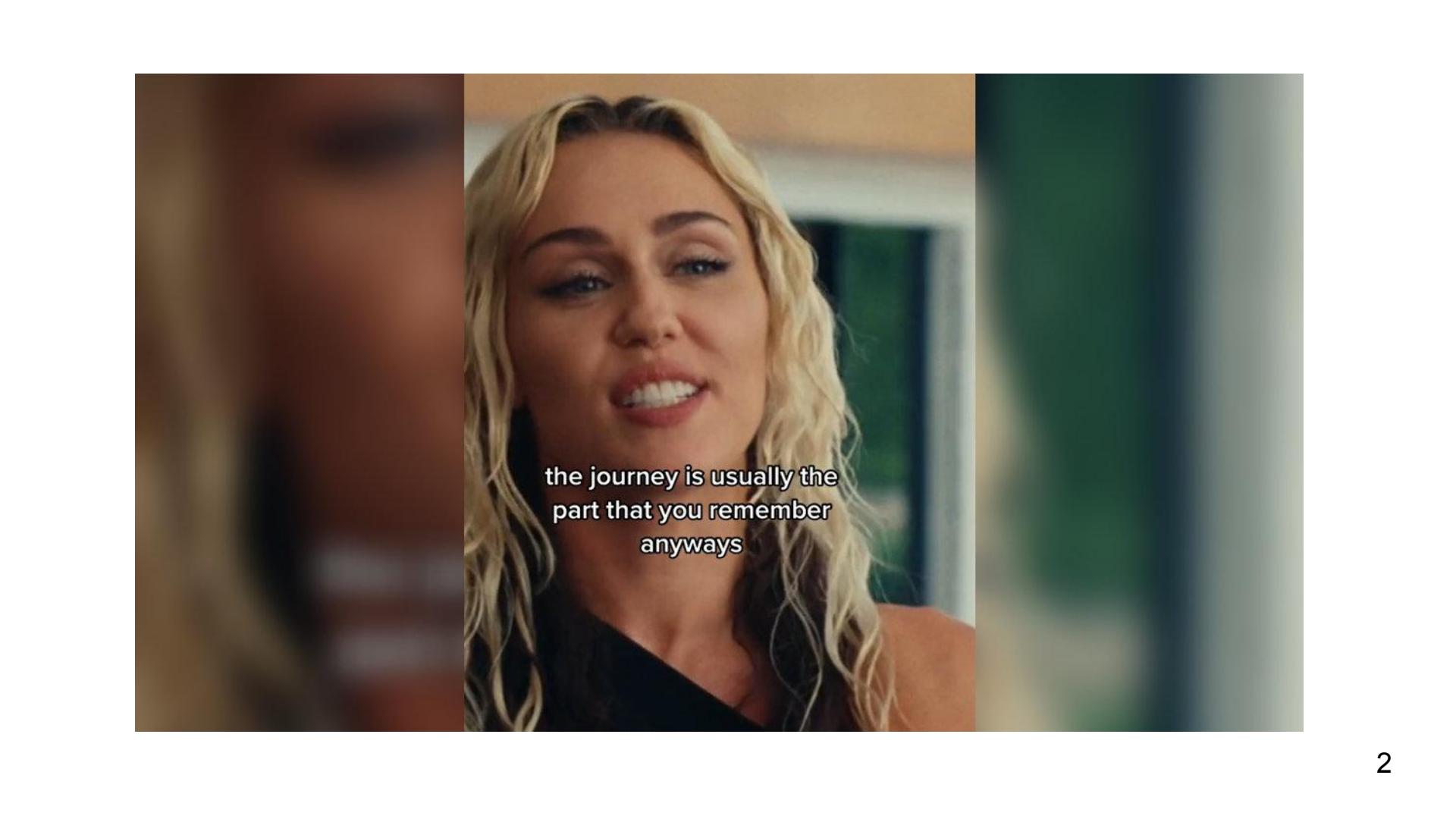


# Regional Income & Expenditure per Household

ITMGT Final Project - Data Track  
Justin Chan, Matthew Ngo, Hannah Perez

A close-up portrait of a woman with long, wavy blonde hair. She is smiling warmly at the camera. The background is slightly blurred, showing what appears to be an indoor setting with warm lighting.

the journey is usually the  
part that you remember  
anyways

# Introduction of Data Set

- Obtained through Kaggle, Justin contacted PSA thrice for more recent
- “Filipino Family Income and Expenditure (2017)“
  - Just below 42000 rows, with 60 columns worth of data
- Potential Points of Interest:
  - Expenditure statistics
  - Regional differences
  - Class differences
  - Comparison to 2020-2021 study of Filipino Family Expenditure
  - Possible explanations to information gathered

# Project Background & Relevance

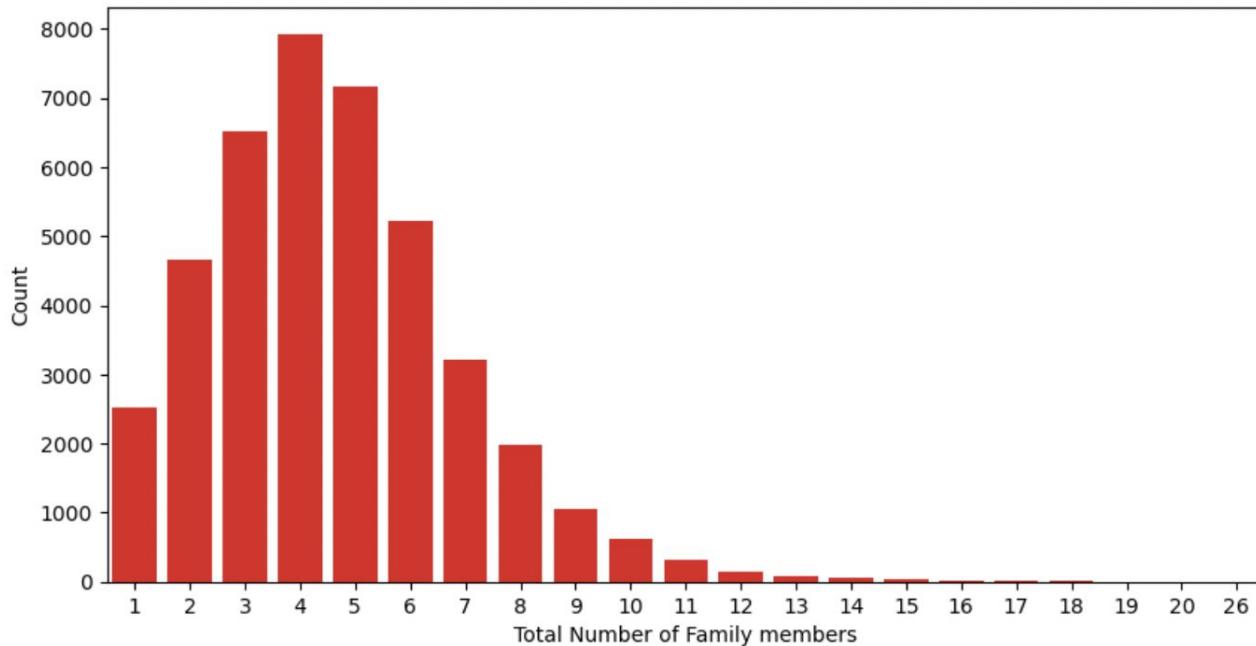
- Exploratory data analysis (EDA):
  - Family Income
  - Expenditure
- Businesses can benefit from the following:
  - Studying consumer behaviour
  - Preference
  - Purchasing power

# Imported Libraries & Accessing the dataframe

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import math
url = 'https://raw.githubusercontent.com/hannuhhprz/itmgt-intersesh/main/Family-Income-and-Expenditure.csv'
pd.set_option('display.max_columns', None)
df = pd.read_csv(url)
df
```

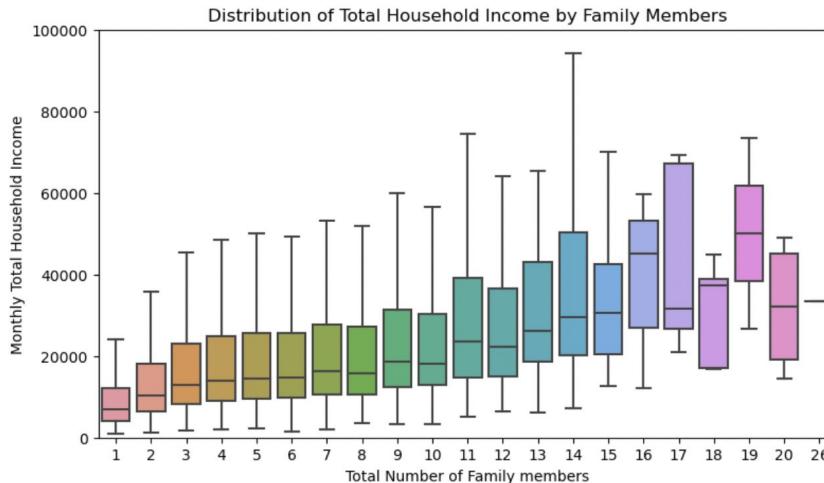
# Total Number of Family Members

```
: fig, ax = plt.subplots(figsize = (10, 5))
family_members = sns.countplot(x = 'Total Number of Family members', data = df, color = 'Red')
plt.ylabel('Count')
plt.show()
```



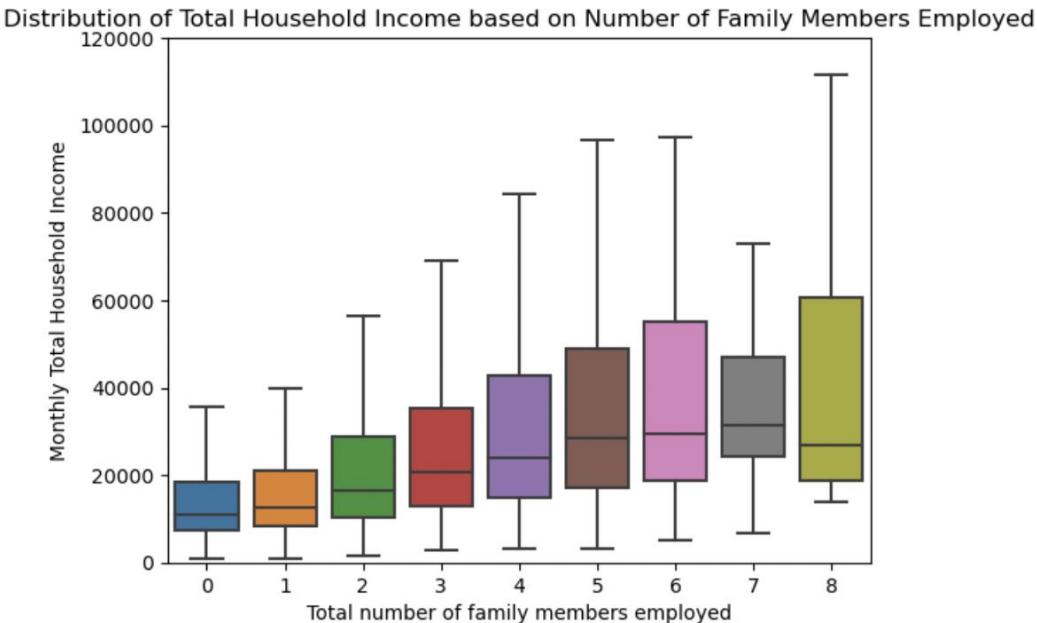
# Total Number of Family Members to Income

```
fig, ax = plt.subplots(figsize = (9,5))
sns.boxplot(y = 'Monthly Total Household Income', x = 'Total Number of Family members', data = df, fliersize = 0)
plt.xlabel('Total Number of Family members')
plt.ylabel('Monthly Total Household Income')
plt.title('Distribution of Total Household Income by Family Members')
plt.ylim(0, 100000)
plt.show()
https://seaborn.pydata.org/generated/seaborn.boxplot.html
#outliers were removed here using fliersize = 0
#we can see that the boxes and their respective means are increasing
#or directly proportional to the number of family members because
#more members are working, contributing to higher total household income
#however because there are only little families with more than 15 members,
#the data on the right side does not fully represent for all families with >15 members
```



# Total EMPLOYED Members to Income

```
sns.boxplot (y = 'Monthly Total Household Income', x = 'Total number of family members employed', data = df, fliersize  
plt.xlabel('Total number of family members employed')  
plt.ylabel('Monthly Total Household Income')  
plt.title('Distribution of Total Household Income based on Number of Family Members Employed')  
plt.ylim(0, 120000)  
plt.show()
```

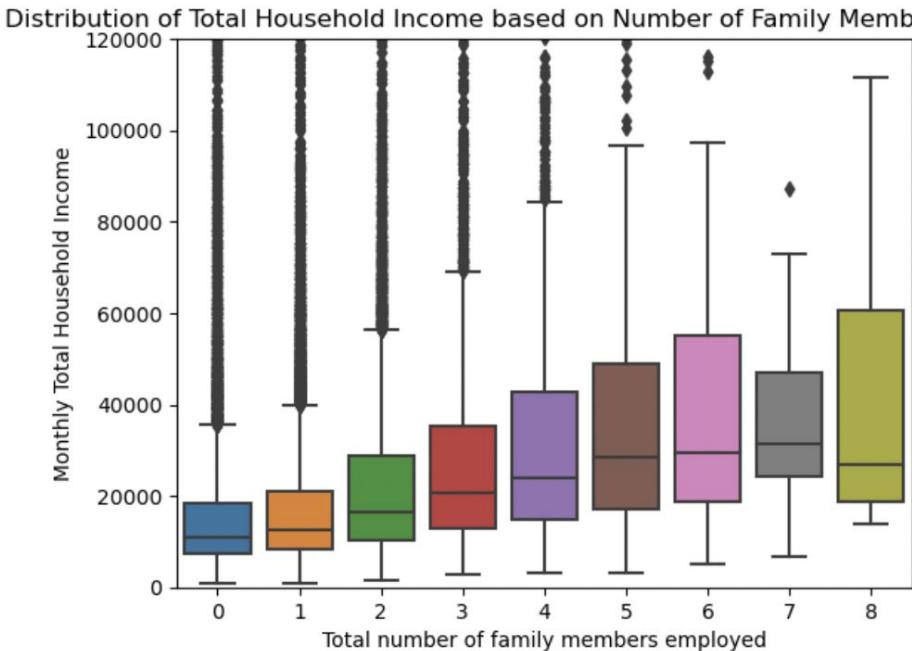


# Failed/ugly outputs: If we didn't remove outliers

```

sns.boxplot(y = 'Monthly Total Household Income', x = 'Total number of fam'
plt.xlabel('Total number of family members employed')
plt.ylabel('Monthly Total Household Income')
plt.title('Distribution of Total Household Income based on Number of Family'
plt.ylim(0, 120000)
plt.show()

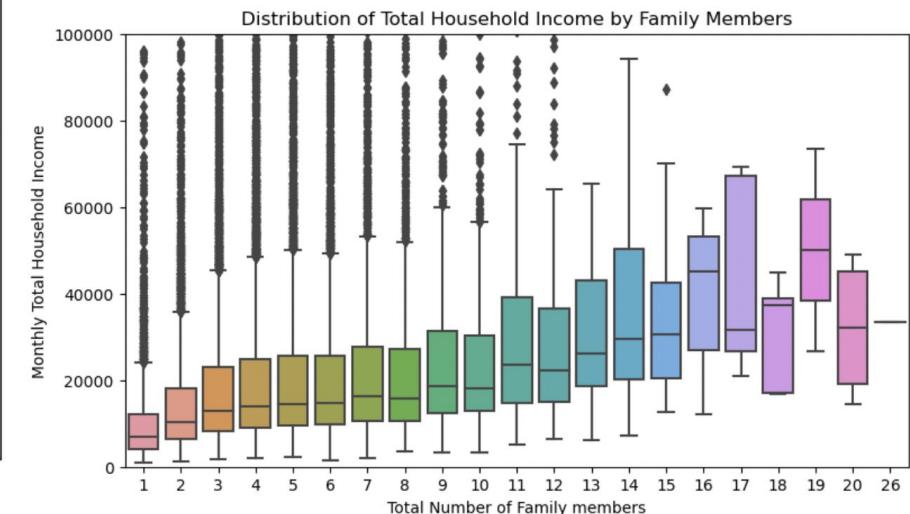
```



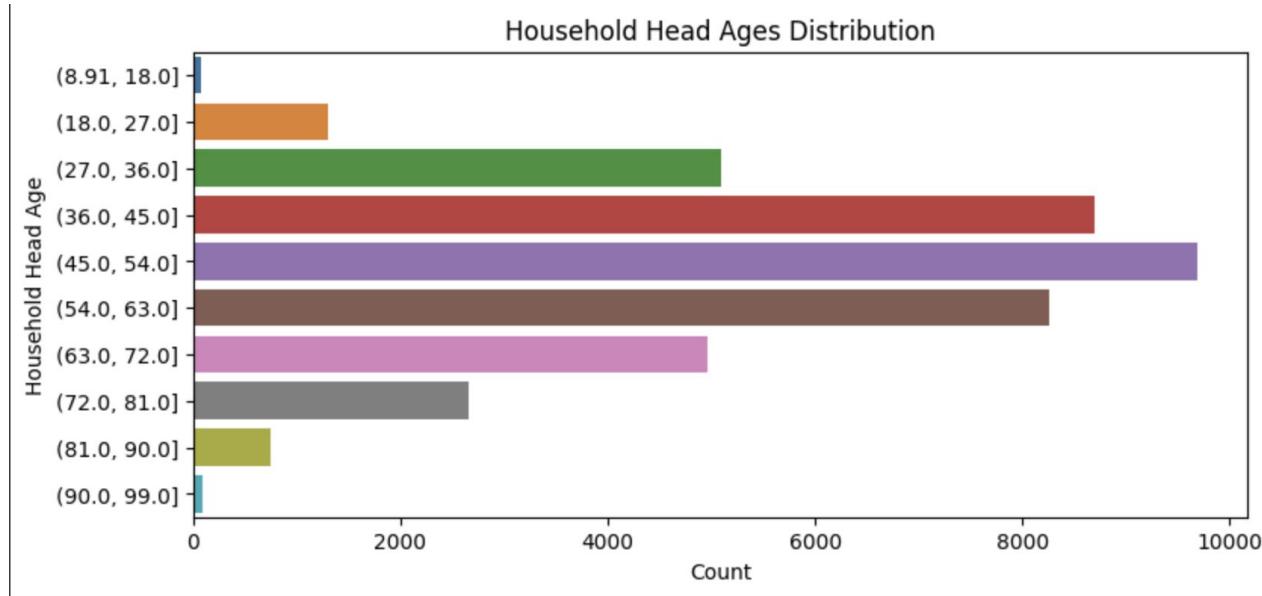
```

fig, ax = plt.subplots(figsize = (9,5))
sns.boxplot(y = 'Monthly Total Household Income', x = 'Total Number of Family members', data = df)
plt.xlabel('Total Number of Family members')
plt.ylabel('Monthly Total Household Income')
plt.title('Distribution of Total Household Income by Family Members')
plt.ylim(0, 100000)
plt.show()
https://seaborn.pydata.org/generated/seaborn.boxplot.html
#outliers were removed here using fliersize = 0
#we can see that the boxes and their respective means are increasing
#for directly proportional to the number of family members because
#more members are working, contributing to higher total household income
#however because there are only little families with more than 15 members,
#the data on the right side does not fully represent for all families with >15 members

```



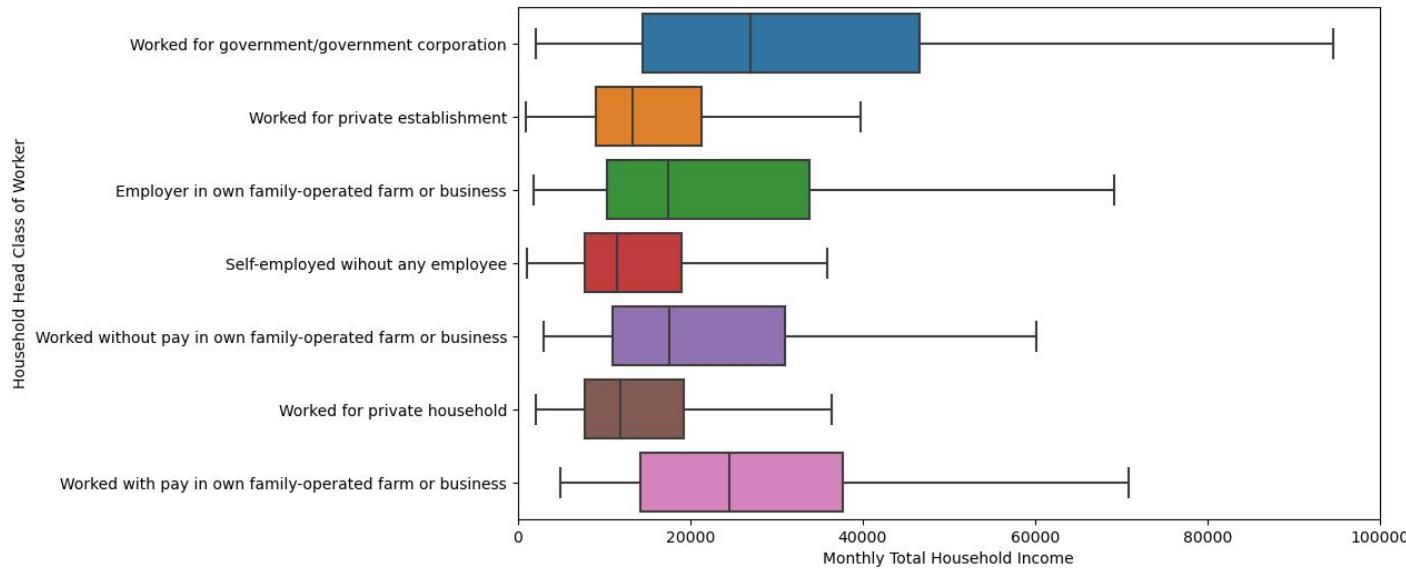
# Distribution of Household Head Ages (Mode)



```
household_head_ages = pd.cut(df['Household Head Age'], 10)
fig, ax = plt.subplots(figsize = (9, 4))
household_head_age_graph = sns.countplot(y=household_head_ages)
household_head_age_graph.set(xlabel = 'Count')
plt.title('Household Head Ages Distribution')
plt.show()
```

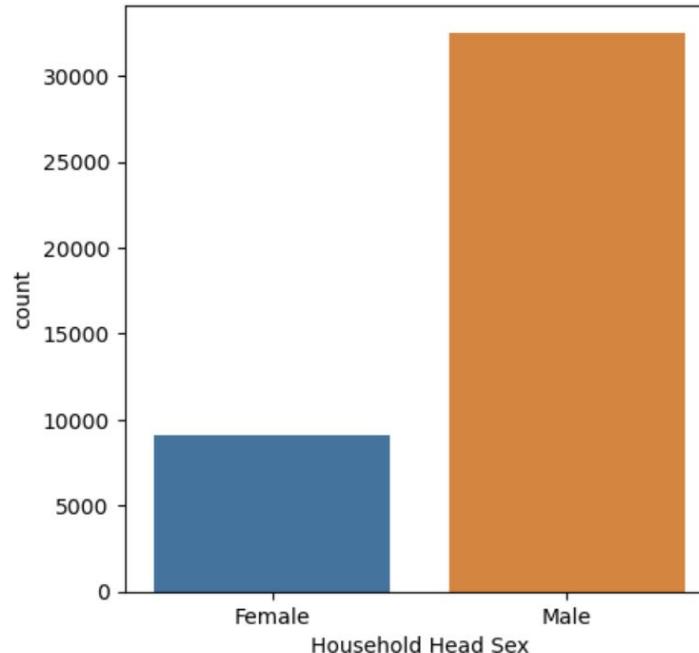
# Household Head Type of Work Plot

```
In [117]: f, ax = plt.subplots(figsize = (10,6))
sns.boxplot(x = df['Monthly Total Household Income'], y = df['Household Head Class of Worker'], fliersize = 0)
plt.ylabel('Household Head Class of Worker')
plt.xlabel('Monthly Total Household Income')
plt.xlim(0,100000)
plt.show()
#outliers were removed
#shows that households with household heads that work for gov/govcorps earn the most compared to other kinds of worker
```



# Sexes of Household Heads

```
fig,ax=plt.subplots(figsize = (5,5))
sns.countplot(x = 'Household Head Sex', data = df)
plt.show()
#ph is a patriarchal society
```



# Gender Occupations (Top 10)

Household Head Occupation	
Farmhands and laborers	3166
Rice farmers	2646
General managers/managing proprietors in transportation, storage and communications	1903
Corn farmers	1607
Building construction laborers	1009
Carpenters and joiners	999
Inland and coastal waters fishermen	996
General managers/managing proprietors in wholesale and retail trade	996
Coconut farmers	948
Car, taxi and van drivers	823
Name: Male, dtype: int64	

Household Head Occupation	
General managers/managing proprietors in wholesale and retail trade	1032
Farmhands and laborers	312
Hand launderers and pressers	264
Market and sidewalk stall vendors	235
Rice farmers	203
Domestic helpers and cleaners	184
General managers/managing proprietors in manufacturing	170
Shop salespersons and demonstrators	161
General elementary education teaching professionals	141
Street ambulant vendors	135
Name: Female, dtype: int64	

# Top 10 Most Common Jobs

Household Head Occupation	
Farmhands and laborers	3478
Rice farmers	2849
General managers/managing proprietors in wholesale and retail trade	2028
General managers/managing proprietors in transportation, storage and communications	1932
Corn farmers	1724
Coconut farmers	1036
Building construction laborers	1014
Inland and coastal waters fishermen	1011
Carpenters and joiners	1000
Car, taxi and van drivers	824
Name: Household Head Sex, dtype: int64	

# Average Annual Income/Region

Region	Total Household Income	Households	Ave. Household Income
NCR	1738159488	4130	420861.86
IVA - CALABARZON	1262586551	4162	303360.54
III - Central Luzon	948328293	3237	292965.18
CAR	464957336	1725	269540.48
XI - Davao Region	582431470	2446	238115.89
I - Ilocos Region	559082478	2348	238110.08
II - Cagayan Valley	525410874	2219	236778.22
VII - Central Visayas	596904567	2541	234909.31
VI - Western Visayas	628592073	2851	220481.26
IVB - MIMAROPA	270639721	1249	216685.12
X - Northern Mindanao	403927030	1887	214057.78
Caraga	350888945	1782	196907.38
VIII - Eastern Visayas	459773390	2337	196736.58
IX - Zamboanga Peninsula	341509624	1788	191000.91
V - Bicol Region	460052778	2472	186105.49
XII - SOCCSKSARGEN	388293751	2122	182984.8
ARMM	302910846	2248	134746.82

- Luzon based regions seem to have the highest income per household, followed by Visayas and Mindanao based regions.
- ARMM has around 50% the households of NCR, but only about a third of the average household income
  - These claims are based solely on dataset, which may or may not complete censuses

# Average Monthly Income/Region

\*similar trend\*

Region	Monthly Household Income	Households	Ave. Monthly Household Income
NCR	144846624.0	4130	35071.82
IVA - CALABARZON	105215545.92	4162	25280.04
III - Central Luzon	79027357.75	3237	24413.77
CAR	38746444.67	1725	22461.71
XI - Davao Region	48535955.83	2446	19842.99
I - Ilocos Region	46590206.5	2348	19842.51
II - Cagayan Valley	43784239.5	2219	19731.52
VII - Central Visayas	49742047.25	2541	19575.78
VI - Western Visayas	52382672.75	2851	18373.44
IVB - MIMAROPA	22553310.08	1249	18057.09
X - Northern Mindanao	33660585.83	1887	17838.15
Caraga	29240745.42	1782	16408.95
VIII - Eastern Visayas	38314449.17	2337	16394.72
IX - Zasmboanga Peninsula	28459135.33	1788	15916.74
V - Bicol Region	38337731.5	2472	15508.79
XII - SOCCSKSARGEN	32357812.58	2122	15248.73
ARMM	25242570.5	2248	11228.9

- Luzon based regions seem to have the highest income per household, followed by Visayas and Mindanao based regions.
- ARMM has around 50% the households of NCR, but only about a third of the average household income
  - These claims are based solely on dataset, which may or may not complete censuses

# Code Snapshots - Annual, Monthly, Food Expenditure

```
# AVERAGE ANNUAL INCOME PER REGION
```

```
annu_df = df.copy()
annu_df['Households'] = 'Yes'
annu_df['Ave. Household Income'] = annu_df['Total Household Income']
annu_df[['Region','Total Household Income','Households',
         'Ave. Household Income']].groupby(['Region']).agg({'Total Household Income':'sum','Households':'count',
                                                               'Ave. Household Income':'mean'}).round(2).sort_values(['Ave. Household Income'],ascending = False)
```

```
# AVERAGE MONTHLY INCOME PER REGION
```

```
mon_df = df.copy()
mon_df['Monthly Household Income'] = df['Total Household Income']/12
mon_df['Households'] = 'Yes'
mon_df['Ave. Monthly Household Income'] = mon_df['Monthly Household Income']
mon_df[['Region','Monthly Household Income','Households','Ave. Monthly Household Income']].groupby(['Region']).agg(
    {'Monthly Household Income':'sum','Households':'count','Ave. Monthly Household Income':'mean'}).round(2).sort_values(['Ave. Monthly Household Income'],ascending = False)
```

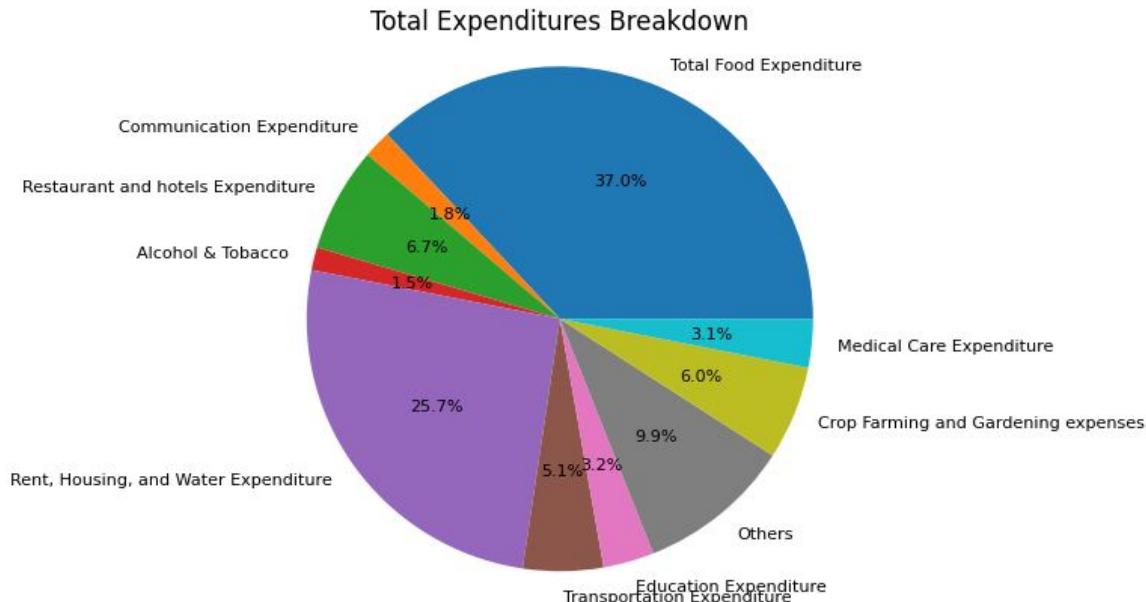
```
# AVERAGE ANNUAL FOOD EXPENDITURE
```

```
food_df = df.copy()
food_df['Ave. Annual Food Expenditure'] = food_df['Total Food Expenditure']
food_df['Households'] = 'Yes'
food_df[['Region','Total Food Expenditure','Households','Ave. Annual Food Expenditure']].groupby(['Region']).agg(
    {'Total Food Expenditure':'sum','Households':'count','Ave. Annual Food Expenditure':'mean'}).round(2).sort_values(['Ave. Annual Food Expenditure'],ascending = False)
```

# Total Expenditure Breakdown (All PH) - 2017 (our code)

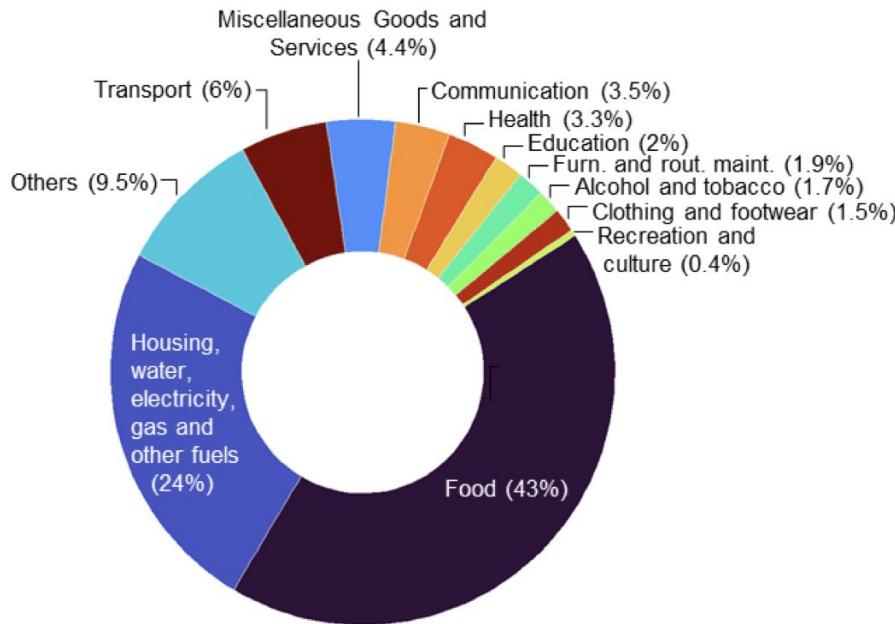
```
simplified_expenditures_df = df.copy()
simplified_expenditures_df['Alcohol & Tobacco'] = simplified_expenditures_df['Alcoholic Beverages Expenditure'] + simpl
simplified_expenditures_df['Rent, Housing, and Water Expenditure'] = simplified_expenditures_df['Housing and water Expe
simplified_expenditures_df['Others'] = simplified_expenditures_df['Special Occasions Expenditure'] + simplified_expendi
```

```
expenditure_category = ['Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expenditure', 'Al
plt.pie(simplified_expenditures_df[expenditure_category].sum(), labels=simplified_expenditures_df[expenditure_category]
plt.axis('equal')
plt.title('Total Expenditures Breakdown')
plt.show()
```

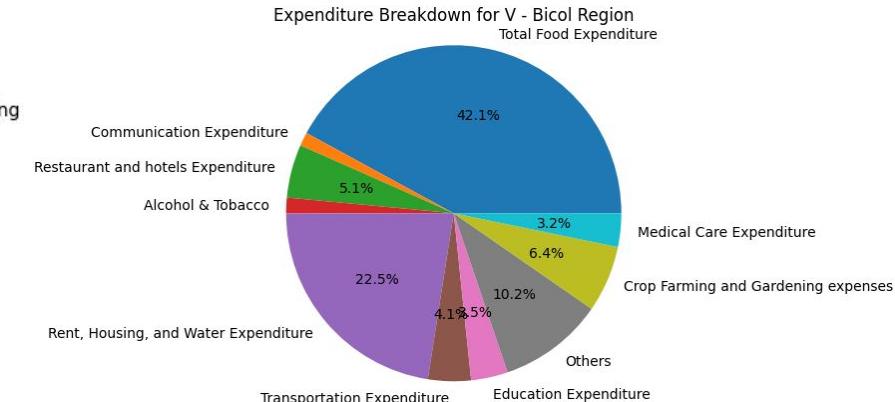
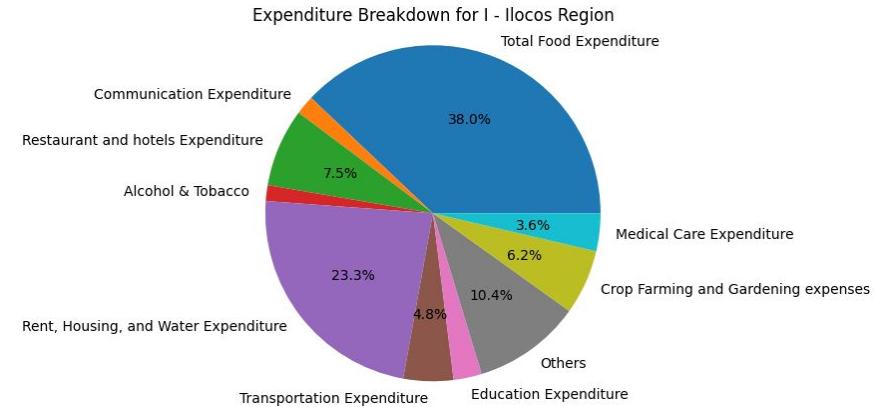
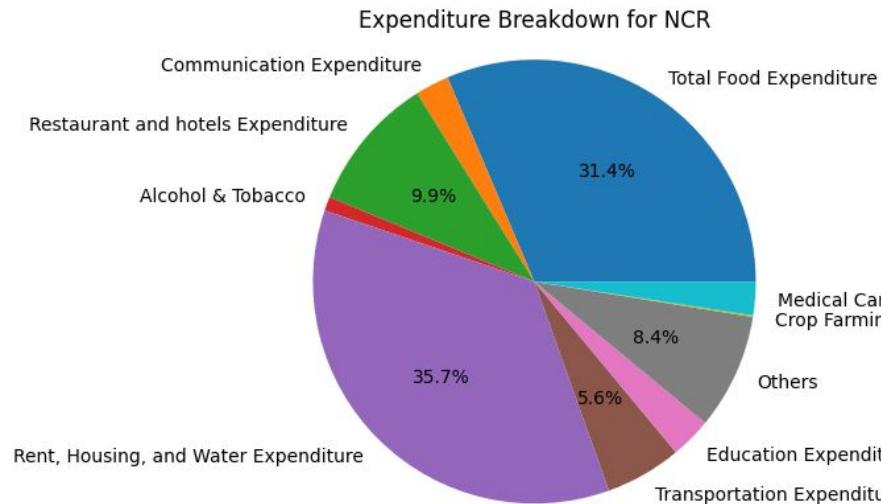


# Total Expenditure Breakdown (All PH) - 2021

**FIGURE 1**  
**AVERAGE SPENDING PATTERN OF FILIPINO HOUSEHOLDS**  
**(IN %), 2021**



# Total Expenditure Breakdown (NCR vs. Other Regions)



We only showed sample pie charts of the regions other than NCR since the regions show more or less the same data. If you want to see the other charts, we can show the colab

# Total Expenditure Breakdown (NCR vs. Other Regions)

```
#PER REGION BREAKDOWN OF EXPENDITURES
#IF THE PERCENTAGE IS LESS THAN 3%, I HID IT TO AVOID OVERLAPPING
#THIS TOOK ME LIKE 90MINS OMG

exp_categories = simplified_expenditures_df[['Region', 'Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expenditure', 'Alcohol & Tobacco Expenditure', 'Entertainment Expenditure', 'Transportation Expenditure', 'Food at home Expenditure', 'Food away from home Expenditure']]
exp_only_categories = simplified_expenditures_df[['Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expenditure', 'Alcohol & Tobacco Expenditure', 'Entertainment Expenditure', 'Transportation Expenditure', 'Food at home Expenditure', 'Food away from home Expenditure']]
region_exp_categories = exp_categories.sort_values('Region')
unique_regions = region_exp_categories['Region'].unique() #sort accdg to name to iterate in for loop

def autopct_format(pct): #this is to hide the percentages on the graph if it is less than 3% so that it won't overlap much with the other percentages
    if pct<3:
        return ""
    else:
        return '%1.1f%%' %pct

for region in unique_regions:
    region_data = region_exp_categories[region_exp_categories['Region'] == region]
    expenditure_sum = region_data.iloc[:, 1: ].sum() #calculate sum of each expenditure for selected columns

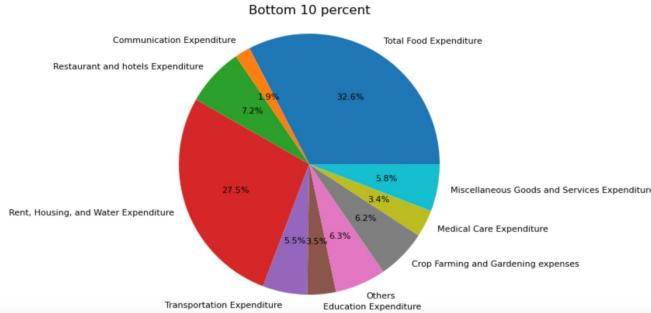
    plt.pie(expenditure_sum, labels=expenditure_sum.index, autopct=autopct_format)
    plt.axis('equal')
    plt.title(f'Expenditure Breakdown for {region}')
    plt.show()
```

# Comparison of Top and Bottom 10% Households (2017)

```
In [268]: bottom_10pct_families = class_df['Monthly Total Household Income'].quantile(0.1)
bottom_10_df = class_df[class_df['Monthly Total Household Income'] <= bottom_10pct_families]
bottom_10_df['Alcohol & Tobacco'] = bottom_10_df['Alcoholic Beverages Expenditure'] + bottom_10_df['Tobacco Expenditure']
bottom_10_df['Rent, Housing, and Water Expenditure'] = bottom_10_df['Housing and water Expenditure'] + bottom_10_df['Imputed Housing Costs']
bottom_10_df['Others'] = bottom_10_df['Special Occasions Expenditure'] + bottom_10_df['Clothing, Footwear and Other Wear Expenses']

expenditure_category = ['Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expenditure', 'Rent, Housing, and Water Expenditure', 'Transportation Expenditure', 'Education Expenditure', 'Crop Farming and Gardening expenses', 'Medical Care Expenditure', 'Miscellaneous Goods and Services Expenditure', 'Others']

plt.pie(bottom_10_df[expenditure_category].sum(), labels=bottom_10_df[expenditure_category].sum().index, autopct='%.1f'
plt.axis('equal')
plt.title('Bottom 10 percent')
plt.show()
```



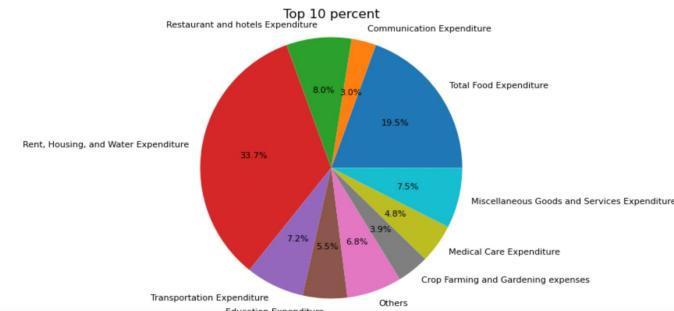
About less than  $\frac{1}{4}$  of total expenditures is allocated towards food, about  $\frac{1}{3}$  allocated towards housing, the rest of the expenditures allocated towards other needs →

← About  $\frac{1}{3}$  of total expenditures is allocated towards food,  $\frac{1}{4}$  allocated towards housing, the rest of the expenditures allocated towards other needs

```
In [284]: top_10pct_families = class_df['Monthly Total Household Income'].quantile(0.9)
top_10_df = class_df[class_df['Monthly Total Household Income'] >= top_10pct_families]
top_10_df['Alcohol & Tobacco'] = top_10_df['Alcoholic Beverages Expenditure'] + top_10_df['Tobacco Expenditure']
top_10_df['Rent, Housing, and Water Expenditure'] = top_10_df['Housing and water Expenditure'] + top_10_df['Imputed Housing Costs']
top_10_df['Others'] = top_10_df['Special Occasions Expenditure'] + top_10_df['Clothing, Footwear and Other Wear Expenses']

expenditure_category = ['Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expenditure', 'Rent, Housing, and Water Expenditure', 'Transportation Expenditure', 'Education Expenditure', 'Crop Farming and Gardening expenses', 'Medical Care Expenditure', 'Miscellaneous Goods and Services Expenditure', 'Others']

plt.pie(top_10_df[expenditure_category].sum(), labels=top_10_df[expenditure_category].sum().index, autopct='%.1f'
plt.axis('equal')
plt.title('Top 10 percent')
plt.show()
```



# Comparison of Top and Bottom 10% Households (2021)

## Bottom 10% (1st Decile)

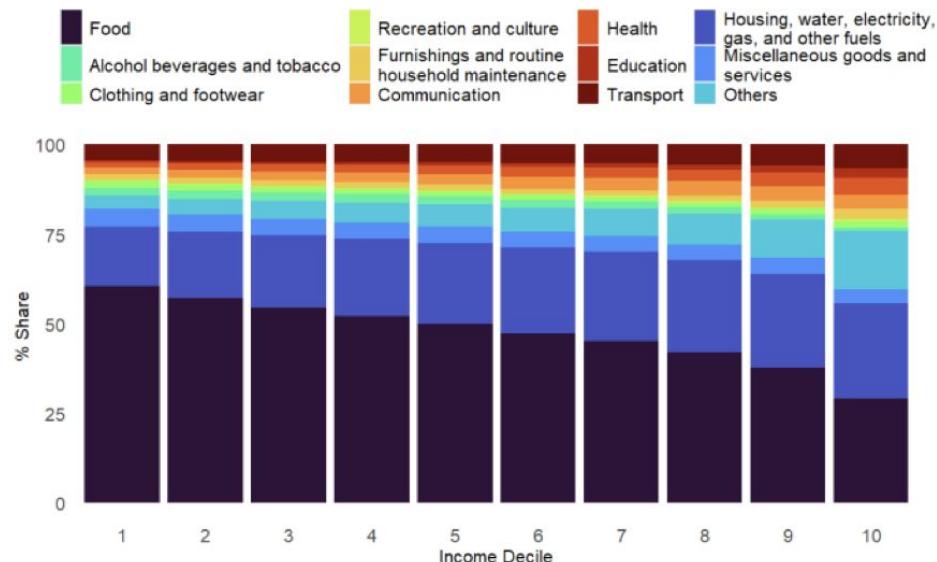
- Around 60% allocated towards food
- Housing and related expenditures have shrunk to less than 25%

## Top 10% (10th Decile)

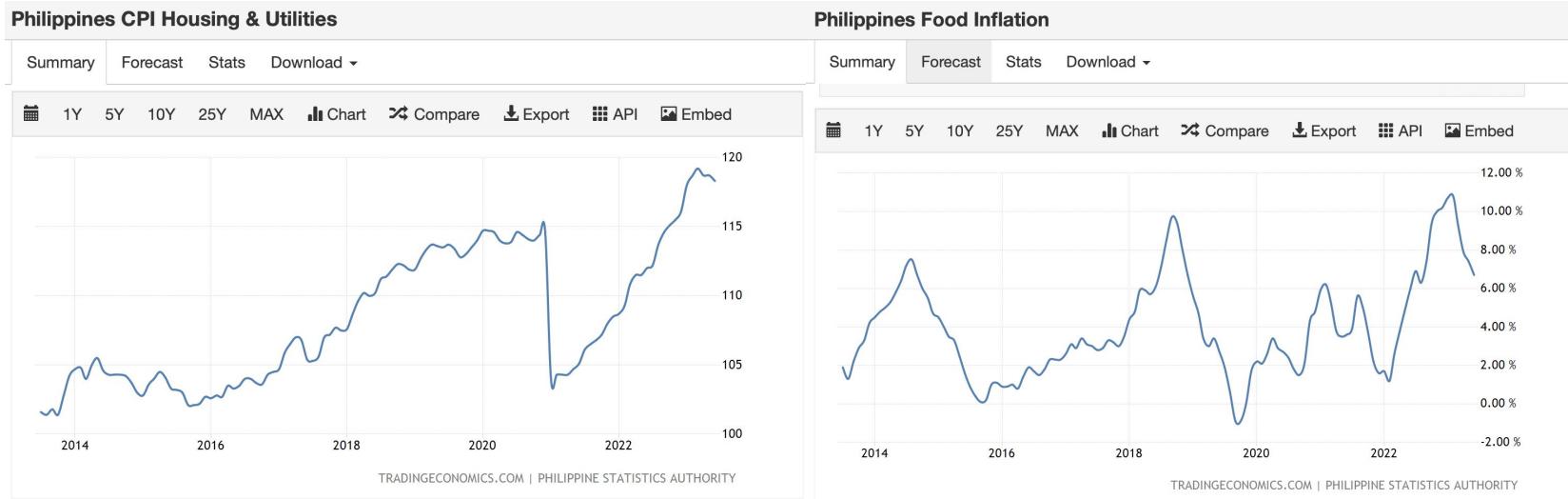
- Food and housing expenditures have increased to about 25% of total expenditure consumption each.

Chart taken from Congressional Policy and Budget Research Dept.

**FIGURE 2**  
**HOUSEHOLD FINAL CONSUMPTION EXPENDITURE BY PER CAPITA INCOME DECILE (IN %), 2021**



# Comparison of Top and Bottom 10% Households



CPI (Consumer Price Index) of Housing/Utilities was noticeably lower 2021 than 2017, which could explain why they comprised less of the total expenditures in 2021. Meanwhile, the CPI of food increased immensely in 2021, which explains why it comprised more of the total expenditures in 2021.

# Social Class Distribution according to PIDS

```
income_ranges = [0, 10957, 21914, 43828, 76666, 131484, 219140]
classes = ['Poor', 'Low-income but not poor', 'Lower middle', 'Middle', 'Upper middle', 'Upper middle but not rich', 'Rich']

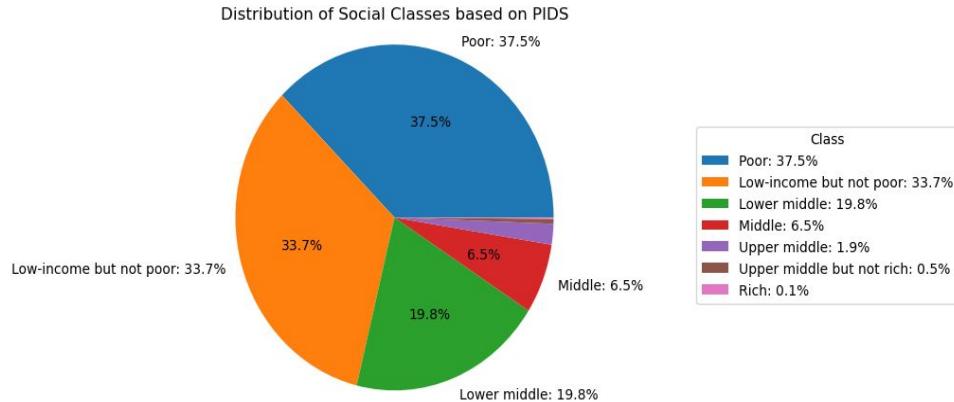
def assign_class(income):
    for i in range(len(income_ranges)-1):
        if income >= income_ranges[i] and income < income_ranges[i+1]:
            return classes[i]
    return classes[-1] # this return is for rich class

# using .apply() in the function
class_df['Class'] = class_df['Monthly Total Household Income'].apply(assign_class)

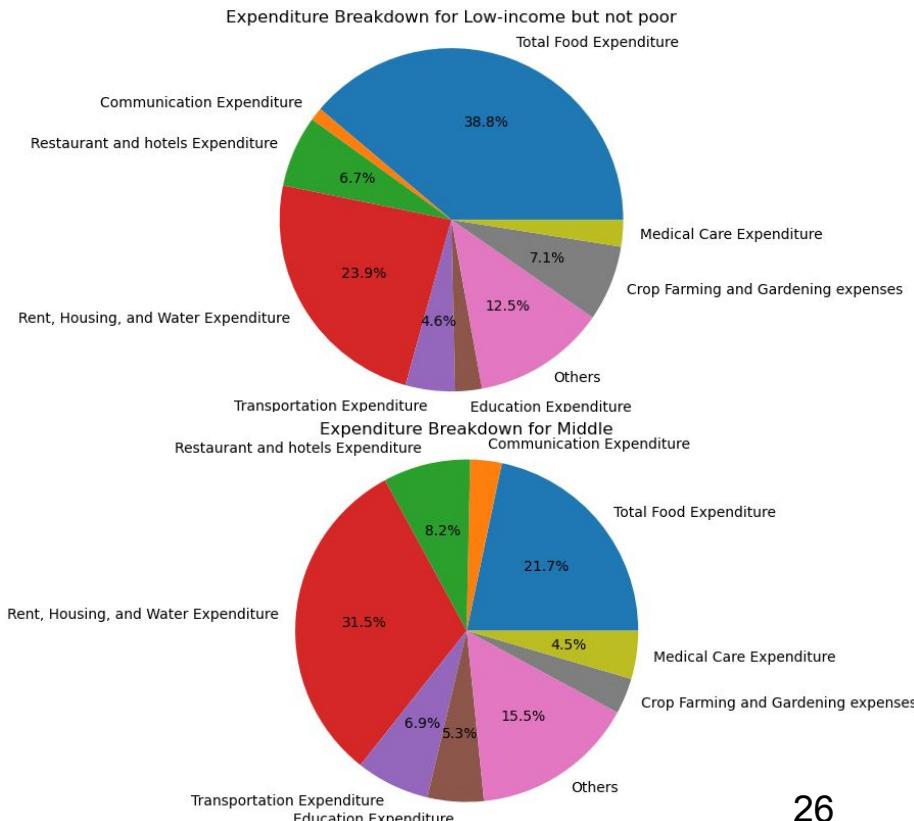
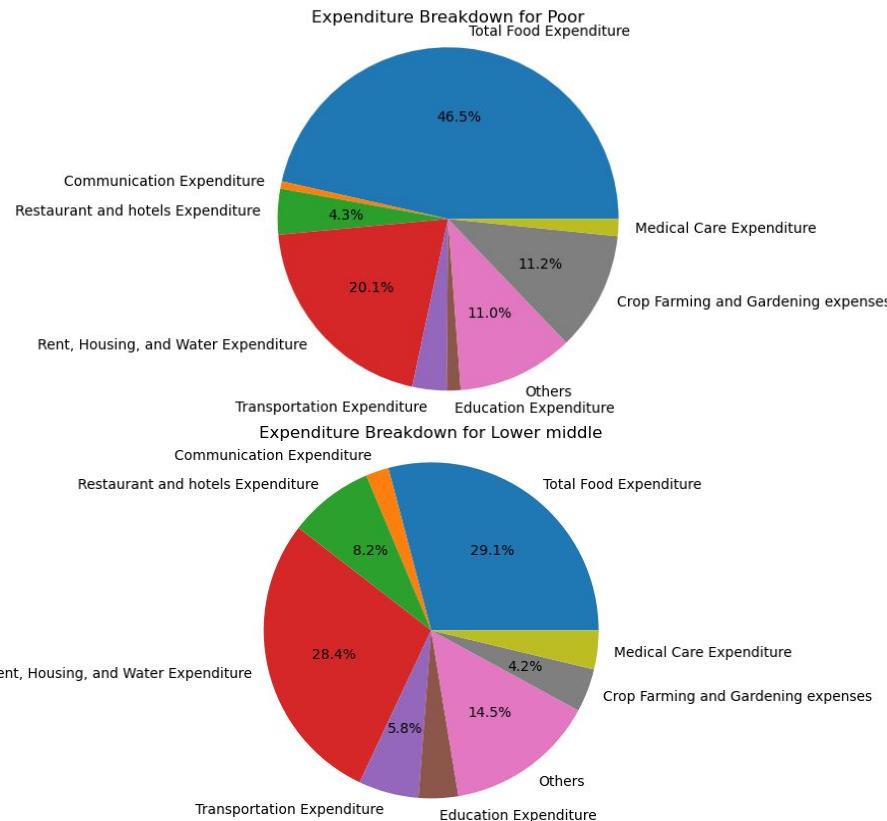
class_count = class_df['Class'].value_counts()
class_percentages = class_count / class_count.sum() * 100
legend_labels = [f'{label}: {class_percentages[label]:.1f}%' for label in class_count.index]
chart_labels = [f'{label}: {class_percentages[label]:.1f}%' if class_percentages[label] >= 5 else '' for label in class_count.index]

def autopct_formatt(pct): #this is to hide the percentages on the graph if it is less than 5% so that it won't overlap
    if pct<5:
        return ""
    else:
        return '%1.1f%%' %pct

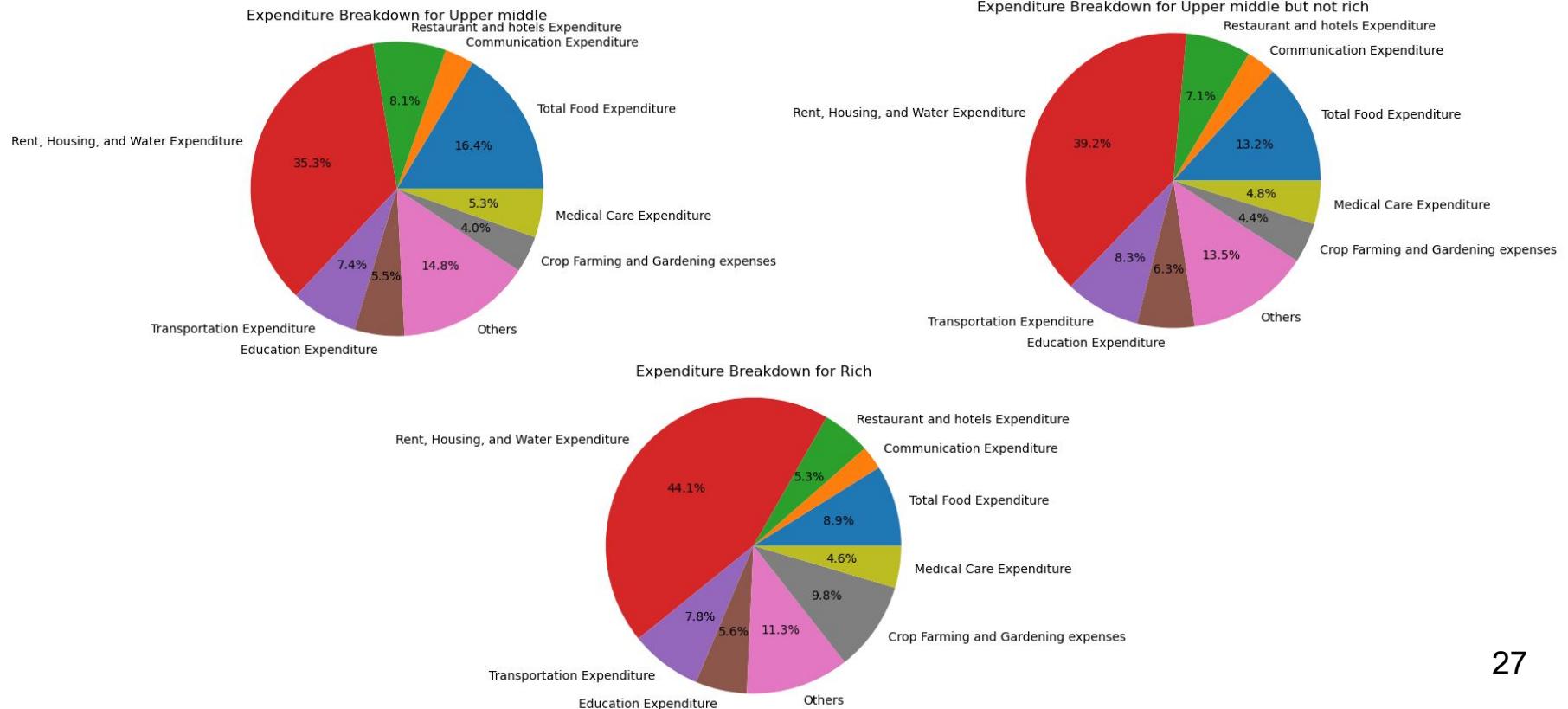
plt.pie(class_count, labels=chart_labels, autopct=autopct_formatt)
plt.axis('equal')
plt.title('Distribution of Social Classes based on PIDS')
plt.legend(legend_labels, title='Class', loc='center left', bbox_to_anchor=(1, 0.5), bbox_transform=plt.gcf().transFigure)
plt.show()
```



# Expenditure Breakdown per Class (Low to Middle)



# Expenditure Breakdown per Class (Upper Classes)



# Expenditure Breakdown per Class (Code)

```
In [282]: class_df['Others'] = class_df['Others'] + class_df['Alcohol & Tobacco']
exp_categories = class_df[['Class', 'Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expen
exp_only_categories = class_df[['Total Food Expenditure', 'Communication Expenditure', 'Restaurant and hotels Expenditu
class_exp_categories = exp_only_categories.sort_values('Class')
unique_class = class_exp_categories['Class'].unique() #sort accdg to name to iterate in for loop

def autopct_format(pct): #this is to hide the percentages on the graph if it is less than 4% so that it won't overlap m
    if pct<4:
        return ""
    else:
        return '%1.1f%%' %pct

for clas in unique_class:
    class_data = class_exp_categories[class_exp_categories['Class'] == clas]
    expenditure_sum = class_data.iloc[:,1:1].sum() #calculate sum of each expenditure for selected columns

    plt.pie(expenditure_sum, labels=expenditure_sum.index, autopct=autopct_format)
    plt.axis('equal')
    plt.title(f'Expenditure Breakdown for {clas}')
    plt.show()

# As we go to higher social classes, the rent, housing, and water expenditure increases proportionally
```

# Average Food Expenditure/Region

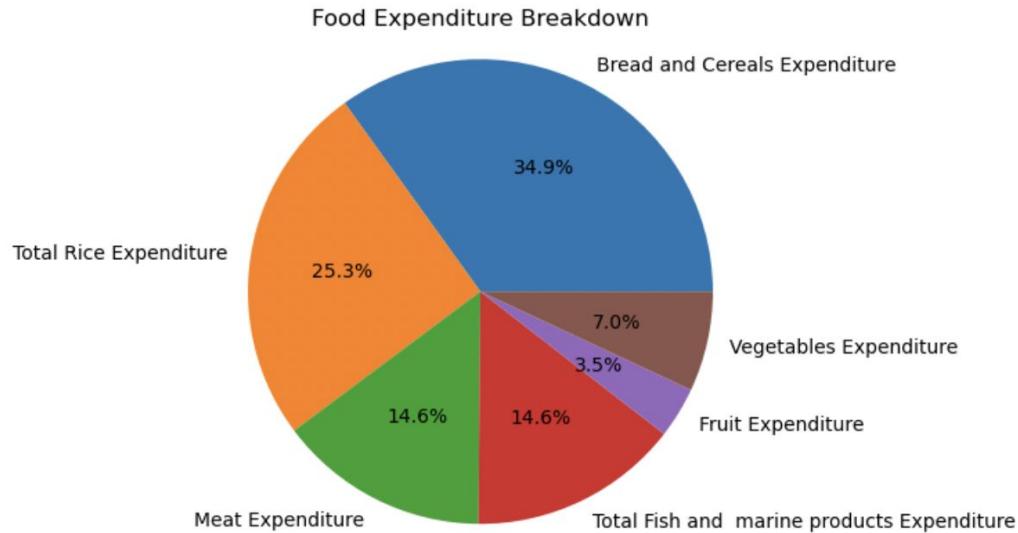
(similar trend to income)

Region	Total Food Expenditure	Households	Ave. Annual Food Expenditure
NCR	524842286	4130	127080.46
IVA - CALABARZON	438399898	4162	105333.95
III - Central Luzon	322815333	3237	99726.7
VII - Central Visayas	214224555	2541	84307.18
XI - Davao Region	198436464	2446	81126.93
I - Ilocos Region	189366054	2348	80649.94
CAR	138608546	1725	80352.78
VI - Western Visayas	227592553	2851	79829.03
V - Bicol Region	189877811	2472	76811.41
II - Cagayan Valley	167766071	2219	75604.36
Caraga	128148360	1782	71912.66
XII - SOCCSKSARGEN	152228224	2122	71738.09
IVB - MIMAROPA	88379607	1249	70760.29
VIII - Eastern Visayas	163201892	2337	69833.93
IX - Zamboanga Peninsula	124525829	1788	69645.32
ARMM	145965496	2248	64931.27
X - Northern Mindanao	120980449	1887	64112.59

- Luzon based regions seem to have the highest food expenditures per household, followed by the regions of Cebu and Davao, then other Visayas and Mindanao based regions.
- This could be attributed to higher incomes. More income = more incentive to spend.

# Food Expenditure Breakdown (All PH)

```
food_category = ['Bread and Cereals Expenditure', 'Total Rice Expenditure', 'Meat Expenditure', 'Total Fish and marine products Expenditure', 'Vegetables Expenditure', 'Fruit Expenditure']
plt.pie(df[food_category].sum(), labels=df[food_category].sum().index, autopct='%.1f%%')
plt.axis('equal')
plt.title('Food Expenditure Breakdown')
plt.show()
#Pie chart for food expenditure breakdown
```



# Food Expenditure Breakdown Code

```
new_df_for_food_expenditure = df.copy() #make new df cuz this might be messy lol
food_categories = new_df_for_food_expenditure[['Region', 'Bread and Cereals Expenditure', 'Total Rice Expenditure', 'Meat Expenditure', 'Other Food Expenditure']]
food_only_categories = new_df_for_food_expenditure[['Bread and Cereals Expenditure', 'Total Rice Expenditure', 'Meat Expenditure', 'Other Food Expenditure']]
region_food_categories = food_categories.sort_values('Region')
unique_regions = region_food_categories['Region'].unique() #sort accdg to name to iterate in for loop

for region in unique_regions:
    region_data = region_food_categories[region_food_categories['Region'] == region]
    expenditure_sum = region_data.iloc[:, 1:].sum() #calculate sum of expenditure for selected columns

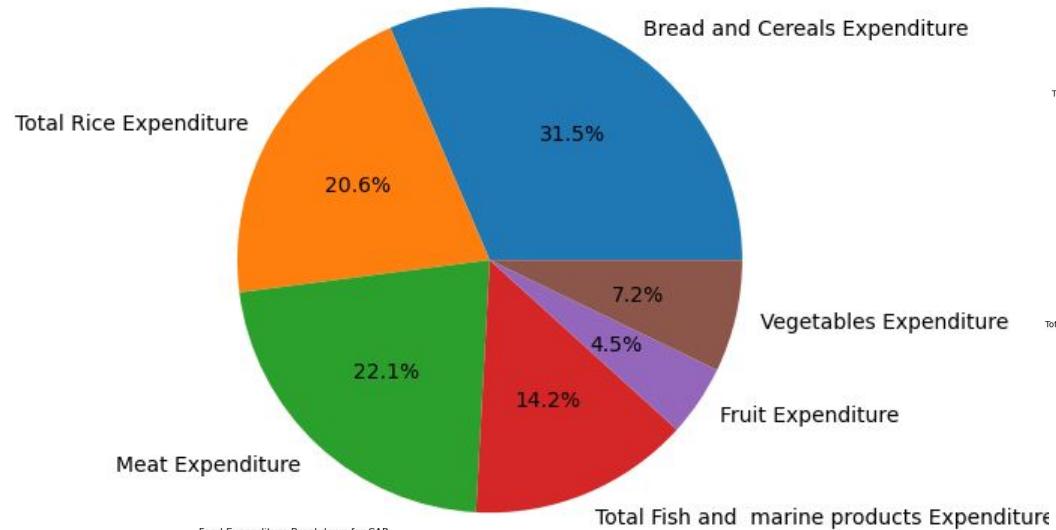
    plt.pie(expenditure_sum, labels=expenditure_sum.index, autopct='%.1f%%')
    plt.axis('equal')
    plt.title(f'Food Expenditure Breakdown for {region}')
    plt.show()
```

# Food Expenditure Breakdown (NCR vs other regions)

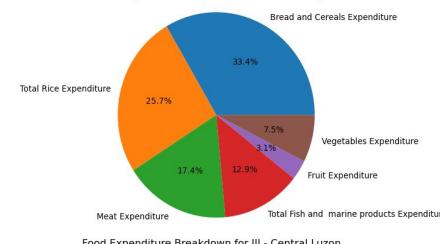
32

There are 17 regions, this slide shows only Luzon

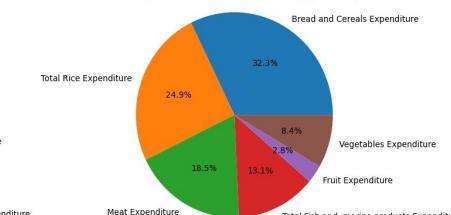
Food Expenditure Breakdown for NCR



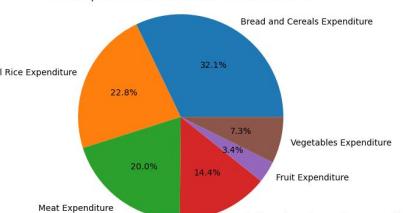
Food Expenditure Breakdown for I - Ilocos Region



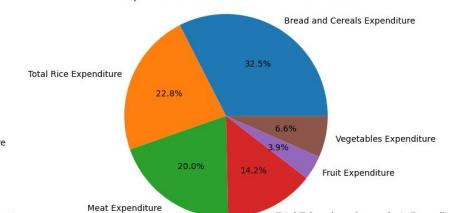
Food Expenditure Breakdown for II - Cagayan Valley



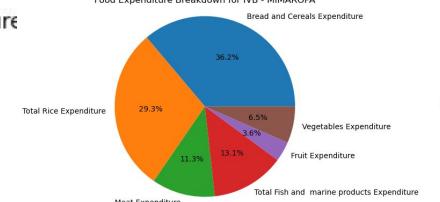
Food Expenditure Breakdown for III - Central Luzon



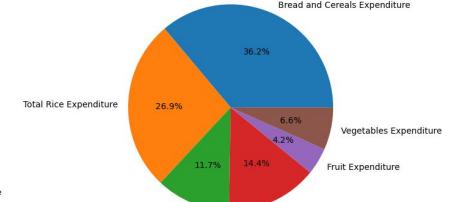
Food Expenditure Breakdown for IV-A - CALABARZON



Food Expenditure Breakdown for IV-B - MIMAROPA

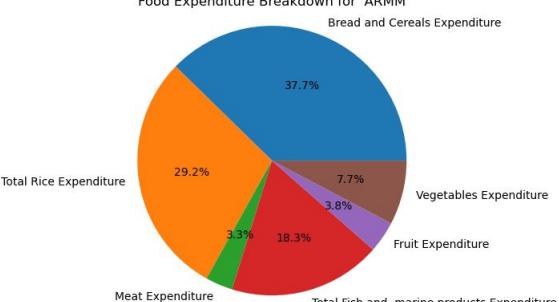
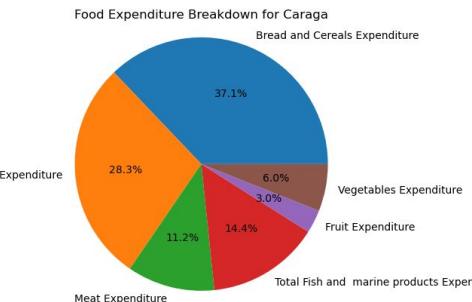
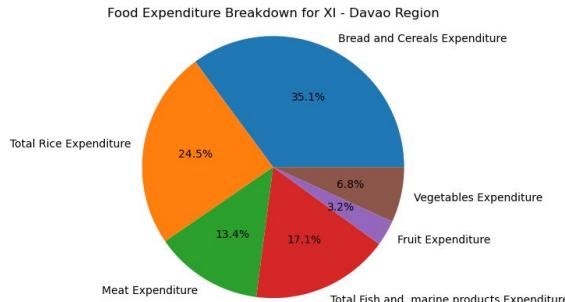
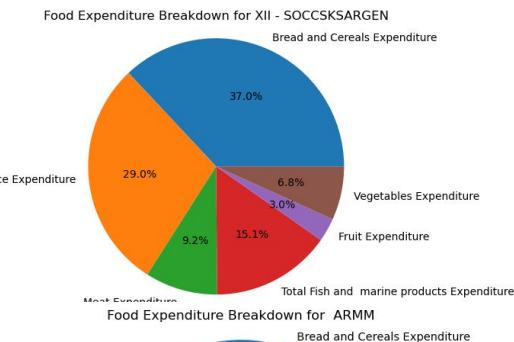
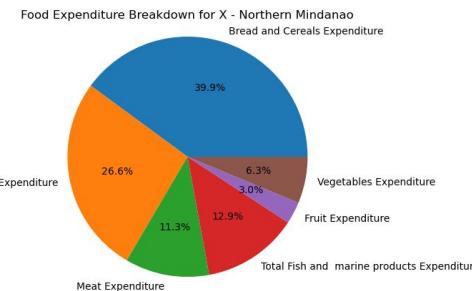
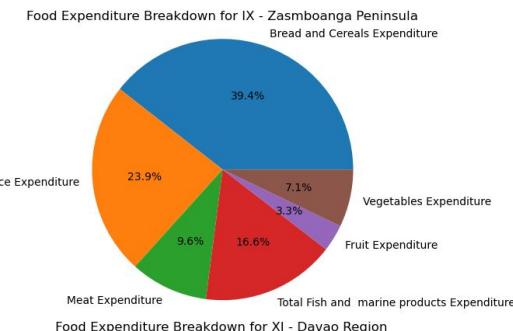
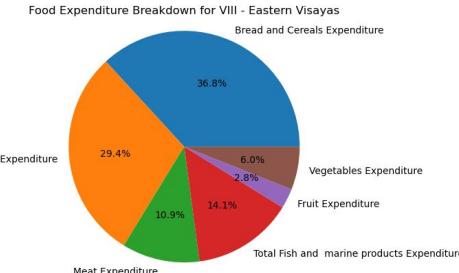
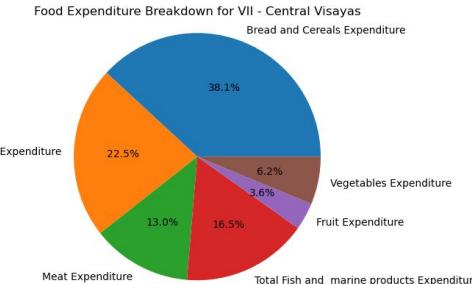
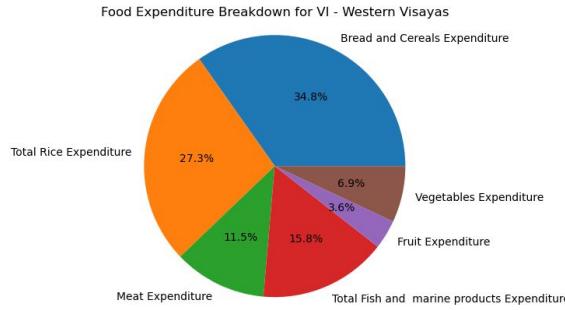


Food Expenditure Breakdown for V - Bicol Region



# Food Expenditure Breakdown (Visayas, Mindanao)

33



```
new_df_for_food_expenditure = df.copy() #make new df cuz this might be messy lol
food_categories = new_df_for_food_expenditure[['Region', 'Bread and Cereals Expenditure', 'Total Rice Expenditure', 'Meat Expenditure', 'Total Fish and marine food_only_categories = new_df_for_food_expenditure[['Bread and Cereals Expenditure', 'Total Rice Expenditure', 'Meat Expenditure', 'Total Fish and marine products']]
region_food_categories = food_categories.sort_values('Region')
unique_regions = region_food_categories['Region'].unique() #sort accdg to name to iterate in for loop

for region in unique_regions:
    region_data = region_food_categories[region_food_categories['Region'] == region]
    expenditure_sum = region_data.iloc[:, 1: ].sum() #calculate sum of expenditure for selected columns

    plt.pie(expenditure_sum, labels=expenditure_sum.index, autopct='%.1f%%')
    plt.axis('equal')
    plt.title(f'Expenditure Breakdown for {region}')
    plt.show()
```

# Conclusion (Brief Summary)

- Filipino households are generally patriarchal in nature (breadwinners)
- Economic prosperity is generally focused towards Northern Philippines (e.g. regions based in Luzon)
- Occupations tend to be more “rural” towards the South
  
- As we climb the social ladder, Filipinos spend proportionately less on Food, and spend proportionately more on Rent & Housing
- #1 Food expenditure of Filipinos is on Breads & Cereals, followed by Rice as #2, except in NCR where meat is #2

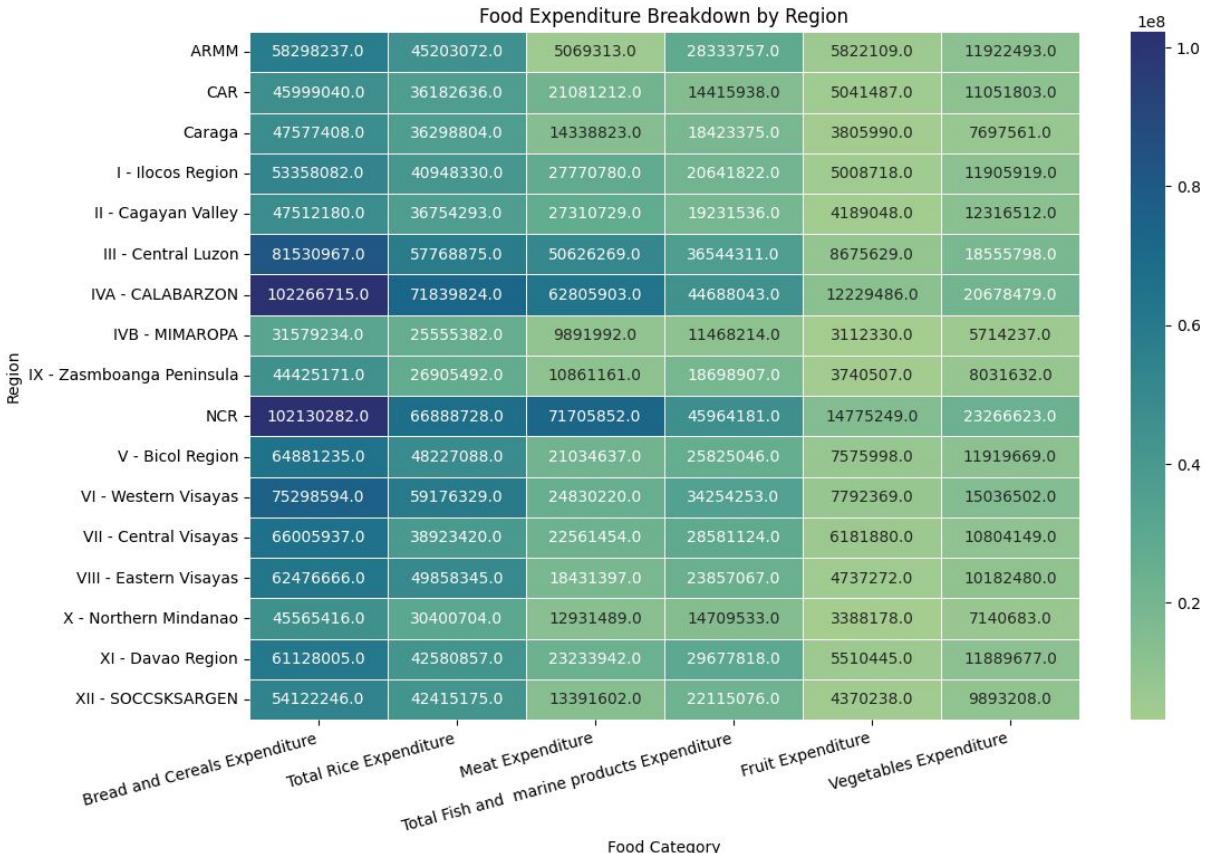
# Conclusion (Usefulness of the EDA)

- Relevant government agencies
  - NEDA: evaluating current and future economic policies/programs, better budget allocations
  - DSWD: “ayuda”, alleviating poverty in the country
- Business & Investors
  - Gain insights in spending habits of target markets (social class or region)
  - Learn demand patterns & regional variations which aid them in market development & research, pricing strategies, and even location targeting
- Analyzing future trends

# Recommendations and Limitations

- Recommendations
  - More recent raw dataset
  - Compare NCR with other capital regions of ASEAN
  - Profiling regions
- Limitations and Challenges
  - Data accessibility
  - Working remotely on the codes (colab sucks lol)
  - Data visualization
  - Predicting household income with sklearn
  - Time constraints

# Post-presentation addition: Heat map of food expenditures



# LINK DUMP

<https://www.kaggle.com/datasets/grosvenpaul/family-income-and-expenditure>

<https://www.pids.gov.ph/details/who-are-identified-rich-poor-gov-t-shows-income-class-brackets-in-ph>

[https://matplotlib.org/stable/plot\\_types/stats/boxplot\\_plot.html#sphx-glr-plot-types-stats-boxplot-plot-py](https://matplotlib.org/stable/plot_types/stats/boxplot_plot.html#sphx-glr-plot-types-stats-boxplot-plot-py)

<https://newsinfo.inquirer.net/1389371/fish-cheaper-best-protein-source-amid-rising-prices-of-meat-poultry-group>

<https://newsinfo.inquirer.net/1652149/filipinos-spend-a-third-of-their-budget-for-food-says-study>

<https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidspn1818.pdf>

<https://www.aseantoday.com/2017/03/education-inequality-poverty-a-paradox-in-the-philippines/>

[https://cpbrd.congress.gov.ph/images/PDF%20Attachments/Facts%20in%20Figures/FF2022-71\\_Consumption\\_Patterns\\_Among\\_Fil\\_Households\\_2021.pdf](https://cpbrd.congress.gov.ph/images/PDF%20Attachments/Facts%20in%20Figures/FF2022-71_Consumption_Patterns_Among_Fil_Households_2021.pdf)

<https://tradingeconomics.com/philippines/cpi-housing-utilities>

<https://tradingeconomics.com/philippines/food-inflation#:~:text=Food%20Inflation%20in%20Philippines%20averaged,percent%20in%20September%20of%202019.>

<https://www.da.gov.ph/facing-the-big-challenges-in-philippine-agriculture/>

<https://psa.gov.ph/content/highlights-preliminary-results-2021-family-income-and-expenditure-survey-fies-visit-1>

<https://www.investopedia.com/terms/s/self-employed.asp>

<https://www.philstar.com/lifestyle/food-and-leisure/2014/07/31/1352071/let-them-eat-healthy-bread>

<https://www.pids.gov.ph/details/who-are-identified-rich-poor-gov-t-shows-income-class-brackets-in-ph>

SNS, Pandas, Matplotlib Documentations