UNIVERSITY OF OULU

521251S-3003: Special Course in Information Technology 3 - Data Mining Project

CUSTOMER PERSONALITY ANALYSIS

Group Members

Arash Nedaei Janbesaraei
Zohreh Yousefi Dahka
Kavinda Kulasinghe
Taufiq Ahmed
Md Mobusshar Islam

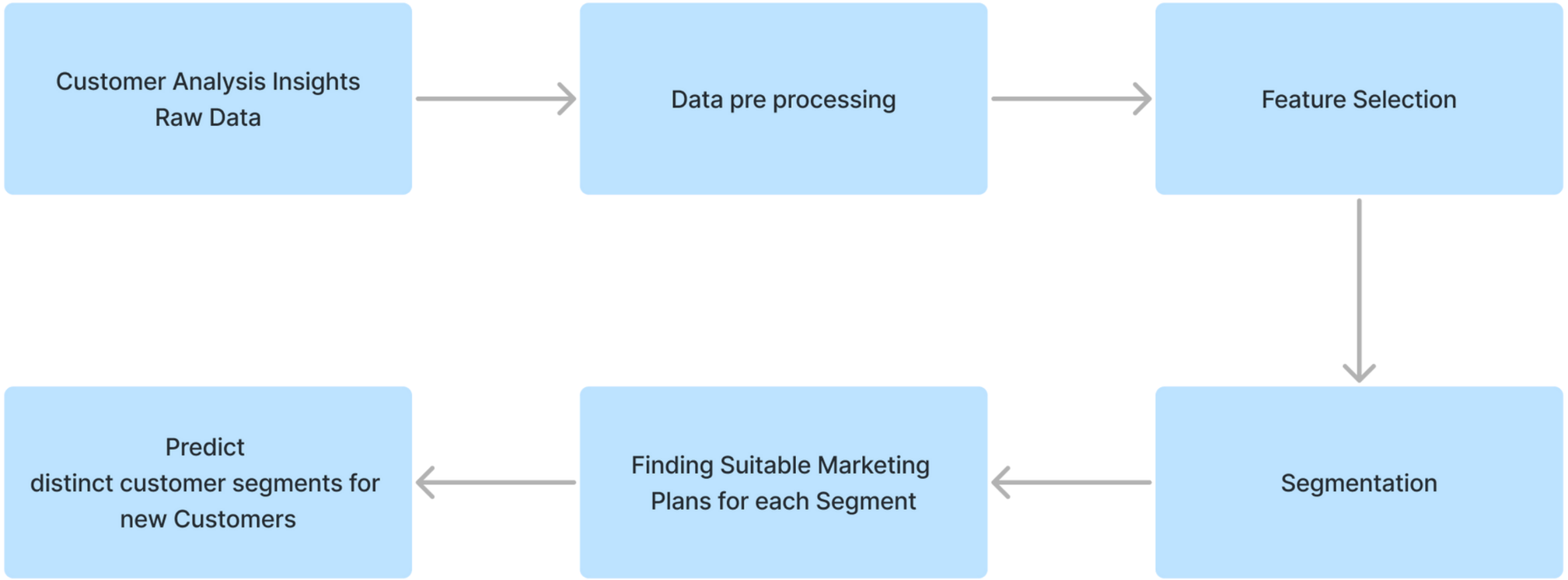
ABOUT THE DATASET

- Customer Personality Analysis is a detailed analysis of a company's ideal customers.
- Dataset has 2240 data points and 29 attributes
- It can be categorized into,
  - Customer's information
  - Products
  - Place
  - Promotioin

RESEARCH IDEA

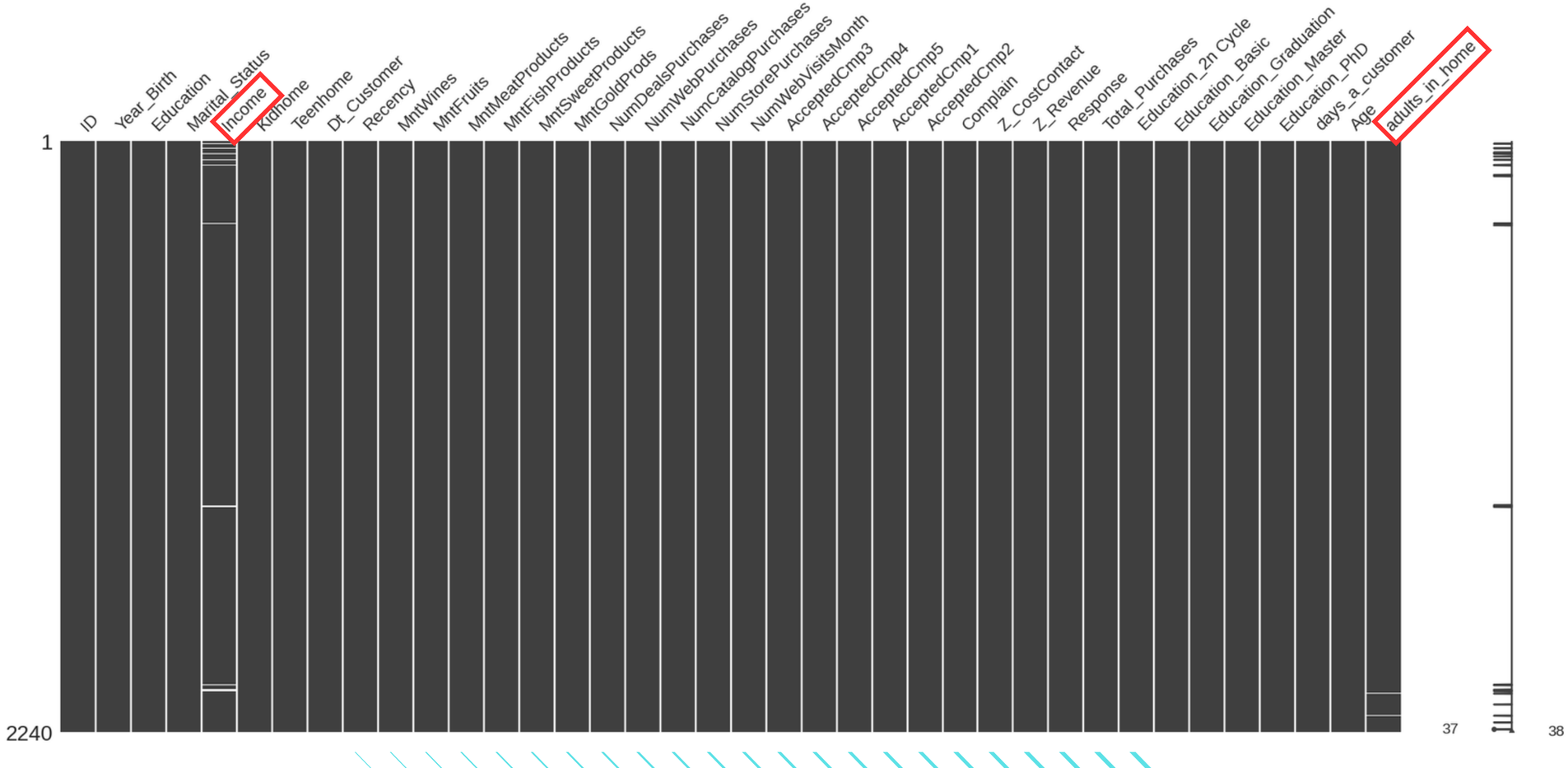
- Identifying how marketing strategies can be optimized to effectively target and engage distinct customer segments.
  - What clustering techniques can be employed to accurately identify distinct customer segments based on relevant demographic, income, ... factors?
  - Predicting the customer segment to which a new customer belongs to devise tailored marketing strategies.

RESEARCH PLAN



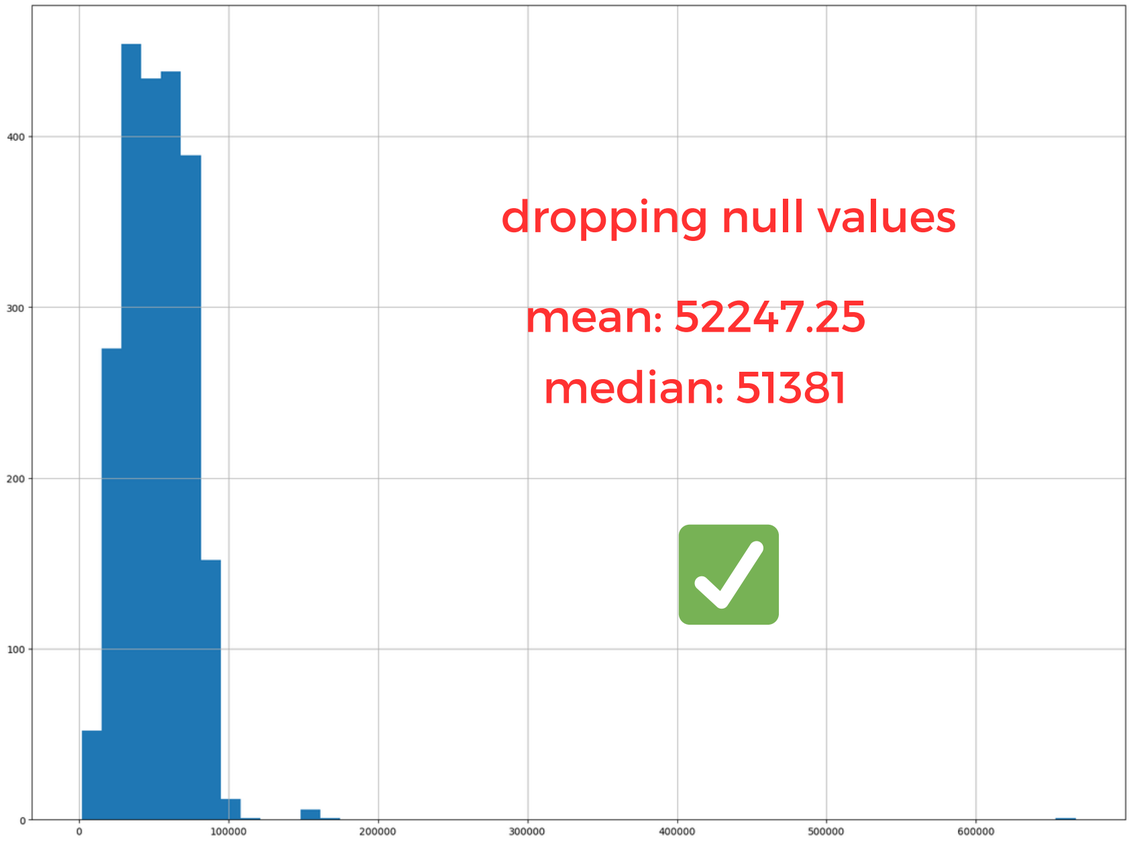
DATA PREPROCESSING

HANDLING MISSING VALUES



DATA PREPROCESSING

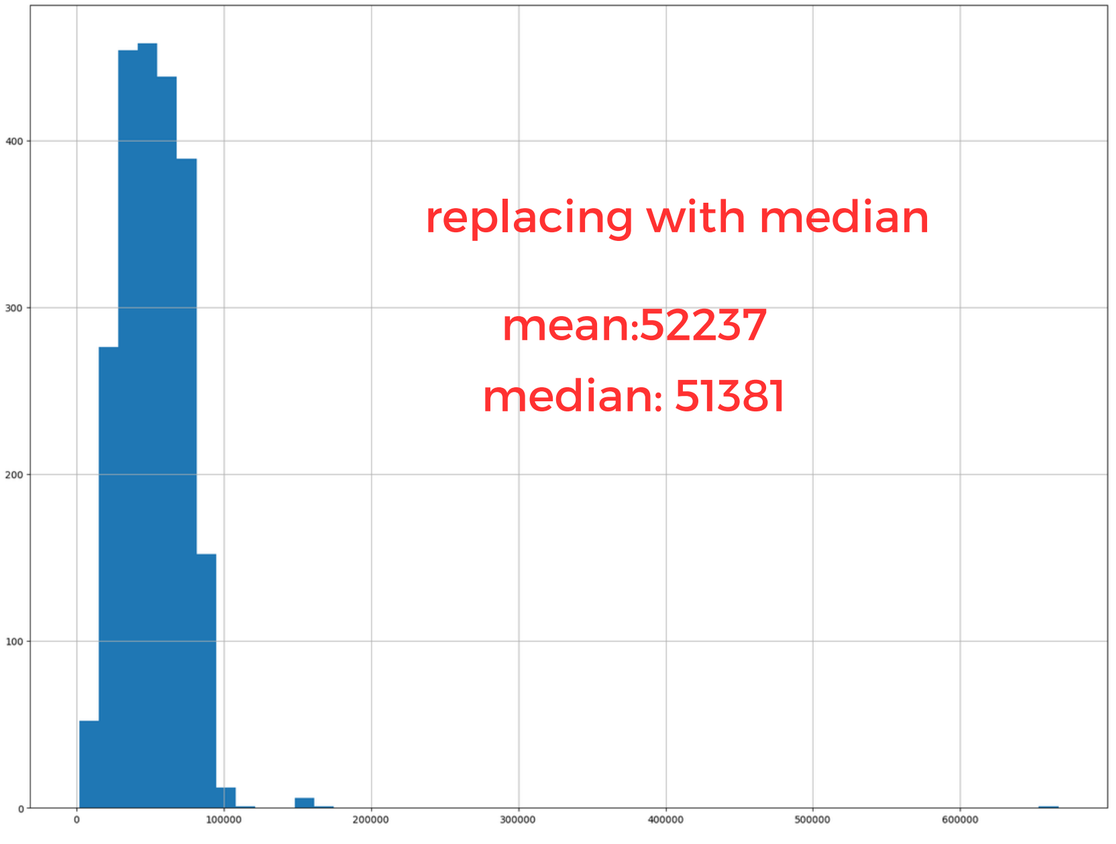
Handling missing values in Income



dropping null values

mean: 52247.25

median: 51381



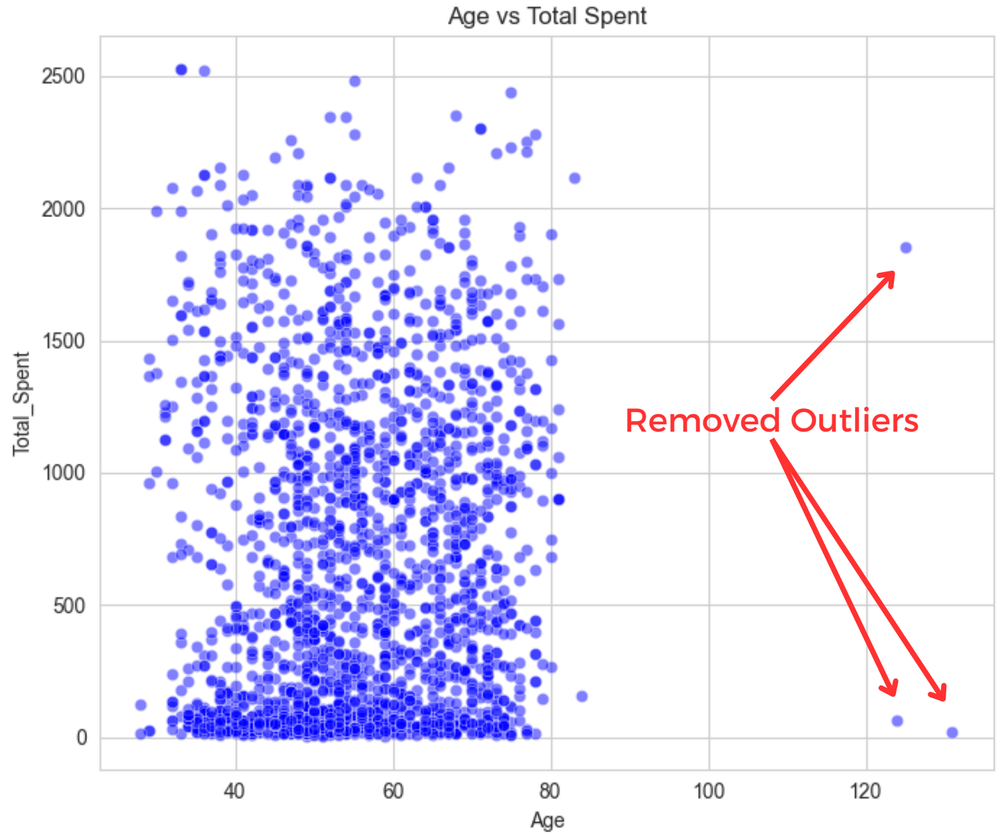
replacing with median

mean:52237

median: 51381

DATA PREPROCESSING

OUTLIER DETECTION



DATA PREPROCESSING

FEATURE CREATION AND REMOVAL

Added

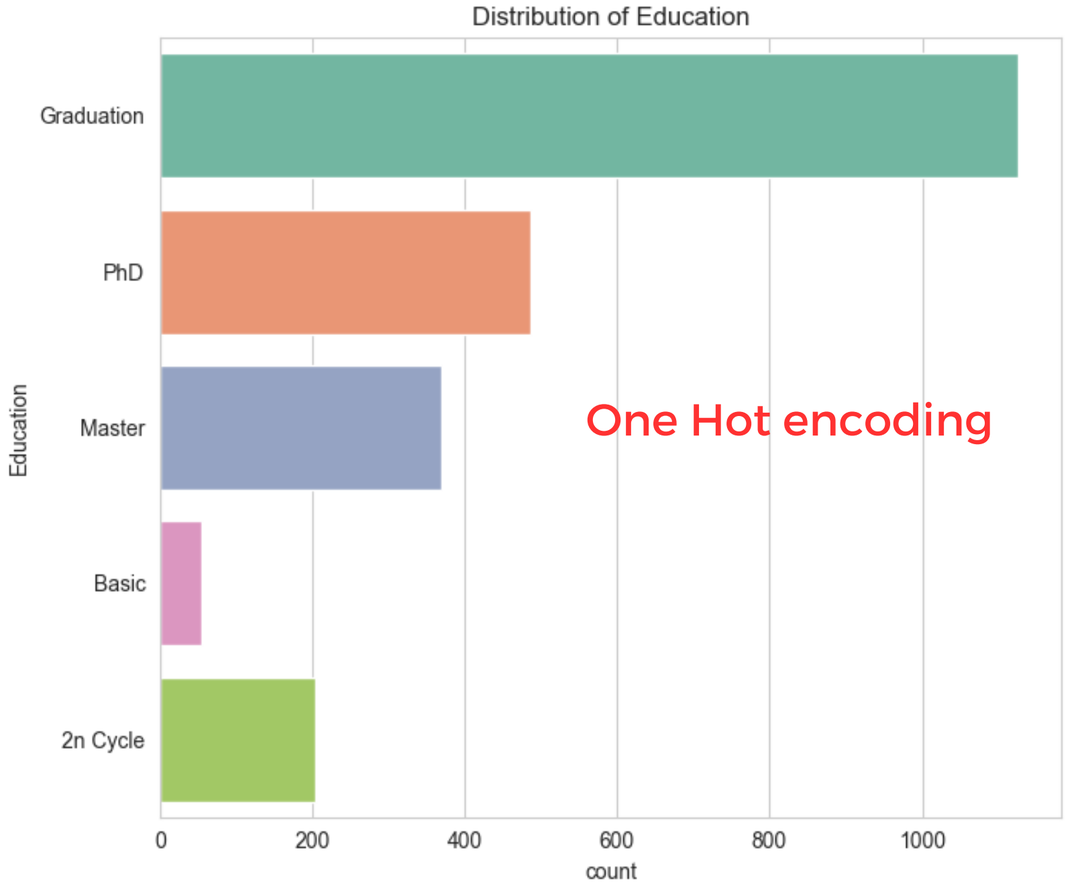
- Age (Based on the latest year in the dataset)
- Spent (Total spending on the products)
- Children (Merged kids and teenagers)
- Adults
- Family Size
- Days_a_customer (Based on the latest date in the dataset)
- Total_Purchases
- Income_per_person (Income divided by family size)

Removed

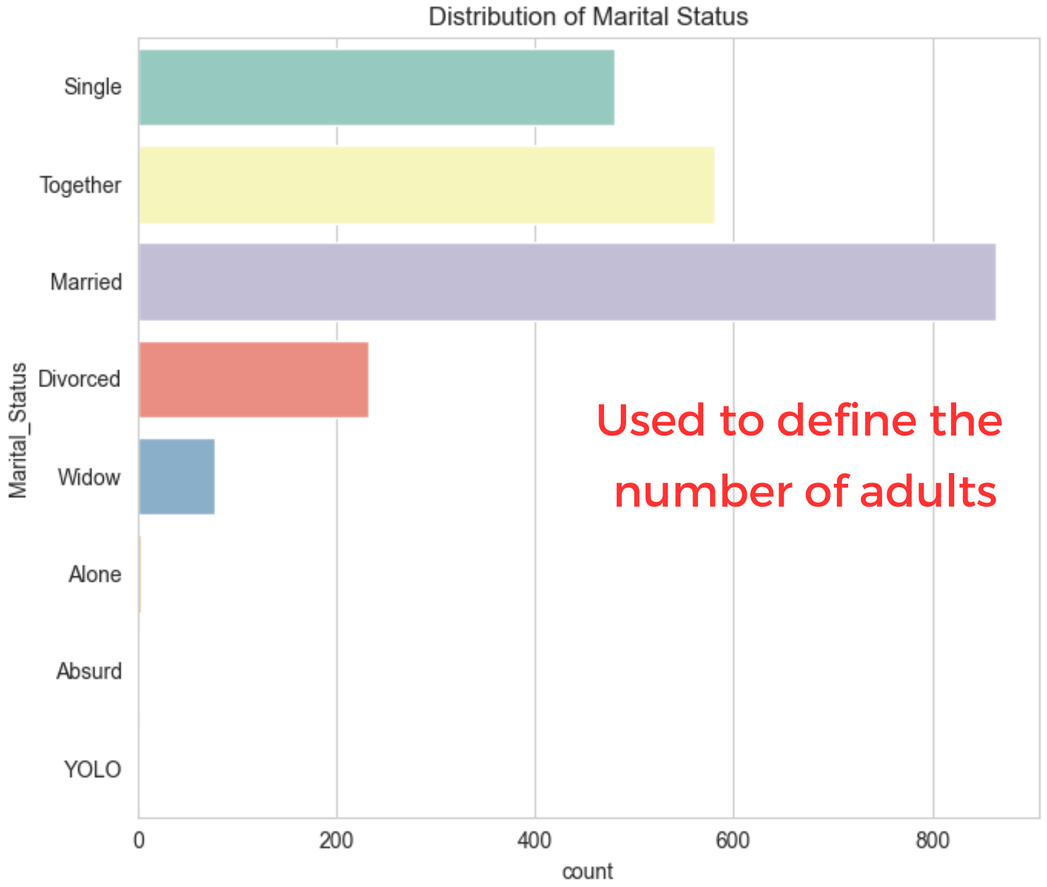
- Dt_Customer
- Marital Status
- Year_Birth
- Z_CostContact
- Z_Revenue

DATA PREPROCESSING

ENCODING CATEGORICAL VARIABLES

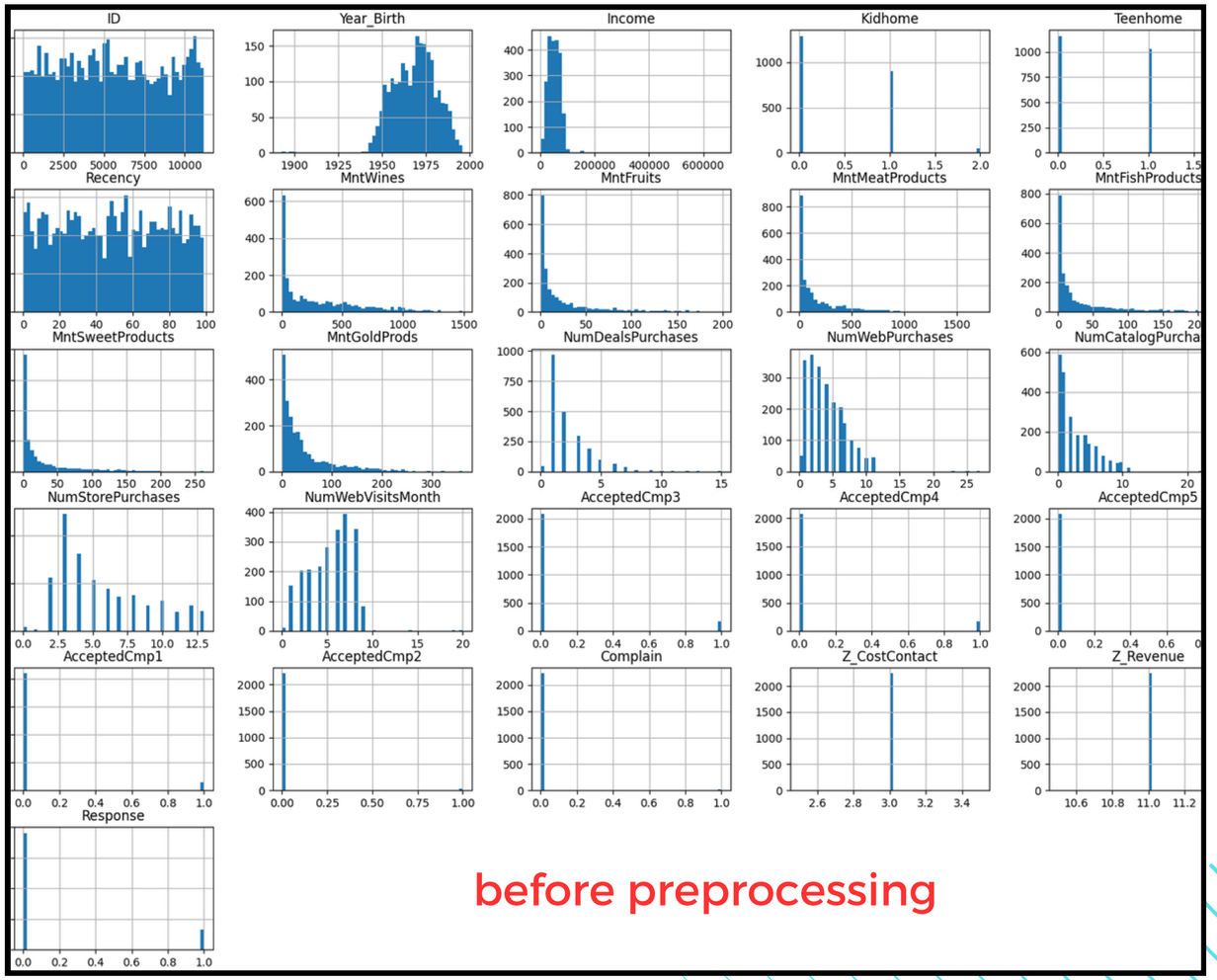


One Hot encoding

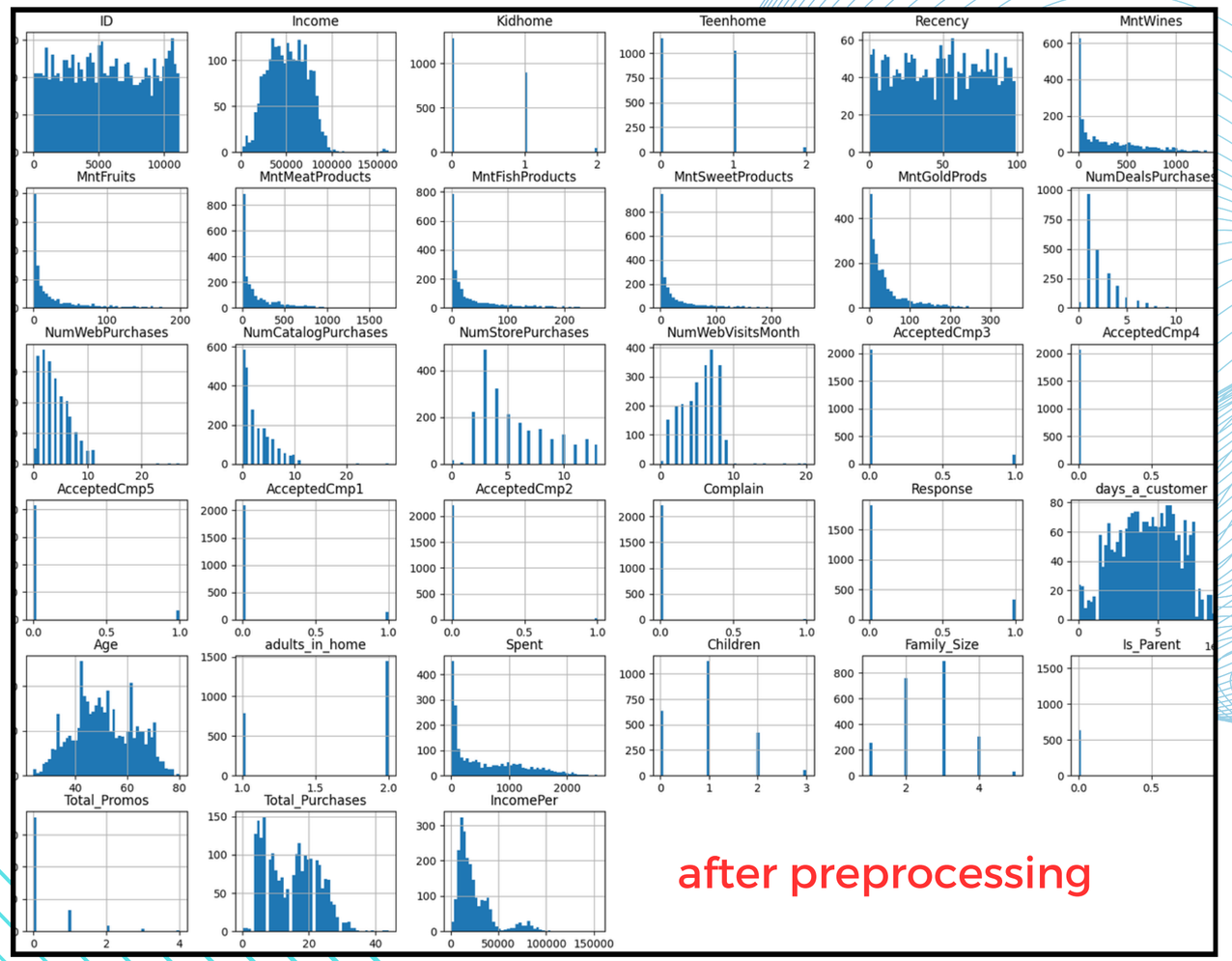


Used to define the number of adults

FINDINGS



before preprocessing



after preprocessing

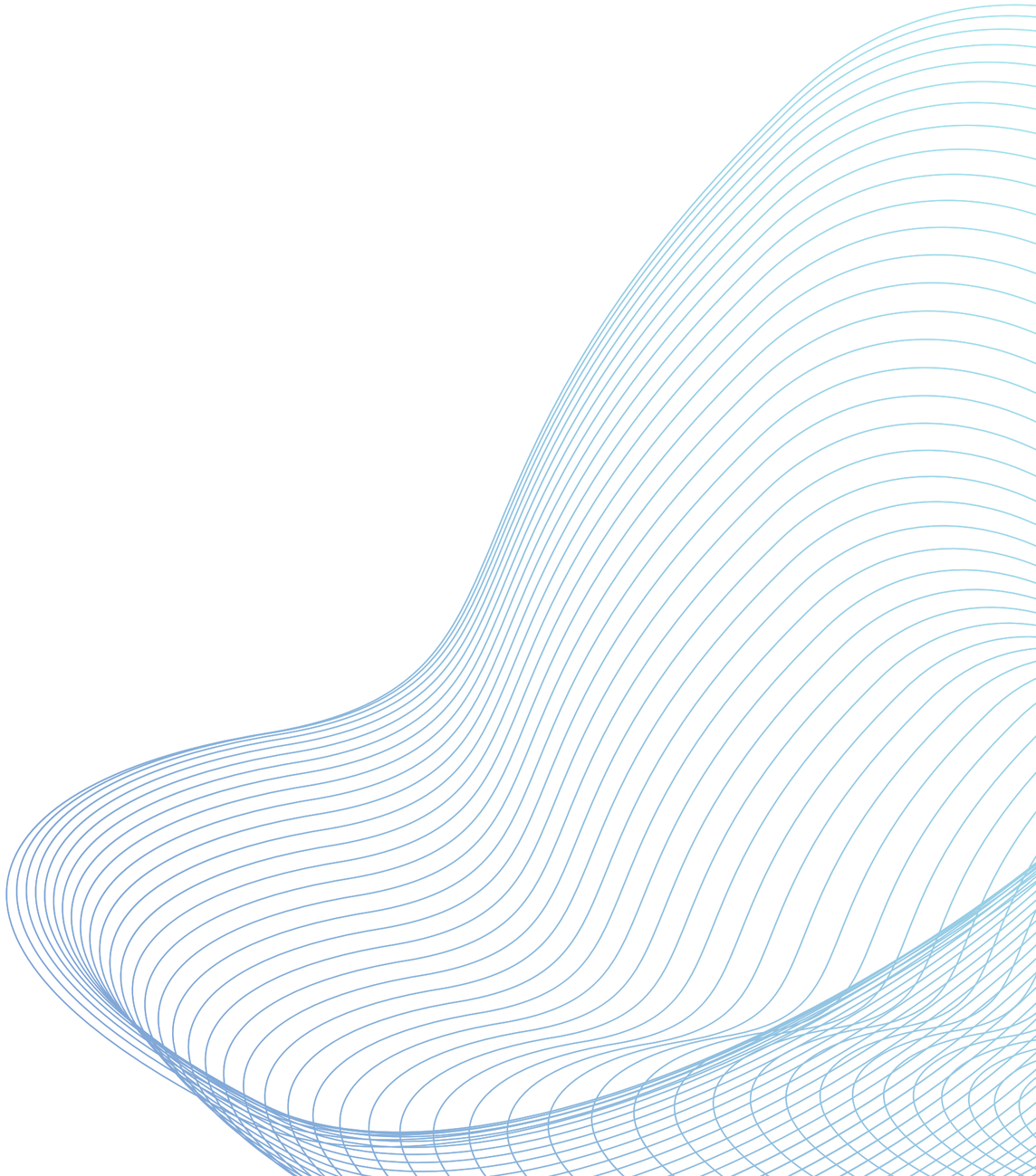
FINDINGS

- The chart for Income per person is different than that of Income
- The chart for total spending is more similar to income per person rather than Income.
- replacing the null values with median can change the mean and so we decided to remove null values

FUTURE STEPS

- Compare and find out the best cluster
- Finding the best marketing strategy for different customer segments
- Predicting the customer segment to which a new customer belongs to devise tailored marketing strategies.

INDIVIDUAL CONTRIBUTIONS

Arash Nedaei Janbesaraei	Data Preprocessing, Peer Review
Zohreh Yousefi Dahka	Data Preprocessing, Research Plan
Kavinda Kulasinghe	Data Preprocessing, Presentation Preparation
Taufiq Ahmed	Data Preprocessing, Peer Review
Md Mobusshar Islam	Data Preprocessing, Presentation Delivery

THANK YOU