Data Analysis on Filipino Family Income and Expenditure

By: Johann Sebastian E. Catalla, BSCS-1

An output for the course CP102: Computer Programming 2

The dataset was sourced through bagoo from the Philippine Statistics Authority's (PSA) Family Income and Expenditure Survey (FIES) nationwide. The survey, which is undertaken every three (3) years, is designed to collect data on family income and expenditure, including, among others, household expenditures by kind of expenditure, sources of income in cash, and related information affecting income and expenditure patterns in the Philippines.

The Dataset contains more than 40k observations and 60 variables which is primarily comprised of the household income and expenditures of that specific household. For this analysis, I will be using Python's **pandas**, **numpy**, and **matplot** libraries.

```
import pandas as pd
import numpy as np
import matplotlib as mpl
%matrplotlib inline
%matplotlib inline(backend='qt5')
%matplotlib inline(backend='inline')
# !ls
```

```
dataset = pd.read_csv("family income and expenditure.csv")
dataset
```

	Total Household Income	Region	Total Food Expenditure	Main Source of Income	Agricultural Household indicator	Bread and Cereals Expenditure	Total Rice Expenditure	Meat Expenditure	Total fish and marine products Expenditure	Fish Expenditure	Number of Refrigerators/Freezers	Number of Washing Machine	Number of Car, Jeep, Van	Number of Land Owned
0	480332	CAR	117648	Wage/Salaries	0	67140	30500	20476	16896	3325	1	1	0	0
1	198235	CAR	31766	Wage/Salaries	0	17670	13500	11436	1323	2016	0	1	0	0
2	107351	CAR	70179	Wage/Salaries	0	44625	24500	19600	4725	4900	1	1	0	0
3	67626	CAR	33048	Wage/Salaries	0	18480	2835	5054	1512	126	0	1	0	0
4	67626	CAR	18522	Wage/Salaries	0	4800	3937	4059	1728	378	0	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
41539	170775	SOCCSKSARGEN	48025	Entrepreneurial Activities	0	22675	27040	1476	6150	1050	1	1	0	0
41540	137028	SOCCSKSARGEN	31102	Entrepreneurial Activities	1	2001	1215	1886	4356	1680	0	1	0	0
41541	153171	SOCCSKSARGEN	45886	Entrepreneurial Activities	2	29646	27559	890	4794	1232	1	1	0	0
41542	235372	SOCCSKSARGEN	81410	Entrepreneurial Activities	1	28686	26040	2583	11556	2002	1	1	0	0
41543	128598	SOCCSKSARGEN	56448	Entrepreneurial Activities	1	4480	41400	1400	7191	2042	0	1	0	0

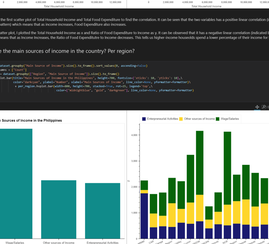
41544 rows × 60 columns

QUESTIONS

1. What is the Average Yearly Food Expenditure and Average Income per region? Which region has the highest and lowest?

2. What is the ratio of Food Expenditure to income per region? Which region has the highest and lowest?

```
# Grouping the regions and calculating the means of the food expenditure and income
dataset['Ratio of Food Expenditure'] = dataset['Total Food Expenditure'] / dataset['Total Household Income']
# Calculate the ratio of food expenditure to income
ave, r = dataset.groupby('Region')['Ratio of Food Expenditure'].mean(), sort_values(ascending=False).to_frame()
total_food = dataset.groupby('Region')['Total Food Expenditure', 'Total Household Income'].mean().round(2), sort_values(by='Total Food Expenditure', 'Total Household Income')
# Plotting the bar chart
total_food.plot(kind='bar', figsize=(10, 6), title='Average Yearly Food Expenditure and Average Yearly Income per Region', ylabel='yearly', xlabel='Region')
    height=18, width=20, alternative=False, line_separator=1,
ave.plot(kind='barh', figsize=(10, 6), title='Ratio of Food Expenditure to Income', xlabel='Ratio of Food Expenditure', ylabel='Region')
    fontsize=7, title='', xticks=' ', ylabel='', xlabel='', legend=False, legend_loc='lower right', alpha=None)
# Save the plot as a PNG file
plt.savefig('ratio_food_expenditure.jpg', dpi=300, bbox_inches='tight')
```



It is evident that the National Capital Region (NCR) has the highest Average Household Income (428K) and the Highest Average Food Expenditure (127K). Moreover, the Autonomous Region of Muslim Mindanao (ARMM) has the lowest Average Household Income (129K) while Northern Mindanao has lowest Average Food Expenditure (64K).

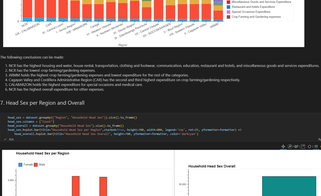
In terms of Ratio of Food Expenditure to Income, SOCCSKSARGEN has the highest ratio with 0.53, meaning 53% of their income goes to food. On the other hand, the region with the lowest ratio is NCR with 0.29. Despite having the highest income and highest food expenditure, NCR has the lowest ratio of food to income, which implies that their other expenses are marginally more prioritized compared to other regions.

It can be observed that lower income regions are more likely to allot a higher percentage of their income to food expenditures. Similarly, higher income regions are more likely to have a lower percentage of their income allotted for food expenditures.

3. Scatter plot of Household Income and Total Food Expenditure in the Philippines

4. Scatter plot of Household Income and Ratio of Food Expenditure to Income

```
dataset['Ratio of Food Expenditure'] = dataset['Total Food Expenditure'] / dataset['Total Household Income']
dataset.hvplot.scatter(x='Total Household Income', y='Total Food Expenditure', alpha=0.5, xlabel='Total Household Income', ylabel='Total Food Expenditure', by='Region', legend=False)
    height=700, xformatter='%.0f', yformatter='%.0f', fontsize={'xticks': 7, 'yticks': 7})
```



We can observe the first scatter plot of Total Household Income and Total Food Expenditure to find the correlation. It can be seen that the two variables has a positive linear correlation (indicated by the upward trend) where particular points which means that as income increases, Total Food Expenditure also increases.

In the second scatter plot, I plotted the Total Household Income on x and Ratio of Food Expenditure to Income on y. It can be observed that it has a negative linear correlation (indicated by the lower inclined trend) which means that as income increases, the Ratio of Food Expenditure to Income decreases. This tells us Higher income households spend a lower percentage of their income for food.

5. What are the main sources of income in the country? Per region?

```
counts = dataset.groupby('Main Source of Income').size().to_frame(), sort_values(ascending=False)
counting columns in a dataset
counts.plot(kind='bar', title='Main Sources of Income', ylabel='in Thousands')
counts.hvplot.barh(xlabel='Main Sources of Income', ylabel='in Thousands', height=700, fontsize={'xticks': 10, 'yticks': 10},)
    stacked=False, legend='top_right', legend_cols=2, legend_offset=(0, 25), xformatter='%.0f', yformatter='%.0f',
    = pd.pivot_table(dataset, values='Total Household Income', index=['Region', 'Main Source of Income'], aggfunc=np.mean).round(2)
    color=['#ff7f0e', '#2ca02c', '#d62728', '#9467bd', '#8c564b', '#e377c2', '#7f7f7f', '#bcbd22', '#17becf'], fontsize={'xticks': 10, 'yticks': 10},
    alpha=None, xlabel='Region', ylabel='Income', fontsize={'xticks': 10, 'yticks': 10}, rot=None, legend=True,
    = dataset.groupby('Region')['Main Source of Income'].value_counts(normalize=True).round(2).unstack()
    = pd.pivot_table(dataset, values='Total Household Income', index=['Region', 'Main Source of Income'], aggfunc=np.mean).round(2)
```