Data Description

The US Bureau survey contains information about the products each household bought, the number of household members, and general information regarding the household. Please see Tables 1, 2, and 3, which represent the household member dataset **(HOUSEHOLD_MEMBERS.csv)**, household dataset **(HOUSEHOLDS.csv)**, and expenditure dataset **(Expenditures.csv)** respectively. The datasets were originally designed to predict whether or not an item was a gift, but the same dataset helped solve our unique and original problem.

The datasets did not have complete definitions and input features marked with an asterisk (*) meant that descriptions were inferred based on exploratory data analysis.

Table 1 shows the household member dataset. This dataset provides us with details about each member within the household.

Table 1 - Shows us the Household member dataset **(HOUSEHOLD_MEMBERS.csv)**

Input Feature	Description
HOUSEHOLD_ID	Unique Identifier for each household
YEAR	Year of survey
MARITAL	Marital status of household member
SEX	Gender of household member
AGE	Age of household member
WORK_STATUS*	Number of jobs the household member is doing.

Table 2 shows the household dataset. This dataset provides us with details about each household. The datasets did not have complete definitions and input features marked with an asterisk (*) meant that descriptions were inferred based on exploratory data analysis.

Table 2 -Shows us the Household dataset **(HOUSEHOLDS.csv)**

Input Feature	Description
HOUSEHOLD_ID	Unique Identifier for each household
YEAR	Year of survey
INCOME_RANK*	Household Quintiles of income before taxes
INCOME_RANK_1*	Household Quintiles of income before taxes
INCOME_RANK_2*	Household Quintiles of income before taxes
INCOME_RANK_3*	Household Quintiles of income before taxes
INCOME_RANK_4*	Household Quintiles of income before taxes

INCOME_RANK_5*	Household Quintiles of income before taxes
INCOME_RANK_MEAN*	Household Quintiles of income before taxes Footnote: This is not the average of the various rankings. A SQL test was conducted to check for this.
AGE_REF	Age of House

Table 3 shows the expenditure dataset. This dataset provides us with details about each item bought by the household. The datasets did not have complete definitions and input features marked with an asterisk (#) meant that descriptions could not be inferred based on exploratory data analysis.

Table 3 - Shows us the Expenditure dataset **(Expenditures.csv)**

Input Feature	**Description**
EXPENDITURE_ID	Unique Identifier for each item bought
HOUSEHOLD_ID	Unique Identifier for each household
YEAR	Year that item was bought
MONTH	Month that item was bought
PRODUCT_CODE	Numerical product code of item bought
COST	Cost of item bought
GIFT	If the item was purchased as a gift
IS_TRAINING#	Unknown#

The aggregated features from the expenditure and household member datasets were then merged with the household dataset to create the final dataset, which was used for the Python EDA. The datasets did not have complete definitions and input features marked with an asterisk (*) meant that descriptions were inferred based on exploratory data analysis. Please review the SQL code to see the final transformations. Table 4 contains the final dataset before Python EDA **(Derived_Household_Features.csv)**.

Table 4 - Final Dataset Before Python EDA

Features	Descriptions
HOUSEHOLD_ID	Unique Identifier for each household
YEAR	Year of survey
INCOME_RANK*	Household Quintiles of income before taxes
INCOME_RANK_1*	Household Quintiles of income before taxes
INCOME_RANK_2*	Household Quintiles of income before taxes
INCOME_RANK_3*	Household Quintiles of income before taxes
INCOME_RANK_4*	Household Quintiles of income before taxes
INCOME_RANK_5*	Household Quintiles of income before taxes
INCOME_RANK_MEAN*	Household Quintiles of income before taxes Footnote: This is not the average of the various rankings. A SQL test was conducted to check for this.
AGE_REF	Age of house
NUMBER_OF_HOUSEHOLD_JOBS	Number of jobs held by household members.
NUMBER_OF_HOUSEHOLD_MEMBERS_WITH_JOBS	Number of household members with jobs.
NUMBER_OF_HOUSEHOLD_MEMBERS	Number of household members.
NUMBER_OF_FEMALES	Number of females in the household.
NUMBER_OF_GIFTS	Number of gifts purchased by the household.
NUMBER_OF_PURCHASES	Number of purchases made by the household.
BAND_ONE_PURCHASES	Number of household purchases made on products that fall into band one (the average purchase price of the product code was above $5000).
BAND_TWO_PURCHASES	Number of household purchases made on products that fall into band two (the average purchase price of the product code was between $500 and $4999.99).
BAND_THREE_PURCHASES	Number of household purchases made on products that fall into band three (the

	average purchase price of the product code was between $250 and $499.99).
BAND_FOUR_PURCHASES	Number of household purchases made on products that fall into band four (the average purchase price of the product code was between $100 and $249.99).
BAND_FIVE_PURCHASES	Number of household purchases made on products that fall into band five (the average purchase price of the product code was between $50 and $99.99).
BAND_SEX_PURCHASES	Number of household purchases made on products that fall into band sex (the average purchase price of the product code was between $10 and $49.99).
BAND_SEVEN_PURCHASES	Number of household purchases made on products that fall into band seven (the average purchase price of the product code was between $0.01 and $9.99).
TOTAL_SPENT	The total spent by the household in the year of the survey.