



Bharatiya Vidya Bhavan's

**Sardar Patel Institute of Technology**

(Autonomous Institute Affiliated to University of Mumbai)

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India

**Name: Varada Khadake**

**Roll No. :2021300059**

**AIM:**

Create advanced charts using Tableau / Power BI / R / Python / Plotly or Chart or D3.js to be performed on the dataset - Socio economic data

- Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, Jitter, Line, Area, Waterfall, Donut, Treemap, Funnel
- Write observations from each chart

**DATASET:**

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

The Philippine Statistics Authority (PSA) spearheads the conduct of the Family Income and Expenditure Survey (FIES) nationwide. The survey, which is undertaken every three (3) years, is aimed at providing data on family income and expenditure, including, among others, levels of consumption by item of expenditure, sources of income in cash, and related information affecting income and expenditure levels and patterns in the Philippines. Inside this data set is some selected variables from the latest Family Income and Expenditure Survey (FIES) in the Philippines. It contains more than 40k observations and 60 variables which is primarily comprised of the household income and expenditures of that specific household

**Columns:**

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 ,Fruit Expenditure ,Vegetables Expenditure ,Restaurant and hotels Expenditure ,Alcoholic Beverages Expenditure ,Tobacco Expenditure ,Clothing, Footwear and Other Wear Expenditure ,Housing and water Expenditure ,Imputed House Rental Value ,Medical Care Expenditure ,Transportation Expenditure ,Communication Expenditure ,Education Expenditure ,Miscellaneous Goods and Services Expenditure ,Special Occasions Expenditure ,Crop Farming and Gardening expenses ,Total Income from Entrepreneurial Activities ,Household Head Sex ,Household Head Age ,Household Head Marital Status ,Household Head Highest Grade Completed ,Household Head Job or Business Indicator ,Household Head Occupation ,Household Head Class of Worker ,Type of Household ,Total Number of Family members ,Members with age less than 5 year old ,Members with age 5 - 17 years old ,Total number of family members employed ,Type of Building/House ,Type of Roof ,Type of Walls ,House Floor Area ,House Age ,Number

of bedrooms ,Tenure Status ,Toilet Facilities ,Electricity ,Main Source of Water Supply ,Number of Television ,Number of CD/VCD/DVD ,Number of Component/Stereo set ,Number of Refrigerator/Freezer ,Number of Washing Machine ,Number of Airconditioner ,Number of Car, Jeep, Van ,Number of Landline/wireless telephones ,Number of Cellular phone ,Number of Personal Computer ,Number of Stove with Oven/Gas Range ,Number of Motorized Banca ,Number of Motorcycle/Tricycle

## **Tables and respective columns:**

### **Fact Table: Fact\_Household\_Income\_Expenditure**

Pirmary key: Household\_Id

- Total\_Household\_Income
- Total\_Food\_Expenditure
- Bread\_and\_Cereals\_Expenditure
- Total\_Rice\_Expenditure
- Meat\_Expenditure
- Total\_Fish\_and\_Marine\_Products\_Expenditure
- Fruit\_Expenditure
- Vegetables\_Expenditure
- Restaurant\_and\_Hotels\_Expenditure
- Alcoholic\_Beverages\_Expenditure
- Tobacco\_Expenditure
- Clothing\_Footwear\_and\_Other\_Wear\_Expenditure
- Housing\_and\_Water\_Expenditure
- Imputed\_House\_Rental\_Value
- Medical\_Care\_Expenditure
- Transportation\_Expenditure
- Communication\_Expenditure
- Education\_Expenditure
- Miscellaneous\_Goods\_and\_Services\_Expenditure
- Special\_Occasions\_Expenditure
- Crop\_Farming\_and\_Gardening\_Expenses
- Total\_Income\_from\_Entrepreneurial\_Activities

## **Dimension Tables**

### **Dim\_Household**

**Primary Key:**

- Household\_ID

**Attributes:**

- Main\_Source\_of\_Income
- Agricultural\_Household\_Indicator (Yes/No)
- Type\_of\_Household
- Total\_Number\_of\_Family\_Members
- Members\_Age\_Less\_Than\_5
- Members\_Age\_5\_to\_17
- Family\_Members\_Employed
- House\_Floor\_Area
- House\_Age
- Number\_of\_Bedrooms
- Tenure\_Status
- Electricity (Yes/No)
- Main\_Source\_of\_Water\_Supply

**Dim\_Household\_Head**

**Primary Key:**

- Household\_ID (or a separate Household\_Head\_ID if one household can have multiple heads over time)

**Attributes:**

- Household\_Head\_Sex
- Household\_Head\_Age
- Household\_Head\_Marital\_Status
- Household\_Head\_Highest\_Grade\_Completed
- Household\_Head\_Job\_or\_Business\_Indicator (Employed/Unemployed)
- Household\_Head\_Occupation
- Household\_Head\_Class\_of\_Worker

**Dim\_Housing\_Characteristics**

**Primary Key:**

- Household\_ID

**Attributes:**

- Type\_of\_Building\_House

- Type\_of\_Roof
- Type\_of\_Walls
- Toilet\_Facilities

**Dim\_Appliances**

Number\_of\_Television

Number\_of\_CD\_VCD\_DVD

Number\_of\_Component\_Stereo\_Set

Number\_of\_Refrigerator\_Freezer

Number\_of\_Washing\_Machine

Number\_of\_Airconditioner

Number\_of\_Car\_Jeep\_Van

Number\_of\_Landline\_Wireless\_Telephones

Number\_of\_Cellular\_Phone

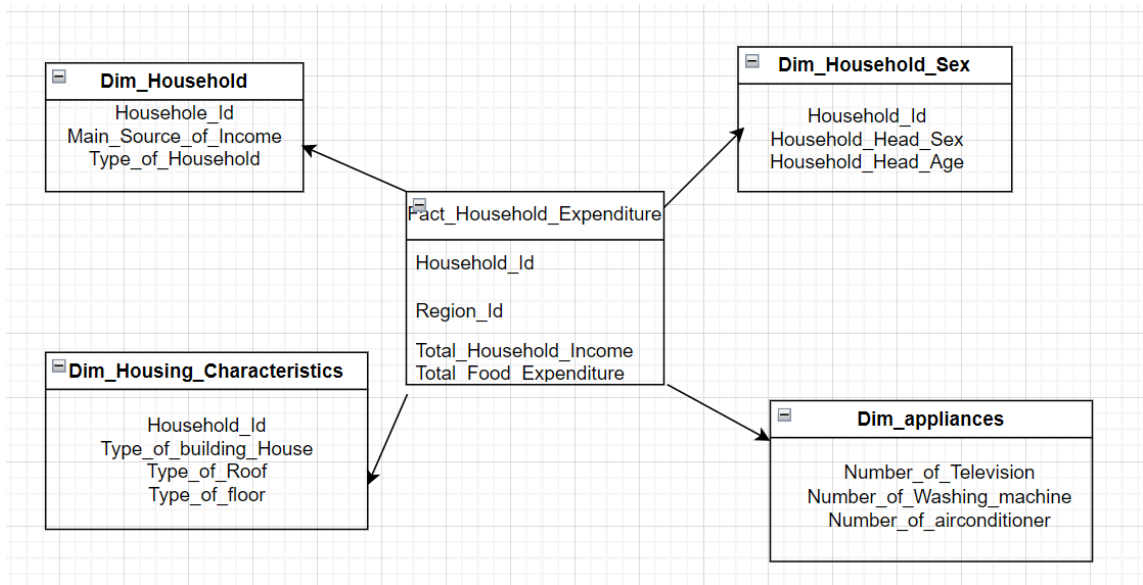
Number\_of\_Personal\_Computer

Number\_of\_Stove\_with\_Oven\_Gas\_Range

Number\_of\_Motorized\_Banca

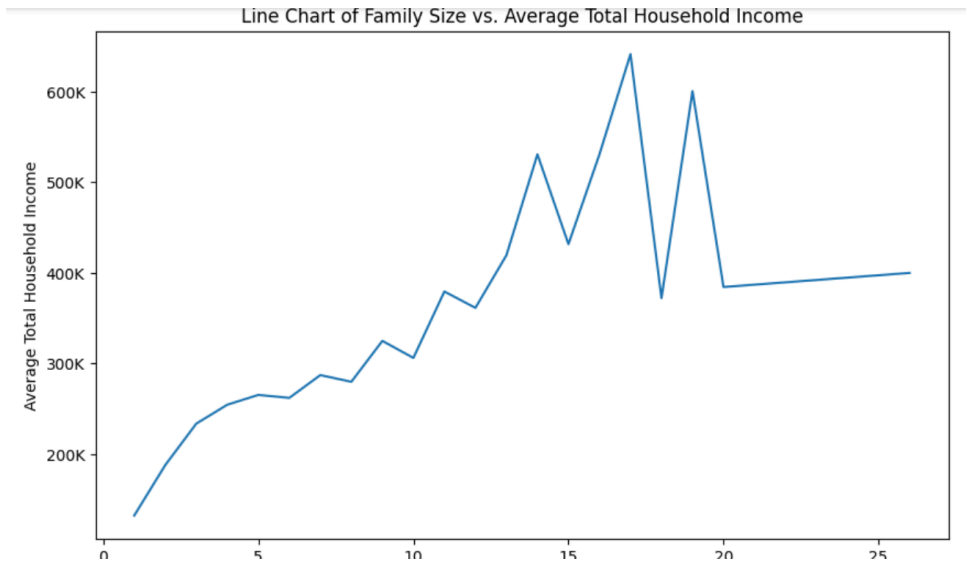
Number\_of\_Motorcycle\_Tricycle

Star Schema:



PROCEDURE DESCRIPTION:

1.Line Chart

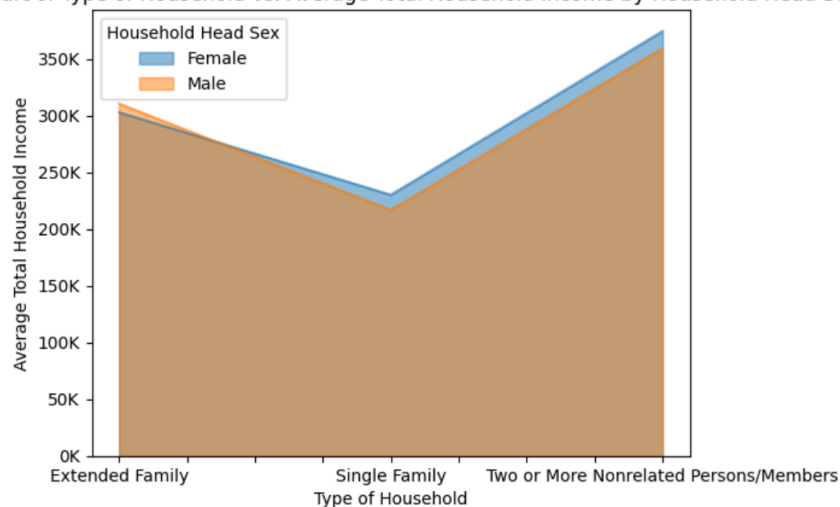


The line chart illustrates the relationship between family size and average total household income. The trend shows that as family size increases from 1 to around 16 members, the

average total household income generally increases, with some fluctuations. The income peaks significantly at certain points, particularly around a family size of 16, where it reaches the highest value. However, after this peak, there's a sharp decline in income, followed by another fluctuation as the family size increases beyond 16 members. This pattern suggests that while larger families may have higher incomes, possibly due to more earning members, the relationship is not entirely linear, indicating that other factors may influence household income beyond just family size. The sharp dips after certain family sizes might be due to increased financial strain or fewer earners contributing relative to dependents.

## 2. Area Chart

Area Chart of Type of Household vs. Average Total Household Income by Household Head Sex

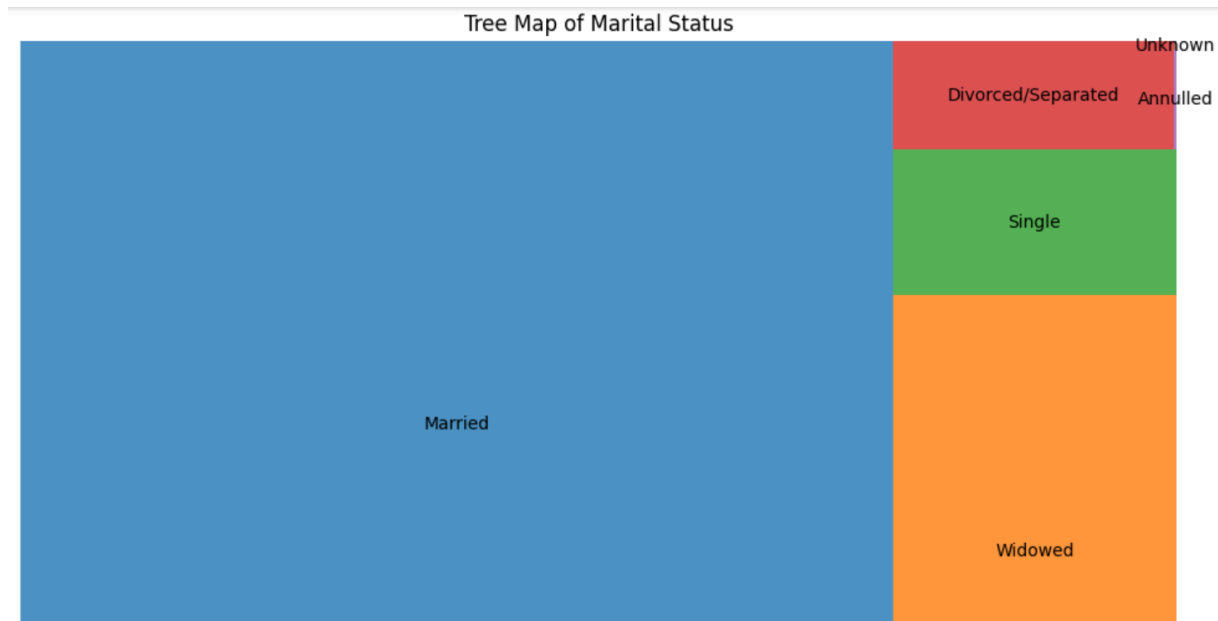


The area chart displays the relationship between the type of household and average total household income, segmented by the sex of the household head. The chart reveals that households headed by both males and females have a similar income distribution across different household types, with male-headed households generally earning slightly more.

For "Extended Family" households, the income is relatively higher compared to "Single Family" households, where the income dips slightly. However, for households with "Two or More Nonrelated Persons/Members," there is a notable increase in income, surpassing both extended and single-family households. This trend suggests that households with more diverse or nontraditional family structures might have multiple income sources, contributing to higher overall income. The slight difference between male

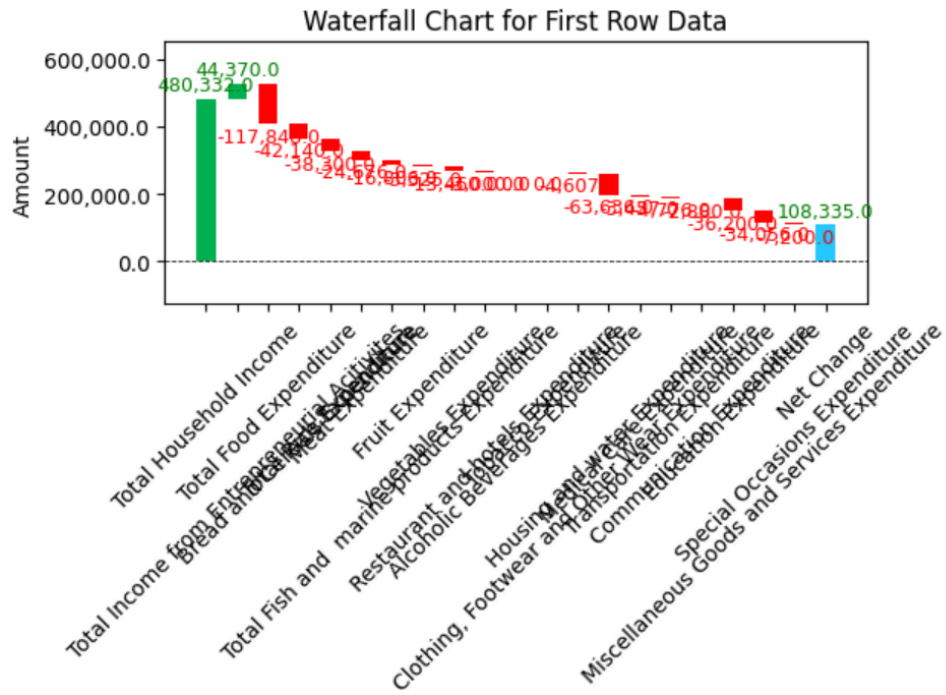
and female-headed households indicates a minor gender gap in income, which seems consistent across the various household types.

### 3. Treemap



The treemap provides a visual representation of the distribution of marital statuses within the dataset. The most prominent category is "Married," which occupies the largest area, indicating that the majority of the population in the dataset is married. Following this, "Widowed" and "Single" also have significant portions, but they are notably smaller than the "Married" category. The categories "Divorced/Separated" and "Annulled" are much smaller, reflecting a lower proportion of individuals in these marital statuses. The "Unknown" category is the smallest, suggesting that very few records have missing or unidentified marital status. This visualization highlights the predominance of marriage within the surveyed population and provides insight into the relative prevalence of other marital statuses.

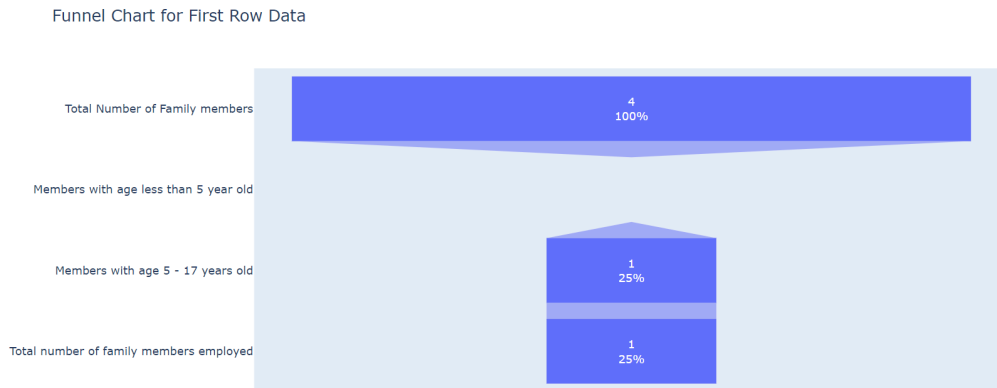
#### 4. Waterfall Chart



- **Total Household Income:** The starting point is a significant amount of income.
- **Non-Food Expenditures:** A large portion of the income is allocated to non-food categories such as housing, clothing, and transportation.
- **Food Expenditures:** Food-related expenses, including fruits, vegetables, and restaurant meals, constitute a significant portion of the budget.
- **Net Change:** The chart indicates a positive net change, suggesting that the total income exceeds the total expenditures.



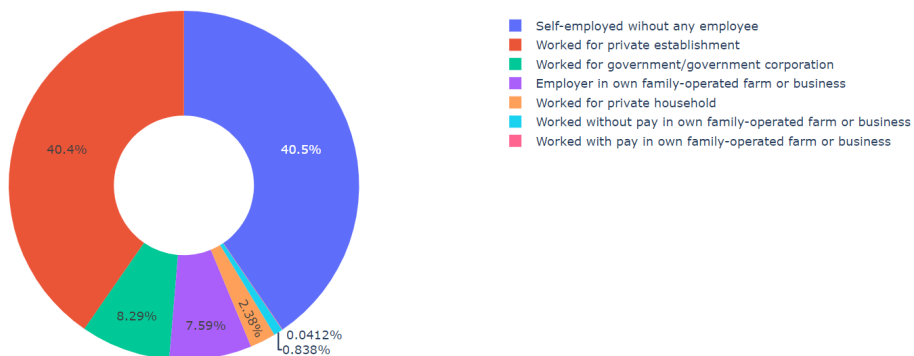
## 5. Funnel Chart



- **Total Family Members:** The chart starts with a total of four family members.
- **Age Group 0-5:** 75% of the family members are over 5 years old.
- **Age Group 5-17:** 25% of the family members fall within this age range.
- **Employed Members:** 25% of the family members are employed.

## 6. Donut Chart

Donut Chart for Household Head Class of Worker



- **Self-Employed without Employees:** A significant portion of household heads (40.4%) are self-employed without any employees.
- **Worked for Private Establishment:** Another large group (40.5%) worked for private establishments.
- **Government/Government Corporation:** A smaller percentage (8.29%) worked for government or government-related entities.
- **Employer in Own Family-Operated Farm or Business:** 7.59% were employers in their own family-operated farms or businesses.
- **Worked for Private Household:** A very small percentage (2.38%) worked for private households.
- **Worked Without Pay in Own Family-Operated Farm or Business:** 0.638% worked without pay in their own family-operated farms or businesses.
- **Worked with Pay in Own Family-Operated Farm or Business:** The smallest category (0.0412%) worked with pay in their own family-operated farms or businesses.