**Hi, my name is Wentao Sun and welcome to this presentation video where I’ll walk you through the steps of this data visualization on Netflix shows.**

**explaining why you chose the dataset**

As a computer science major, my fascination isn't confined to the code I write or the algorithms I study; it extends to the tales told by data. To me, datasets are like novels waiting to be read, each number a character, each trend a plot twist. My leisurely dives into Netflix's vast array of shows have always been my escape from the university grind, but these shows mean more to me than just screen stories. They're a constellation of data points, each holding a secret to viewer behaviors, industry shifts, and cultural preferences.

This blend of personal interest and academic zeal drives me to create a visualization on Netflix shows data concerning countries, titles, directors, and genres to better understand how the shows are rendered. Beginning with the extraction of data as a CSV file from the website, the task for me is to translate this raw data into a visual narrative using PowerBI. The process is initiated with loading the data into the software, followed by customizing the aesthetic elements, choosing graphs, such as the background color and transparency, to ensure clarity and visual appeal.

**walk you through the dashboard**

The first step in my data visualization journey was the implementation of a map graph. By meticulously selecting the distinct countries from the 'country' column within the data table, I could harness the geographical distribution of Netflix's content. The map was crafted with a gradient color scheme to represent the intensity of the show's presence in each region. Shades of deep red illustrated countries with a higher density of shows, while lighter color indicated fewer. This feature not only provided an immediate visual representation of the data but also included interactive elements, if you hover the cursor over any nation would reveal the precise number of shows available there. The map's interactive zoom functionality was vital, by zooming in/out, this allows a transition from a broad overview to a detailed inspection of the data. We can see in this chart that the US hosts the most shows, followed by India, then UK, this can potentially suggest audience in these countries use Netflix the most when it comes to watching shows/movies.

I then shifted focus to a donut chart, a powerful tool for displaying part-to-whole relationships. This chart was instrumental in segmenting the dataset into movies and TV shows, visually articulating the proportion of each category held within the entire video list. We can derive from this chart that movies takes up to 80% of the whole video set and TV shows span the rest.

Following this, an area chart served as a window into the historical progression of content publication spanning a century, from 1920 to 2021. This visualization, rich with information, was interactive, providing details of the number of shows or movies released in any given year upon cursor interaction. A noticeable trend was the exponential growth in content starting around the year 2000, an insight that perhaps indicated this growth is technological advances in the film industry.

Stacked bar charts were utilized to delineate the various ratings and genres of the shows and movies. This form of visualization was particularly effective in highlighting the distribution of content across different age suitability categories and the prevalence of various genres, thereby providing a snapshot of Netflix's content diversity. We can see from these two stacked bar charts that ratings with 13+ takes the most shows, followed by 16+ and 18+. This makes sense since the lower this age limit, there should be more shows available to watch. For genres, the top three are Drama, Comedy, and Drama Suspense, this indicates the audiences’ taste or the overall film industry’s trend in making shows in these genres. This data is important because Netflix can integrate these genres and ratings with their promotion algorithm to recommend these kind of shows at the start page.

Finally we go back to the top, this provides an overview of the whole report so people can get a general understanding before looking at the charts. By selecting distinct titles to get total number of titles in the table, distinct ratings to get the total number of ratings and same thing goes for total number of genres and directors. For start year, we select the minimum release date and for end year we select the maximum release date in the table

**Suggestions for Extensions or Deeper Investigations:**

To dive deeper and extract further value from this dataset, I would recommend a longitudinal study on genre trends with predictive analytics to forecast shifts in viewer preferences. Incorporating demographic data could enrich the dataset, allowing for a more segmented analysis of viewing patterns. A sentiment analysis of user reviews could provide a deeper layer of understanding regarding viewer reception. Performance metrics like retention rates and average watch times would add significant value to content strategy insights.

**Personal Challenges and Solutions:**

Navigating the sheer volume of data was initially daunting. Developing a systematic approach to cleaning and categorizing the data was my first hurdle, which I overcame by leaning on the robust data transformation features within PowerBI and educating myself through community forums and tutorials.

Choosing the right visualizations was also a challenge. I needed them to be not only accurate but also intuitive for the audience to understand. To tackle this, I relied heavily on feedback from my friends who use Netflix, ensuring the visualizations I chose conveyed the intended story clearly and effectively.

Lastly, the task of constructing meaningful narratives from the data required a balance between technical accuracy and creative storytelling. To sharpen these skills, I constantly sought inspiration from industry-leading data visualizations, taking notes of how they distilled complex data into compelling stories.

By addressing these challenges and limitations, my PowerBI project evolved from a simple class assignment into a comprehensive analysis that brought the data behind Netflix's content to life.

窗体底端