CS5630: Data Visualization Group Project Proposal

Basic Info:
- **Project Title:** Asian Americans' Usage of Tinder in the U.S.
- **Team Members:**
  Tailang Cao, u1480633@umail.utah.edu, u1480633
  Yijun Zhan, u1475897@umail.utah.edu, u1475897
  Tung hua Chen, u1484206@umail.utah.edu, u1484206
- **Link to Repository:**
  https://github.com/dataviscourse2024/group-project-tinder-visualization

Background and Motivation:

Online dating platforms, such as Tinder, have revolutionized how people form relationships and interact, yet research often overlooks how different ethnic and cultural groups experience these platforms. Asian Americans, as a diverse and rapidly growing population in the U.S., offer a compelling case for understanding how identity, cultural values, and racial perceptions influence their experiences in online dating.

Our motivation stems from a personal curiosity about how digital spaces can both reflect and challenge societal norms, especially with regard to race and dating preferences. Research has shown that racial biases and stereotypes persist on dating apps, affecting users' interactions and outcomes. As Asian Americans navigate a complex cultural landscape—balancing traditional values with modern American norms—it becomes important to examine how they engage with platforms like Tinder, which emphasize quick judgments and visual attraction.

Additionally, this project aligns with broader discussions about representation, diversity, and inclusion in the tech world. By visualizing patterns and trends in Asian Americans' usage of Tinder, this research aims to contribute to a deeper understanding of how race and technology intersect in personal and social spheres.

Project Objectives:

Analyze and visualize patterns in Asian Americans' usage of Tinder in the U.S. to better understand how identity, cultural values, and racial perceptions influence their experiences in online dating. Provide applicable and profound data based experience for real life usage.

Data:

Data source: Pew Research Center, Survey of U.S. adults conducted July 5-17, 2022.

Link: **https://www.pewresearch.org/dataset/american-trends-panel-wave-111/**

Data Processing:

1.Data Cleaning:
- The raw data collected from Tinder needs to be cleaned to ensure consistency and accuracy. This includes handling missing values, removing duplicates, and addressing any outliers or erroneous entries.
- Special attention will be given to formatting issues, such as normalizing the date format, converting categorical variables into appropriate formats, and encoding text fields (if necessary).

2. Data Transformation:
- We will transform raw data into a more usable format by converting key features such as location (latitude and longitude) into geographical regions or city names.
- We'll also create new features, such as matching frequency or the number of messages sent, by combining multiple columns in the dataset.

3. Data Aggregation:
- For visualization purposes, data will be aggregated at various levels, such as user age groups, gender, and geographic location. We will also calculate key statistics like average match rate, like rate, and message response rate.

4. Data Filtering:
- Filters will be applied to focus specifically on Tinder usage by Asian Americans in the U.S., discarding irrelevant records from the dataset.

Visualization Design:

1. Overview Visualizations:
    - Demographic Overview: A demographic breakdown of Tinder users in the dataset (age, gender, and location) using pie charts and bar graphs.
    - Geographical Heatmap: A heatmap showing Tinder activity levels across different regions of the U.S., with a specific focus on areas with a high population of Asian Americans.
2. Interaction Patterns:
    - Line Charts: Depict trends in activity levels over time, such as the number of matches or messages per day or week.
    - Relationship Visualization: Network graphs showing the interconnections between users (e.g., who liked whom, mutual matches) to visualize social interactions.
3. User Preferences:
    - Bar Graphs: Visualize the age and gender preferences of Asian American users when swiping on Tinder, showing which age ranges or demographics are liked more frequently.
    - Like and Match Rates: Use stacked bar charts to visualize the comparison of like rates and match rates across different user categories.
4. Interactive Features:
    - Allow users to filter the visualizations by factors such as age group, gender, and region, to explore different aspects of Tinder usage among Asian Americans

Must-have Features:

1. Basic Data Visualization:
   • Display basic demographic information and interaction patterns of Asian Americans on Tinder through bar charts, line charts, and heatmaps.
2. Geographical Analysis:
   • A heatmap or choropleth map showing Tinder usage patterns and match activity in different U.S. regions, highlighting where Asian American users are most active.
3. Filtering & Interactivity:
   • Allow the user to apply filters based on key demographic factors such as age, gender, and region to explore different subsets of the data.
4. User Preferences & Match Rate:
   • Include visualizations of user preferences (like age and gender) and the match rate across different groups to provide deeper insights into dating behaviors.

Optional Features:

1. Sentiment Analysis:
   • Use natural language processing (NLP) techniques to analyze the sentiment of messages exchanged between Tinder users, and visualize the results (e.g., positive, negative, neutral sentiments).
2. Predictive Models:
   • Implement machine learning models to predict the likelihood of a match based on user profiles, and visualize the outcomes in an interactive dashboard.
3. Time Series Animation:
   • Create an animated time series that shows changes in Tinder activity patterns over time, highlighting peak usage periods or trends.
4. Advanced Interactivity:
   • Add more advanced interactive features such as tooltips, draggable sliders for time ranges, or hover effects to reveal additional data insights.

Project Schedule:

Week 1 (Sept 16 – Sept 22)	**Data Collection & Preliminary Research**	☐ **Divide tasks for collecting, cleaning, and preparing the data.** ☐ **Start initial data cleaning (remove duplicates, handle missing values, etc.).**
Week 2 (Sept 23 –	**Data Cleaning & Processing**	☐ **Start sketching initial**

Sept 29)		visualizations and decide on tools (D3.js).
Week 3 (Sept 30 – Oct 6)	**Initial Visualization & Prototyping**	☐ **Develop basic visualizations and begin building a prototype.** ☐ **Gather feedback from peers and the instructor on initial prototypes.**
Week 4 (Oct 7 – Oct 13)	**Finalize Visualizations & Integrate Interactivity**	☐ **Finalize all visualizations and add interactive elements (e.g., filters, hover effects).** ☐ **Build the project website and integrate the visualizations.** ☐ **Ensure all visual elements are responsive and functional on multiple devices.** ☐ **Draft a screencast video showcasing the key findings and interactive elements.**
Week 5 (Oct 14 – Oct 20)	**Refinements & Testing**	☐ **Review the visualization for any errors, refine them, and make necessary adjustments.** ☐ **Ensure visualizations effectively communicate your key findings.**
	Final Submission	☐ **All website files and libraries (code).** ☐ **Data (or cloud storage links if necessary).** ☐ **Process Book (PDF).** ☐ **README file (including links to the project site and screencast).**