CSE: 578 Data Visualization Project Report

Sahil Yogesh Hadke
Email: shadke1@asu.edu
ASU ID: 1229679960

Abstract

This project report outlines the development of visualizations for user stories to create targeted marketing profiles for UVW College. The primary objective is to increase enrollment by tailoring marketing efforts to specific demographic groups based on income as the key demographic. Leveraging data from the United States Census Bureau, key demographic factors influencing income, such as age, gender, education status, marital status, and occupation, are analyzed. The report discusses the challenges faced in efficiently analyzing the Census Bureau data, including data preprocessing and adjusting chart readability. Progress made includes learning advanced Python libraries, finalizing two charts (histogram and stacked bar chart), and devising a plan to address challenges. The report concludes with a plan to refine existing charts, explore additional chart types, identify more user stories, and document the entire process for future reference, laying a solid foundation for successful data analysis and strategy development.

Problem Statement

The problem statement revolves around creating visualizations for user stories to develop targeted marketing profiles for UVW College, ultimately increasing enrollment by tailoring marketing efforts to specific demographic groups based on income as the key demographic. The college seeks to leverage data from the United States Census Bureau to identify key demographic factors that influence an individual's income, including age, gender, education status, marital status, and occupation. By analyzing these factors, the goal is to understand characteristics and plot charts to analyze the earnings of individuals earning less than and more than $50,000 annually and tailor marketing efforts accordingly.

The challenge lies in efficiently analyzing the vast amount of Census Bureau data to extract meaningful insights that can inform marketing strategies. This requires a comprehensive understanding of statistical analysis methods, and data visualization techniques to identify trends and patterns in the data.

Progress Made

The progress made in the project so far is substantial and demonstrates a solid foundation for further development. By learning advanced Python libraries such as Pandas, NumPy, and Matplotlib, the groundwork for efficient data manipulation and visualization has been laid. Installing these libraries on the local machine signifies readiness to begin hands-on work with the dataset.

Understanding the different types of data, including nominal, ordinal, categorical, and non-categorical, is crucial for proper data preprocessing and analysis. Moreover, comprehending the problem statement and dataset columns is pivotal for identifying relevant factors affecting income and devising targeted marketing strategies.

By finalizing two charts, namely the histogram and stacked bar chart, along with corresponding user stories, the project is set to gain valuable insights into income distribution among different demographic groups. These visualizations will serve as a cornerstone for developing tailored marketing profiles for UVW College, aimed at increasing enrollment by effectively targeting specific income brackets.

Overall, the progress made reflects a methodical approach and a solid understanding of the project's objectives, laying a strong foundation for successful data analysis and strategy development.

Issues or Challenges Encountered

The project encountered several challenges that required innovative solutions and continuous learning. Data preprocessing posed a significant hurdle, particularly in handling missing values and determining which attributes to retain for analysis. This step is critical as it directly impacts the quality and reliability of the insights derived from the data.

Adjusting figure size, colors, and xticks to enhance the readability of charts was another challenge. While seemingly trivial, these adjustments are crucial for ensuring that the visualizations effectively convey the intended message and are easily understood by the audience.

The categorical nature of many variables in the dataset also presented a challenge. Categorical variables require specialized plotting techniques to visualize effectively, which required time to research and implement.

Plan to Resolve Issues

To address the challenges encountered, a structured plan has been devised. For data preprocessing, the focus will be on exploring various techniques to handle missing values and selecting the most appropriate attributes for analysis. This will involve further research and experimentation to ensure that the data is cleaned effectively and only relevant variables are retained.

Improving the readability of charts will require a deeper understanding of visualization principles and continued experimentation with different color combinations, figure sizes, and axis settings. This will involve studying the theory behind effective chart design and applying these principles to create visually appealing and informative visualizations.

In terms of choosing user stories and plotting charts, the plan is to continue with the iterative process of trial and error until the most suitable plots are finalized. This will involve refining the user stories to ensure they align with the project's objectives and selecting charts that effectively communicate the insights derived from the data.

Finally, studying the theory behind advanced plots such as pie charts, doughnut charts, mosaic plots and heatmaps will be a priority. This will involve researching these techniques in depth and exploring how they can be applied to the project to gain deeper insights into the data.

Itemized Tasks Completed

- Established a strong foundation in data manipulation and visualization by learning advanced Python libraries like Pandas, NumPy, and Matplotlib, essential for handling and analyzing the dataset efficiently.
- Demonstrated readiness to commence practical work with the dataset by successfully installing the necessary libraries on the local machine, ensuring a seamless transition to data analysis.
- Developed a comprehensive understanding of different data types, such as nominal, ordinal, categorical, and non-categorical, which is vital for accurate data preprocessing and insightful analysis.
- Achieved a clear understanding of the project's objectives and dataset columns, enabling the identification of key demographic factors affecting income and the formulation of targeted marketing strategies for UVW College.
- Finalized two key visualizations, the histogram, and stacked bar chart, along with user stories, indicating readiness to gain valuable insights into income distribution among various demographic groups, crucial for developing tailored marketing profiles.
- Positioned the project to leverage these insights to increase enrollment by effectively targeting specific income brackets, demonstrating a strategic approach to data-driven decision-making in marketing.

Itemized Tasks To Be Completed and Initial Plan to Complete

- Refine existing charts by plotting subplots of histograms for all numerical values in the dataset, providing a comprehensive overview of the distribution of these variables and their relationship with income.
- Explore additional chart types such as mosaic plots, doughnut charts, and various bar chart variations (e.g., grouped bar, stacked bar) to visualize categorical variables and their impact on income, enhancing the depth and breadth of insights gained from the data.
- Conduct further study and research to identify additional user stories that can provide valuable insights into income trends and demographics, enabling the development of more targeted marketing strategies for UVW College.
- Develop a detailed plan to document the entire process, including data preprocessing steps, chart development, user story creation, and insights gained, ensuring that the project is well-documented for future reference and replication.
- Implement the refined charts, explore additional chart types, and document the entire process according to the detailed plan, ensuring that all aspects of the project are completed in a timely and efficient manner.