



Descriptive Data Analysis

Customer Personality Analysis

Group – 7

Revanth Vemula

Raghava Sagar

Harshil Chauhan

Problem statement

Distribution of
different
personality
traits.

Most prevalent
customer
segments.

Dataset

- This dataset is on customer personality analysis acquired from Kaggle.
- It explains about the which consumer category is most likely to purchase the product, and then market the product exclusively to that group.
- Link to dataset: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>

ID	Year_Birth	Education	Marital_St...	Income	Kidhome	Teenhome
5524	1957	Graduation	Single	58138	0	0
2174	1954	Graduation	Single	46344	1	1
4141	1965	Graduation	Together	71613	0	0
6182	1984	Graduation	Together	26646	1	0
5324	1981	PhD	Married	58293	1	0
7446	1967	Master	Together	62513	0	1
965	1971	Graduation	Divorced	55635	0	1
6177	1985	PhD	Married	33454	1	0
4855	1974	PhD	Together	30351	1	0
5899	1950	PhD	Together	5648	1	1
1994	1983	Graduation	Married		1	0
387	1976	Basic	Married	7500	0	0
2125	1959	Graduation	Divorced	63033	0	0
8180	1952	Master	Divorced	59354	1	1
2569	1987	Graduation	Married	17323	0	0
2114	1946	PhD	Single	82800	0	0
9736	1980	Graduation	Married	41850	1	1
4939	1946	Graduation	Together	37760	0	0
6565	1949	Master	Married	76995	0	1
2278	1985	2n Cycle	Single	33812	1	0

Description of Dataset

Data	Description
Marketing_campaign.csv (2240 rows, 29 columns)	The dataset has information about People, Products, Promotions, and Place.
	People: Includes Id, birth date, education, income, marital status, Customer data, etc.
	Products: Includes wines, fruits, meat, fish, sweet and gold products
	Promotions: Deals purchased, coupon used and response
	Place: where the item was purchased like store, web, catalog and no. Of web visits

Data Cleaning

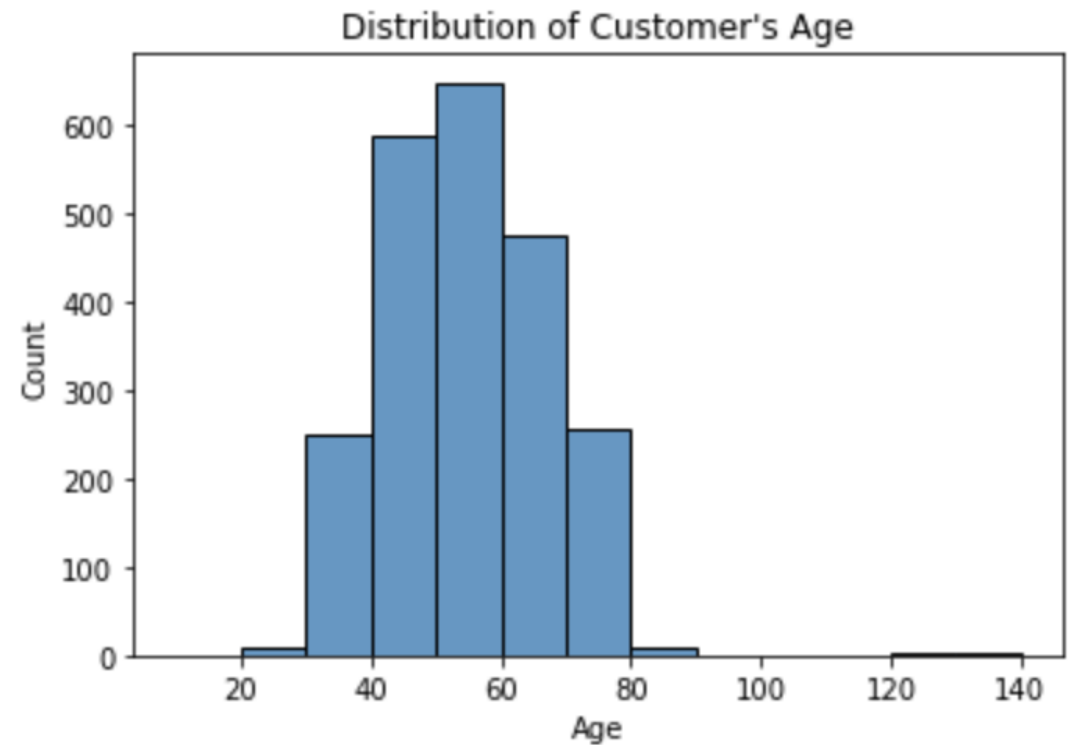
- The entire data was fitted in a single column hence we can use Seperator which divides each attribute into separate column.
- `df = pd.read_csv("data/marketing_campaign.csv", sep="\t")`
`df.head()`

Code:

[Customer Personality Analysis.ipynb](#)

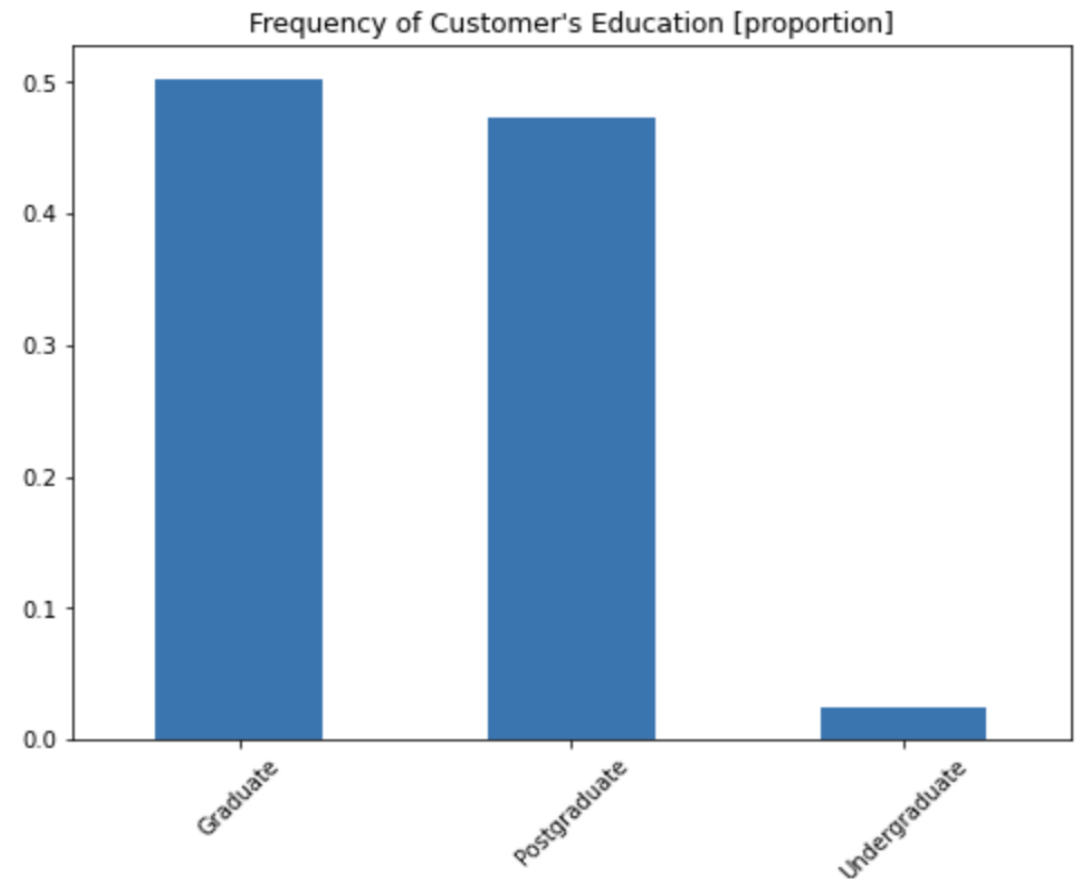
Descriptive Data Analysis – Age Range

- Year_Birth is converted into an age feature.
- Distribution of age is considered.
- The greatest number of customers are in the age range of 40 to 60



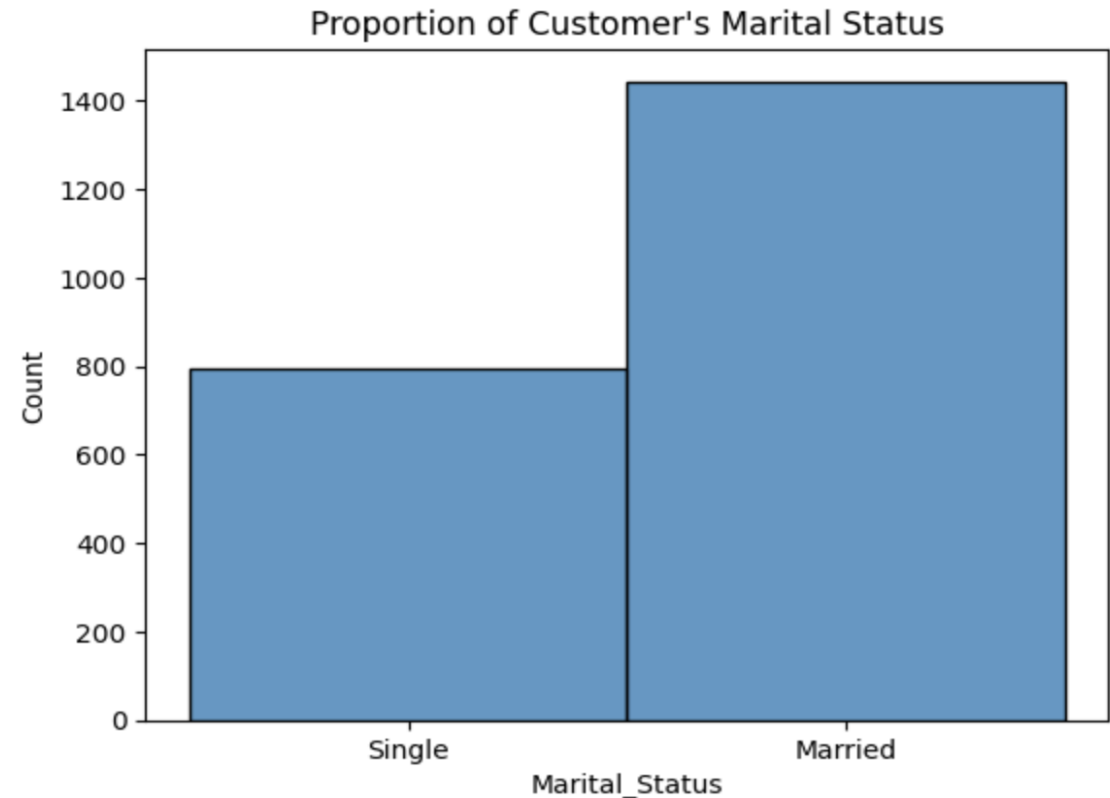
Descriptive Data Analysis – Education

- It is seen from the summary that half of our customer's highest education level is first degree graduation.
- About 50% of customers have their education level at bachelor's degree which is then followed by customers with postgraduate level of education.



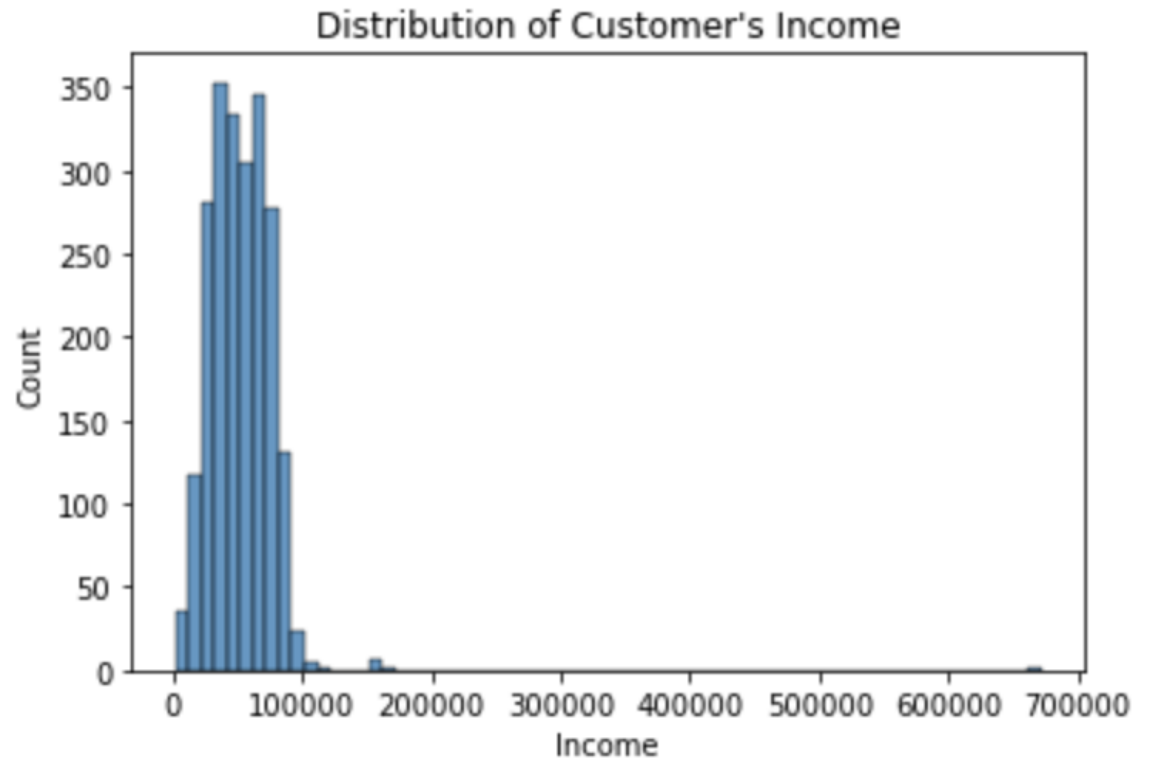
Descriptive Data Analysis – Marital Status

- It is seen from the summary that close to 65% of customers are married while the remaining close to 35% are single.



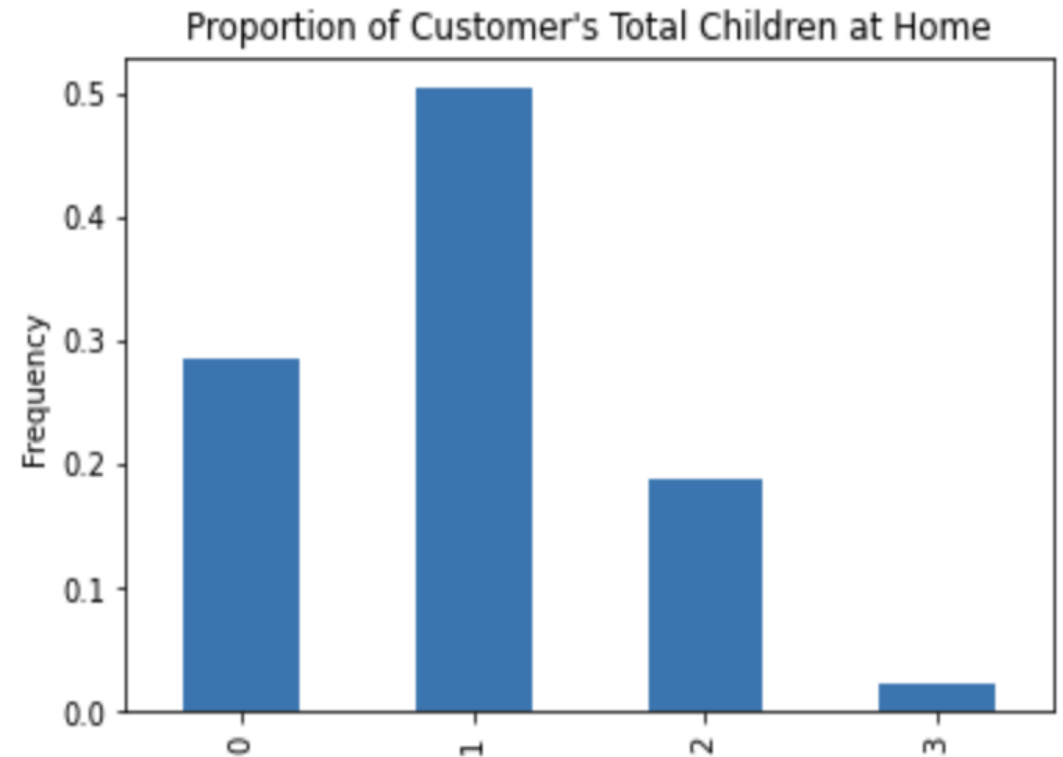
Descriptive Data Analysis – Income

- It is seen that majority of customer's income is within 0-\$100k.
- However, we have another customer's that earn way more than that (above \$600k) like an outlier.



Descriptive Data Analysis – Total Children

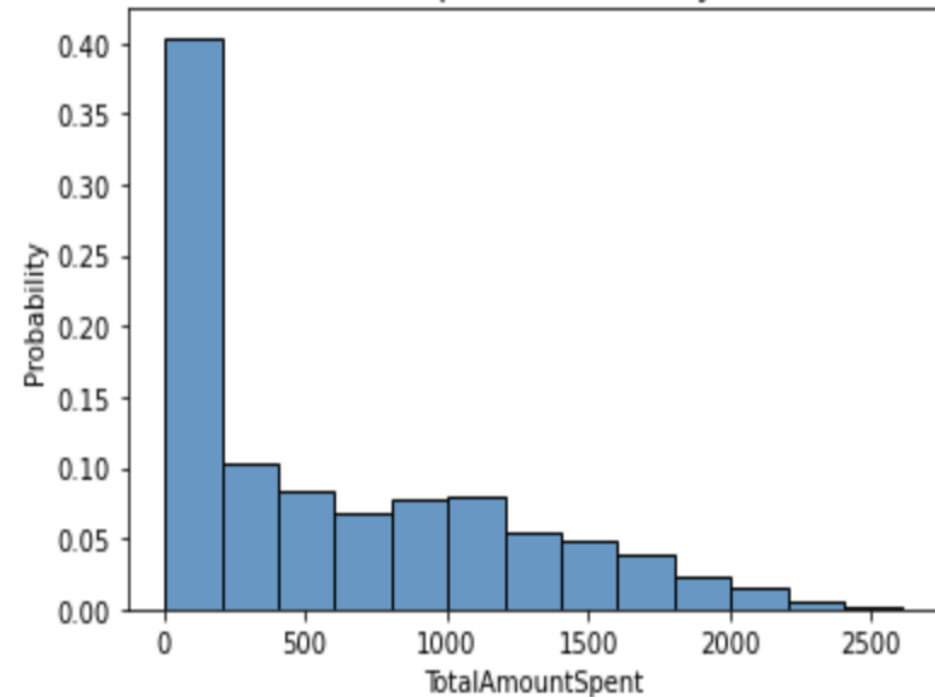
- Close to half of entire customers have the total number of children as 1, while the remaining half (about a quarter) of customers have no children at all.



Descriptive Data Analysis – Total Spent

- It can be seen from the summary that close to half of customer's total amount spent on the company's product is between 0 to 200.

Distribution of Total Amount Spent on Product by Customers [Proportion]





Conclusion

- Income is a key indicator in determining the amount a customer spends.
- Customers with higher education level spends more considering they earn more.
- Married customers spends more considering they have a family to take care of.
- Average spending range for each product is about 0-200.