



Presented By:  
Group 6 Bashee  
AR, Bernard, Enzo, Hector, Sandy

7 October 2023

# LET IT FLOW (WISELY):

A Research Study on Philippine  
Household Cashflow Management  
in 2021

”

**The Philippines is at  
the **bottom 30** out of  
144 countries in terms  
of financial literacy.**



Bangko Sentral ng Pilipinas, July 2023



# Presentation Content

## Content List

**01.**

Introduction

**02.**

Objective

**03.**

Preprocessing

**04.**

EDA & Outliers

**05.**

ML Models

**06.**

Analysis

**07.**

Conclusion

# INTRODUCTION





# Decent Work and Economic Growth

**Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all**

Achieving Goal 8 will require a wholesale reform of the financial system to tackle rising debts, economic uncertainty and trade tensions, while promoting equitable pay and decent work for young people.





OBJECTIVES



# Objectives

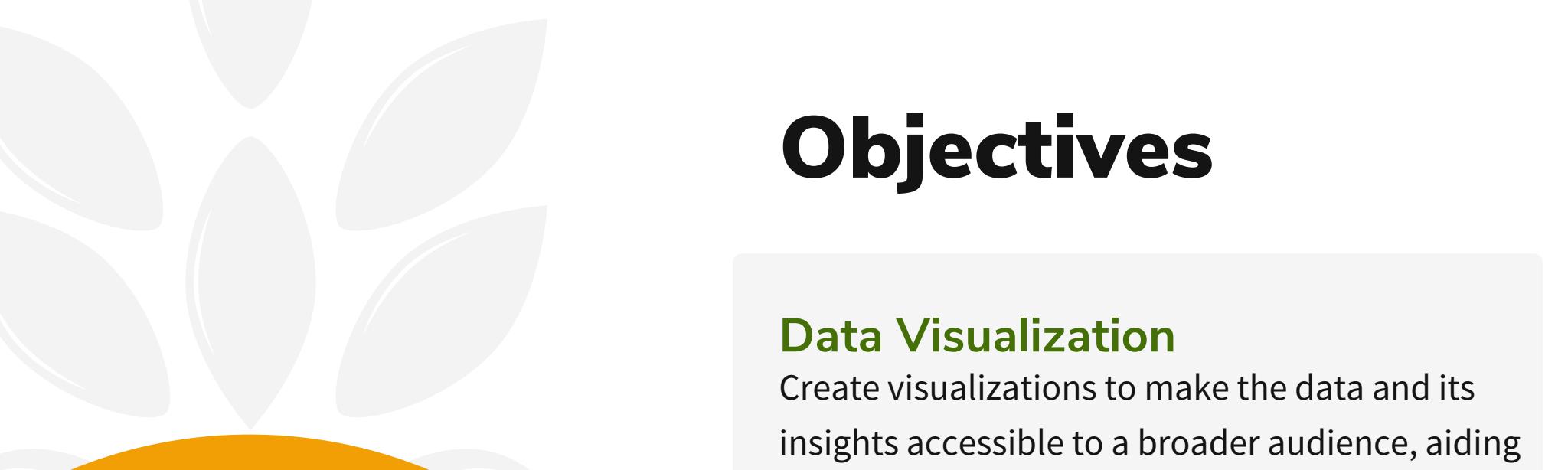


## Predictive Modeling

Create a model that can predict if a Filipino household has a positive or negative cashflow.

## Identify Key Drivers

Determine the primary factors and features that contribute to the positive and negative cash flow within households.



# Objectives



## Data Visualization

Create visualizations to make the data and its insights accessible to a broader audience, aiding policymakers, financial institutions, and individuals in making informed decisions.

## Policy Insights

Provide insights for policymakers, industrialists, and supporting agencies on how to improve financial stability of Filipino households.



**PREPROCESSING**

# 2021 Family Income and Expenditure Survey

The FIES is a nationwide survey of households undertaken **every 3 years** by the Philippine Statistics Authority (PSA).

# 2021 Family Income and Expenditure Survey

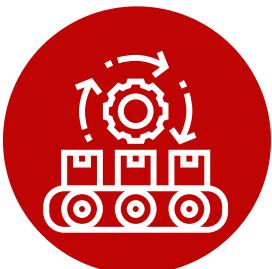
**165,029 survey respondents**

**88 numeric variables**

**58 features selected**

**All 17 regions of the country**

# 2021 Family Income and Expenditure Survey



# Cashflow

## Positive Cashflow

Income is greater than Expenses



## Negative Cashflow

Income is less than Expenses



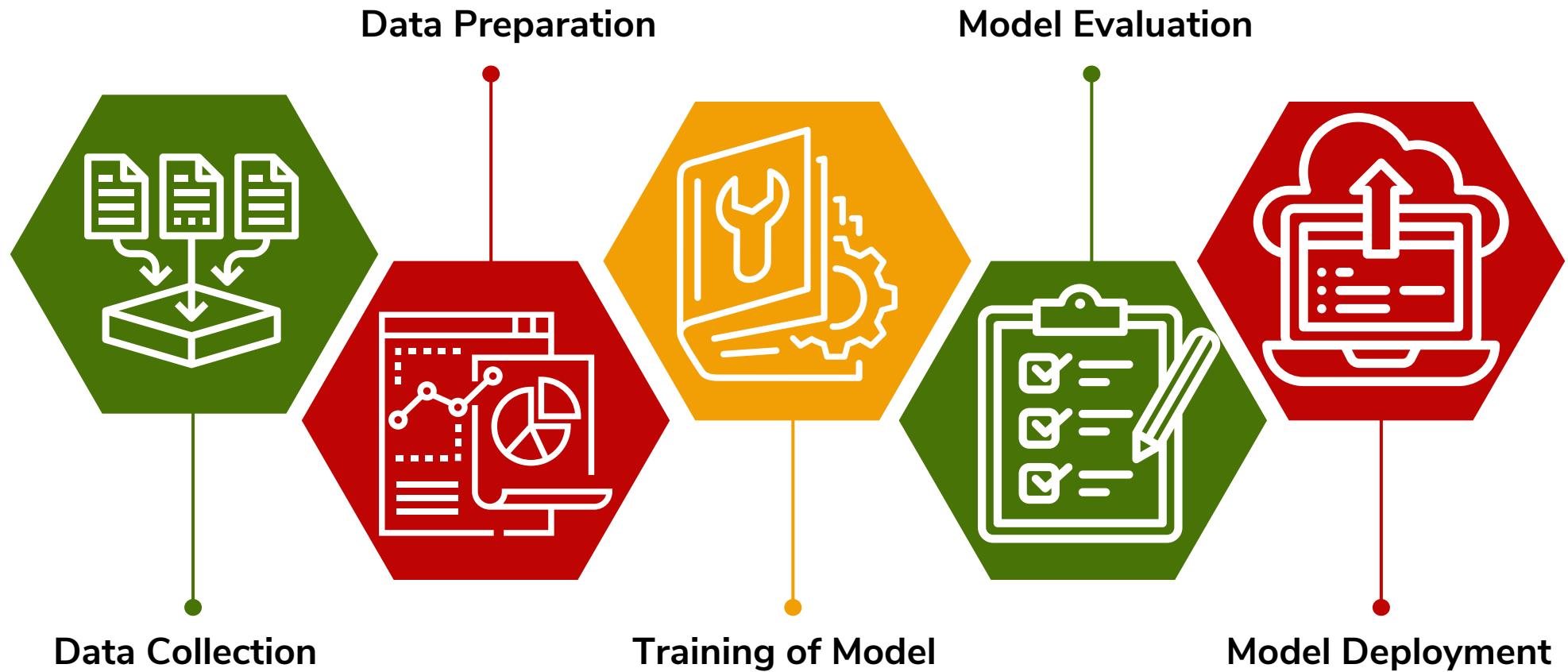
# Limitations

## Several data from previous surveys were removed

Interesting information (such as type of house, water supply source, and number of certain household appliances) which were included in the 2015 and 2018 survey, were not included in the 2021 survey.



# Machine Learning Pipeline

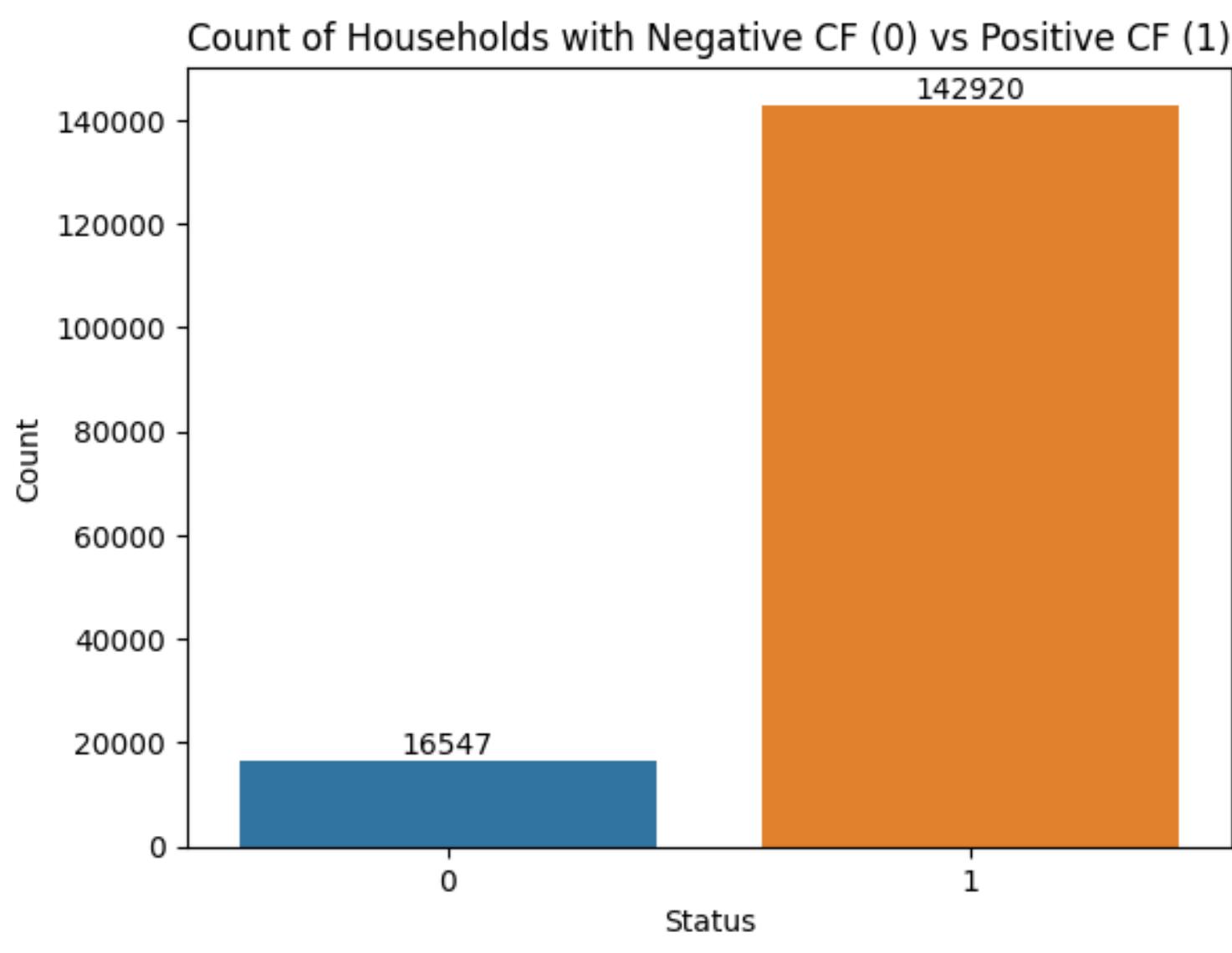




# EDA & OUTLIERS

# Comparison of Households

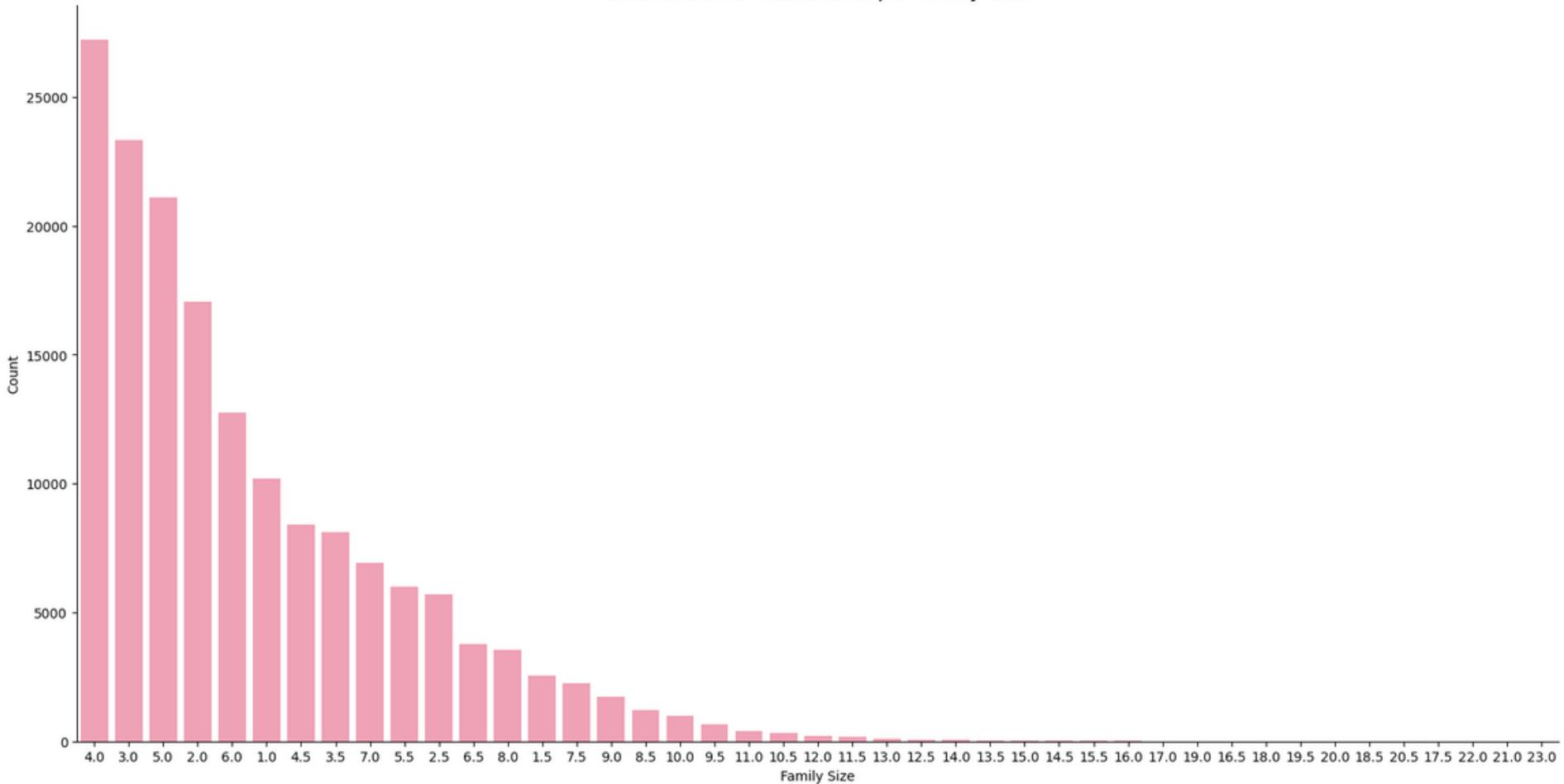
## Positive vs. Negative Cashflow



# Households Per Family Size

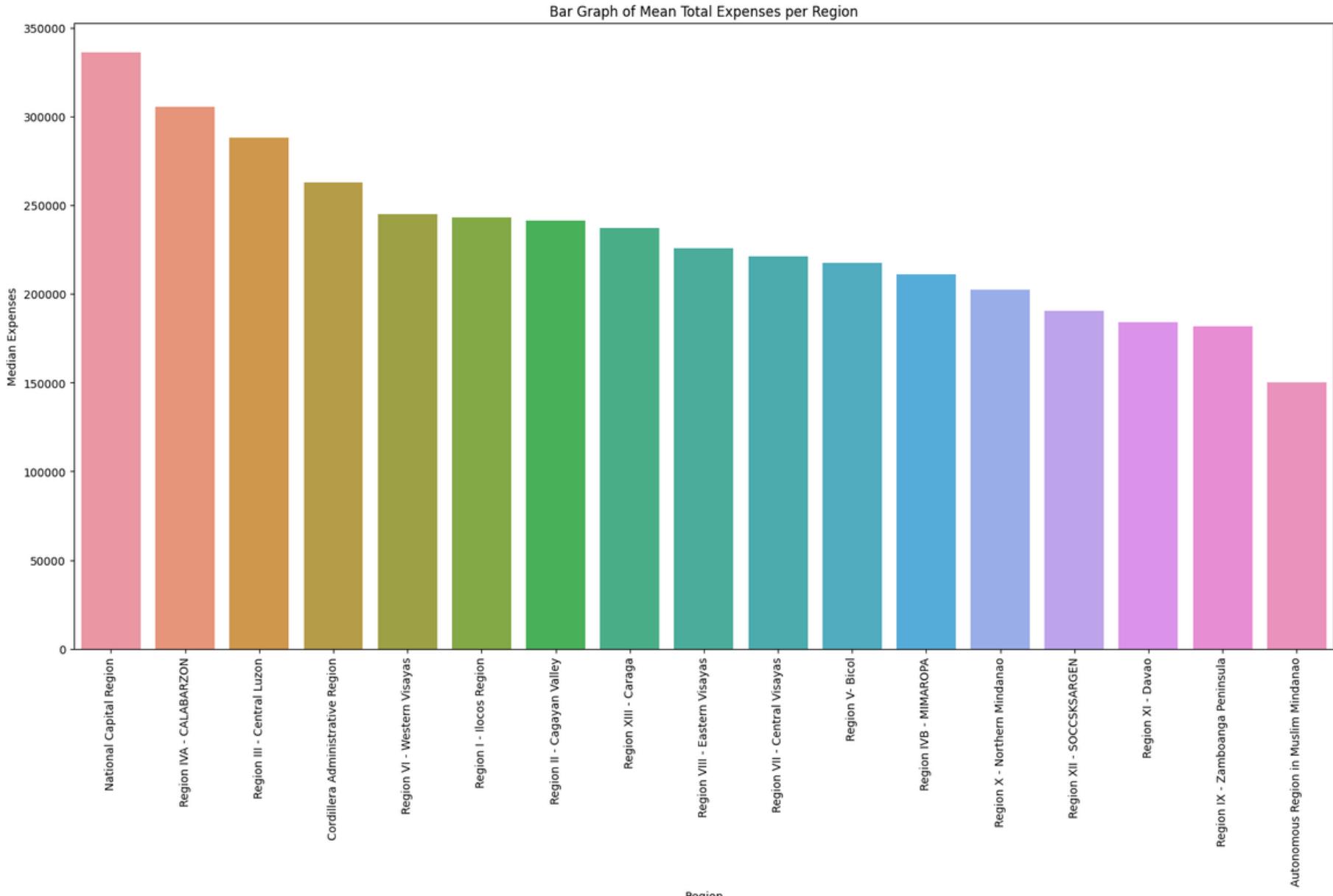
As the family sizes increase beyond 7, the frequencies decrease rapidly.

Distribution of Households per Family Size



# Mean Total Income

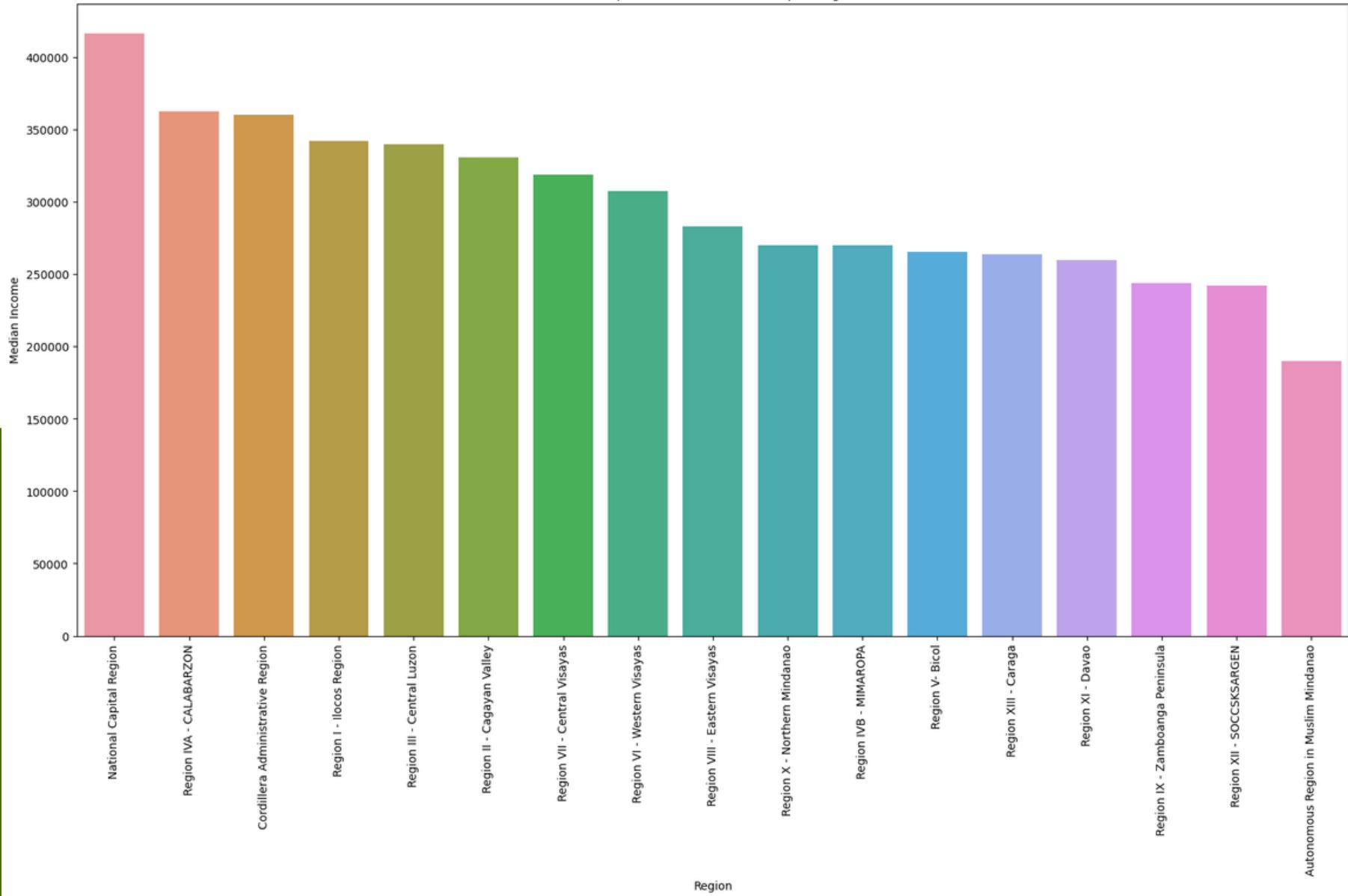
NCR leads amongst all regions on mean total income



# Mean Total Expenses

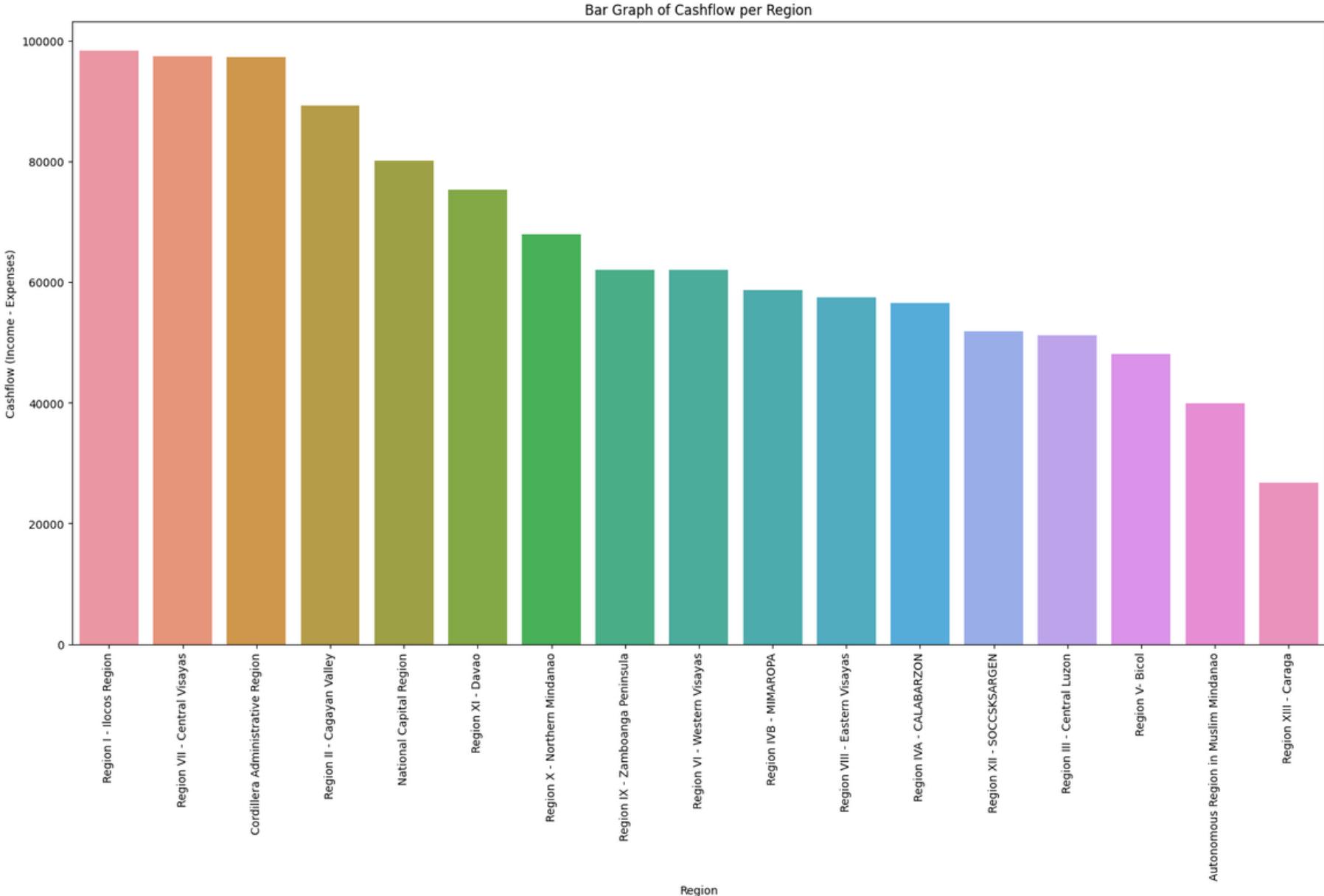
NCR leads the region with the mean total expenses

Bar Graph of Mean Total Income per Region



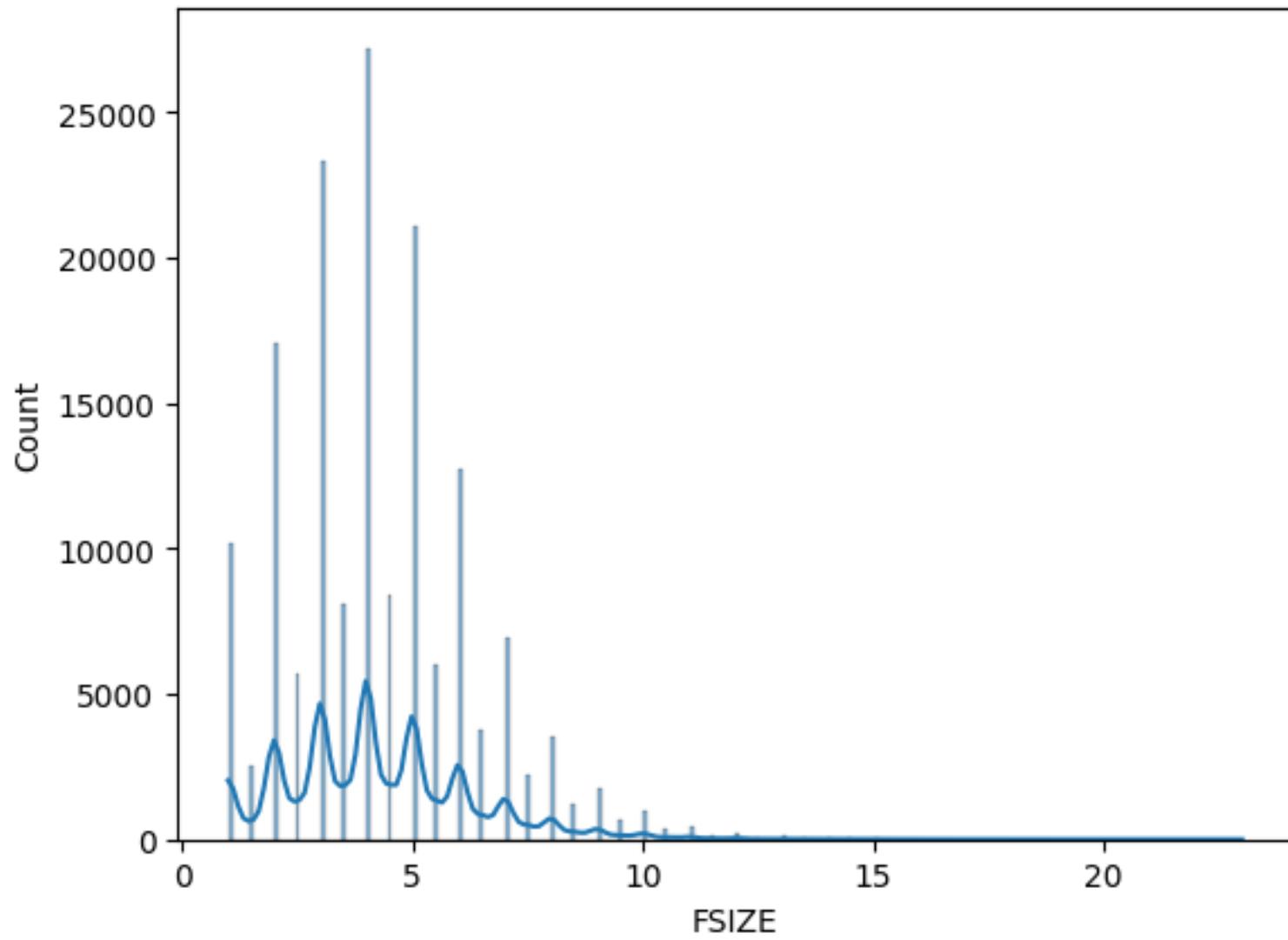
# Mean Net Income

Ilocos Region has the highest mean net income, with NCR at fifth



# Outlier Detection

Family Size is considered for outlier detection and treatment



# MACHINE LEARNING



# Model Building

	Parameters					Sampling Method	
Models	Learning Rate	n_neighbors	n_estimators	max_depth	Oversampling	Undersampling	
KNeighbors Classifier	-	2, 3, 4, 5, 6, 7, 8, 9, 10	-	-			
Logistic Regressor	0.1, 0.3, 0.5, 0.7, 0.9, 1.0	-	-	-			
Decision Tree Classifier	-	-	1,10		SMOTE , ADASYN	TomekLinks, NearMiss	
Random Forest Classifier	-	-	100, 200, 300	5, 10, 15, 20			
Gradient Boosting Classifier	1, 0.5, 0.1	-	1, 10, 50				

# Baseline Performance

Model	Train Recall	Val Recall
KNeighbors Classifier	98.94%	97.97%
Logistic Regressor	99.62%	99.63%
Random Forest Classifier	100.00%	99.95%
Decision Tree Classifier	100.00%	99.09%
Gradient Boosting Classifier	100.00%	99.99%

# Oversampling Summary

	SMOTE		ADASYN	
	Train Recall	Val Recall	Train Recall	Val Recall
KNeighbors Classifier	85.64	85.64	84.84	80.14
Logistic Regressor	76.18	76.27	76.18	76.23
Random Forest Classifier	100.00	97.81	100.00	97.60
Decision Tree Classifier	100.00	95.18	100.00	95.26
Gradient Boosting Classifier	94.75	94.51	94.84	94.61

# Oversampling Summary

	SMOTE		ADASYN	
	Train Recall	Val Recall	Train Recall	Val Recall
KNeighbors Classifier	85.64	85.64	84.84	80.14
Logistic Regressor	76.18	76.27	76.18	76.23
Random Forest Classifier	100.00	97.81	100.00	97.60
Decision Tree Classifier	100.00	95.18	100.00	95.26
Gradient Boosting Classifier	94.75	94.51	94.84	94.61

# Undersampling Summary

	TomekLinks		NearMiss	
	Train Recall	Val Recall	Train Recall	Val Recall
KNeighbors Classifier	98.87	97.04	90.21	30.48
Logistic Regressor	99.53	99.51	82.6	41.89
Random Forest Classifier	100.00	99.93	100.00	53.68
Decision Tree Classifier	100.00	99.08	100.00	77.70
Gradient Boosting Classifier	99.99	99.99	95.96	67.85

# Undersampling Summary

	TomekLinks		NearMiss	
	Train Recall	Val Recall	Train Recall	Val Recall
KNeighbors Classifier	98.87	97.04	90.21	30.48
Logistic Regressor	99.53	99.51	82.6	41.89
Random Forest Classifier	100.00	99.93	100.00	53.68
Decision Tree Classifier	100.00	99.08	100.00	77.70
Gradient Boosting Classifier	99.99	99.99	95.96	67.85

# Resampling + Hypertuning

		Train Recall	Val Recall	Holdout Recall	Holdout Run Time
KNN	SMOTE + Hypertuned	72.80	73.15	73.37	4.32
Logistic Regression	SMOTE + Hypertuned	67.44	69.52	69.33	9.55
Random Forest	TomekLinks + Hypertuned	100.00	99.91	99.92	594.56
Decision Tree	TomekLinks + Hypertuned	99.30	98.20	98.30	132.33
Gradient Boost	TomekLinks + Hypertuned	99.45	100.00	100.00	~1200

# Final Model

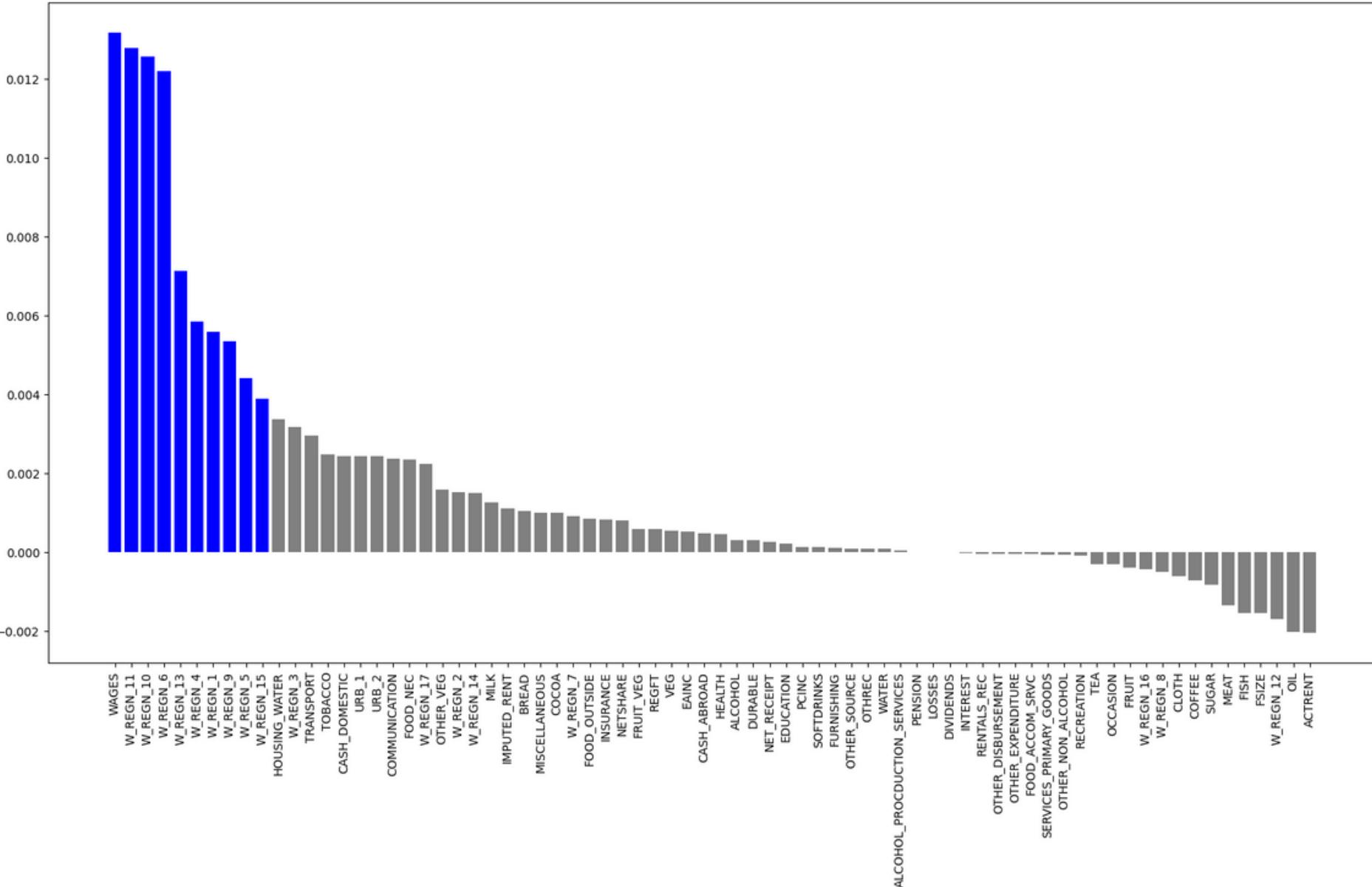
PCC\_Recall = 10.19%

		Train Recall	Val Recall	Holdout Recall	Holdout Run Time
KNN	SMOTE + Hypertuned	72.80	73.15	73.37	4.32
Logistic Regression	SMOTE + Hypertuned	67.44	69.52	69.33	9.55
Random Forest	TomekLinks + Hypertuned	100.00	99.91	99.92	594.56
Decision Tree	TomekLinks + Hypertuned	99.30	98.20	98.30	132.33
Gradient Boost	TomekLinks + Hypertuned	99.45	100.00	100.00	~1200

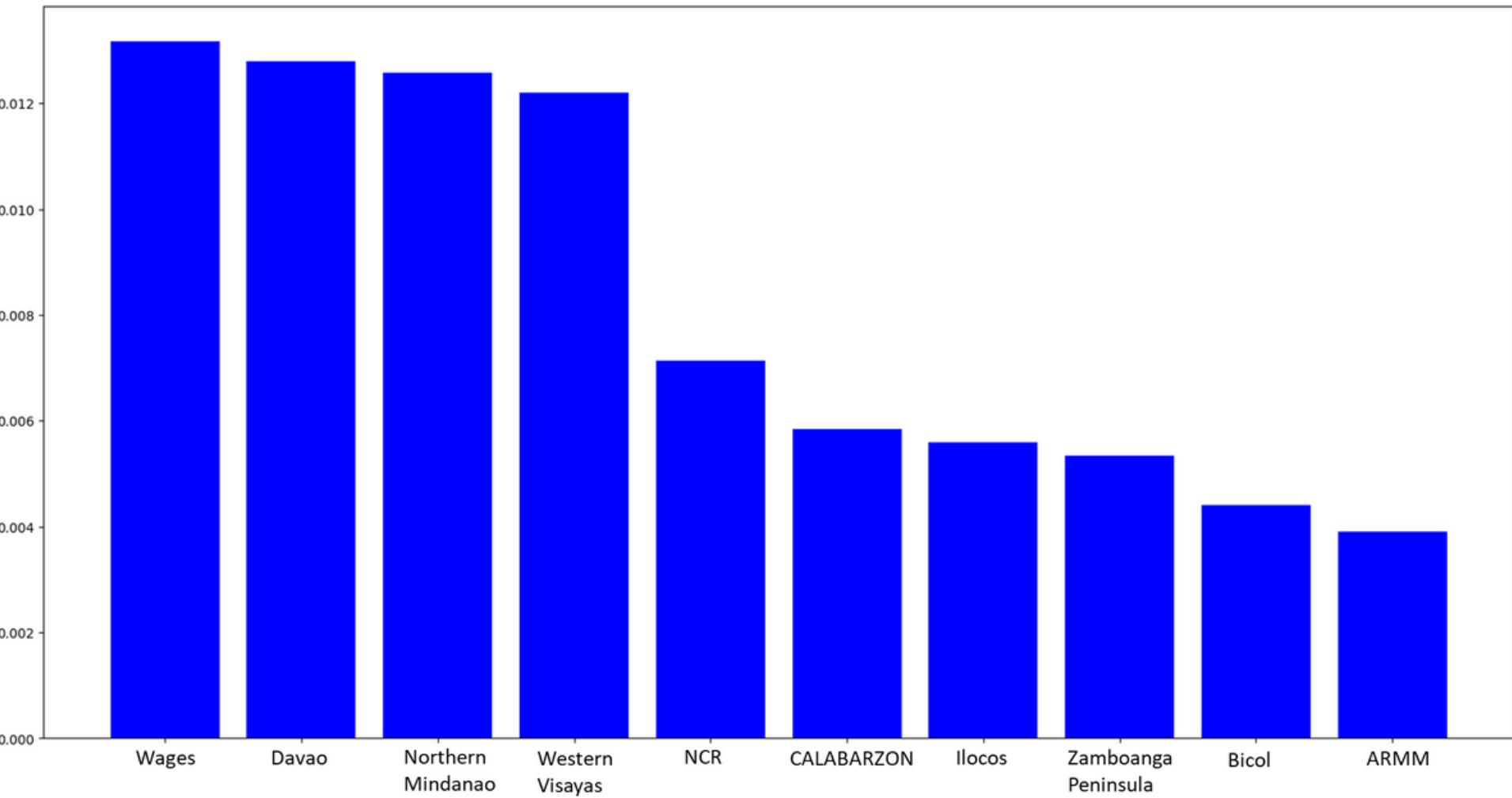


# ANALYSIS

# Feature Importance

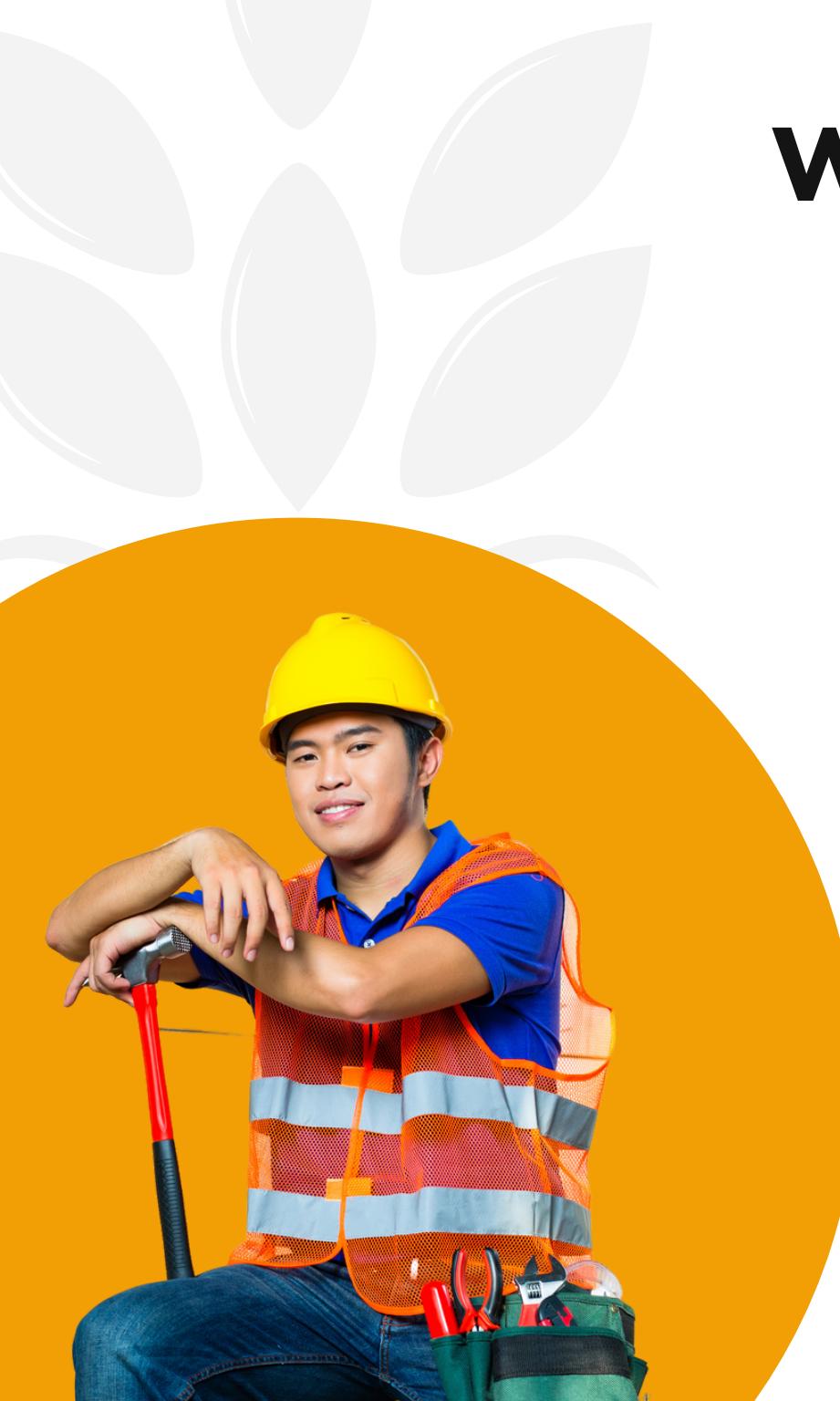


# Feature Importance



A group of four young people are lying on a green lawn, looking up at the camera. They are all making heart shapes with their hands over their eyes. The person on the far left is wearing a white shirt and a gold chain. The person next to them is wearing a striped shirt and a brown watch. The person in the center is wearing a red and orange striped shirt. The person on the far right is wearing a blue and white striped shirt and a wooden bracelet.

**CONCLUSION**



# WAGE IS POWER

The number one contributing feature is 'Wage'

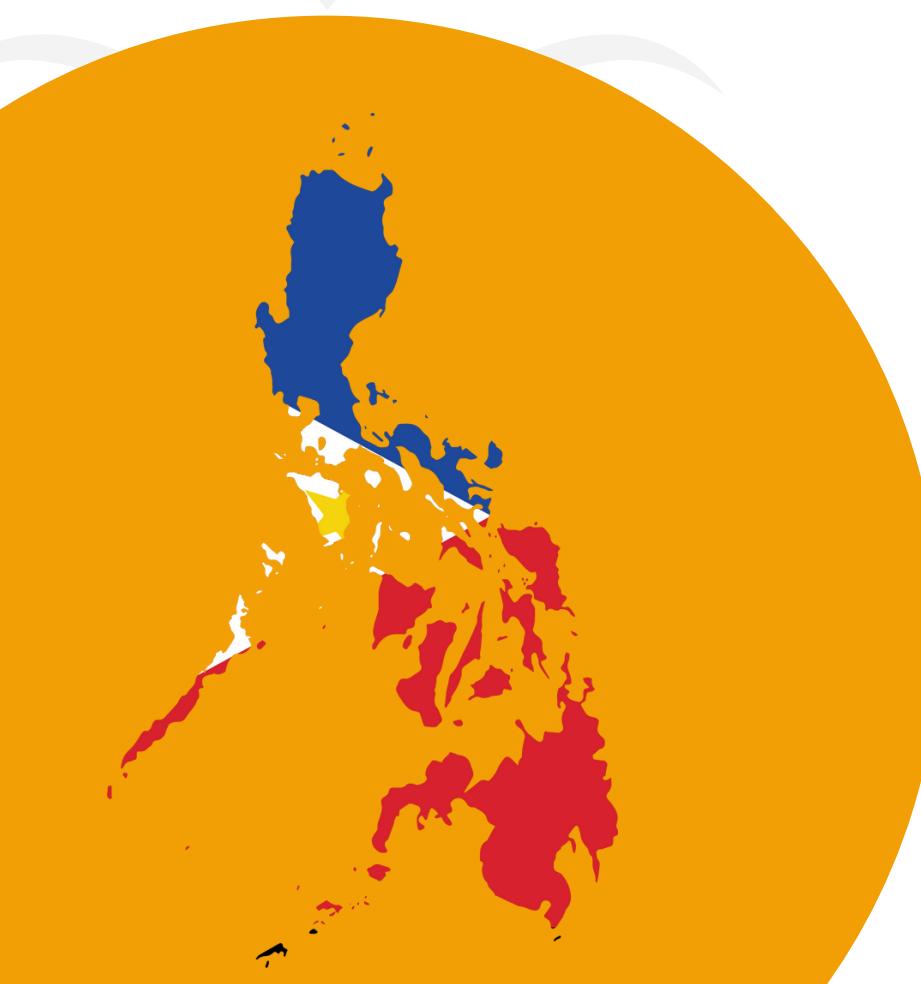
As expected,

a household's total wages is a huge determining factor whether or not a household will be able to save.





# Regional Differentiation



Interestingly, regions appear to contribute greatly to a household's saving capacity.

Occupying 9 out of 10 spots in the top 10 contributing features, it appears that regions play an important factor in a household's potential to save.

Let's dive deeper into this.





# Economic Growth in the Regions

The top 3 regions in feature importance are Davao Region, Northern Mindanao, and Western Visayas.

These regions are seeing a great deal of economic growth around 2022\*, inferring an increase in economic activity in the regions in 2021.

More economic activity  
=  
More opportunity to save  
and spend

\*Sources: [Statista](#), [PSA](#), [BusinessWorld](#)



# On Expenses

For expense features with non-zero feature importance measures, a handful are completely avoidable.

These expenses features are the following:

- Tobacco
- Alcohol
- Softdrinks

Finding ways to encourage limited expenditures on these would be beneficial.

# Recommendations

## 1. The importance of high wages is reiterated.

Thus, it is important to give people the chance to get higher wages.

## 2. Education on financial literacy is a must

To effectively encourage people to watch out their saving and spending habits, education on its importance and how to do it is necessary.



# Recommendations

## 3. The battle field is on the regions with high-economic growth statuses

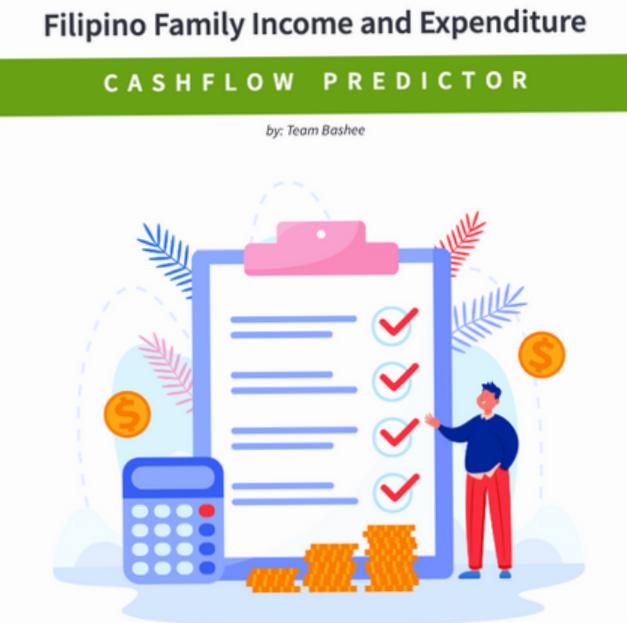
Pareto principle - focus financial literacy education on regions with high economic growth/activities.

## 4. Cutting down on high-model contribution but unnecessary expenses could be looked into

Such are expenses that include tobacco, alcohol, and softdrinks.



# Streamlit App



<https://fies-ph-cashflow-predictor.streamlit.app>



A close-up photograph of five young children, likely of Asian descent, laughing joyfully. They are crowded together, with one child in the center wearing a pink top and another in the foreground wearing a red and white striped shirt. The background is blurred.

7 October 2023

# THANKS

Top 10 Household Expense by Amount

