**Overview**

This project aims to leverage data visualization techniques to tell a compelling story about the Netflix userbase. By analyzing and visualizing this dataset, we aim to uncover insights into user behavior, preferences, and trends, which can inform strategic decisions for content creation, marketing, and customer engagement.

**Dataset**

**Dataset:** Netflix Userbase  
**Size:** 1000+ records  
**Source:** Provided CSV file

**Tools and Libraries**

* **Python Libraries:** Matplotlib, Pandas, Plotly (new library not covered in class)
* **JavaScript Libraries:** Plotly.js
* **Database:** SQLite (for data storage and extraction)
* **Web Framework:** Flask (for interactive API routes)

**Project Components**

1. **Data Storage and Extraction**
   * Store the Netflix userbase data in an SQLite database.
   * Use SQLAlchemy to interact with the database and extract necessary data.
2. **Data Cleaning and Preparation**
   * Use Pandas for data cleaning and preprocessing.
   * Ensure the dataset is formatted correctly for analysis and visualization.
3. **Data Analysis and Visualization**
   * **Visualization 1: User Demographics**
     + Bar charts and pie charts to display the distribution of users by age, gender, and location.
   * **Visualization 2: Viewing Habits**
     + Line charts to show trends in viewing hours over time.
     + Heatmaps to visualize peak viewing times during the day and week.
   * **Visualization 3: Content Preferences**
     + Stacked bar charts to display genre preferences by demographic segments.
     + Treemaps to show the most popular content categories.
4. **Interactive Features**
   * **HTML Menus and Dropdowns:** Allow users to filter visualizations by demographic attributes (age, gender, location).
   * **Flask Backend:** Create interactive API routes to serve data and visualizations dynamically based on user input.
   * **User-Driven Interaction:** Enable users to select and filter data to generate custom visualizations.
5. **Ethical Considerations**
   * Ensure data privacy by anonymizing user data.
   * Address potential biases in the dataset and analysis.
   * Provide a clear explanation of ethical considerations in the README.md file.
6. **Output and Deliverables**
   * **Summary Table:** Provide a summary of key findings in tabular format.
   * **Data Visualizations:** Include interactive visualizations embedded in a web interface.
   * **Analysis and Recommendations:** Offer insights and actionable recommendations based on the visualized data.
   * **README.md:** Document the project overview, usage instructions, ethical considerations, and references.

**GitHub Repository**

* **README.md:** Outline of the project including:
  + Project overview and purpose
  + Instructions on how to use and interact with the project
  + Ethical considerations
  + References for data sources and external code

**Implementation Plan**

1. **Data Storage and Setup:**
   * Import CSV data into SQLite database.
   * Set up Flask backend with API routes.
2. **Data Cleaning and Analysis:**
   * Clean and preprocess data using Pandas.
   * Conduct exploratory data analysis (EDA) to identify key trends and patterns.
3. **Visualization Development:**
   * Create visualizations using Matplotlib, Plotly, and Plotly.js.
   * Integrate visualizations into a web interface with HTML and JavaScript.
4. **Interactivity and User Interface:**
   * Develop interactive features using Flask and JavaScript.
   * Test and refine user interactions to ensure a seamless experience.
5. **Documentation and Finalization:**
   * Write README.md with detailed project documentation.
   * Review and finalize all components for submission.

This proposal provides a comprehensive plan to leverage data visualization techniques to analyze and present insights from the Netflix userbase dataset, incorporating interactive elements and ethical considerations to ensure a robust and user-friendly project.