

Using Machine Learning to predict yearly income above 50,000 in the USA.

Presentation by Peter George Ngugulu

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

Peru

Canada

United States

Mexico

Spain

Turkey

India

China

Japan

Indonesia

Algeria

Egypt

China

Japan

Indonesia

Colombia

Brazil

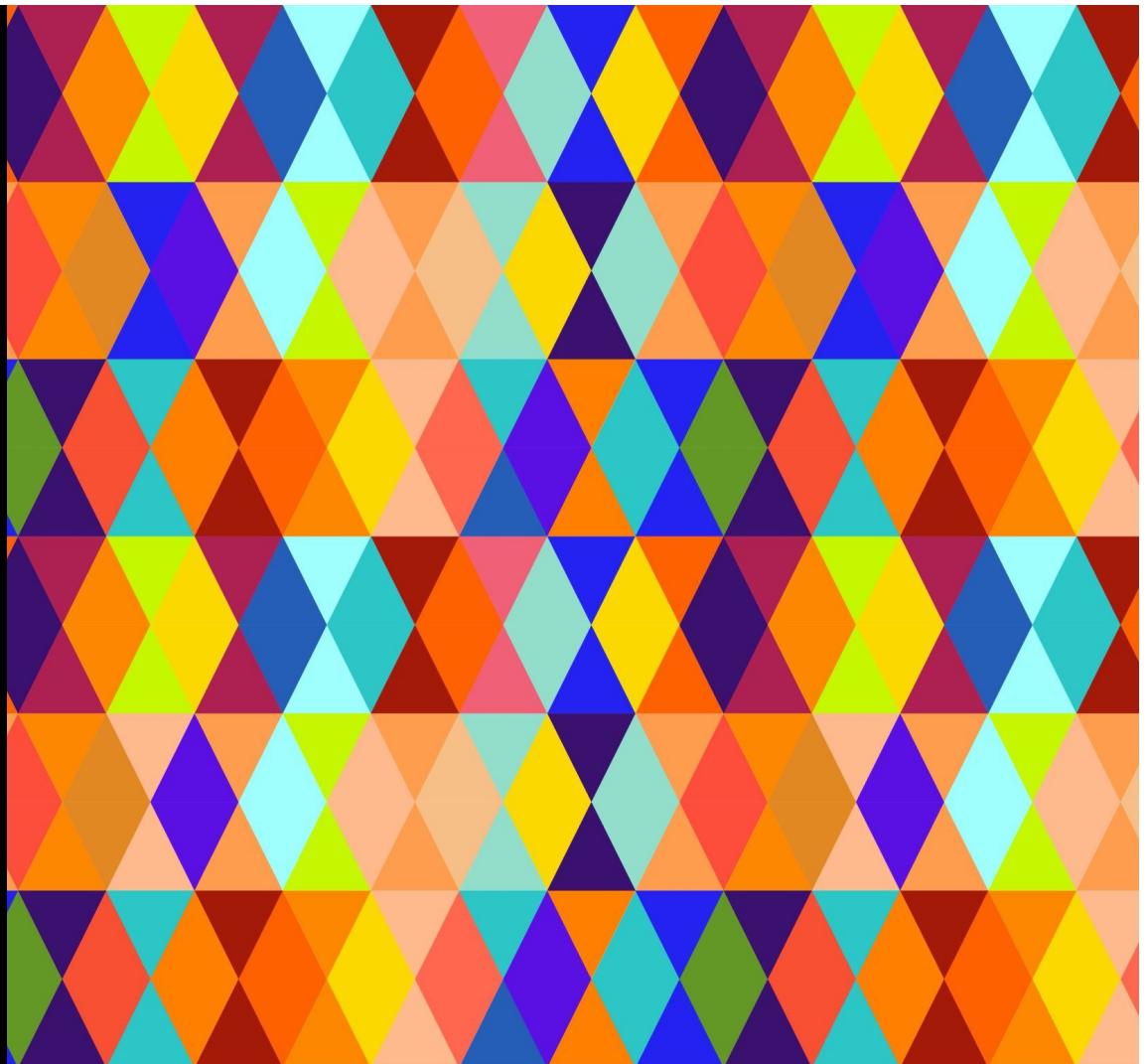
Peru

Canada

United States

Mexico

**Introduction
Questions
Results
Insights
Conclusion**



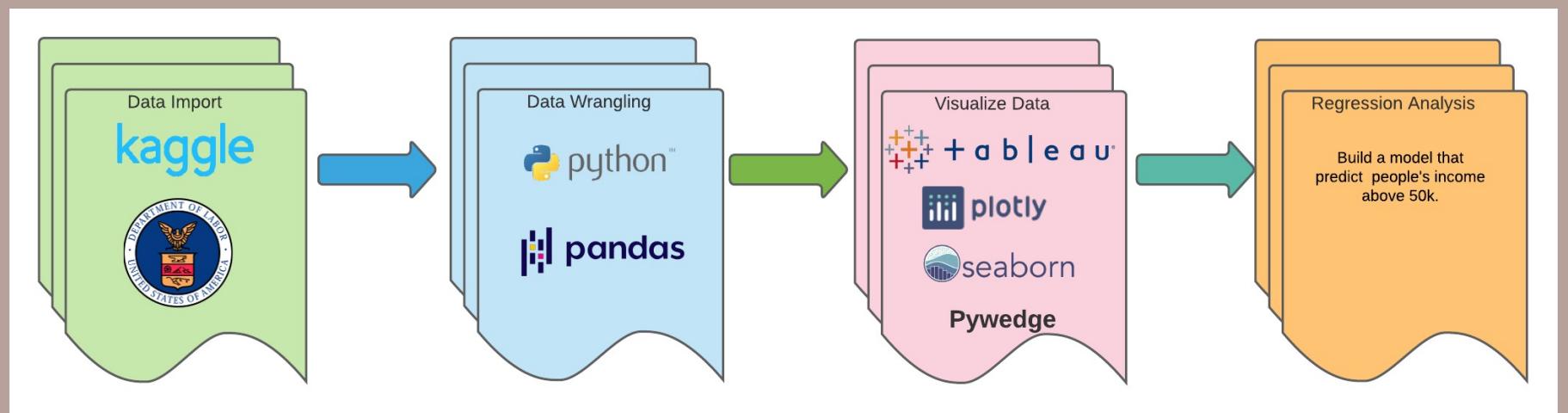
Introduction

- Prediction aim is to determine if a person makes over \$50K a year based on the features.
- The dataset include 14 variables and observations 32561.
- Enhanced data with 4 more data sets.

Questions

- Can features predict if someone earns annual Income of above 50k?
- Is number of male and female equally earns above 50k?
- What is the income differences between races?

Analysis process



Findings

Logistic Regression (84%)

Confusion matrix (Accuracy)

Smote (Imbalance)

ROC (Accuracy)

Decision Tree (86%)

Confusion Matrix (Accuracy)

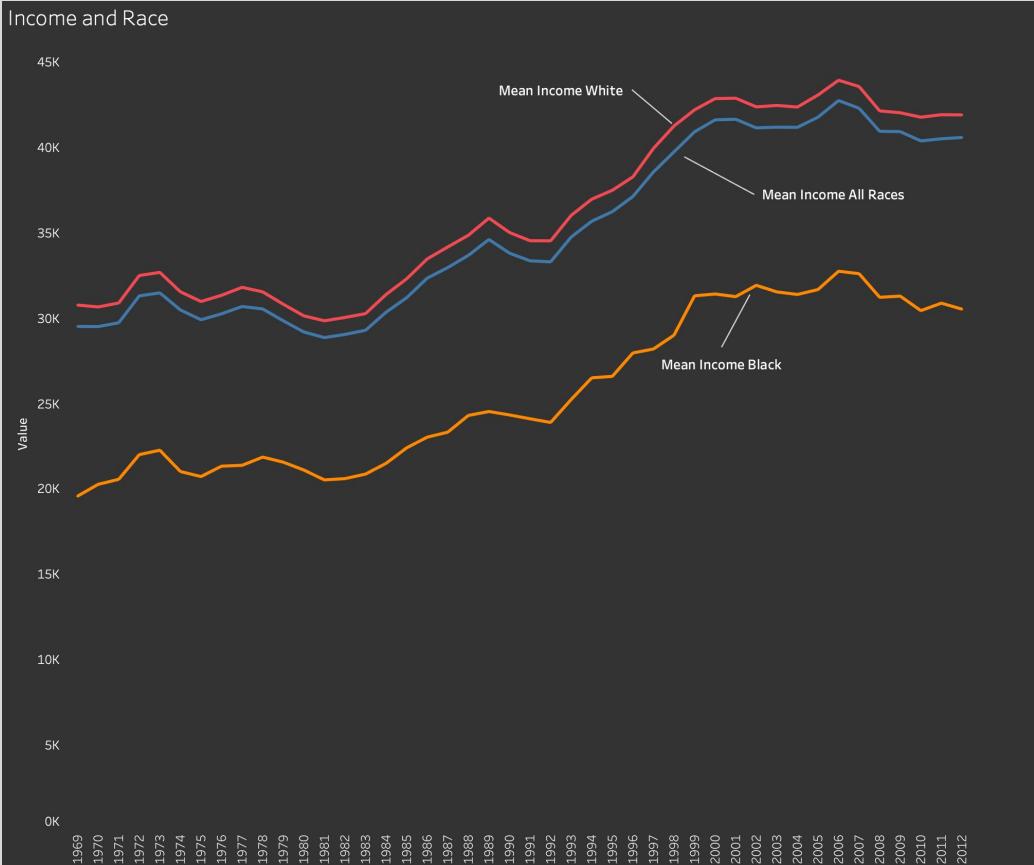
ROC (Accuracy)

Insights Income, Race and Gender

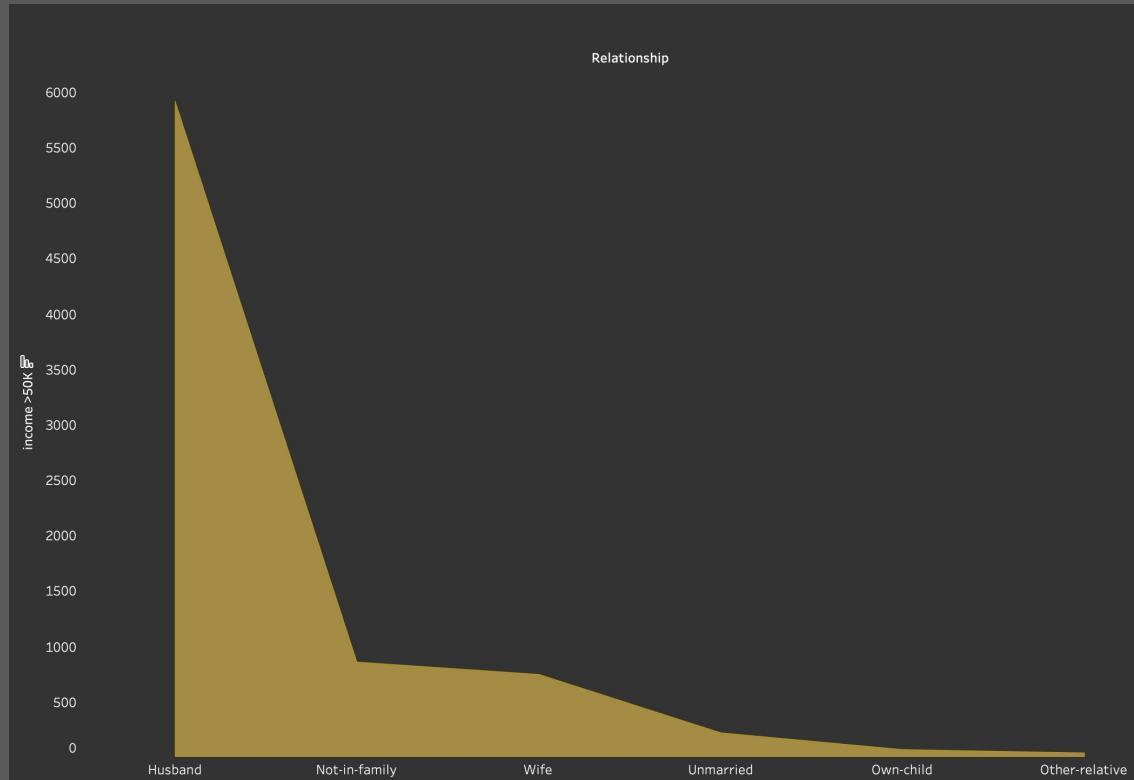


Income differences across races

∞

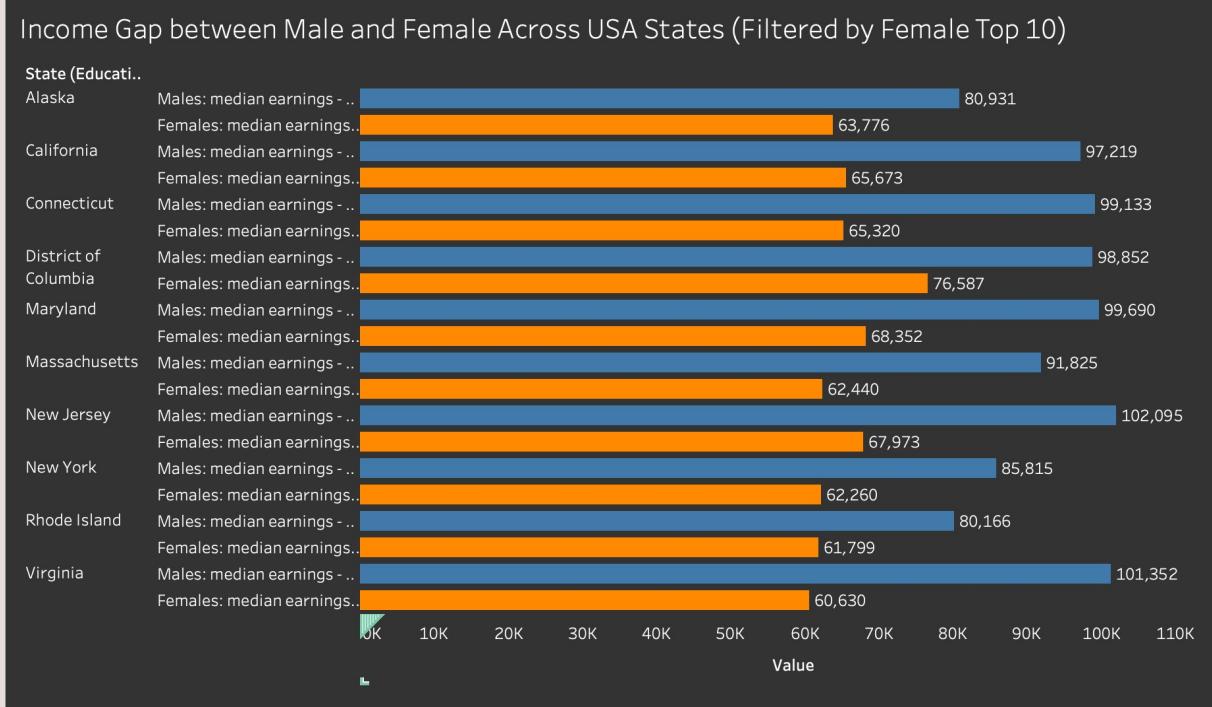


Gender Income Across US cont...

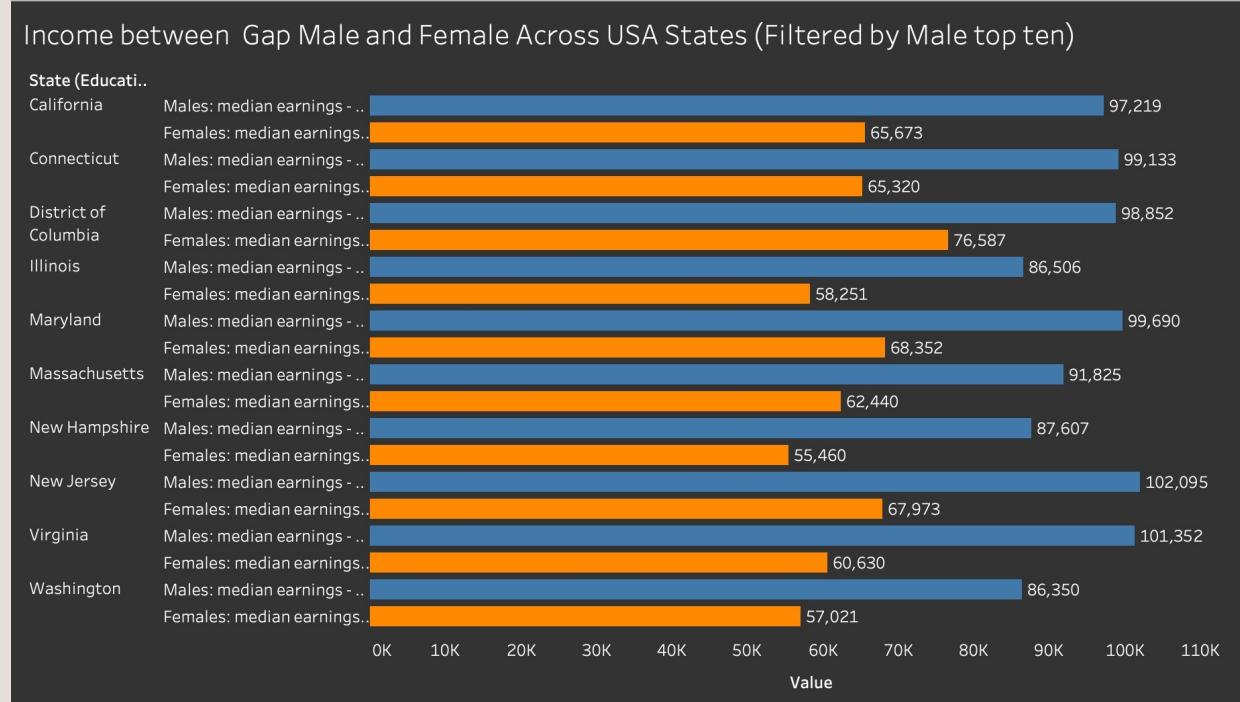




Gender Income Across US...



Gender Income Across US cont...



An aerial photograph of the Hollywood sign in Los Angeles, California. The sign itself is white and mounted on a steep, rocky hillside. In the background, the city of Los Angeles sprawls across a valley, filled with numerous houses, streets, and larger commercial buildings. To the right of the sign, there is a large red metal telecommunications tower with multiple white and grey satellite dishes attached. The foreground consists of the dark, eroded earth of the hillside where the sign is built.

Case Study: California

HOLLYWOOD

California cont...

Atherton 525,324\$ (Annual Household Median Income).



Little River CDP 3,000\$ (Annual Household Median Income).



Challenges



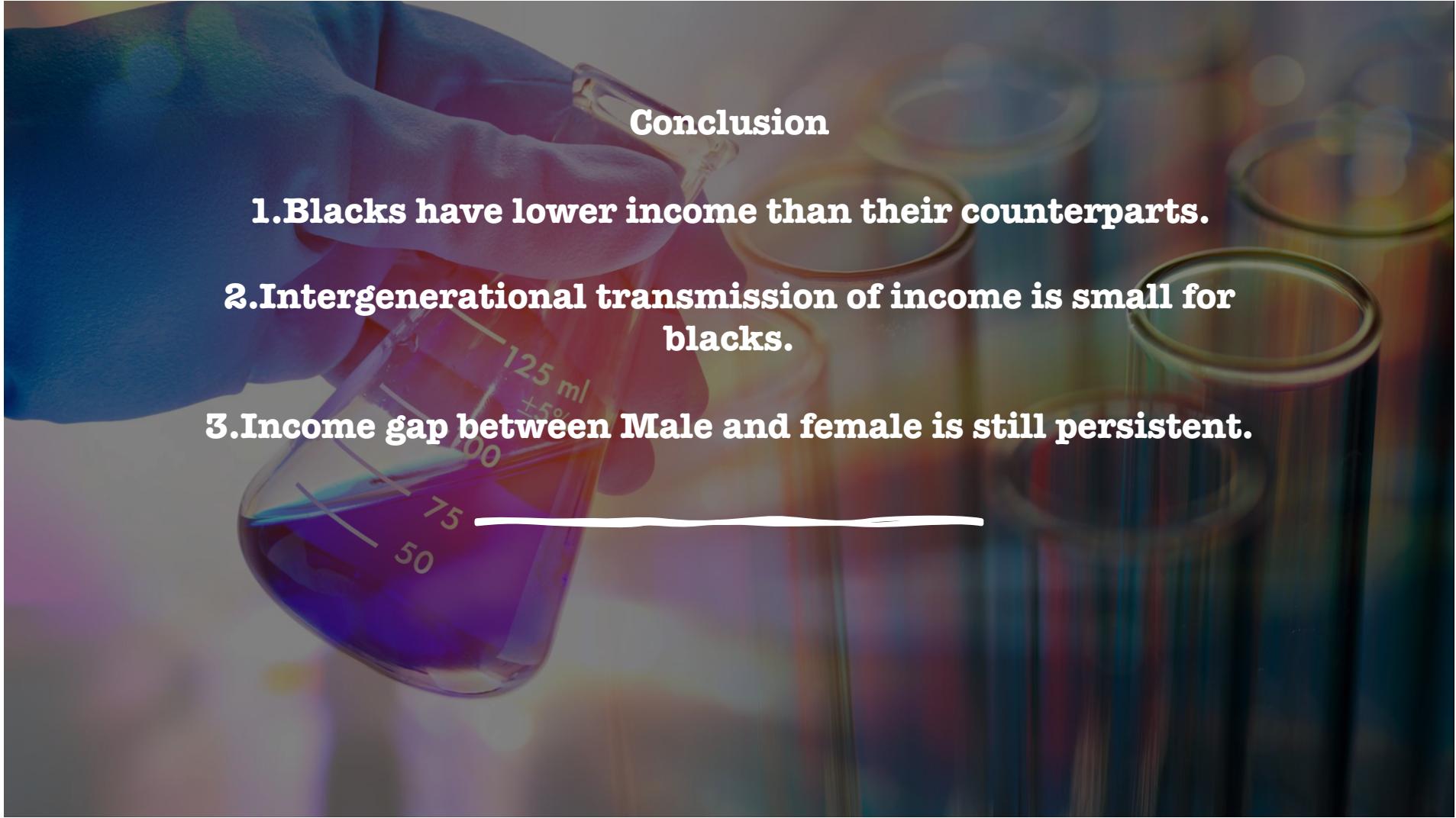
IMBALANCED DATA SET



FINDING DATA



ENHANCED THE DATA SET



Conclusion

- 1. Blacks have lower income than their counterparts.**
- 2. Intergenerational transmission of income is small for blacks.**
- 3. Income gap between Male and female is still persistent.**



Recommendations

Thank you

