students tend to have success with the following format.

* Title Slide

* Include the name of the Project and Group Members

* Motivation & Summary Slide

  * Define the core message or hypothesis of your project.

  * Describe the questions you asked, and _why_ you asked them

  * Describe whether you were able to answer these questions to your satisfaction, and briefly summarize your findings
 * Questions & Data
   * Elaborate on the questions you asked, describing what kinds of data you needed to answer them, and where you found it
 * Data Cleanup & Exploration
 * Describe the exploration and cleanup process
  * Discuss insights you had while exploring the data that you didn't anticipate
  * Discuss any problems that arose after exploring the data, and how you resolved them
  * Present and discuss interesting figures developed during exploration, ideally with the help of

Jupyter Notebook
 * Data Analysis
  * Discuss the steps you took to analyze the data and answer each question you asked in your proposal
  * Present and discuss interesting figures developed during analysis, ideally with the help of Jupyter Notebook
 * Discussion
  * Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?
 * Post Mortem
  * Discuss any difficulties that arose, and how you dealt with them

  * Discuss any additional questions that came up, but which you didn't have time to answer:

What would you research next, if you had two more weeks?

* Questions

  * Open-floor Q&A with the audience