Gender Pay Gap and Socioeconomic Factors Analysis

Exploring the Gender Pay Gap: An Analysis of Socioeconomic Factors Affecting Income Disparity

Sanchit Tomar
IST 692: Responsible AI

GENDER PAY GAP

Overview of the Gender Pay Gap

Analyzing Gender Disparities in Earnings

Introduction to the Gender Pay Gap

The gender pay gap signifies the disparity in earnings between women and men within the workforce, revealing critical insights into economic inequality.

Key Objective 1

Explore the impact of income, occupation, hours worked, and education level on gender disparities in pay to understand underlying factors.

Key Objective 2

Predict gender based on socioeconomic features using analytical methods to uncover trends and patterns that influence earnings.

Importance of the Study

Understanding the gender pay gap is vital for developing policies that promote equity amongst genders in workplaces.

DATASET OVERVIEW

Gender Pay Gap Dataset Description

Key Features and Dataset Overview

Dataset Source

The dataset is sourced from Kaggle, offering valuable insights into the gender pay gap across various demographics.

Income

Includes annual salaries and hourly wages, providing a clear picture of compensation disparities between genders.

Occupation

Categorizes job titles and sectors, allowing for targeted analysis of pay gaps within specific professions.

Hours Worked per Week

Records the total hours worked weekly, essential for understanding the relationship between work hours and pay.

Education Level

Captures the highest degree attained, highlighting the impact of education on income levels across genders.

Other Factors

Includes various demographic and employment-related variables that may influence income, such as age.

Dataset Size

With a total of 350,000 entries and over 150 features, it offers a robust foundation for analysis.

Data Accessibility

Publicly available on Kaggle, making it accessible for further research and analysis by interested parties.

DATA ANALYSIS METHODOLOGY

Approach to Data Analysis

Exploring Gender and Socioeconomic Factors

Systematic Approach

The analysis employs a systematic approach to examine the relationship between gender and socioeconomic factors.

Feature Selection

Selecting relevant features for predictive modeling is crucial to enhance the model's performance.

Data Cleaning

This step involves removing duplicates and addressing missing values to ensure data accuracy.

Exploratory Data Analysis (EDA)

EDA includes visualizing data distributions and identifying correlations between features to uncover insights.

Predictive Modeling

Utilizing machine learning algorithms like Logistic Regression and Decision Trees to predict gender outcomes.

Model Evaluation

Model performance is measured using metrics such as accuracy, precision, and recall to ensure reliability.

GENDER INSIGHTS

Insights from Exploratory Data Analysis

Key Findings on Gender Disparities

01

Income Distribution

Men tend to have higher average incomes compared to women, showcasing a significant disparity in earnings across genders.

02

Occupation Analysis

Certain occupations are dominated by one gender, which significantly affects overall income levels and opportunities for advancement.

03

Hours Worked

On average, men generally work more hours than women, directly influencing their earnings and career growth.

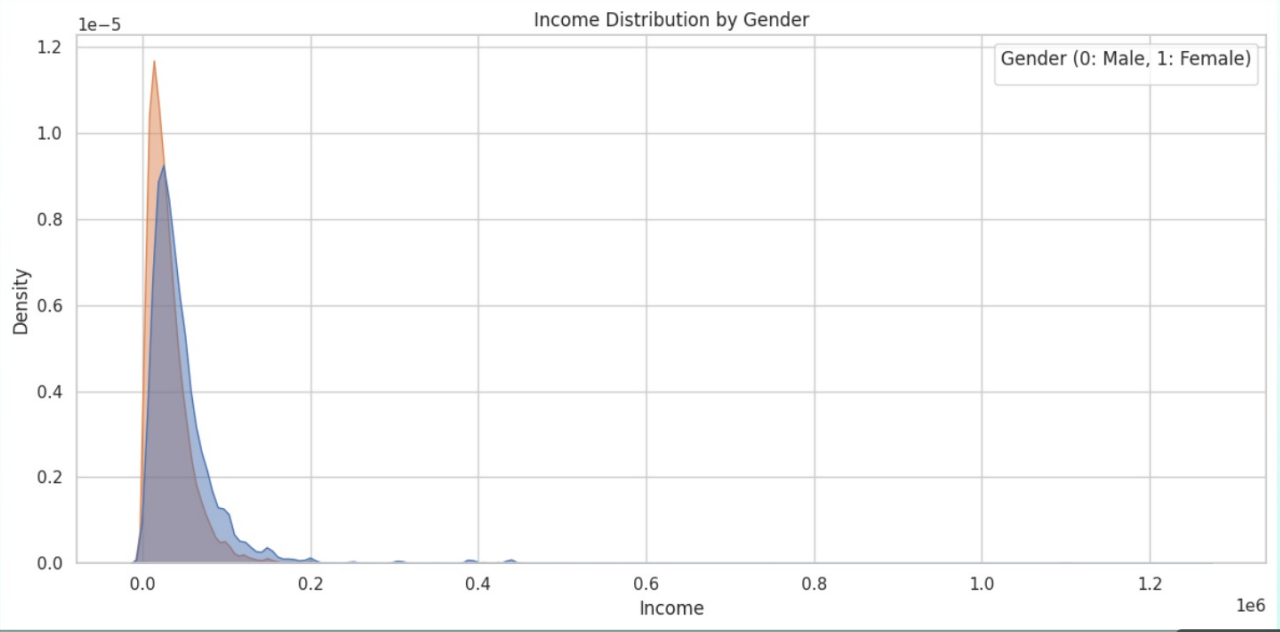
04

Education Level

Higher education levels correlate with increased income; however, disparities in income persist between genders despite education.

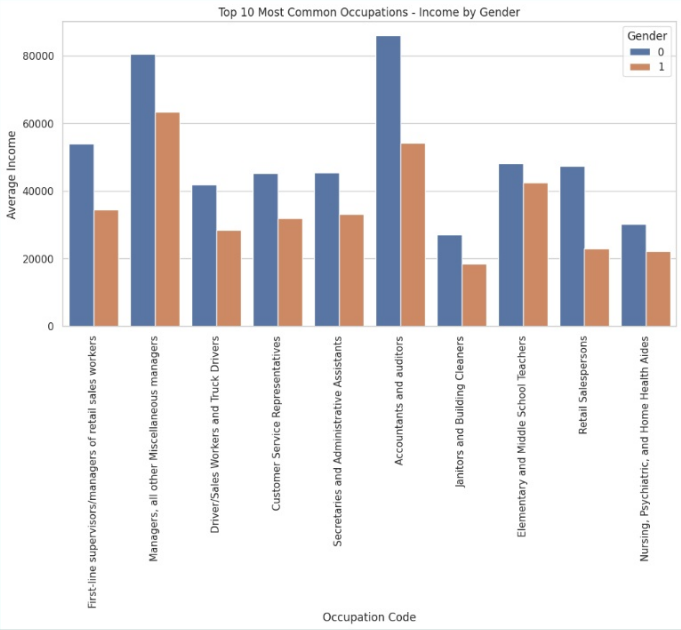
Income Distribution

Distribution of Income by Gender



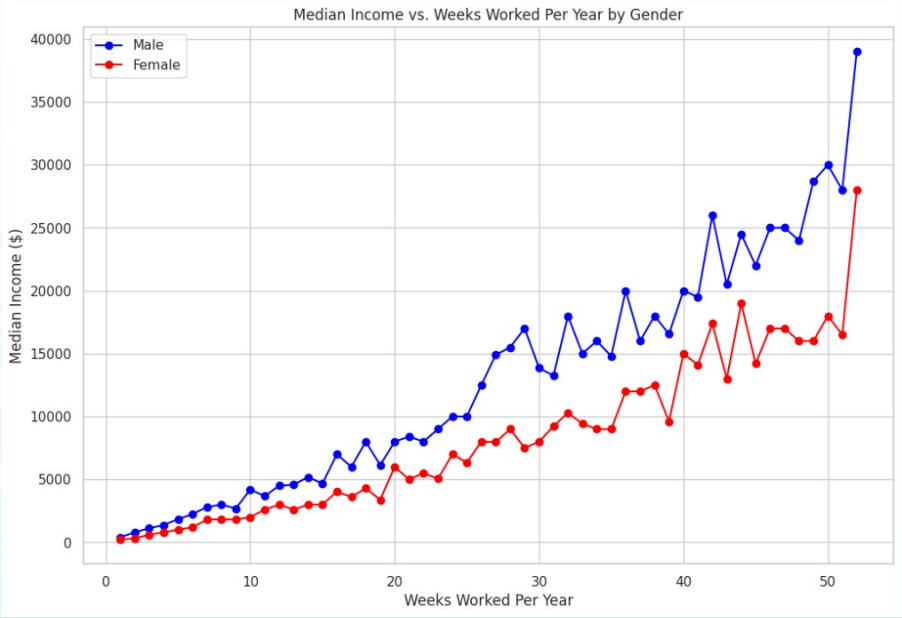
Occupation Analysis

Average Income of Top 10 Occupation by Gender



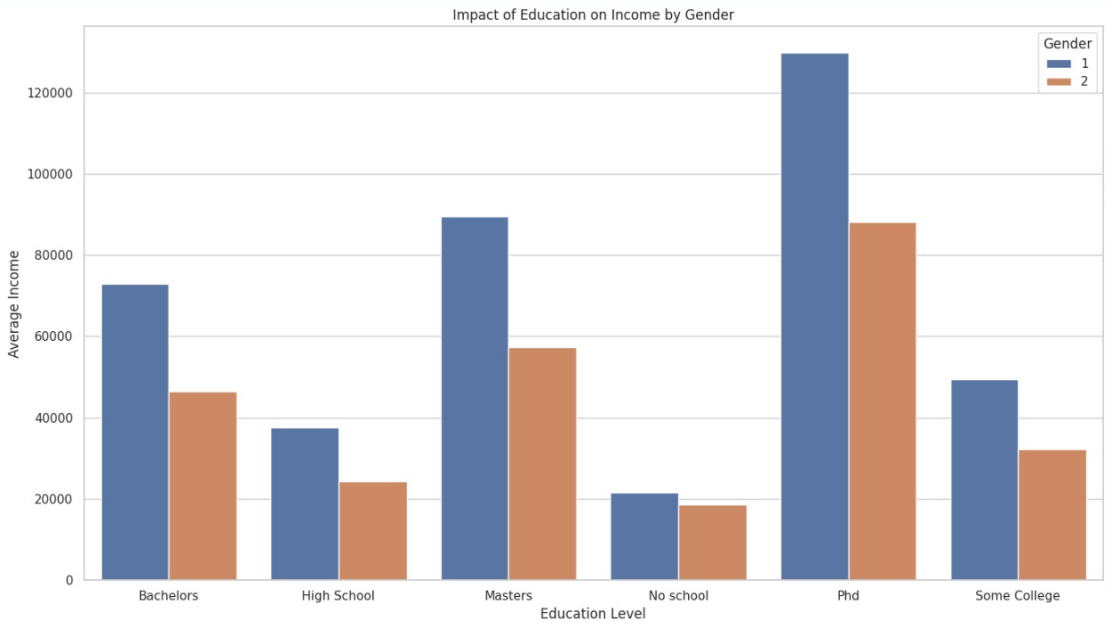
Weeks Worked

Median Income by number of Weeks Worked



Education Level

Impact of Education on Income by Gender



Algorithms Implemented

The analysis employs several machine learning algorithms to effectively predict gender based on selected features.

MODEL EVALUATION

Results of Model Evaluation

Comprehensive Analysis of Predictive Model Performance Metrics

Model	Accuracy	Precision	Recall
Logistic Regression	78%	77%	79%
Decision Tree	70%	70%	70%
Random Forest	77%	77%	77%

LIME, SHAP & Evidently

LIME

LIME explains what factors significantly impact the prediction for a particular observation.

SHAP

SHAP explains what factors overall impacts the models prediction

Is there data drift ?

It is used to determine if a model on data from a specific time is effective for data from another time.

GENDER PAY GAP INSIGHTS

Conclusion and Key Insights on the Gender Pay Gap

Analyzing Gender and Socioeconomic Factors

01

Key Takeaways

The gender pay gap is influenced by multiple factors, including occupation, education level and hours worked.

02

Predictive Modeling

Predictive modeling can identify patterns that may not be immediately apparent through traditional analysis.

03

Future Research Directions

Further research could explore the impact of additional variables such as race and geographic location on the gender pay gap.

04

Call to Action

Implement strategies like Enhance Pay Transparency, Equal Pay Audits, Improve Recruitment Practices, Fairness in Algorithms

Acknowledgement

- **https://www.kaggle.com/datasets/fedesoriano/gender-pay-gap-dataset/data**

  Gender Pay Gap Dataset

- **https://usa.ipums.org/usa/volii/occ_acs.shtml**

  ACS Occupation Codes (OCC)

- **https://www.kaggle.com/code/hephzibahakintunde/gender-pay-gap-visualisation**

  Reference for Visualizations