Final Project Instruction

MSDS 593 Summer 2024

For the final project, you will choose to work on one of the choices below:

- Most Streamed Spotify Songs 2024
  This dataset presents a comprehensive compilation of the most streamed songs on Spotify in 2024. It provides extensive insights into each track's attributes, popularity, and presence on various music platforms, offering a valuable resource for music analysts, enthusiasts, and industry professionals. The dataset includes information such as track name, artist, release date, ISRC, streaming statistics, and presence on platforms like YouTube, TikTok, and more.

- Compare Netflix and Hulu Shows
  Under the page, it provides the data cards from several different streaming platforms. We will focus on the data from Netflix and Hulu.

- Police Arrest from 2021-2023
  This dataset contains information about individuals who have been arrested by police officers in Montgomery County. The data includes demographic information about the individuals arrested, such as their race, gender, age, and ethnicity.

You can sign up for the project you want to do here. All group members need to agree before you sign up for the project. Each data has 6 spots to fill per session, first come, first serve. **The deadline for submitting your choice is at midnight on Sunday, July 28. Your choice will be final.**

For each project, the goals are the same. The audience (including myself) is from the business team of the company or residence of the county. We know the data was collected but have never seen the data, or any descriptive information about this data before. It's your job to provide

- A high-level summary of the data through EDA.
- Some highlight facts about the data through well-designed visualizations.
- Some ideas about how to use this data in business or policy from what you have learned.
- Present the above to the audience.

Your grade will be based on the following rubric:

Signup for a data	**Due Sunday 07/28 at midnight**	**Failing to do so will cause the missing submission for the whole project**
Peer Review	**3 pts** **Due Tuesday 08/06 at midnight**	**Each individual student is required to give a peer review to your groupmate.** With a total score of 4, your final grade of this part will be based on the average of the peer scores your teammates give to you. If you give a groupmate with score <4, please include why you think your teammate should not have a score 4. To avoid conflict, please do not communicate this part with your groupmates.
Presentation Submission	**8 pts** **Due Tuesday 08/06 at midnight** ***Group Submission***	Performed EDA and provided summary descriptions of the data or findings (2 pts) Provided 3 to 6 visualizations meeting requirements below (5 pts) The presentation is organized nicely and easy to read. (1 pts) *Detailed requirements for the analysis are below*
Attendance of all presentation sessions	**1 pts**	
6-minute Oral Presentation	**8 pts** **In-class on Thursday 08/08** ***Group Presentation***	You are going to present your work to the class in 6 minutes. You can not make changes to your slide anymore after the submission. Every group member needs to present a part. The students from other data projects will vote on the presentations by choosing 2 top favorites from each data. I will rank the teams

		by the number of votes: **Top 2: get full 8 points** **3-4: get 7 points** **5-6: get 6 points**

Requirements for analysis included in your presentation submission:

- You will be only graded on EDA and visualization results. It's your freedom to run any predictive models or statistical tests if you feel it helps with your analysis, but that would not be part of the grading rubric.
- You will be submitting a file of your choice for the slides, which should NOT include any code; you will also submit a python script/notebook(s) separately. Failing to submit either will result in losing points.
- You have all the freedom about how to do your EDA and you need to include some summary findings from EDA.
- You need to include between **3 to 6 visualizations** to present ***different key*** information you want the business team to know. Each visualization should include a brief summary of information presented.
- For any visualization you choose to include in the report, it needs to
    - Have a nice color scheme
    - Use the correct data for the information
    - Have a 6-12 word descriptive title
    - Have horizontal or vertical texts
    - Use the Gestalt principle of simplicity, avoid clusters
    - Apply some other Gestalt principles to highlight key information
- You can design the presentation however you see fit, but it needs to have an organized outline.
- This data is public, so some work was done on it. It's okay to read them to find inspiration but you are strictly forbidden to present a similar graph with similar codes (more than 50% similarity).
- It's your freedom to add supplemental data/information that is related to your topic, but that would not be required.