Cafe, Chua, Gamboa, Guian
December 4, 2017
CS 129.1 A
Report Paper

Big Data Problem:
Given the family income and expenditure data of the Philippines, what are the spending behaviors of different socioeconomic classes and what are the implications of these spending behavior? Results will be shown in different charts.

Here is the list of expected output from the Filipino Family and Income and Expenditures:

- Sorting and counting families per socioeconomic class
- Average food expenditure per socioeconomic class
- Average meal cost per socioeconomic class
- Average miscellaneous expenditure per socioeconomic class
- Correlation of food types and medical expenses
- Correlation of income and food expenditures per socioeconomic class
- Correlation of income versus miscellaneous expenditures
- Total excess/deficit per socioeconomic class

Description of the Data Source:

Data source is from a website called Kaggle, a website specializing in datasets and data analysis. By simply searching for Filipino Family Income and Expenditure, one can download the dataset in the links provided. The dataset is available in csv format. The procedures that the team did to obtain the dataset are as follows:

1. Go to Kaggle
2. Search for Filipino family income and expenditures
3. Download the dataset in csv format

Context

According to the Kaggle dataset provider, Francis Paul Flores, that the Filipino Family and Income Expenditure data was headed by the Philippine Statistics Authority (PSA). The survey is undertaken every three (3) years that aims to provide data not only on family income and expenditure, but also levels of consumption by item of expenditure, sources of income and other related information affecting the income and expenditure of the Filipino households.

Content

The Philippine Family Income and Expenditure dataset contains 41,544 entries which includes data regarding a household's:

1. Total Monthly Income
2. Location/Region

3. Total Food Expenditure
4. Total Income and its Source
5. Food Expenditure Breakdown (Bread, Rice, Meat, Fish, Vegetables, etc.)
6. Other Expenditures (Alcohol, Medical, Education, Communication, Clothing, etc.)
7. Household Information (Family Head, Age, Number, Occupation, Facilities, etc.)

Here is a preview of dataset:

Total Hous	Region	Total Food	Main Source of	Agricultura	Bread and	Total Rice E	Meat Expe	Total Fish a	Fruit Exper	Vegetables	Restaurant	Alcoholic B	Tobacco Ex	Clothing, F
480332	CAR	117848	Wage/Salaries	0	42140	38300	24676	16806	3325	13460	3000	0	0	4607
198235	CAR	67766	Wage/Salaries	0	17329	13008	17434	11073	2035	7833	2360	960	2132	8230
82785	CAR	61609	Wage/Salaries	1	34182	32001	7783	2590	1730	3795	4545	270	4525	2735

Note that this dataset does not categorize each family entry to its own socioeconomic class. Because of this, the group used data sourced from Rappler where a table was shown of the various socioeconomic status and their characteristics such as the range of monthly income of each income class. Rappler calculated this data from a similar dataset the group used; the 2012 Family Income and Expenditure Survey (FIES), Philippine Statistics Authority. They used the current standard of the poverty line to indicate the range of monthly income for each socioeconomic class.

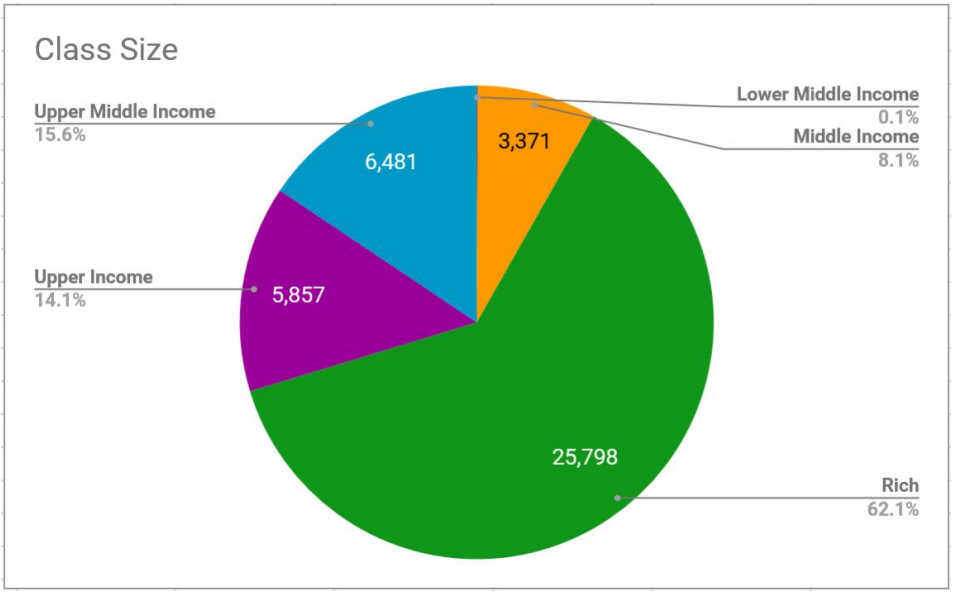
Thus, the group used the table shown below to categorize each household to its corresponding socioeconomic class.

Socioeconomic Class	**Range of Monthly Income (Family Size of 5)**
Poor	Less than PHP7,890 per month
Low	PHP 7,890 to 15,780
Low Middle	PHP 15,780 to 31,560
Middle	PHP 31,560 to 78,900
Upper Middle	PHP 78,900 to 118,350
Upper	PHP 118,350 to 157,800
Rich	At least PHP 157,800

Description, Explanation, and Discussion of the Output

Output 1: Sorting and counting families per socioeconomic class

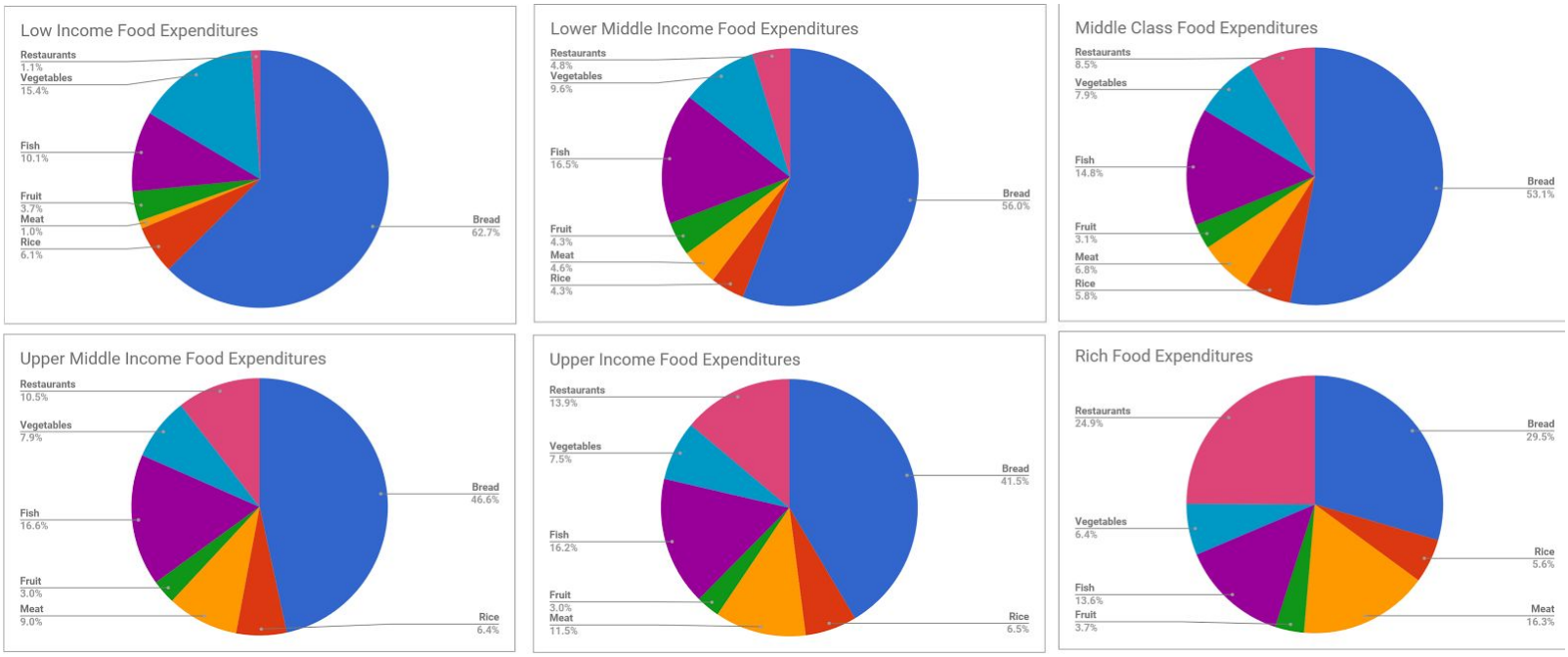
In this that set, most of the household entries fall under the category “rich” which is at 62% of the total entries while the combination of the Poor, Low, and Low Middle is only 10%. This indicates that the dataset is not entirely accurate because the poor is said to be more than the rich.



Output 1

Output 2: Average food expenditure per socioeconomic class

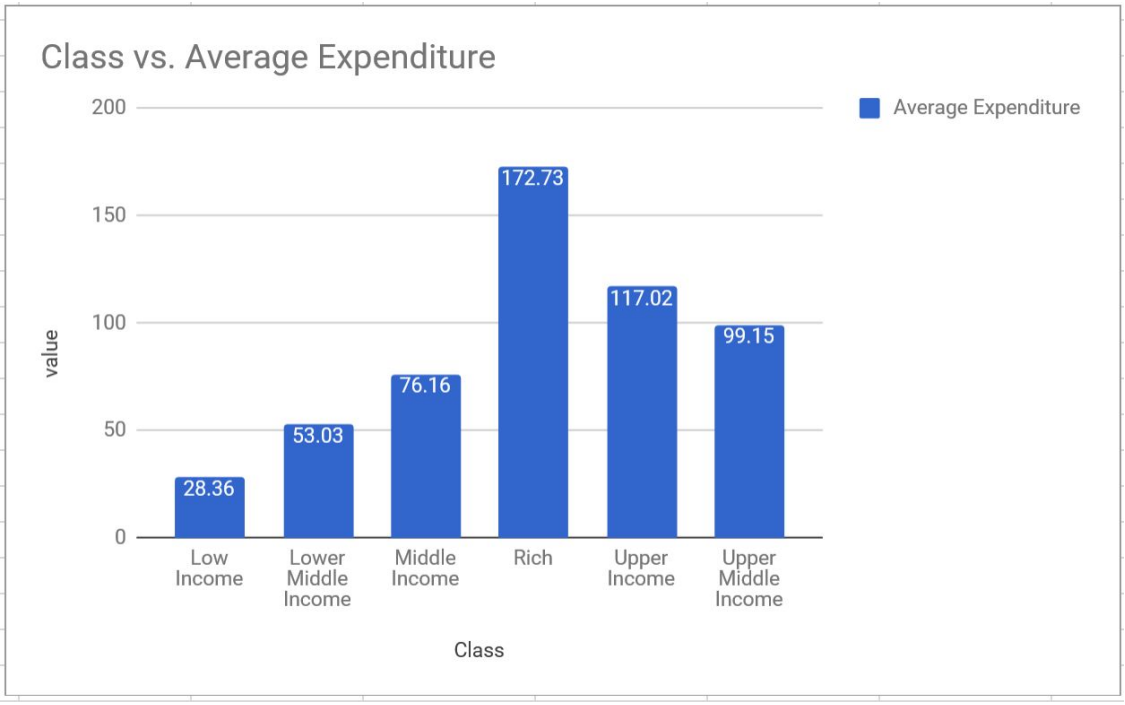
Despite the love of Filipino for rice, Bread is the number 1 food expenditure that Filipino’s spend on. It reaches up to 62.7% of an average “poor” family. However, the for bread decreases as income increases. Meat, on the other hand, has the opposite pattern for bread where the budget increases as income increases. Aside from the differences in price, some food types are harder to store because they need to be refrigerated and is not readily accessible in small stores.



Output 2

Output 3: Average meal cost per socioeconomic class

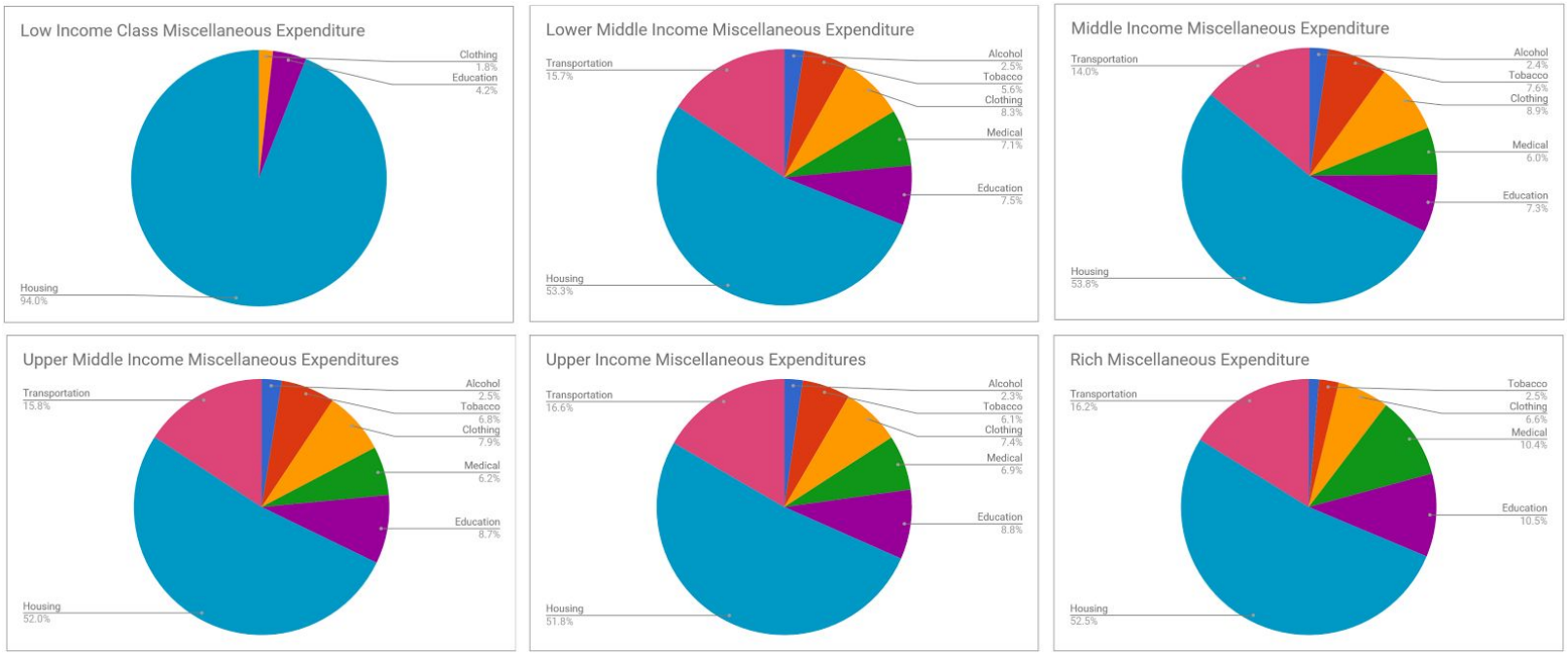
Straight-forward result where more income means more budget for meals. See Chart below to see actual values/costs.



Output 3

Output 4: Average miscellaneous expenditure per socioeconomic class

As for Miscellaneous Expenditures, most Filipinos regardless of socioeconomic class spend most of their budget to Housing, followed by Transportation. The Low Income Class as compared to other socioeconomic class could not even afford to buy luxury products such as alcohol or tobacco, nor do they have the income to afford medical services or products. Percentages on different miscellaneous expenditures are similar regardless of socioeconomic class.



Output 4

Output 5: Correlation of food types and medical expenses

Using Pearsons' Correlation Score, there is weak to no correlation in all food types vs. medical expenditure in all socioeconomic classes except that there is a correlation (score of 0.32) between the fruit and medical expenditures of the class, "rich".

Possible reason is that the "rich" can afford to consult with a doctor, and the doctor recommends them to spend more on fruits in order to stay healthy.

_id.class	value.medical	value.bread	value.rice	value.meat	value.fruit	value.fish	value.veggie
low income	[0.0]	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]
lower middle income	[35531.0]	[0.07578758951071249]	[-0.01652695778894512]	[-0.12049172964424075]	[-0.1424993166240283]	[-0.1181562660382058]	[0.14628946676019774]
middle class	[295623.0]	[0.10631615041347282]	[0.11011960573911118]	[0.016934944255452805]	[0.020103001454382107]	[0.16407496700253776]	[0.11631444335657595]
rich	[2.52462e+06]	[0.05499283195716369]	[-0.056806950526501455]	[0.10797006546970898]	[0.32416595397943865]	[0.055152903885916434]	[0.033279995960603166]
upper income	[536089.0]	[0.12114442933615376]	[0.06627565358924047]	[0.05857013673727686]	[0.161466179981073]	[0.05196140720993903]	[0.11830553688108054]
upper middle income	[508619.0]	[-0.027964683449687527]	[-0.06176882735808164]	[-0.052225142954405876]	[0.013760840001201214]	[-0.042753059789210054]	[-0.022392125617372163]

Output 5

Output 6:Correlation of income and food expenditures per socioeconomic class

There is a huge variation in the correlations in this section. The group concluded that there is no specific pattern in the correlation of income and food expenditures.

Class	Income	Bread	Rice	Meat	Fruit	Fish	Vegetables	Restaurants
Low Income	18,395.00	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]
Lower Middle Income	1,145,089.00	[0.7209067603034415]	[0.6959879839515012]	[0.03680909050678119]	[0.15014259813731495]	[0.06700969981982433]	[0.2565740440752314]	[0.16129835438333104]
Middle Income	19,622,273.00	[0.6965924234045566]	[0.6147137895950135]	[0.557575094347729]	[0.22581132952616895]	[0.60459164811057]	[0.5892362705366438]	[0.11198766897752074]
Rich	87,952,729.00	[0.3876470498915926]	[0.25836268537711315]	[0.5562477422425717]	[0.4688448915186685]	[0.4144856896154958]	[0.2462392350057111]	[0.5978788381813696]
Upper Income	31,807,744.00	[0.7272777853451874]	[0.6782339474056228]	[0.5095140536670997]	[0.40764806619369204]	[0.4503226976695389]	[0.42537698492883874]	[0.3948883162316718]
Upper Middle Income	31,170,098.00	[0.5928418345159039]	[0.5179310364471952]	[0.5075025344157763]	[0.28278938157842176]	[0.45543535556117304]	[0.31647021258892466]	[0.30155688245846557]

Output 6

Output 7:Correlation of income versus miscellaneous expenditures

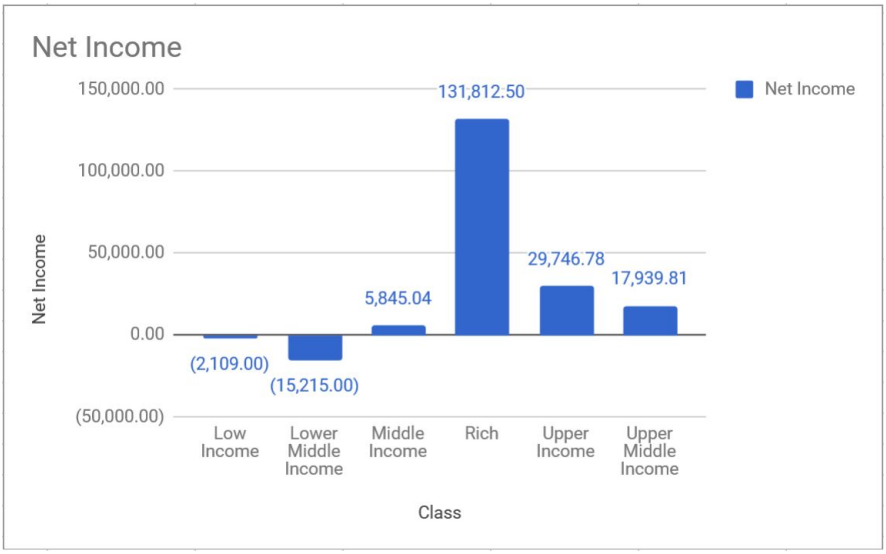
Similar to the previous output, there is a huge variation between the correlations in this output. Noticeable patterns are the strong correlation between income and housing expenditures for all socioeconomic classes, and the strong correlation of income and clothing for Middle, Upper and Rich classes.

Class	Income	Alcohol	Tobacco	Clothing	Medical	Education	Housing	Transportation
Low Income	18395	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]	[NaN]
Lower Middle Income	1,145,089.00	[-0.24474125211363928]	[0.2858528893200227]	[0.20575794578927473]	[0.15740623800608708]	[0.3795078986908859]	[0.534490477227653]	[-0.02135226598681479]
Middle Income	19,622,273.00	[0.18992174169220383]	[0.2594249374193584]	[0.44819642765194084]	[0.2648445642574371]	[0.17402012116110405]	[0.4755404363215061]	[0.39270191443633234]
Rich	87,952,729.00	[0.11126207599122571]	[0.08413438369460206]	[0.5727189211861549]	[0.4413115243716328]	[0.4412383197735635]	[0.6601944254557752]	[0.6361723124185468]
Upper Income	31,807,744.00	[0.2162102894865735]	[0.3105695690160298]	[0.44914162078721587]	[0.24025861060806794]	[0.21706949423469005]	[0.4971766163190154]	[0.4501164588239292]
Upper Middle Income	31,170,098.00	[0.17385907239260862]	[0.2055494658436379]	[0.24534513247396234]	[0.007890766655761935]	[0.13890956530560392]	[0.6024257077229812]	[0.3825115789860795]

Output 7

Output 8: Total excess/deficit per socioeconomic class

Subtracting the average total expenditures from the income of each socioeconomic class, this is the net income/loss of each socioeconomic class.



Output 8

To summarize, the lower the income, the higher the percentages for cheap necessities. And as the income gets higher, there is more budget allocated to luxuries such as clothes and higher costing food types such as meat or even dining in a restaurant. Even with a higher income, budgets on specific food types and other expenditures do not vary often. There are very few patterns that can be seen in using Pearson's Correlation Score with income versus food expenditure and income versus other expenditures.

Reference Links

https://www.kaggle.com/grosvenpaul/family-income-and-expenditure/data

https://www.rappler.com/thought-leaders/98624-who-are-middle-class