Adult
Annual Income
Prediction

...

Michael Menjares

Project Description

The purpose of this project is to predict whether an adult individual would be classified into either annual income class: above $50,000 or below $50,000.

This would be based on factors such as education level, age, gender, occupation, etc.

$50k

Adult Income Dataset

- The dataset is hosted on kaggle.com but is originally from the 1994 US Census database and was extracted by Barry Becker.

- There are 14 columns of demographic features for predicting the income outcome of either '>50K' or '<=50K'

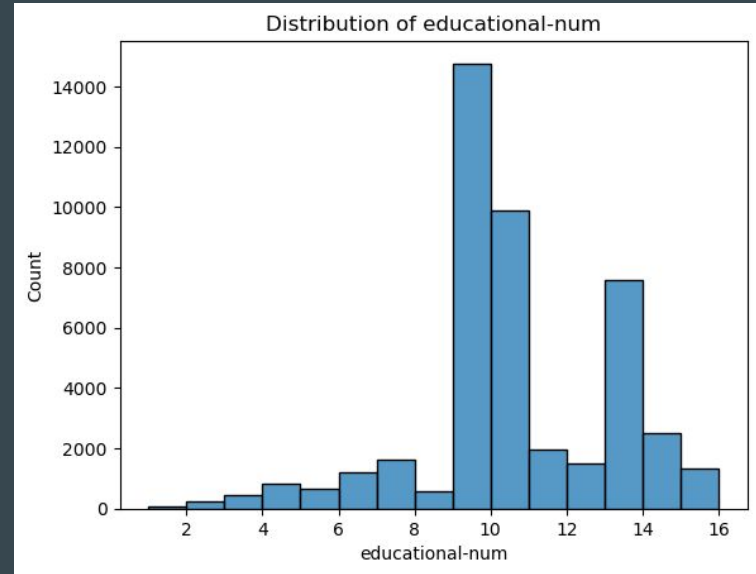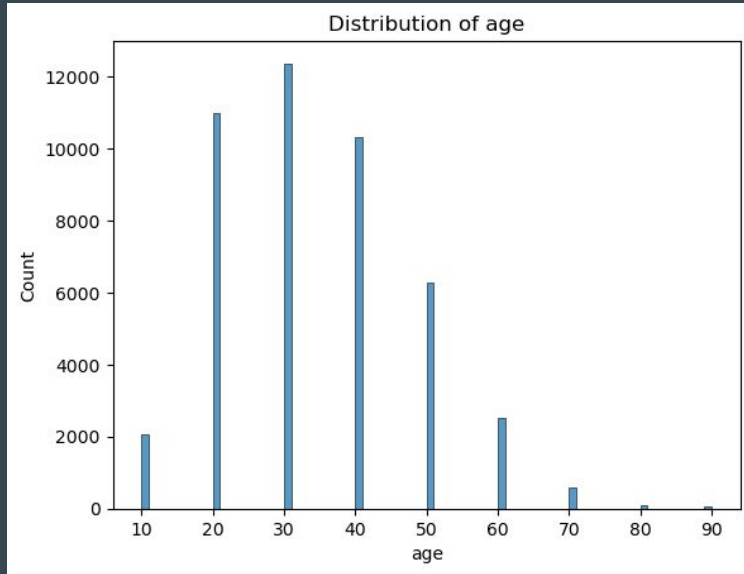- The dataset contains about 49,000 entries

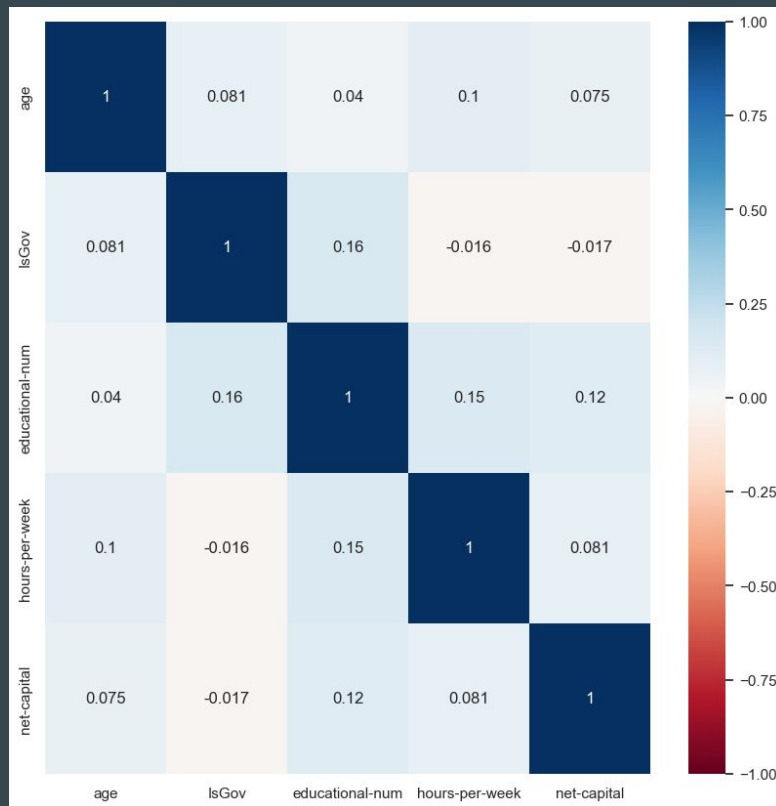| Age	Workclass	Final Weight	Education	Education-Num
Marital Status	Occupation	Relationship	Race	Gender
Capital-Gain	Capital-Loss	Hours-per-week	Native-Country	Income

Stakeholders

The stakeholders can be:

-historians, considering the age of the dataset

-non-profit or local organizations that want to know factors affecting lowering income communities

-government agencies to track economic trends
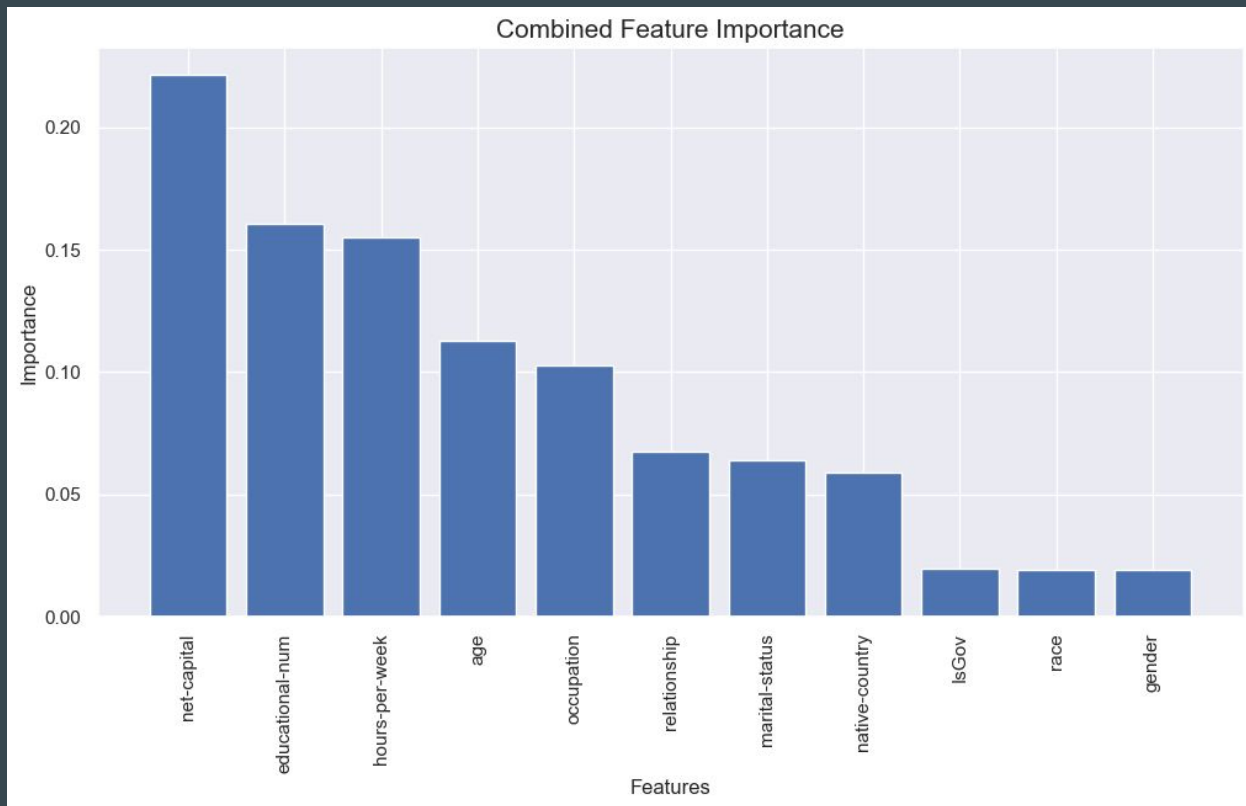
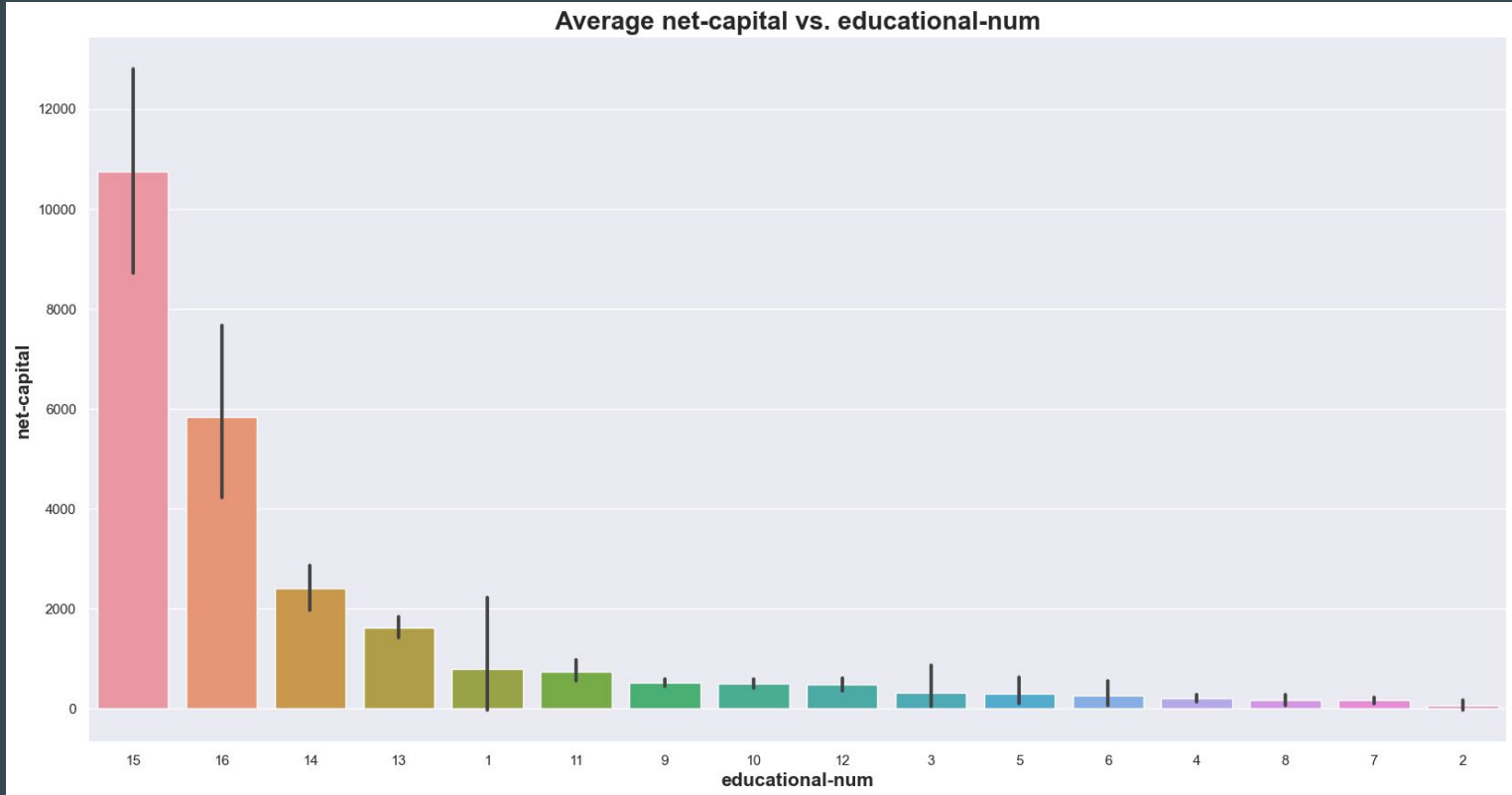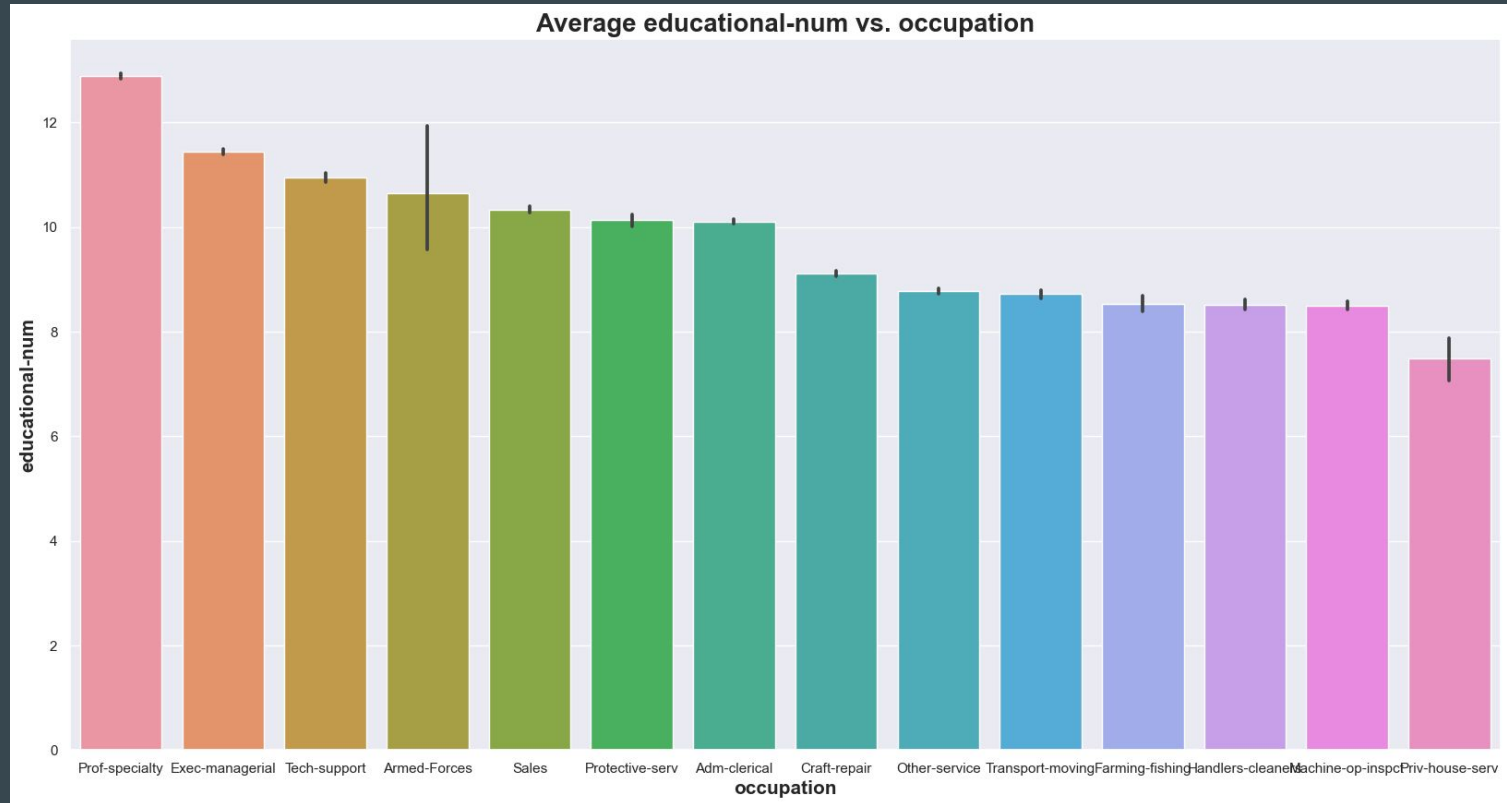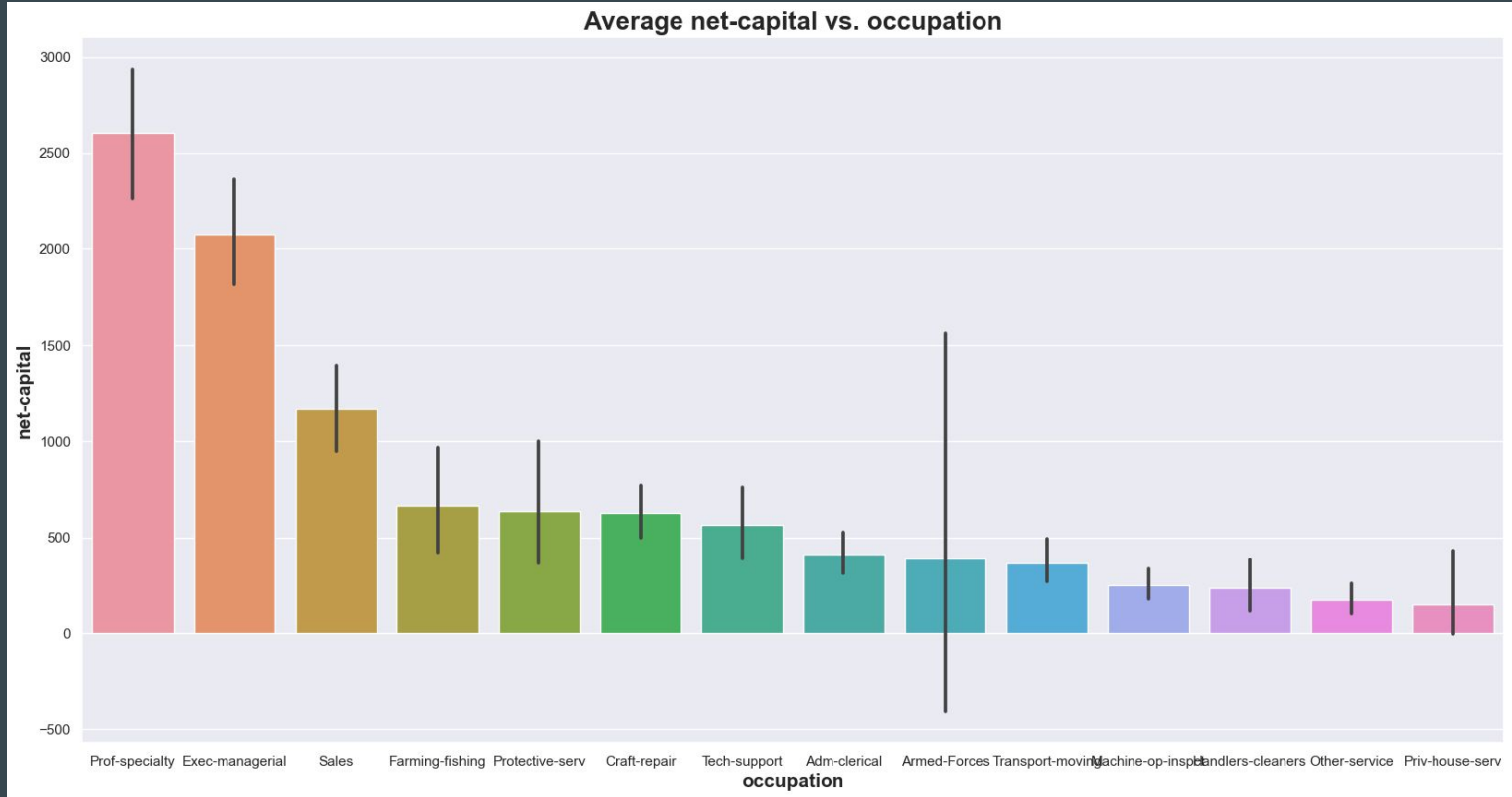-business to learn who to cater their products to

Findings

Findings

Findings

Combined Feature Importance

Findings

Average net-capital vs. educational-num

Findings

Average educational-num vs. occupation

Findings

Average net-capital vs. occupation

Findings

Occupation Counts for Individuals with Income <= 50K
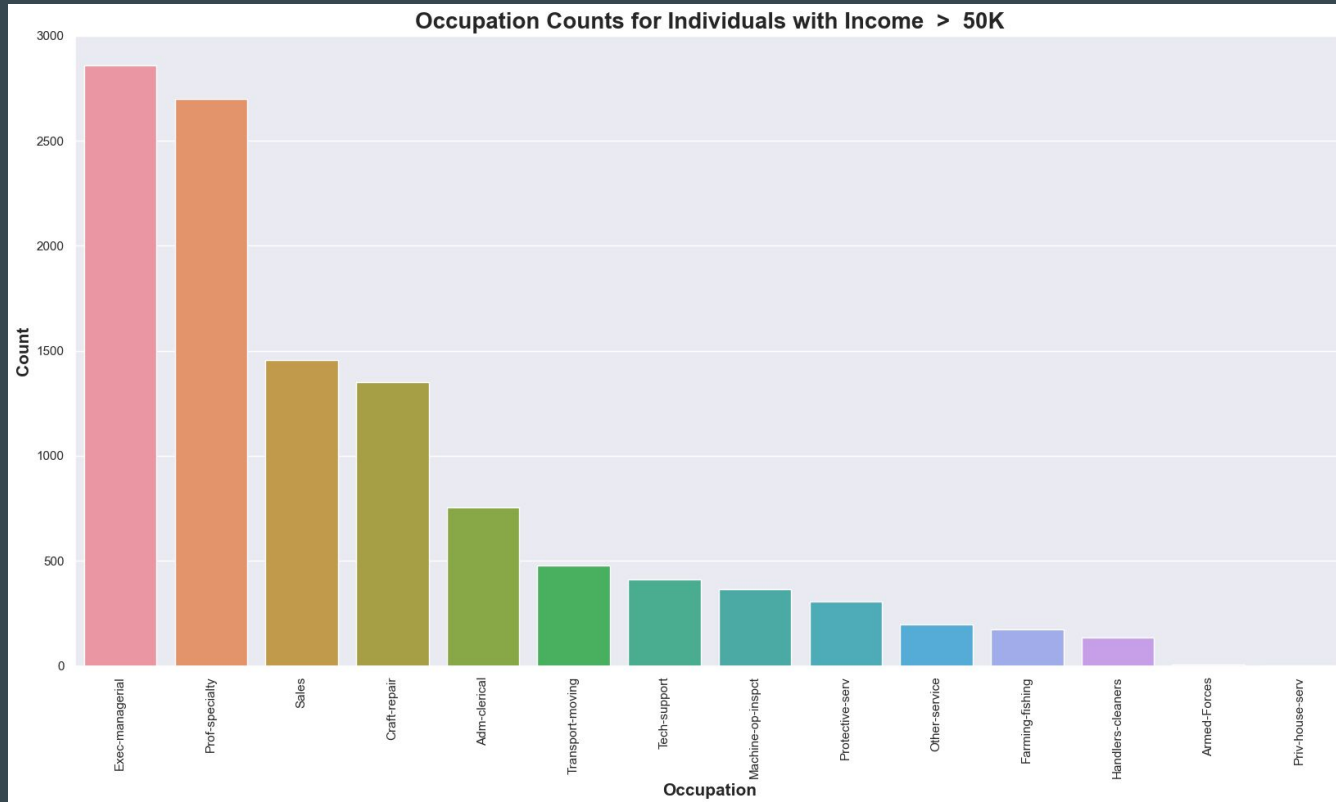
Findings
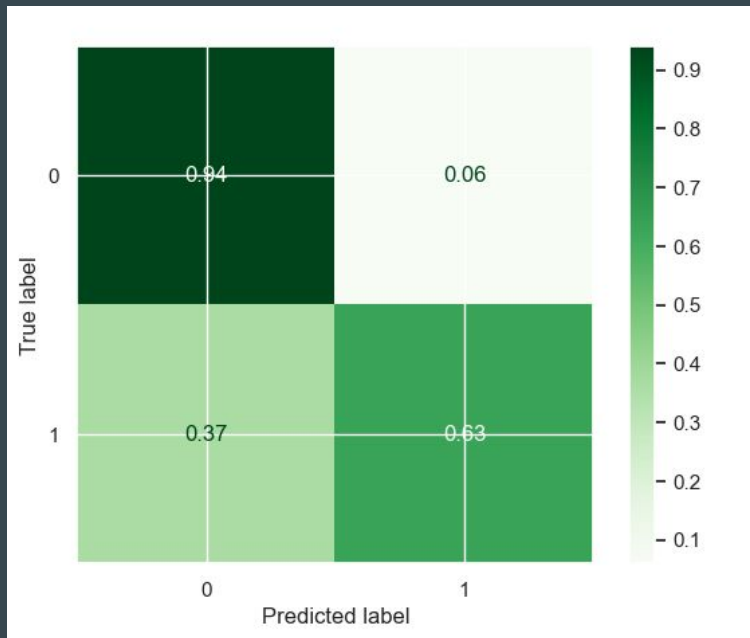

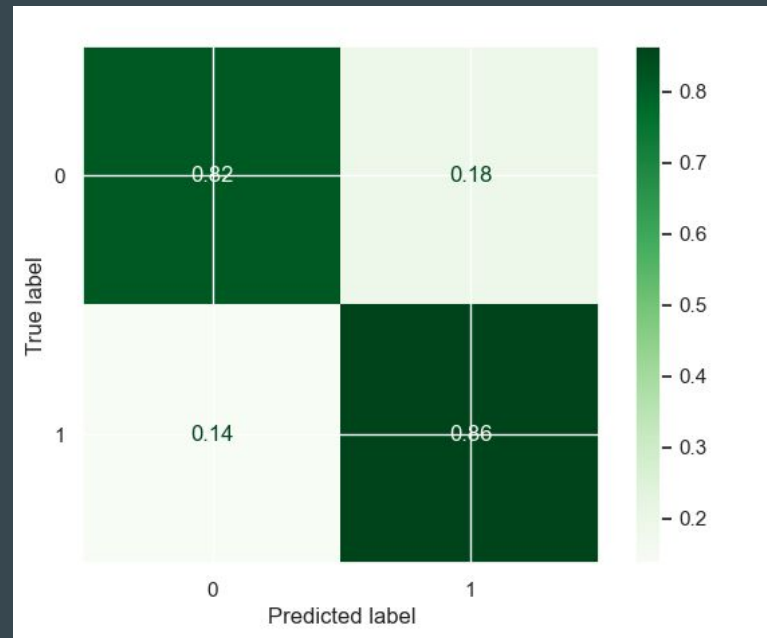Occupation Counts for Individuals with Income > 50K
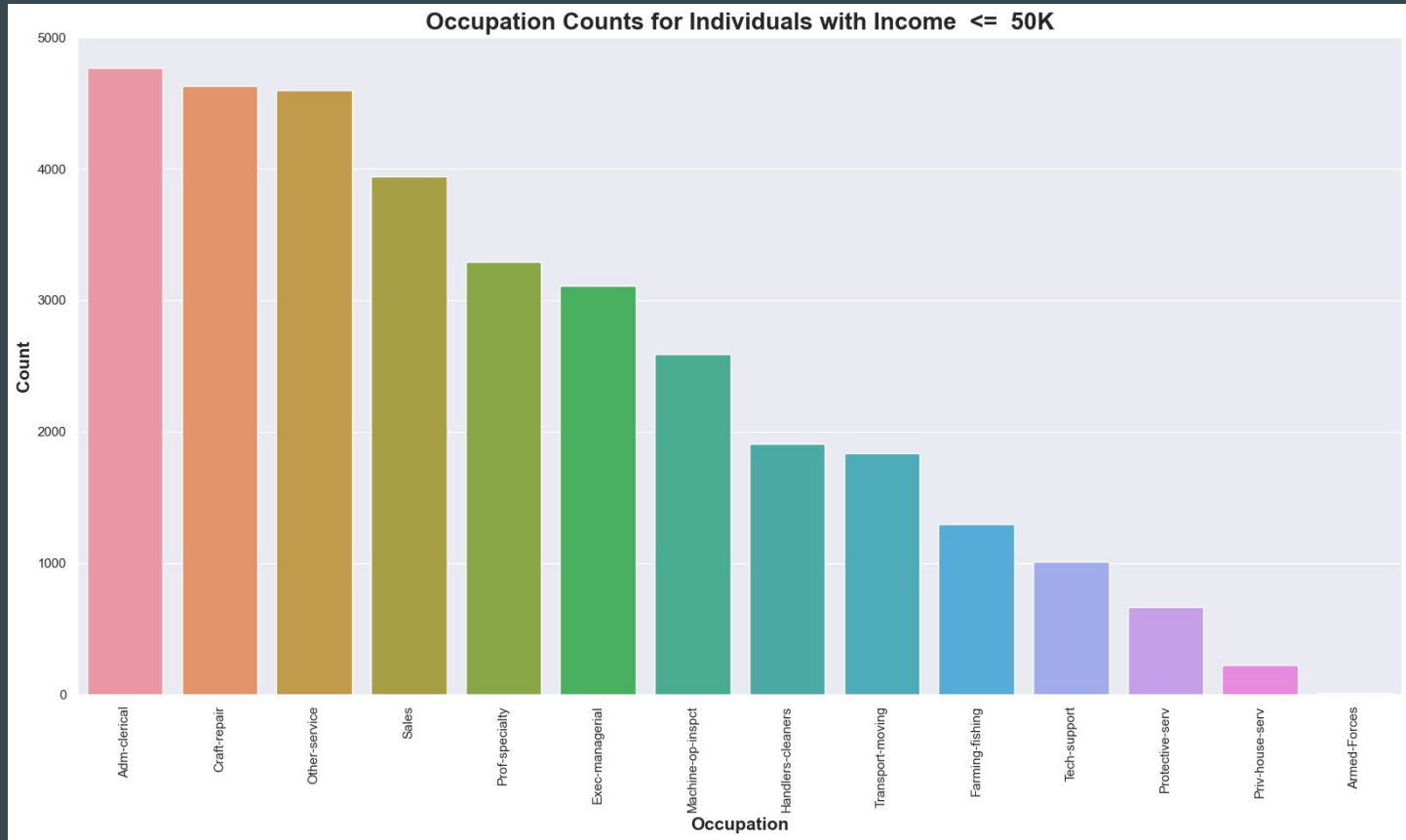
Best Model: LGBM

Not Weighted

Not Weighted

Recommendations

I would like to see a current dataset to compare changes since the last 30 years

I would also like to recommend more attention to the factors that may lead to a lower income to find patterns to assist lower income areas