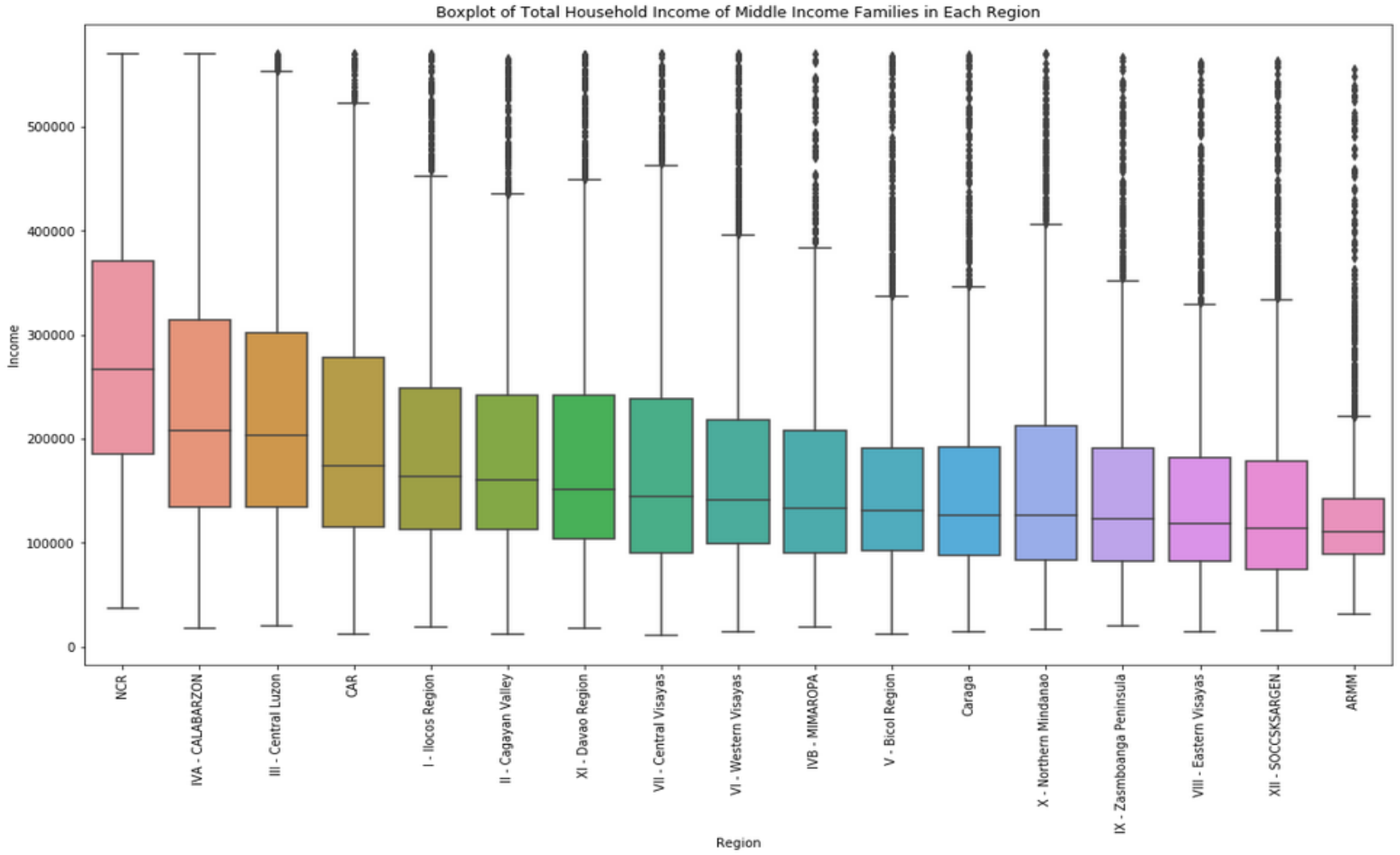
EDA on Filipino Family Income and Expenditure Dataset

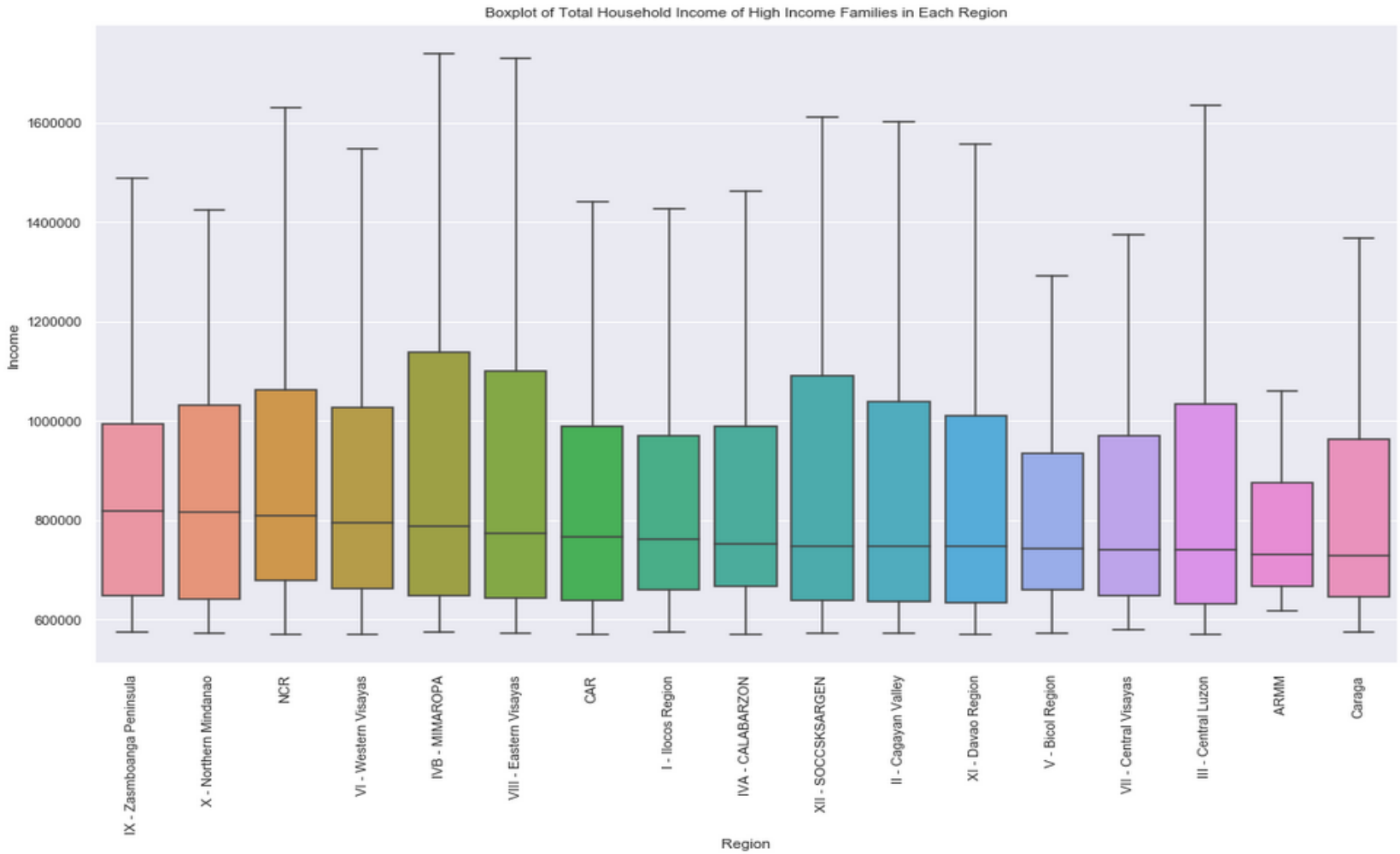
- Anirudh Ayya (PES1201800338)
- Ashish Shenoy (PES1201801447)

Middle Income Families by Region



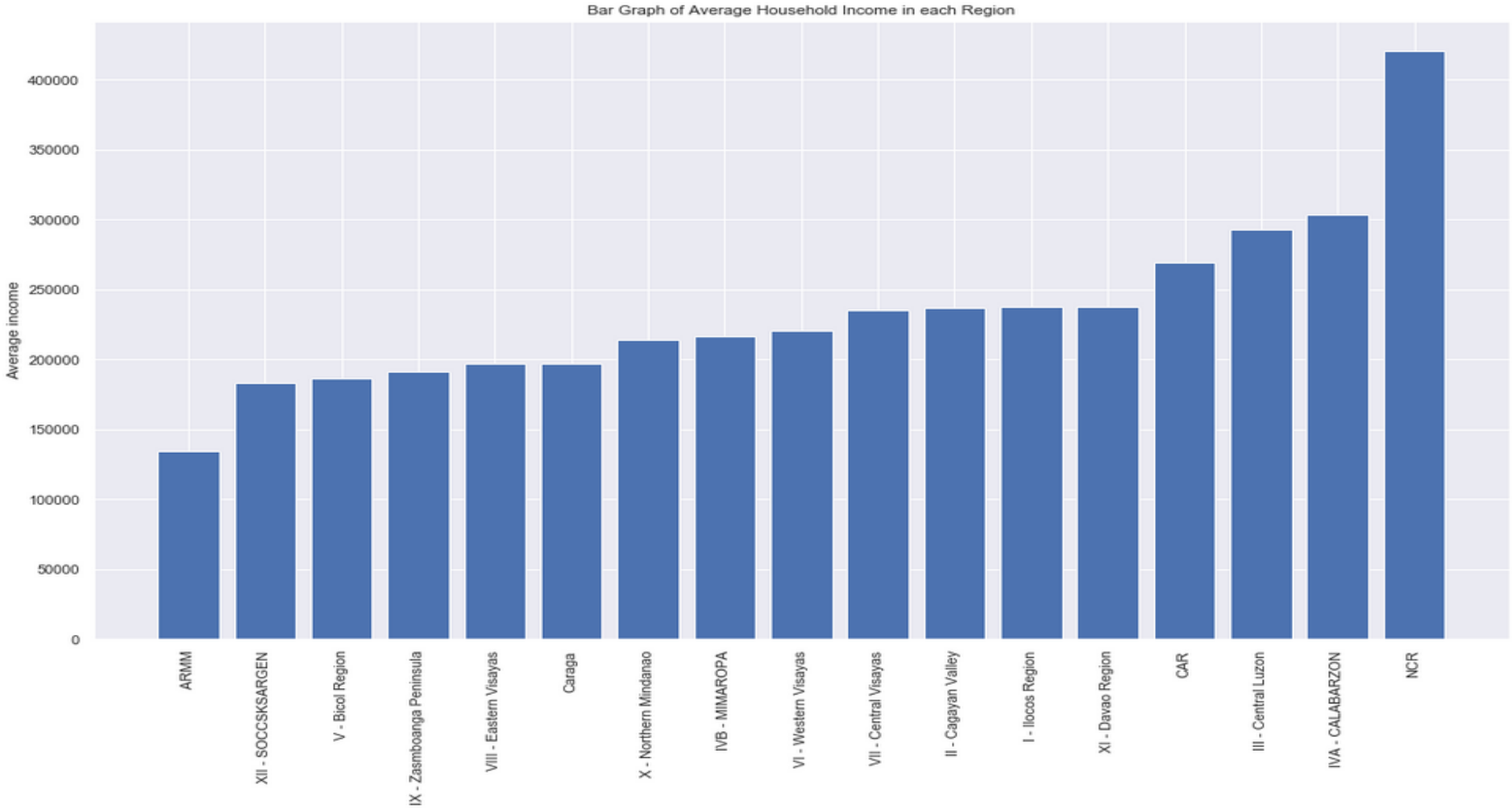
Thus our middle income category represents the weathier developed regions such as **NCR** and **IVA-CALABARZON** well, but doesn't represent poorer agricultural regions such as **ARMM** well.

High Income Families by Region



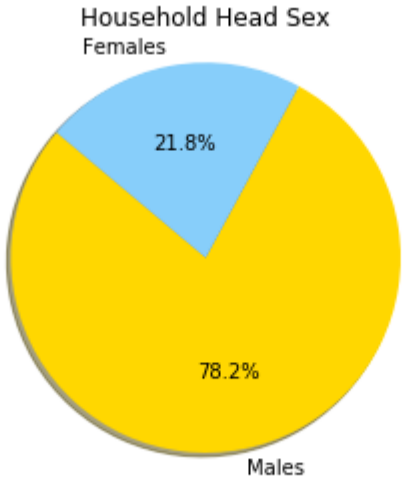
Outliers have been ignored here, but high income category represents all regions extremely well.

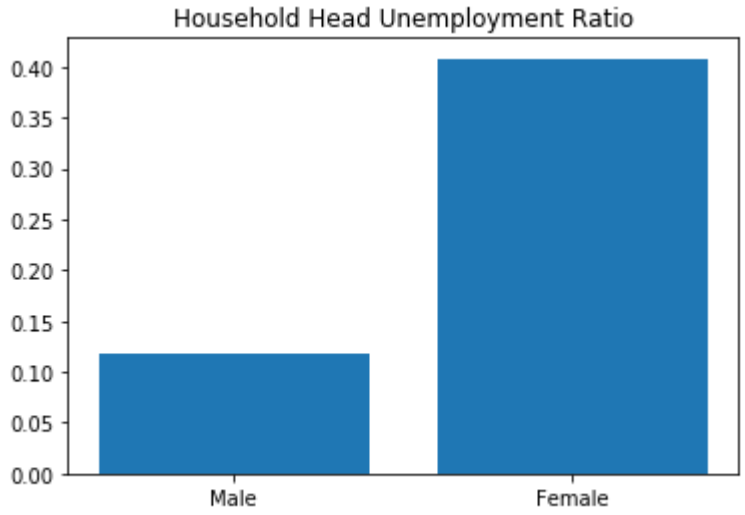
Average Household Income in each Region



We observe people staying in **NCR region have the highest value of average income**, NCR region is the capital region and the population density is likely higher compared to other cities. This graph shows that the metropolitan NCR is the richest region of all ,and the agricultural **ARMM region is the poorest region**.

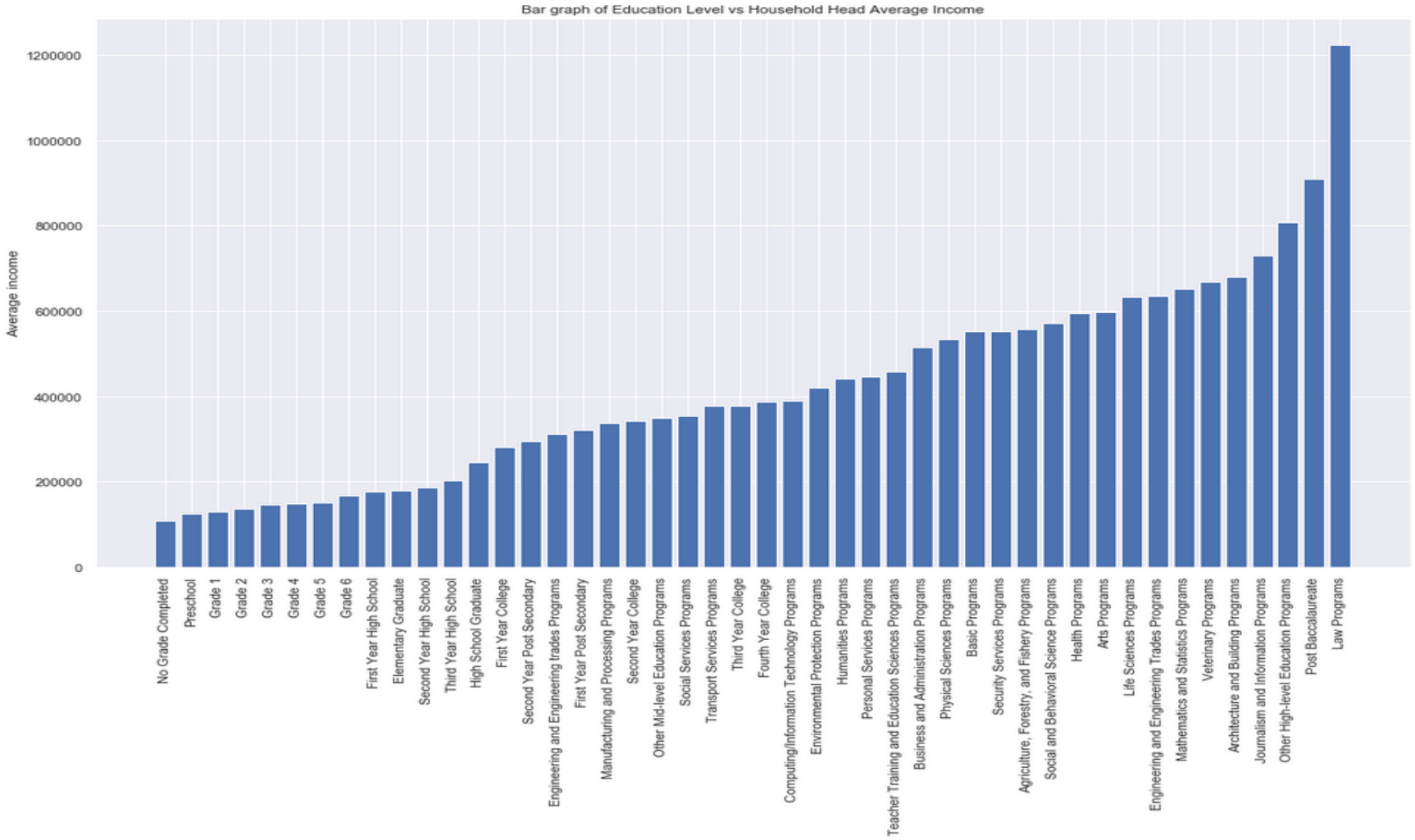
Household Head Sex Unemployment Ratio





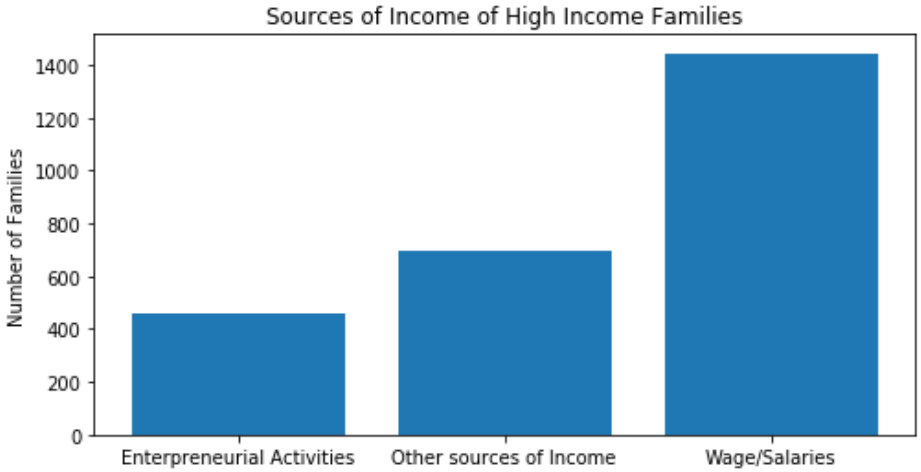
Thus even though the proportion of males as family heads were greater than females, the **unemplyment ratio among females was far higher** than that of males.

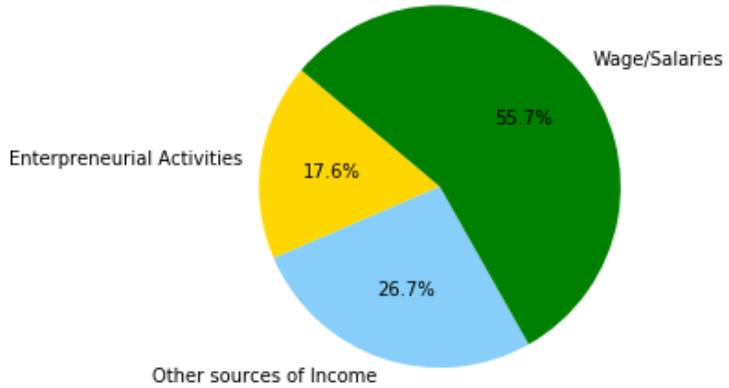
Educational Level vs Average Household Income



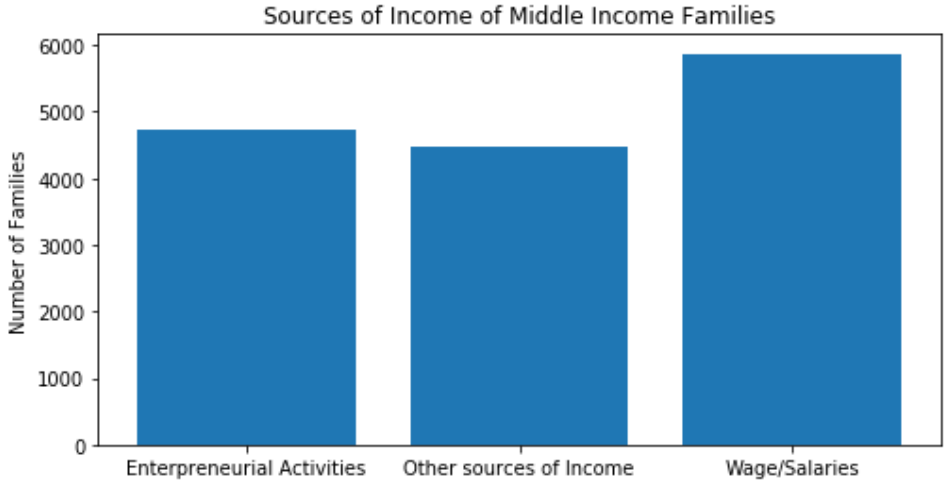
Therefore , we observe that people who have studied **law programs** have the highest average salary. People with **no grade completed** have the lowest average income .

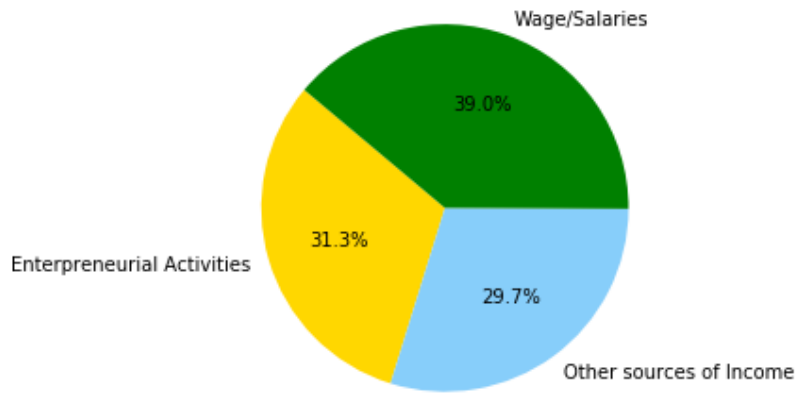
Sources of Income





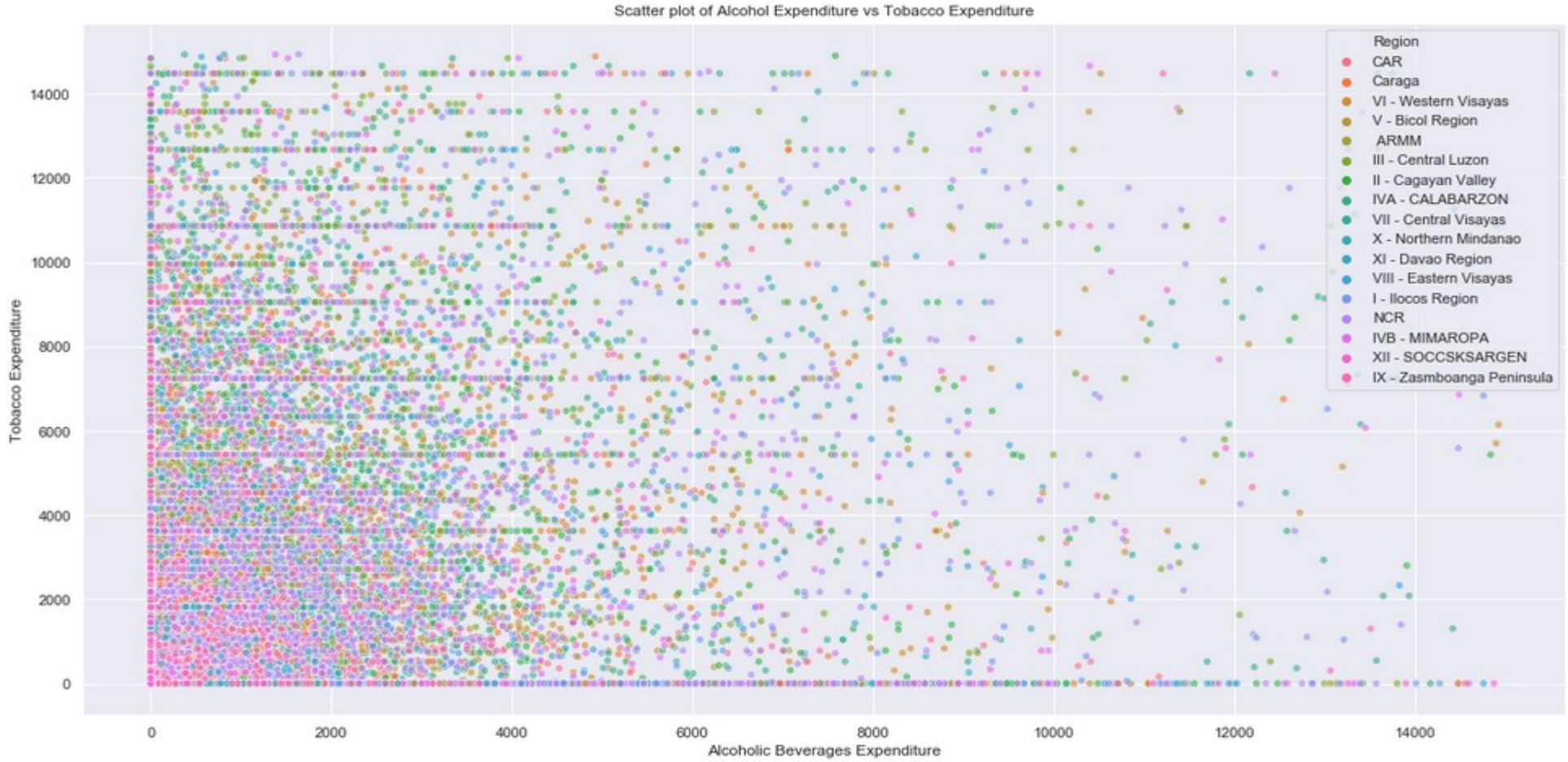
Therefore most of the high income families have their main source of income as **Wage/ Salaries**. There is a low proportion of enterpreneurs. People tend to go for safer options.





We observe that the main source of income for both is wages.
But 31.3**% of the middle income** families become enterpreneurs compared to that of only **17.6% of high income** families.

Correlation between Alcohol and Tobacco Expenditure




```
Correlation = 0.25690251014422383
```

Thus we can see that Alcoholic Beverages Expenditure and Tobacco Expenditure **do not** have a very strong correlation.

Correlation between Household Income and House Floor Area

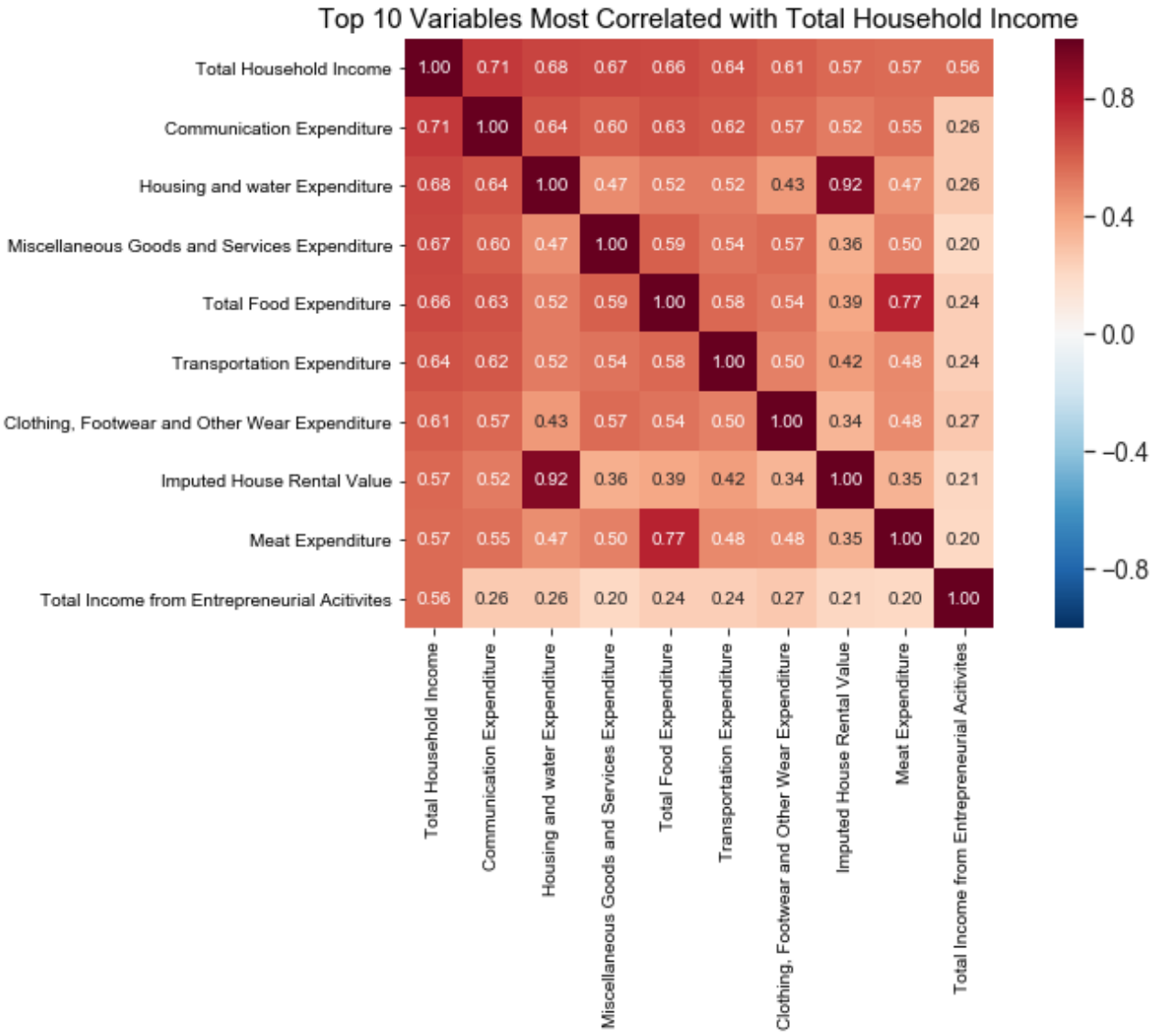
```
Correlation = 0.3432089548446198
```

As the total household income increases, the house floor area **does not** tend to increase. Thus people prefer to buy **smaller homes in well developed regions** rather than large homes in less developed regions.

Correlation Matrix

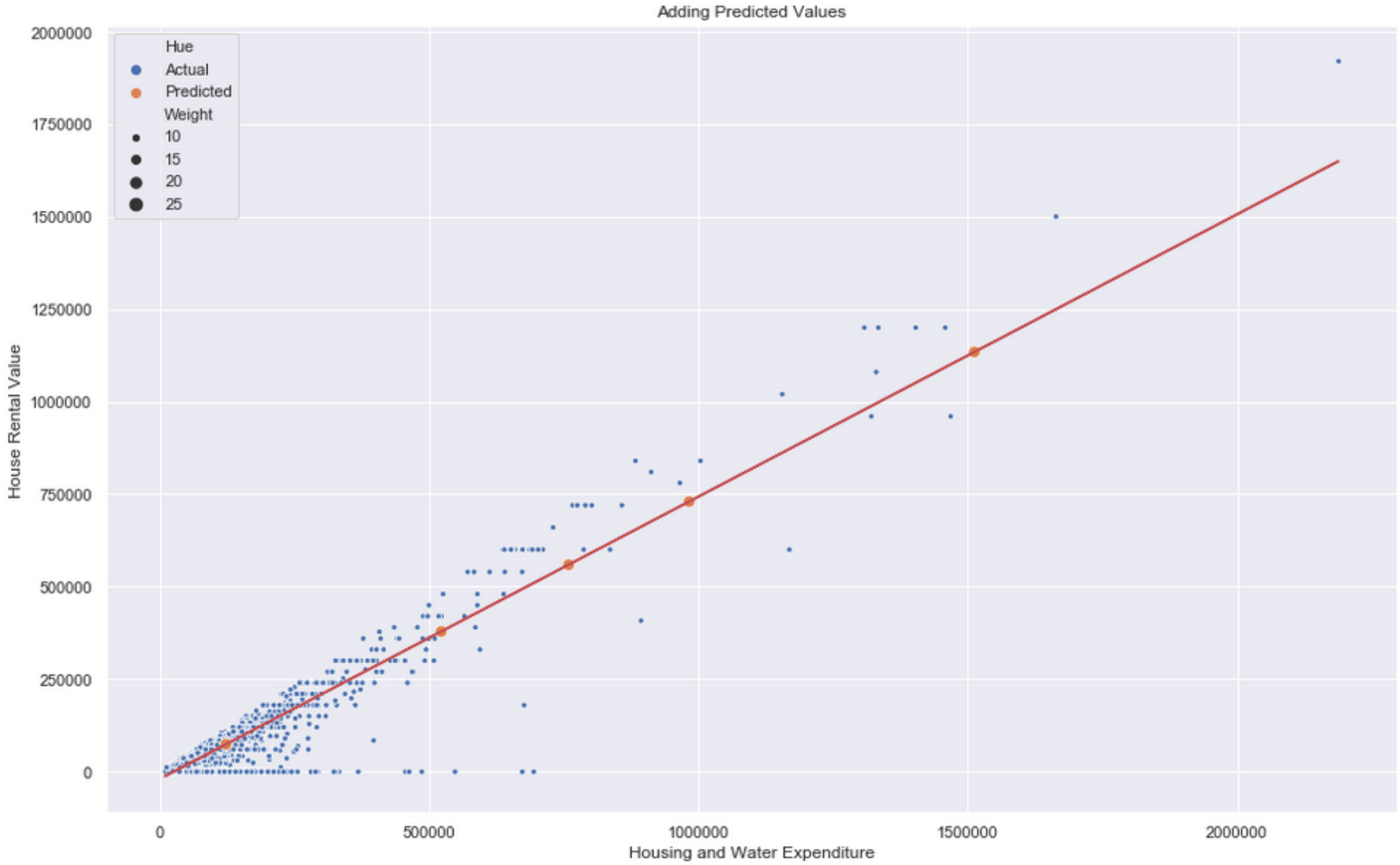
We observe that **Total Household Income** and **Total Food Expenditure** are the variables that correlates most well with all the other variable.

Top 10 Variables Most Correlated with Household Income



Thus we observe that **communication**, **housing** and **water expenditure** are most strongly correlated to Total Household Income. Thus more wealthier families are more connected by means of Internet , Mobile Phones, TVs etc and they also have more well furnished and maintained homes, with adequate water supply.

Predicting House Rental Value from Housing and Water Expenditure



Thank You