

Predict Wage from the Current Population Survey (CPS)



Overview & Business Problem

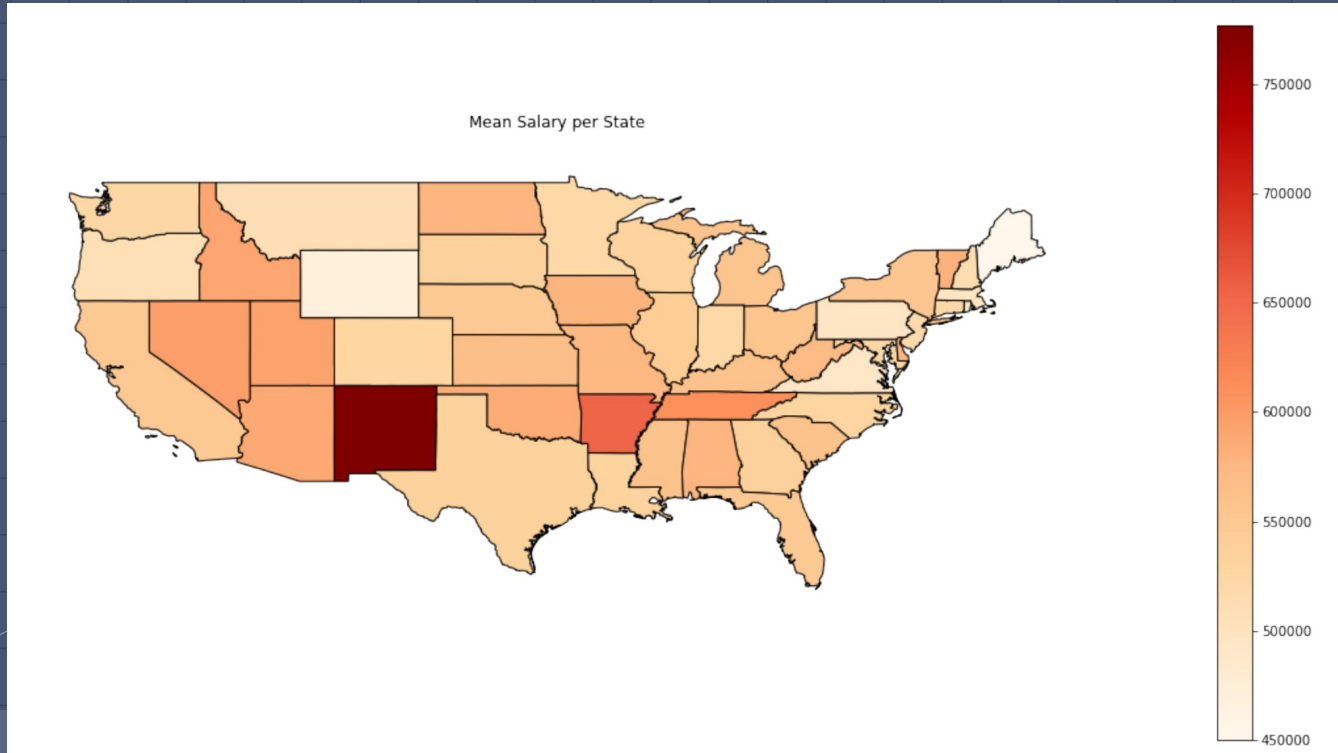
Using the customized data pulled out from CPS.IMPUMS.ORG, we retrieved various demographics data such as housing status, state, age, sex, race, marital status, veteran status, number of children, education level, occupation, industry of occupation, disability status, and health insurance status.

Business Problem

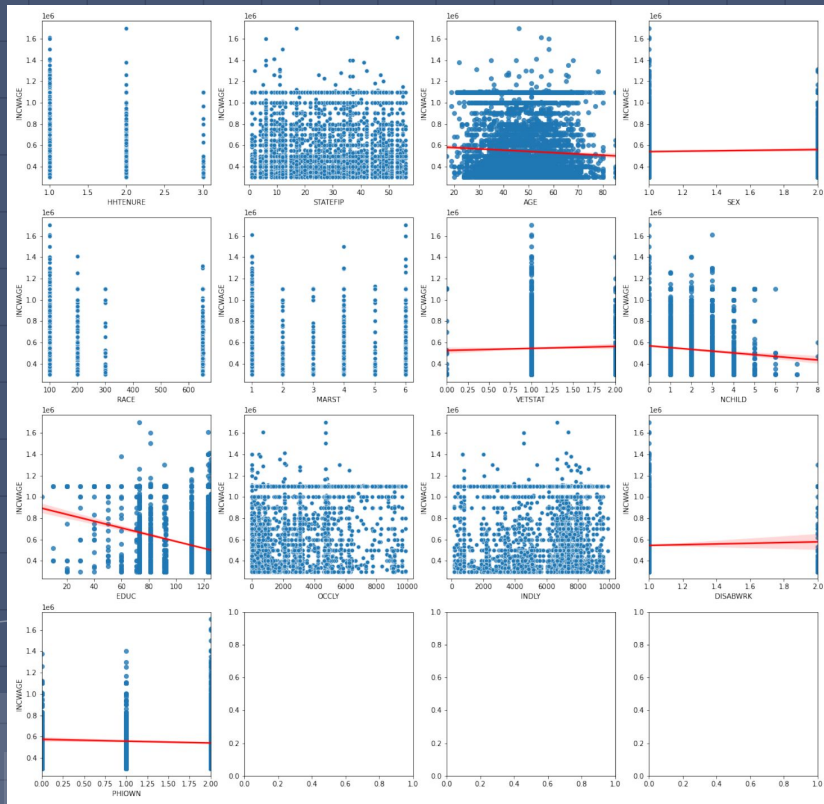
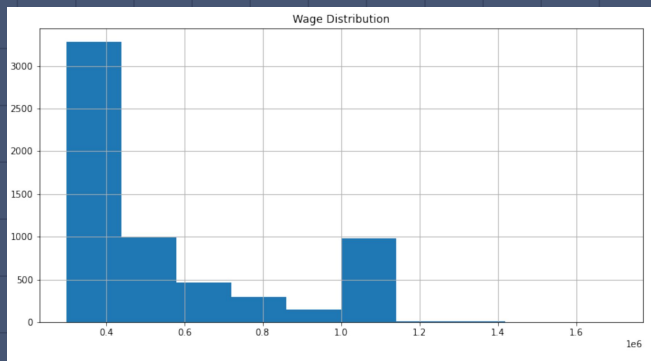
Is the potential client wage more than \$ 1 Million?



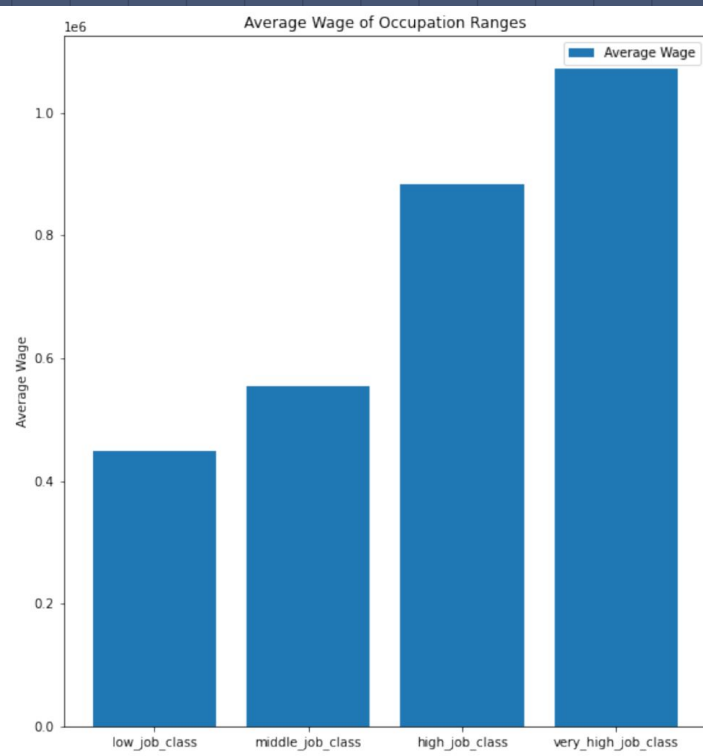
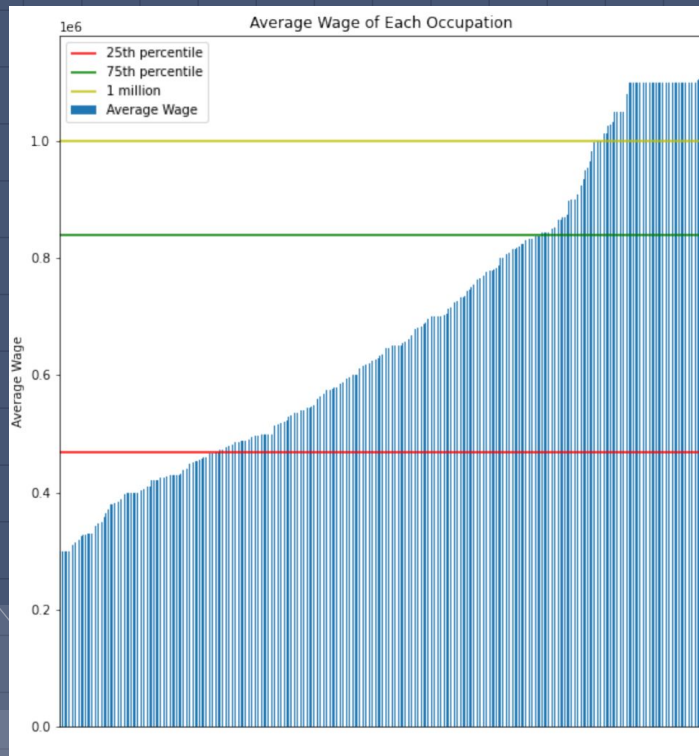
Average Wage of States



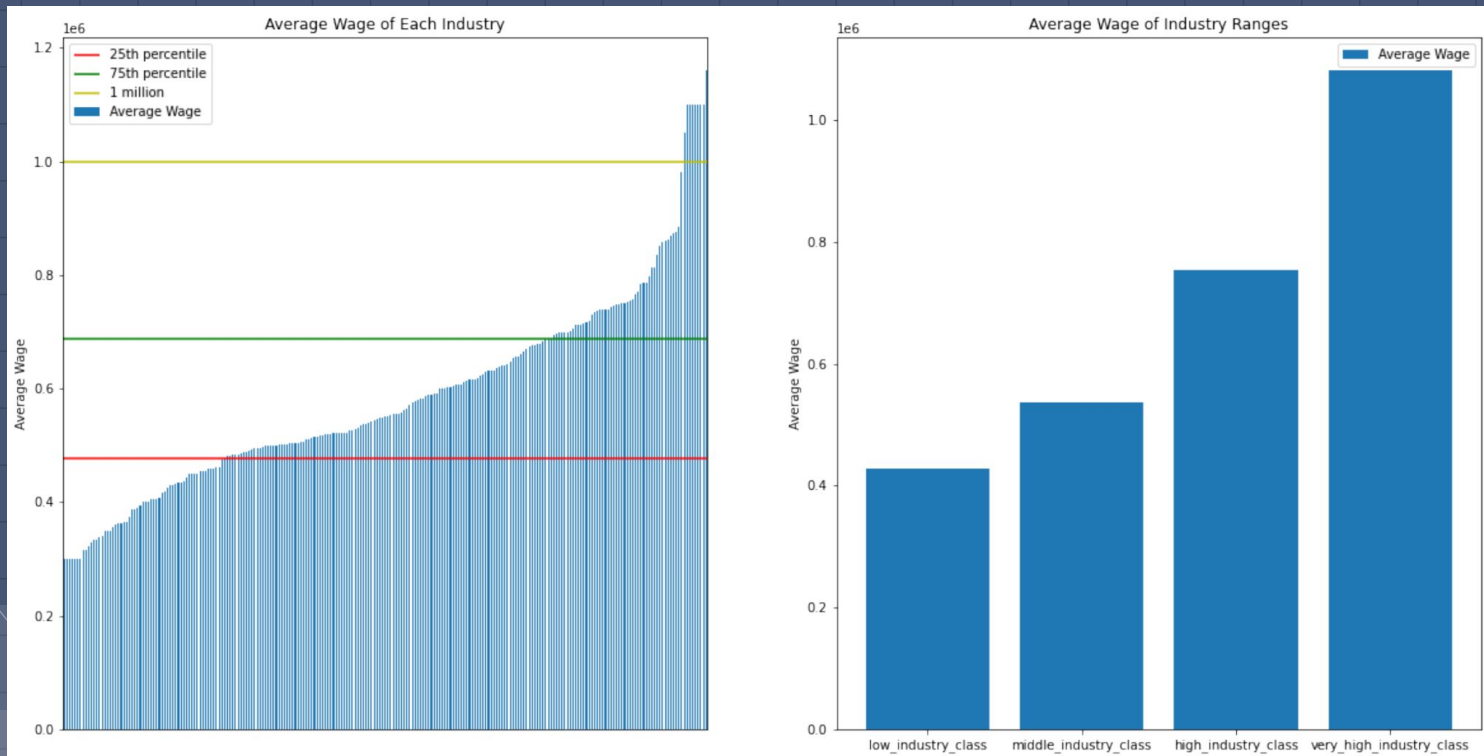
Data Exploration



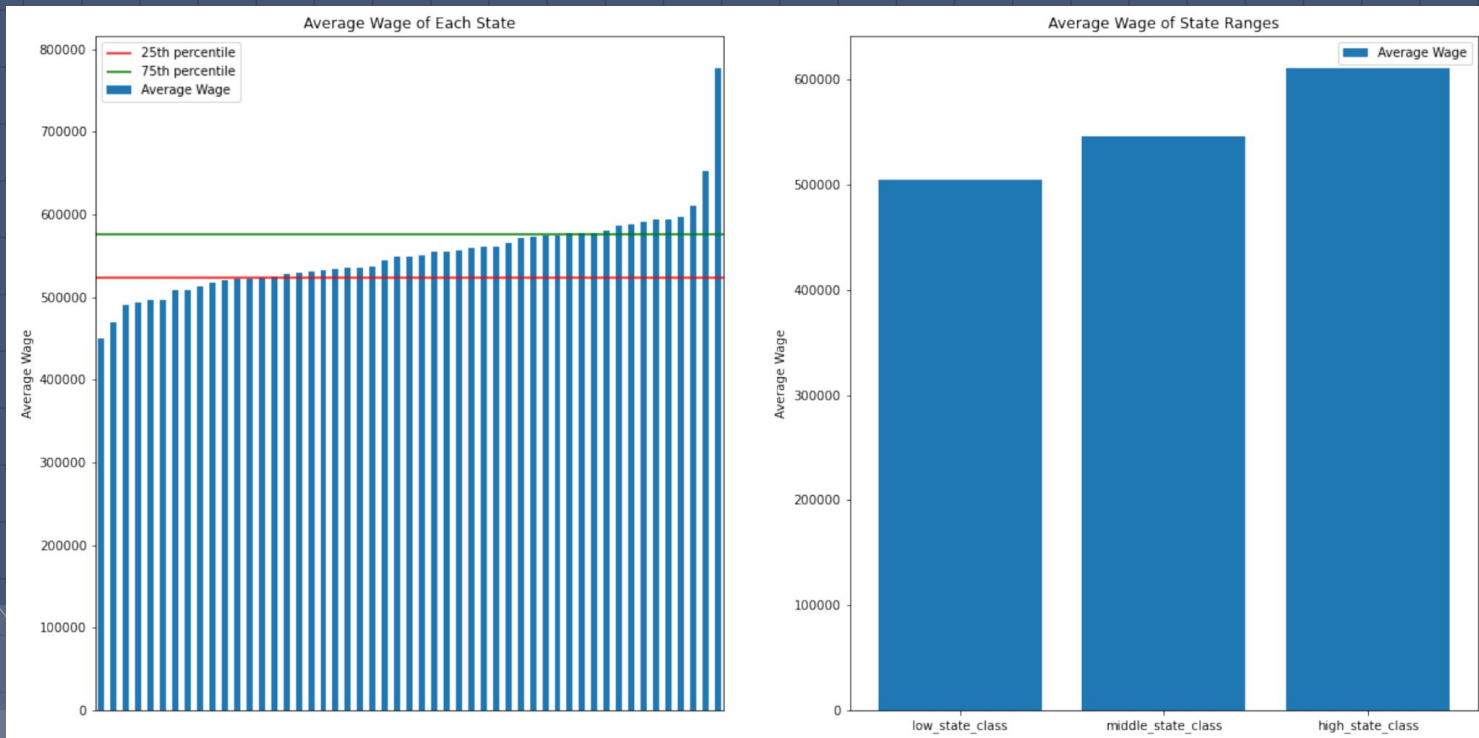
Feature Engineering - Occupation



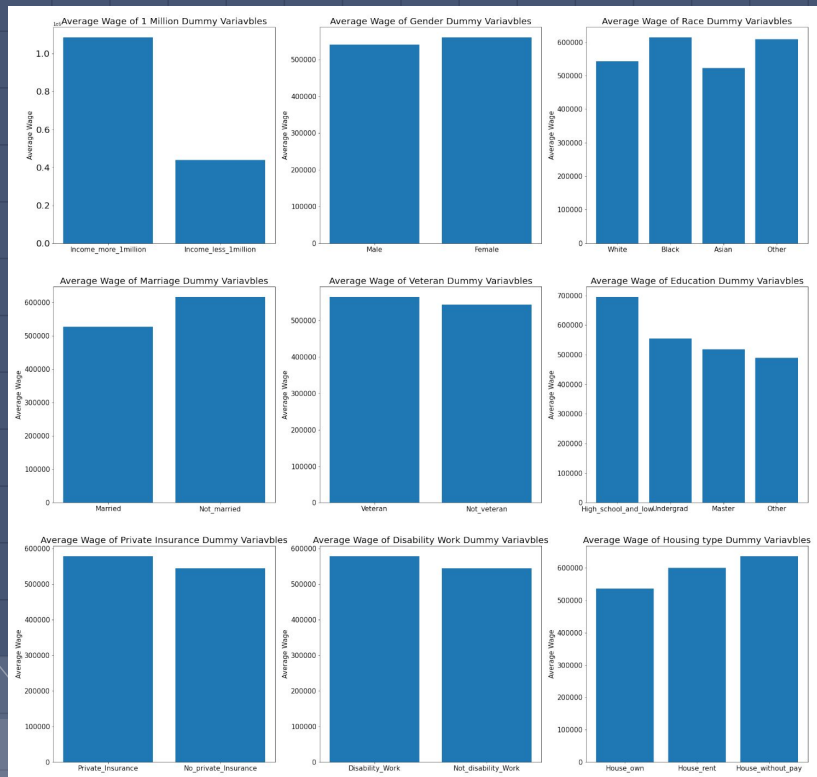
Feature Engineering - Industry



Feature Engineering - State



Feature Engineering - Others

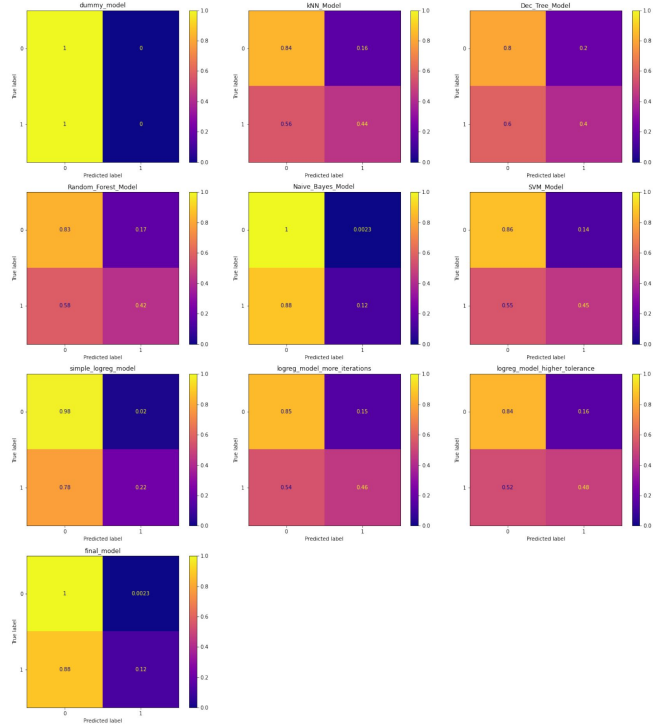


Dummy Variables

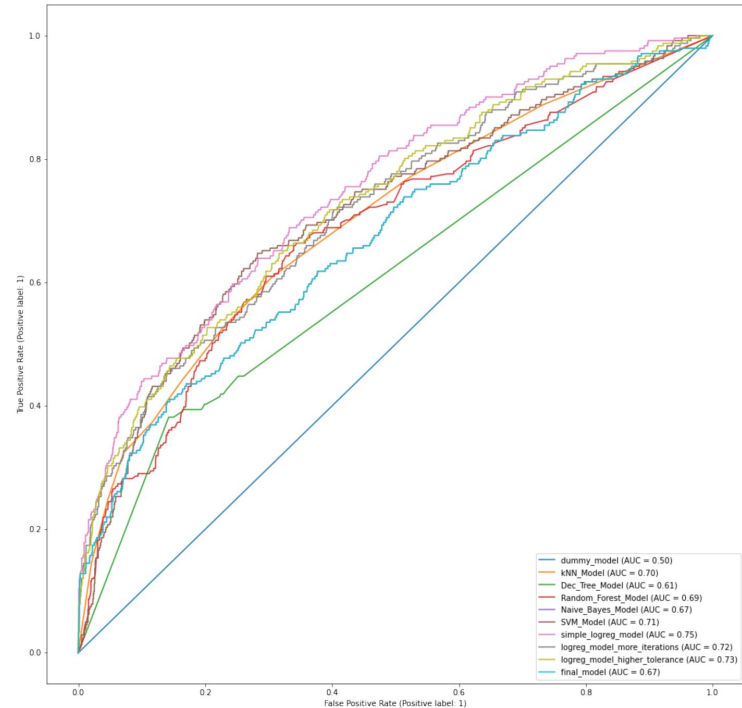
- Gender
- Race
- Married
- Veteran
- Education
- Disability to work
- Private insurance
- Housing type
- Wage more than \$ 1M

Model Comparison

Confusion Matrix Normalize Comparison



ROC Curve for All Models



Conclusion

Two Best Models

	Log-Regression	Naive-Bayes
Accuracy	0.86	0.86
Precision	0.67	0.91
Recall	0.22	0.12
F1-score	0.33	0.21

Conclusion

- We chose Naive-Bayes model because its precision score is higher than Logistic regression model.
- JP Morgan can find more suitable qualified future clients.

Next Step

- Find other independent variables which can make our model more accurately
- Find different method of data cleaning
- Find the way to improve precision from our final model

