

# CSE 578: Data Visualization Course Project Progress Report

**Student Name:** Ravi Tej Chaparala

**Date:** 04/14/2024

**Project Title:** Visualization Analysis of Income Prediction for UVW Marketing Campaigns

## 1. Introduction

This progress report outlines the current state of my individual project for the CSE 578 Data Visualization course, where the primary objective is to assist UVW College's marketing efforts by providing data-driven insights into income distributions. The project leverages demographic data from the United States Census Bureau to develop a series of visualizations that explore how various factors influence income levels.

## 2. Project Objectives

The main goal of this project is to deliver actionable insights through visual data analysis, which UVW College can utilize to tailor their marketing strategies effectively. By analyzing and visualizing the relationships between income and various demographic attributes, the project aims to identify key demographics that UVW can target to enhance their enrollment strategies.

## 3. Progress Made So Far

Data Acquisition and Cleaning:

- Retrieved the dataset from the UCI Machine Learning Repository and performed initial data cleaning, which involved handling missing values and formatting issues.
- Variables relevant to the marketing needs of UVW, such as education, marital status, occupation, and age, were particularly focused on for the analysis.

Development of Visualizations:

- **Education vs. Income:** Visualized the relationship between educational attainment and income levels, which showed a clear trend that higher education correlates with higher income.
- **Marital Status vs. Income:** Explored how different marital statuses are associated with varying income brackets.

- **Multivariate Analysis of Race, Sex, and Income:** Produced a visualization that illustrates the complex interactions between race, gender, and income.
- **Age, Hours Worked, and Income:** Developed a scatter plot that investigates how age combined with hours worked per week impacts income levels.
- **Occupation and Education vs. Income:** Created a visualization showing how different occupations and educational achievements affect income.

#### 4. Challenges Encountered

- **Data Quality Issues:** The dataset contained a significant number of entries with ambiguous or missing data, particularly in the workclass, occupation and native country attributes, which required careful handling to ensure the reliability of the visualizations.
- **Complexity in Multivariate Analysis:** Designing visualizations that effectively represent the interaction between more than two variables was challenging and required several iterations to refine.

#### 5. Next Steps

- **Enhance Visualizations:** Based on feedback from this progress report, I plan to refine the visualizations to make them more intuitive and informative.
- **User Story Expansion:** Additional user stories will be developed to further explore the data, particularly focusing on variables such as capital gain and native country.
- **Report Compilation:** The final report will synthesize all findings and visualizations into a coherent narrative that addresses the project's objectives and provides clear, actionable insights for UVW College.

#### 6. Conclusion

To date, the project has made significant strides in addressing the needs of UVW College by identifying demographic factors strongly correlated with income. The visualizations developed thus far have laid a robust foundation for targeted marketing strategies that could potentially increase college enrollments.