

Income Prediction Project Presentation

Leveraging Data to Predict Income Thresholds

Stakeholders and Problem Statement

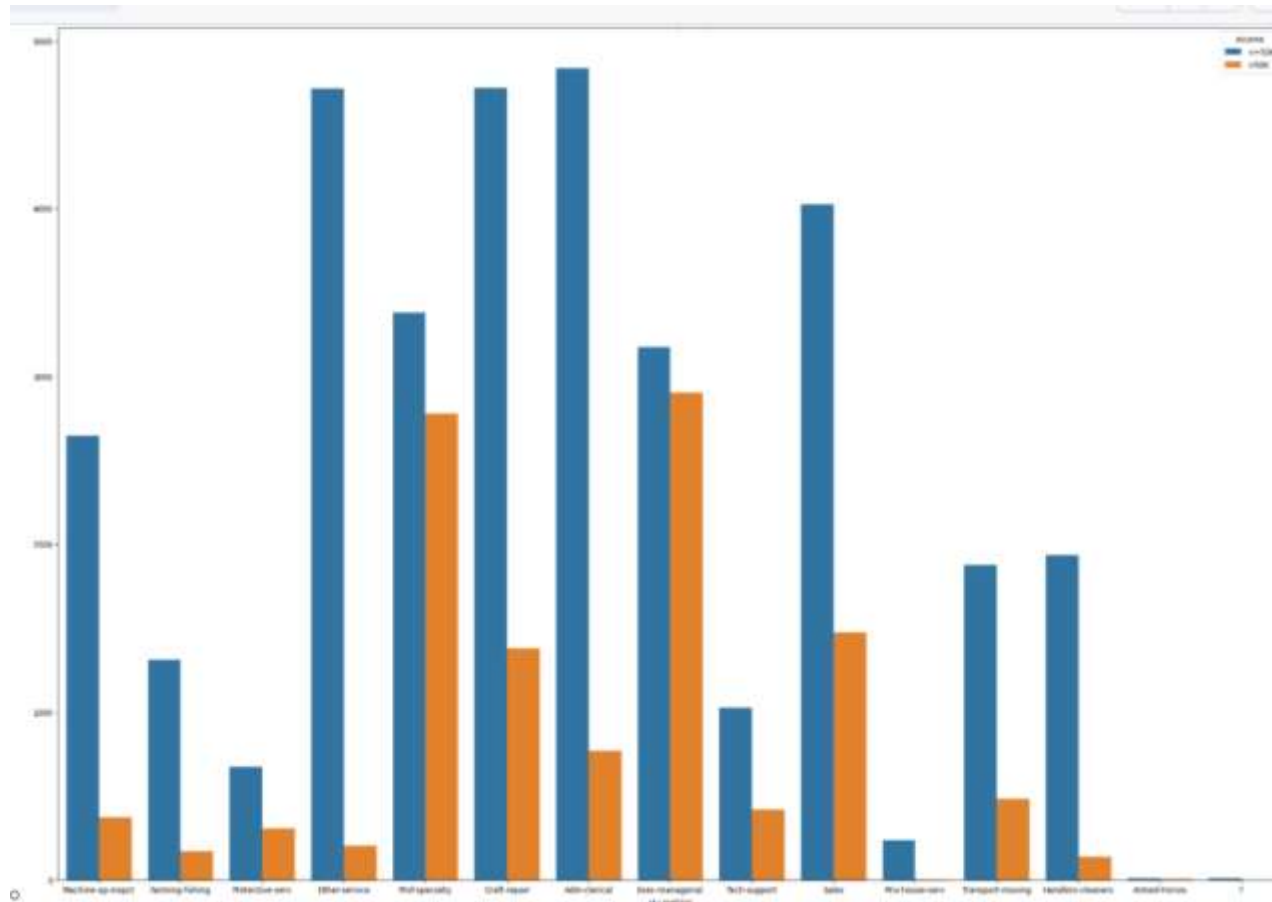
- The goal of this project is to predict whether an individual's income is above or below a certain threshold.
- The stakeholders include organizations or individuals who are interested in understanding the factors that influence an individual's income and want to use this prediction for decision-making.

Data Overview

- The dataset includes various features such as education level, age, gender, occupation, and more, which are believed to influence an individual's income.
- The target variable is whether the income is above or below a threshold.
- **Source of data:**

The data can be downloaded from [Data Dictionary](#)

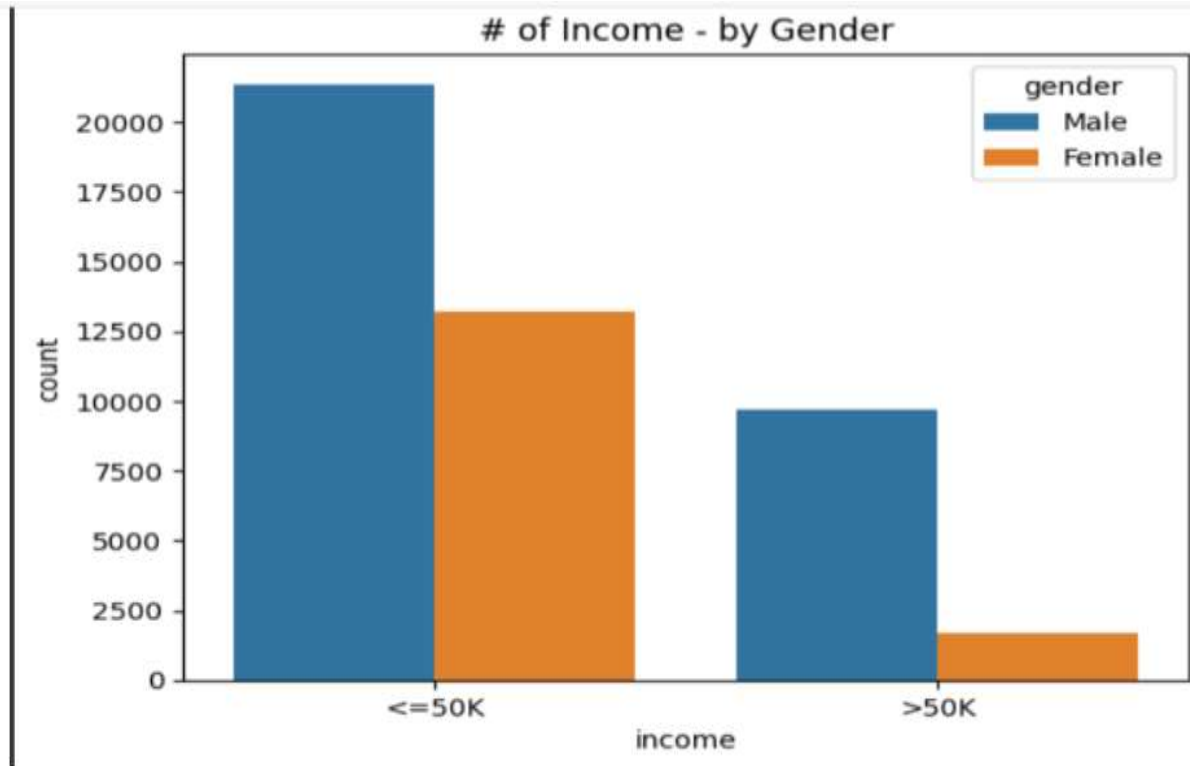
Key Finding 1



- The distribution of educational-num is shifted to the right for individuals earning > \$50K, indicating a higher educational level.

Education Matters: Higher Education Drives Income: Those with advanced education are more likely to earn over \$50K annually.

Key Finding 2



Visualizing the income distribution showed that a significant portion of individuals have incomes below the threshold, which could impact the model's

Age and Income: Among males, higher age is associated with incomes below \$50K

Model Performance

- **DecisionTreeClassifier**
- Training Data:
 - Accuracy: 1.00
- Test Data:
 - Accuracy: 0.81
- **LogisticRegression**
- Training Data:
 - Accuracy: 0.85
- Test Data:
 - Accuracy: 0.85
- **Model with PCA (Logistic Regression)**
- Training Data:
 - Accuracy: 0.83
- Test Data:
 - Accuracy: 0.84

Model Evaluation

- **Decision Tree Model:**

Training Data: Imagine teaching a robot to do a task perfectly when it already knows all the answers. It gets everything right in practice (100%).

Test Data: But when the robot faces new questions it hasn't seen before, it doesn't do as well (81%). It's like a student who aces homework but struggles on a surprise test.

- **LogisticRegression Model:**

Training Data: This model is like a student who studies and scores 85% on practice questions.

Test Data: When faced with new questions, it still gets 85% correct. So, it's consistent and reliable.

- **PCA Model with Logistic Regression:**

Training Data: Here, it's like a student who scores 83% on practice questions.

Test Data: When given new questions, this student gets 84% correct. So, it's pretty good at handling surprises.

Which Model

- Our main aim is to have a model that works well with new, unknown data. So, we want a model that's like a good student, performing consistently well in real-world situations.
- Considering this, the **LogisticRegression model** is our best pick. It's like a dependable friend who's good at handling both practice and real-life challenges. It's a safe bet for solving our income prediction problem effectively.

Recommendations

- **Investing in Education**

- One standout factor is education. It's like the key ingredient in a successful dish. So, consider investing in educational programs to help people improve their skills and potentially earn more.

- **Targeting the Right Age Group**

- We've also pinpointed an age range where people are more likely to earn higher incomes. It's like knowing the right audience for a product. You can focus your efforts on providing them with financial products or services tailored to their needs.

Conclusion

- This project aimed to predict an individual's income using various features from the Adult Income Dataset.
- The analysis provided insights into the relationships between features and income, and the predictive models helped in achieving accurate income predictions.
- The recommendations provided can guide stakeholders in making informed decisions to address income-related challenges.

Q&A