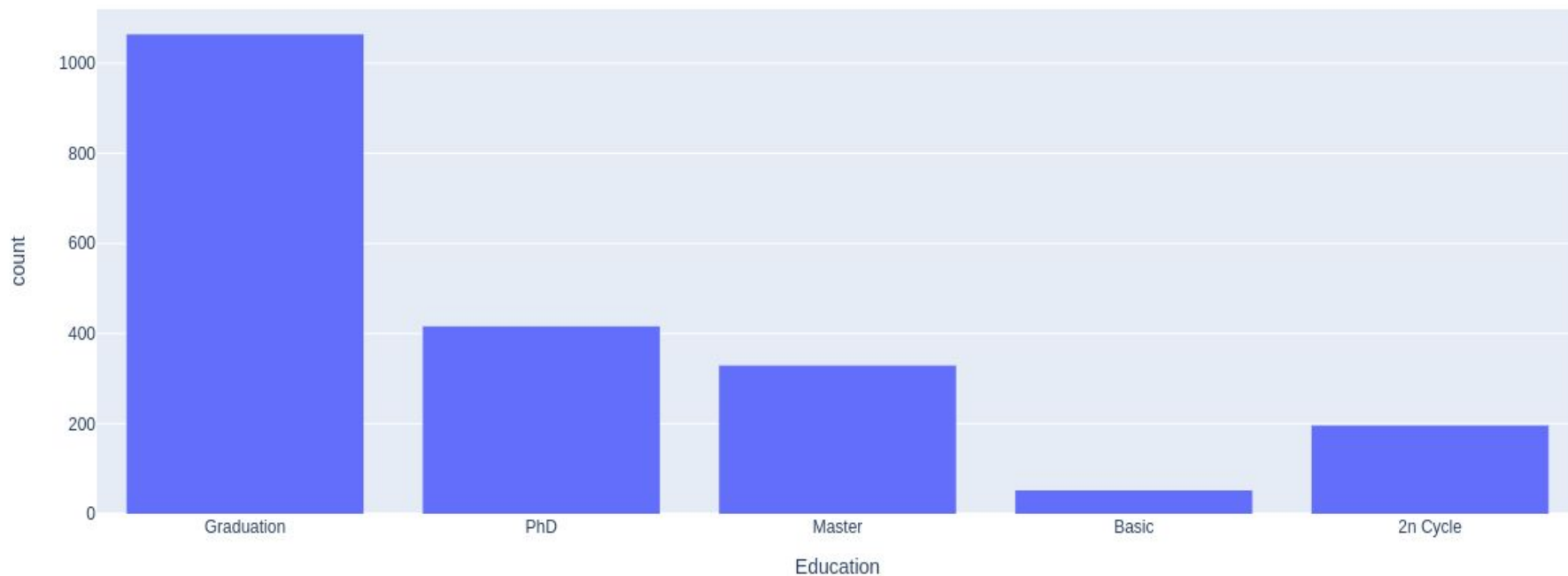


Data Advanced Analytics Test

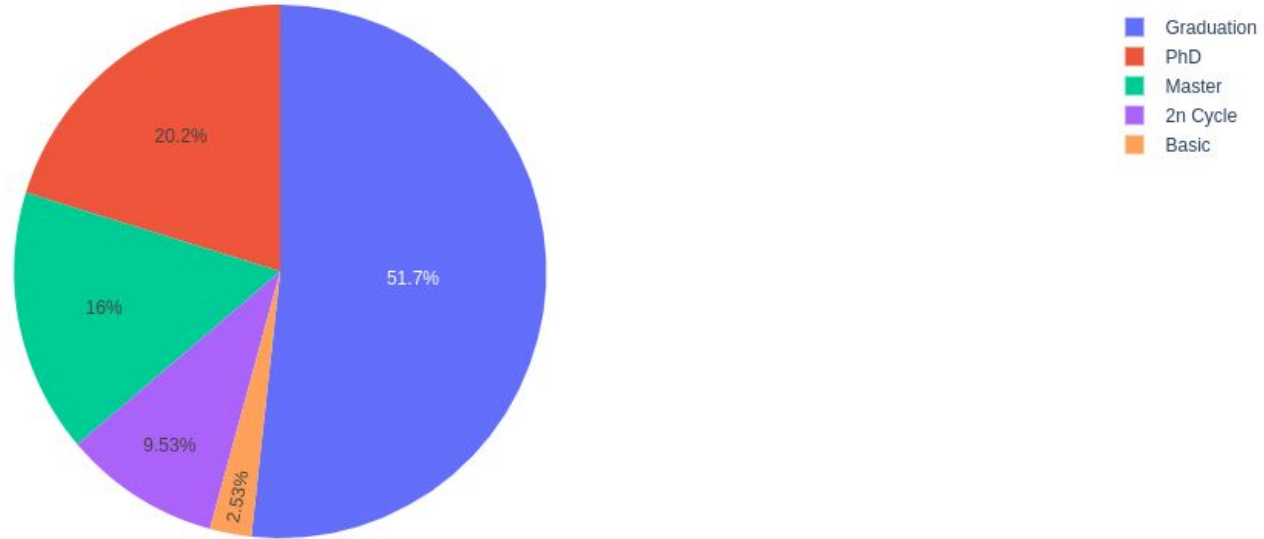
Social Analysis

Total customers by graduation



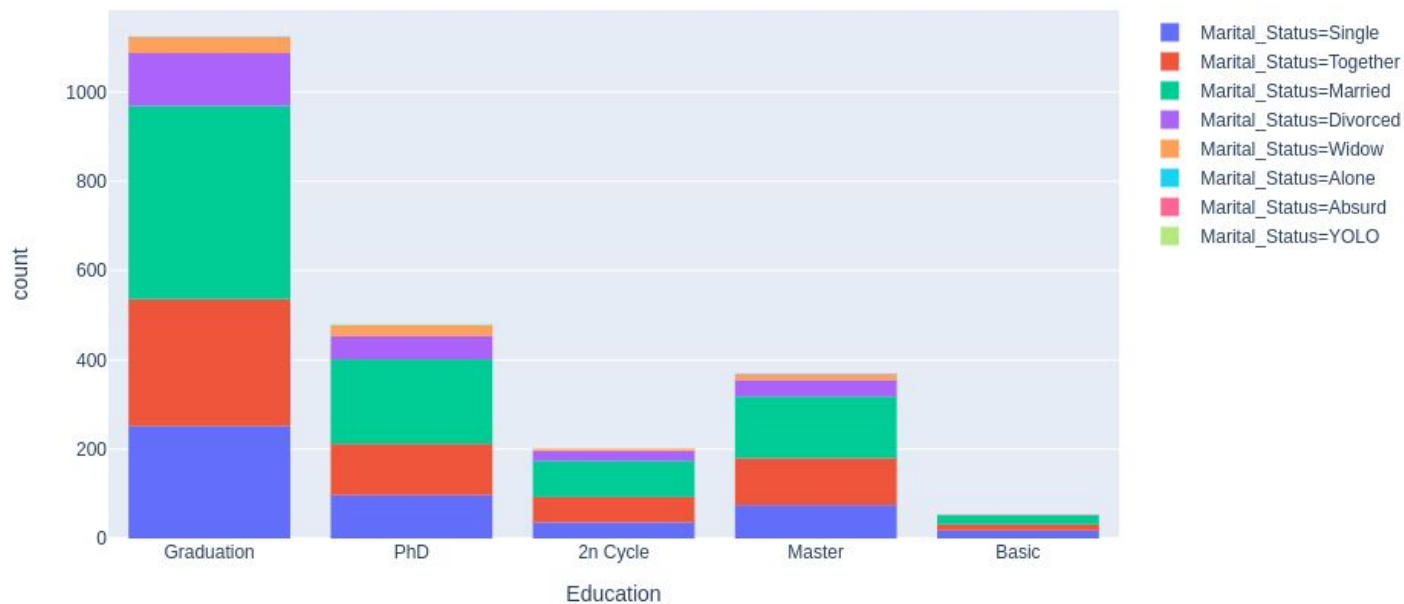
Social Analysis

Total customers% by graduation

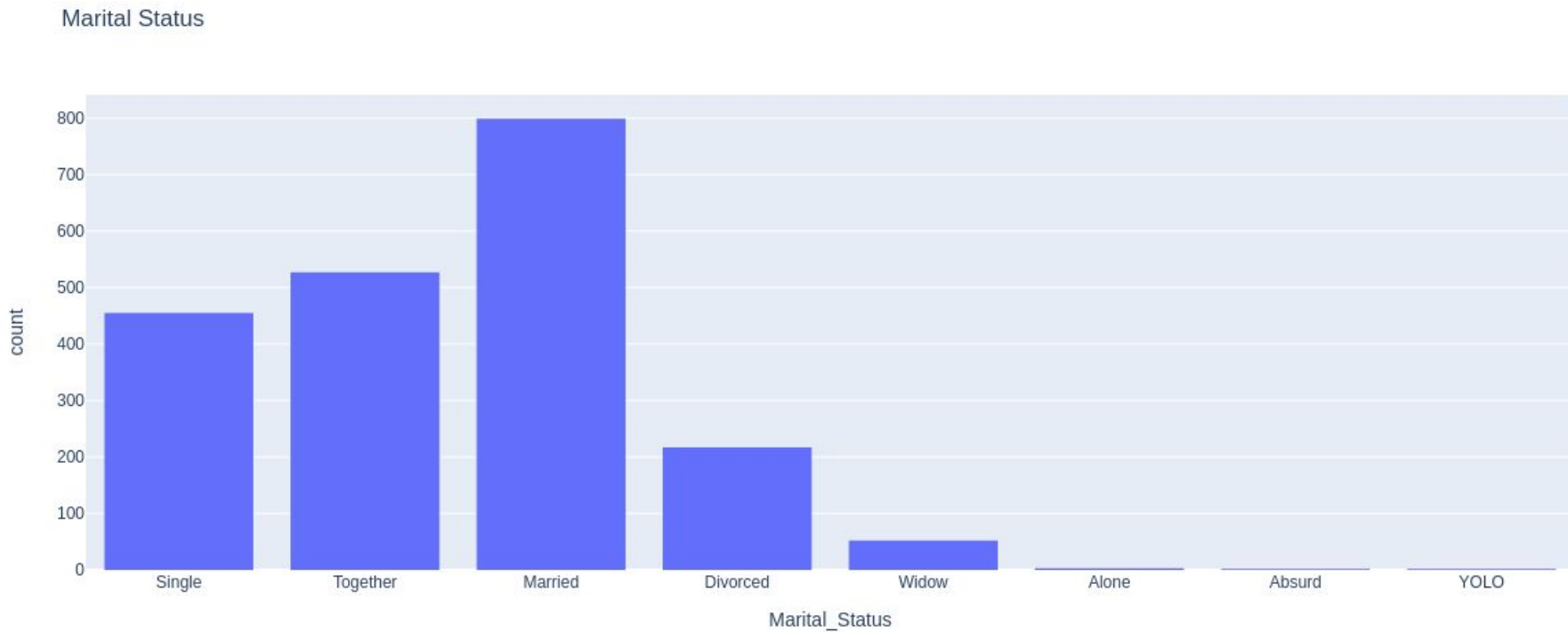


Social Analysis

Education per Marital Status

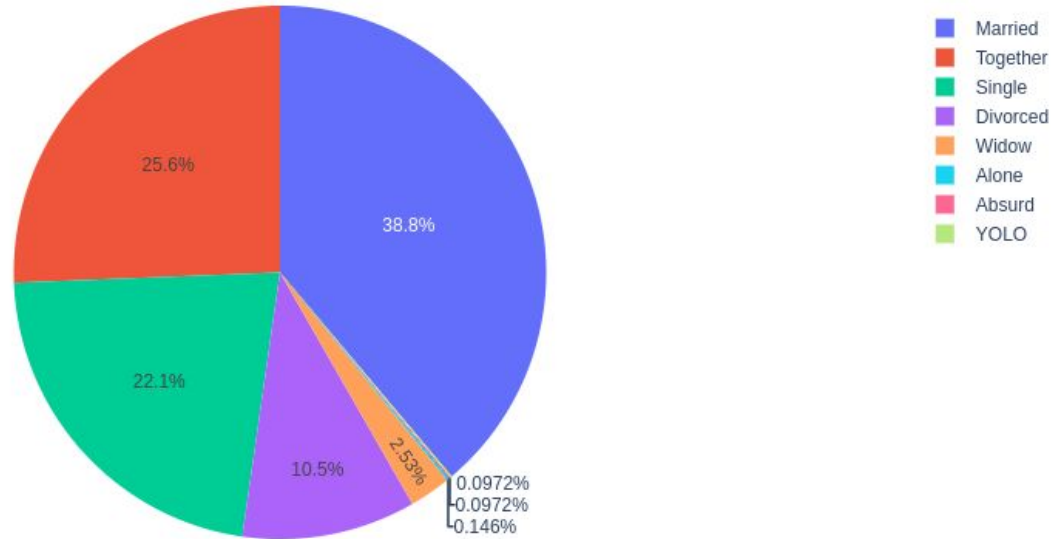


Social Analysis



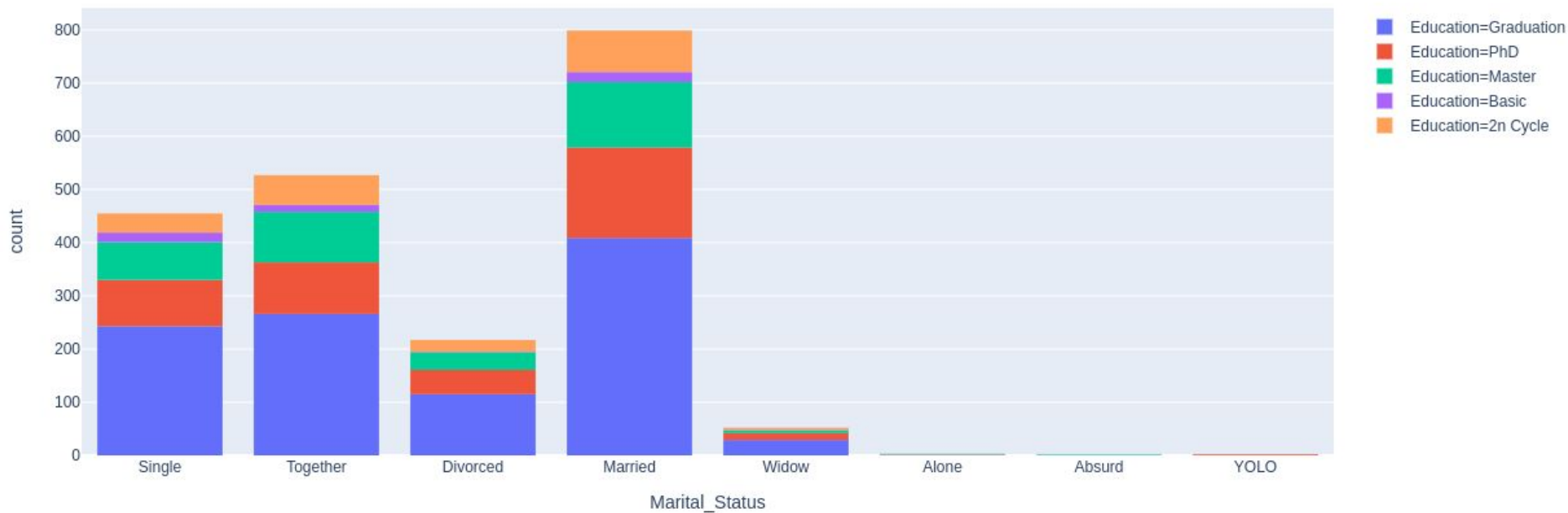
Social Analysis

Marital Status%



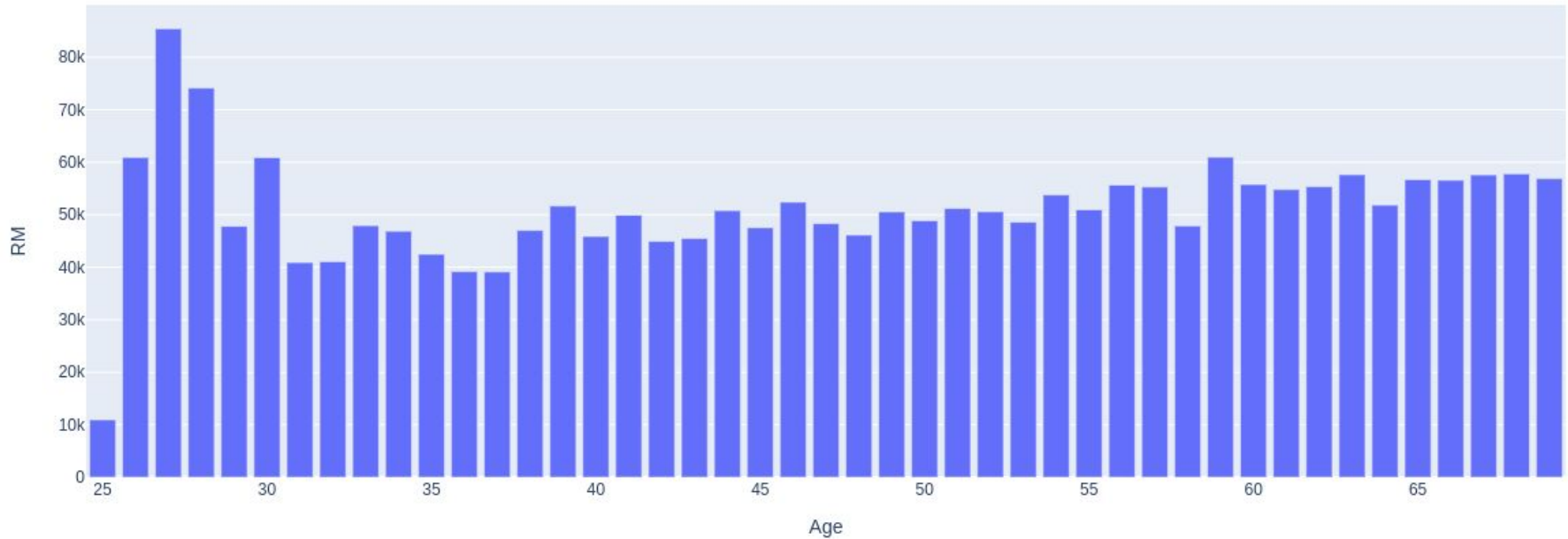
Social Analysis

Marital Status per Education



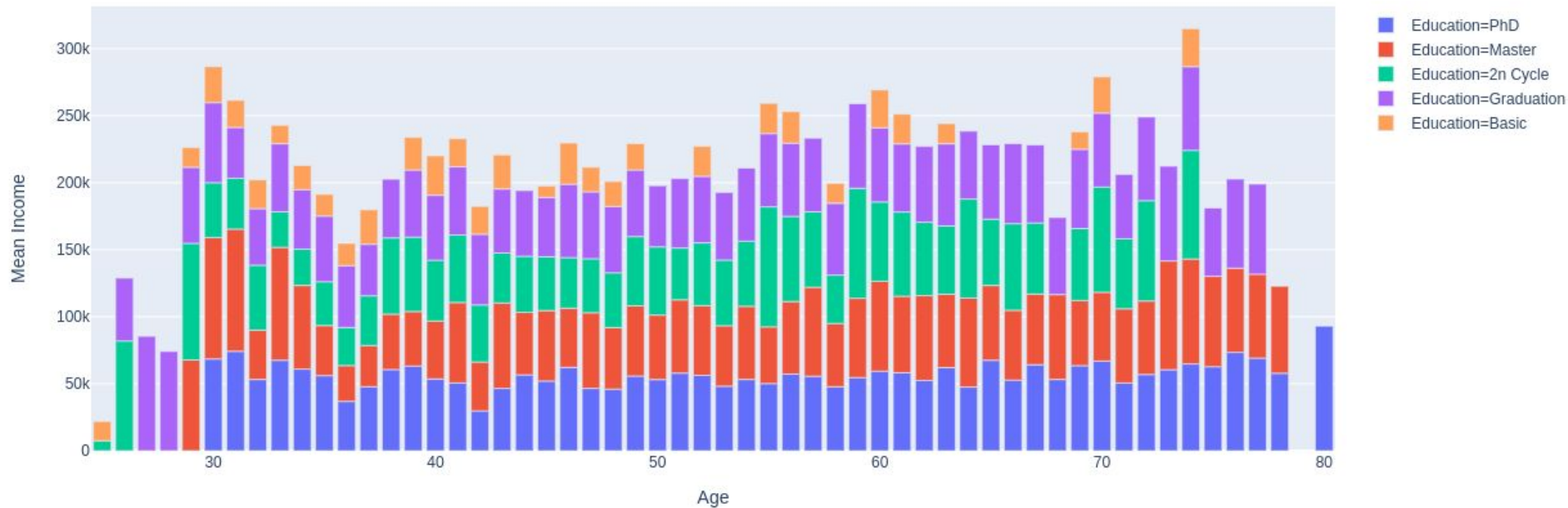
Social Analysis

Mean Income per Age

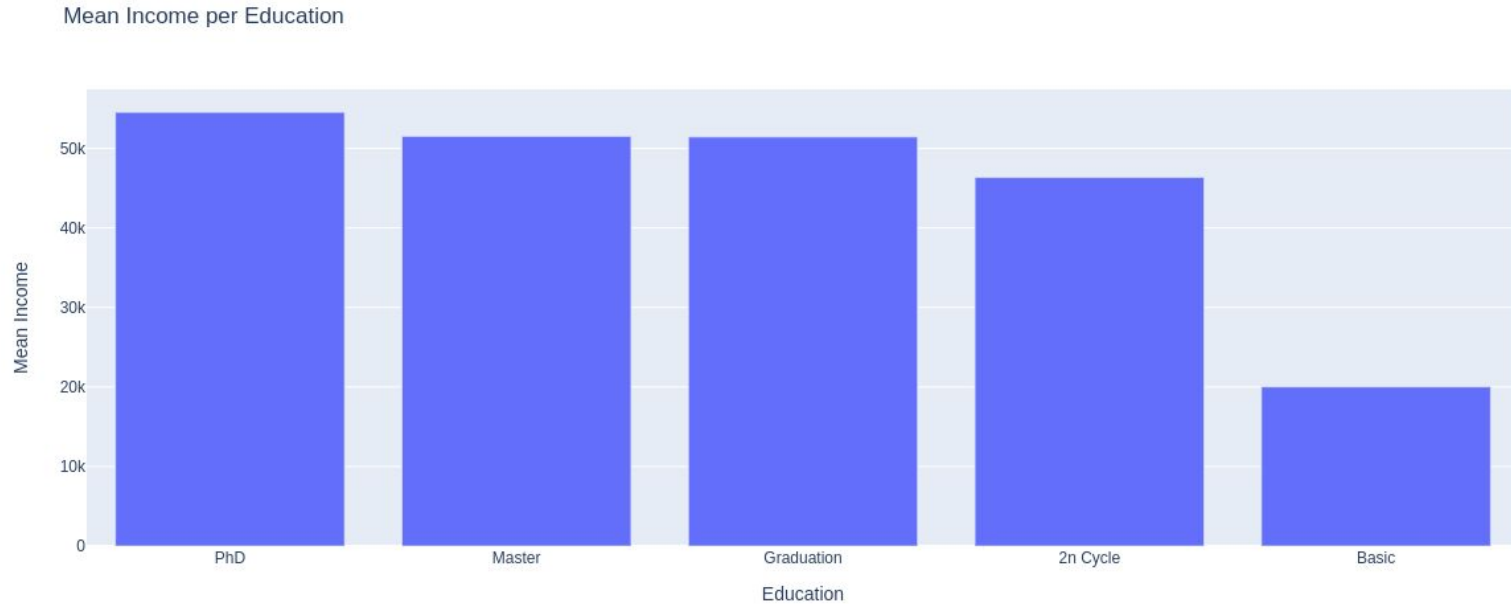


Social Analysis

Mean Income per Education



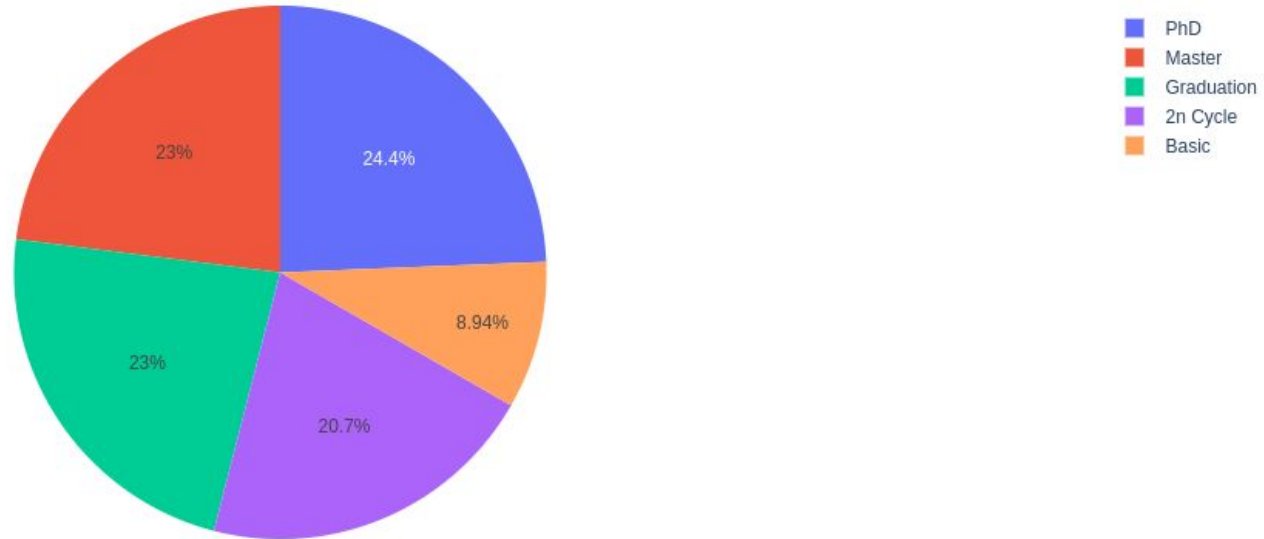
Social Analysis



The average income does not have much variation, only the basic one (2.5% of the dataset)

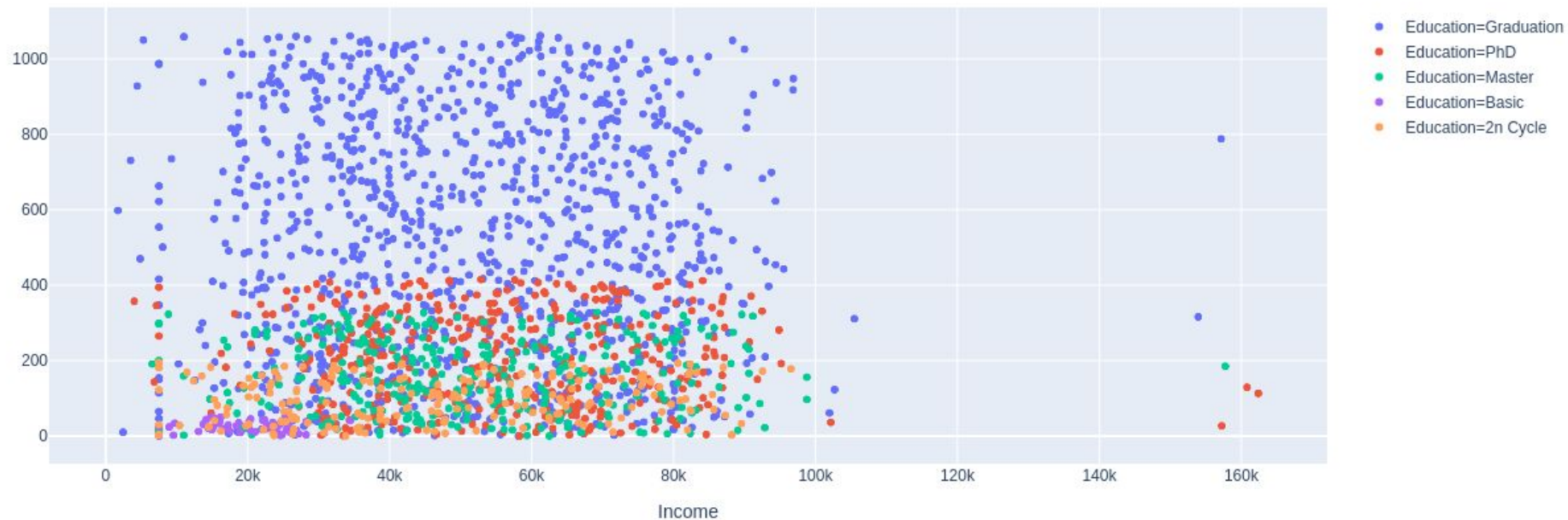
Social Analysis

Mean Income per Education%



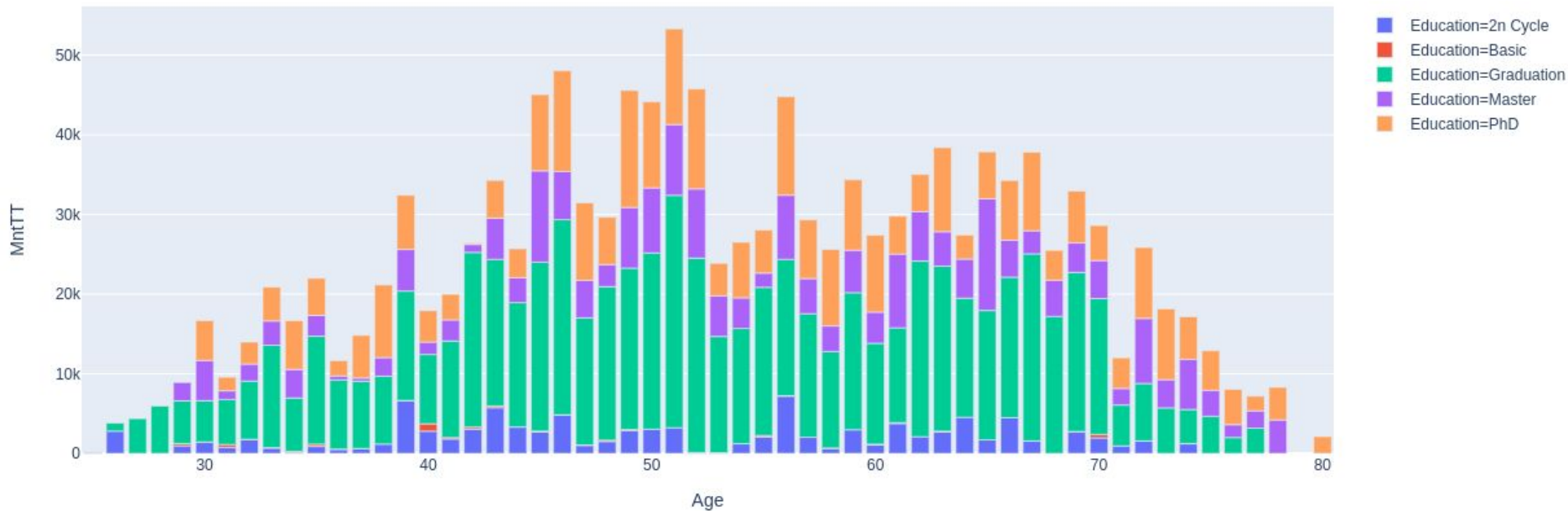
Social Analysis

Income per Education



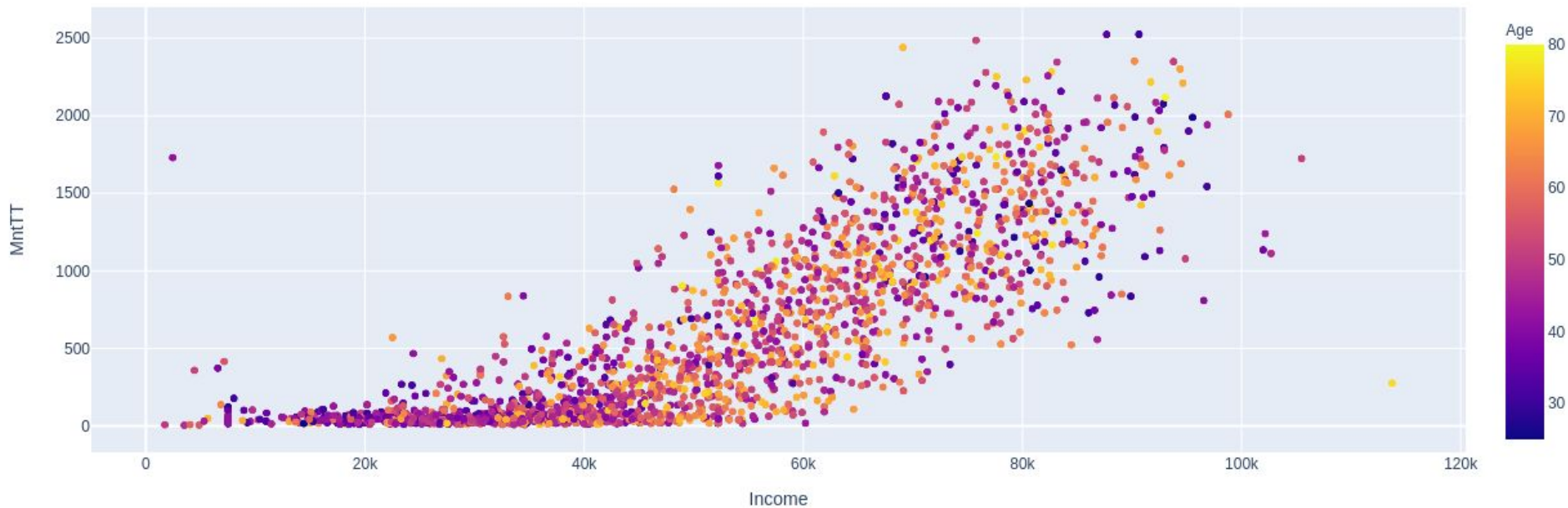
Purchase Analysis

Total Spent per Education



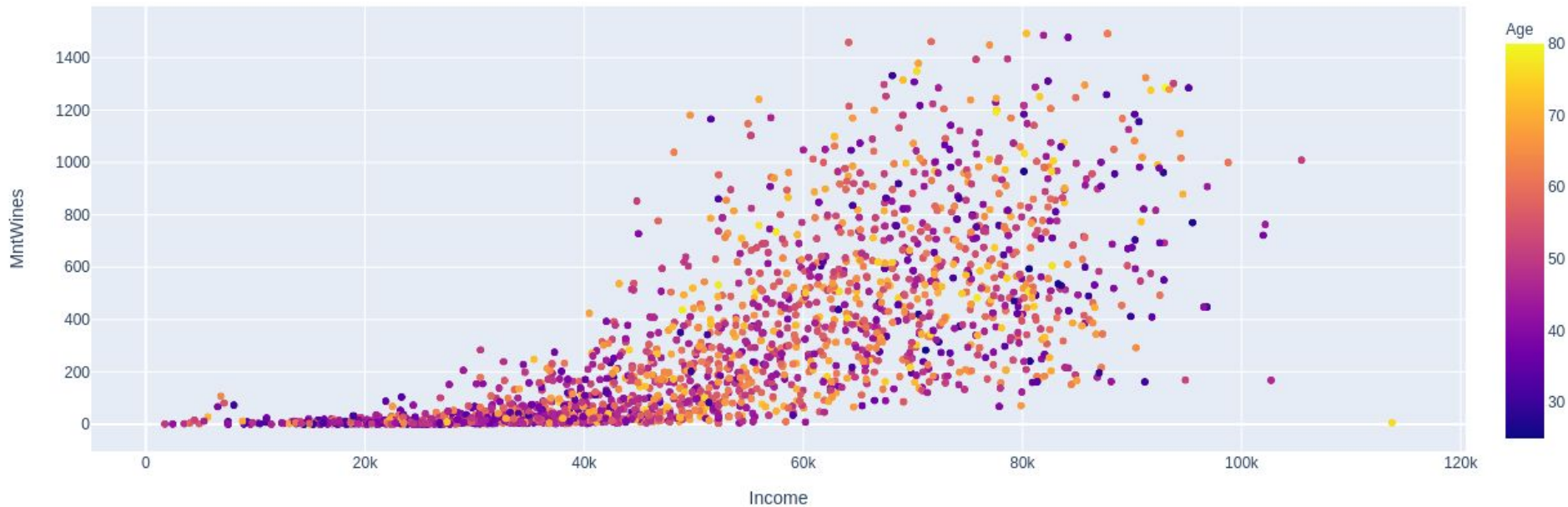
Purchase Analysis

Total Spent per Income



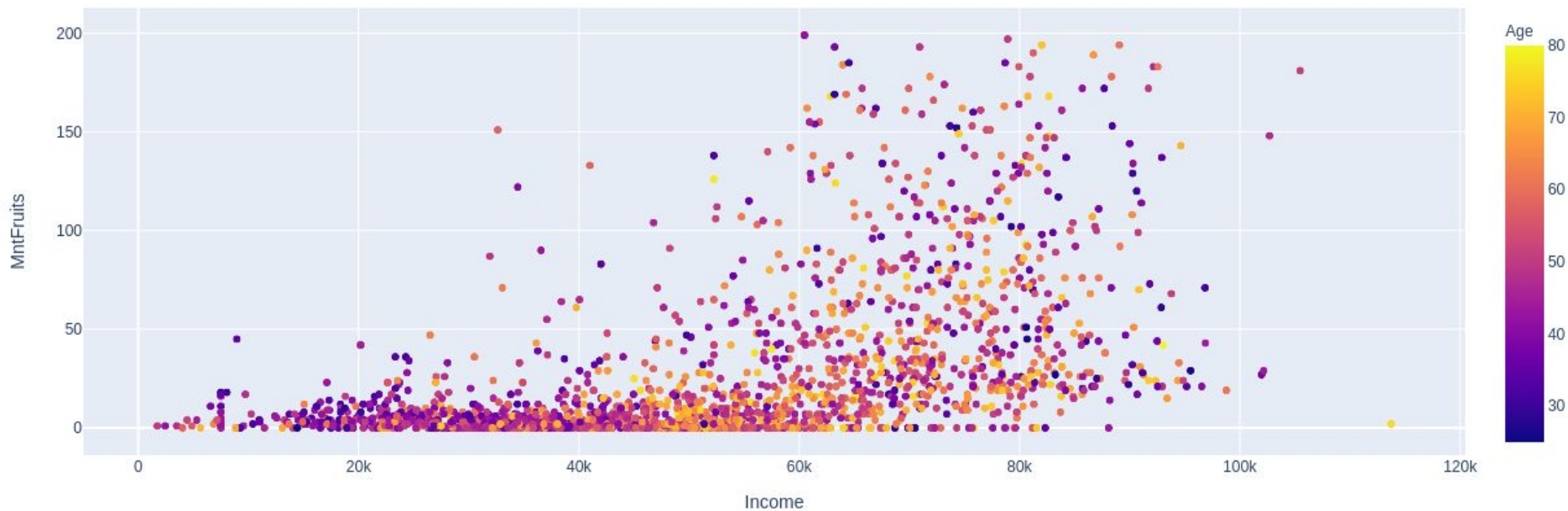
Purchase Analysis

Total Spent with Wine per Income



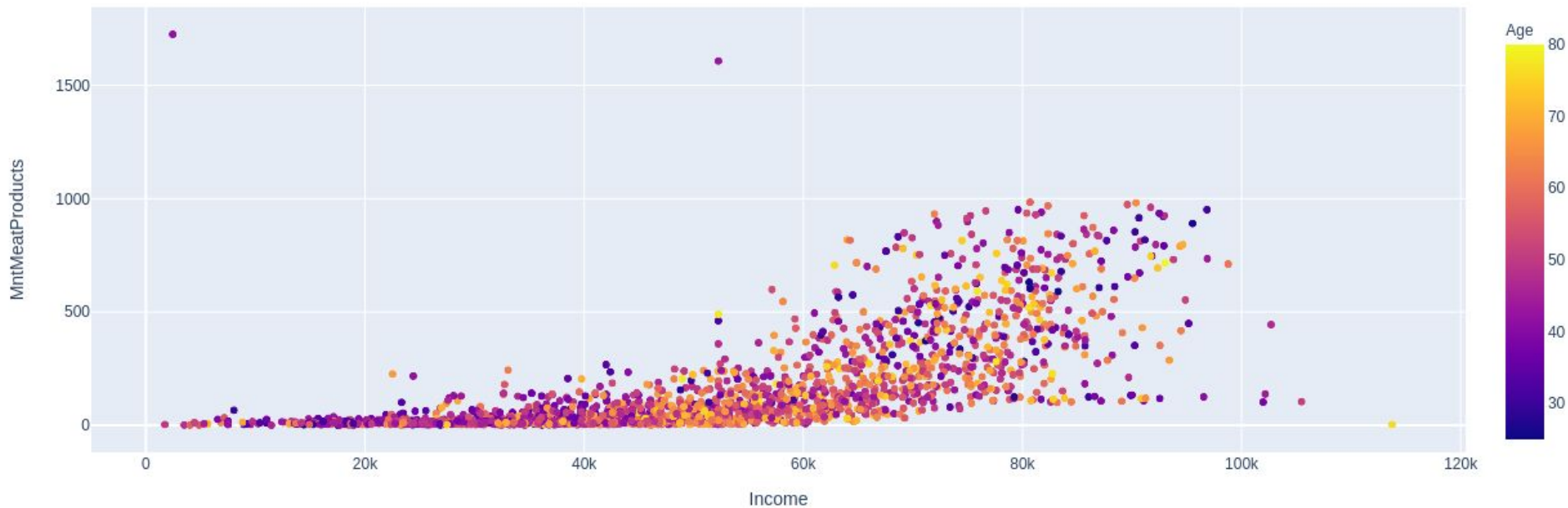
Purchase Analysis

Total Spent with Fruits per Income



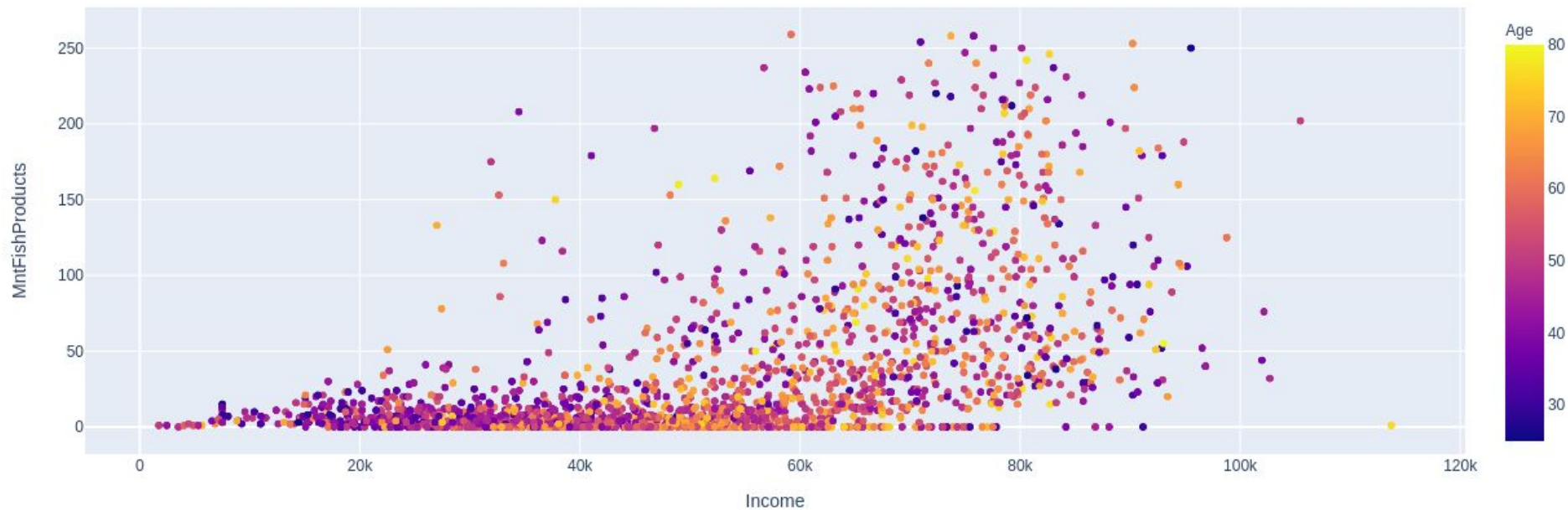
Purchase Analysis

Total Spent with Meat Products per Income



Purchase Analysis

Total Spent with Fish Products



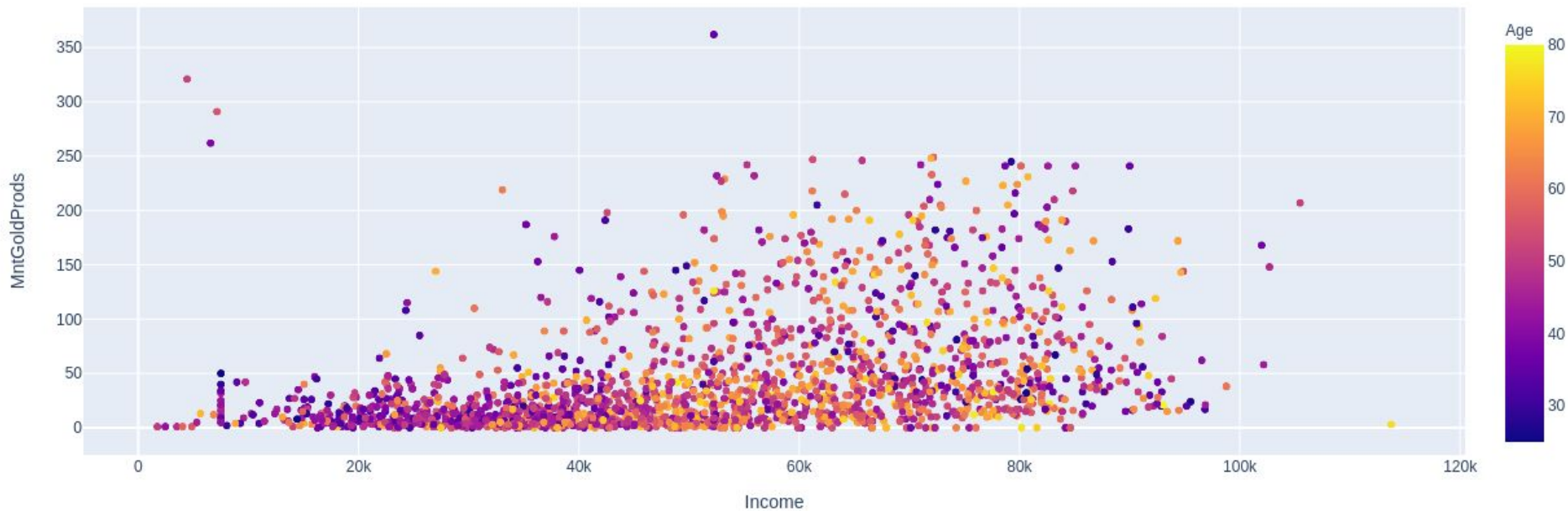
Purchase Analysis

Total Spent with Sweet Products per Income



Purchase Analysis

Total Spent with Gold Products per Income



Purchase Analysis

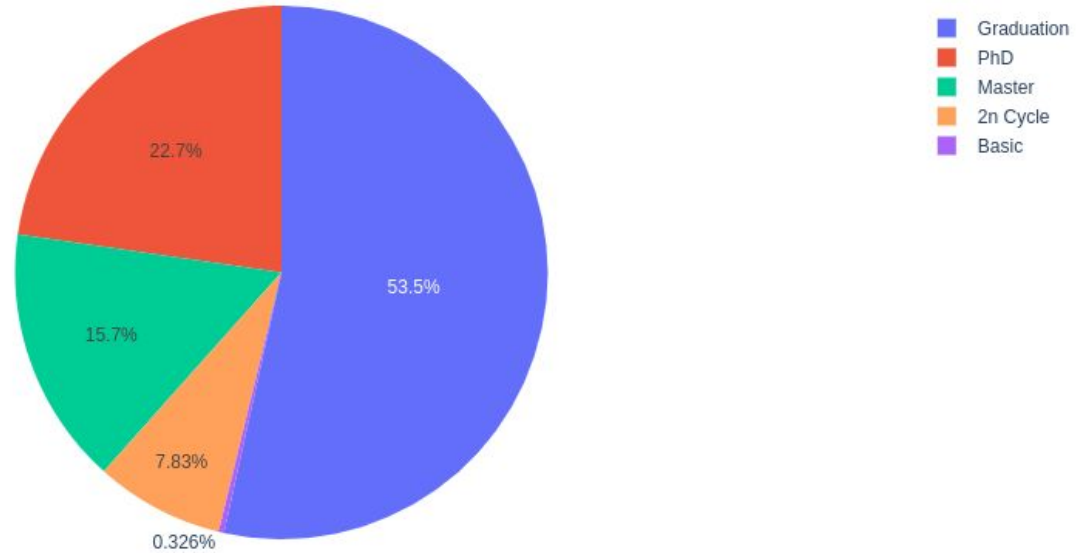
Pearson Correlation of Income per Category of Products

*MnTT = Total Spent



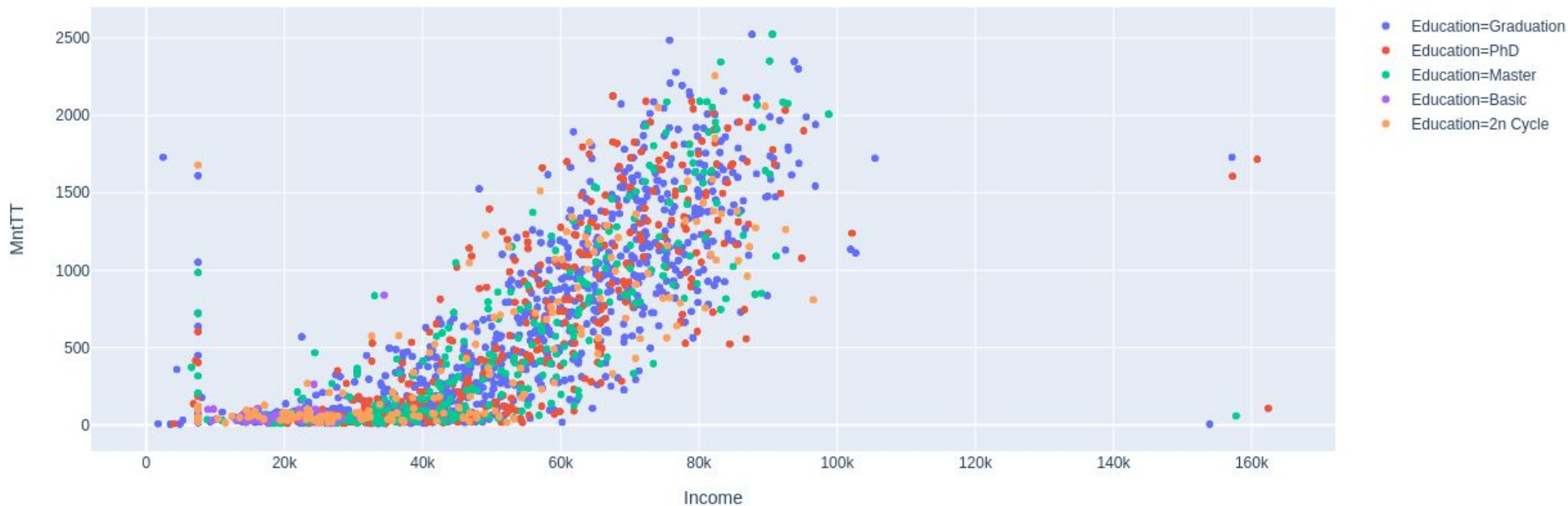
Purchase Analysis

Total Spend% per Education



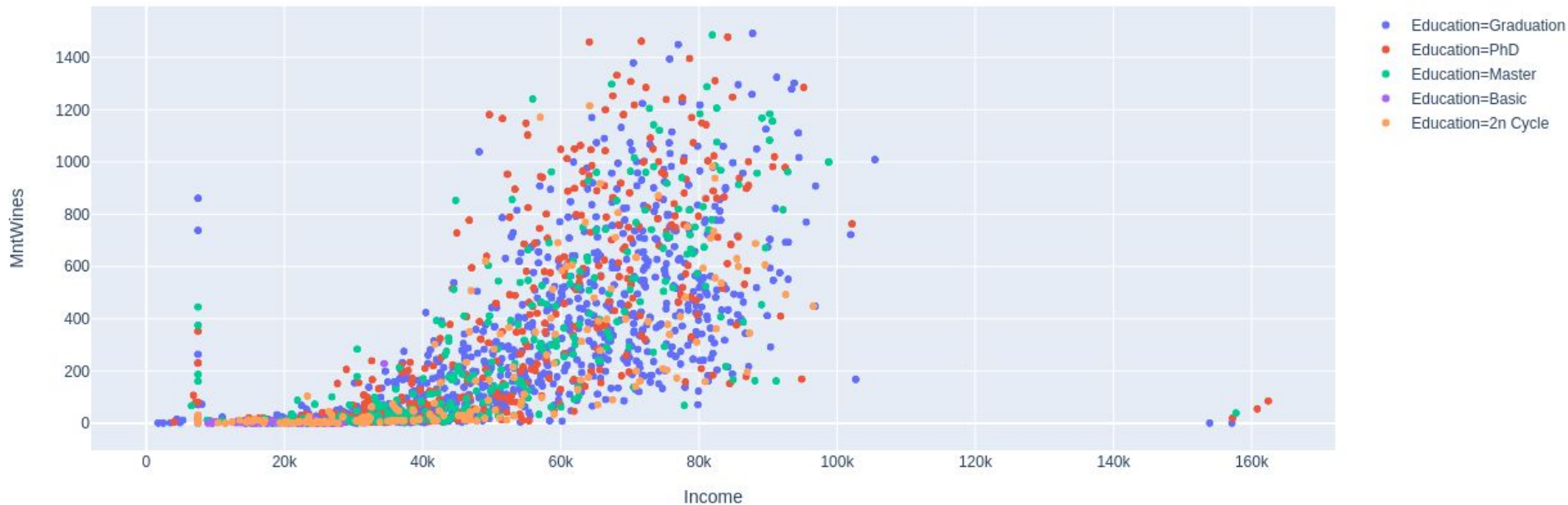
Purchase Analysis

Total Spend per Education



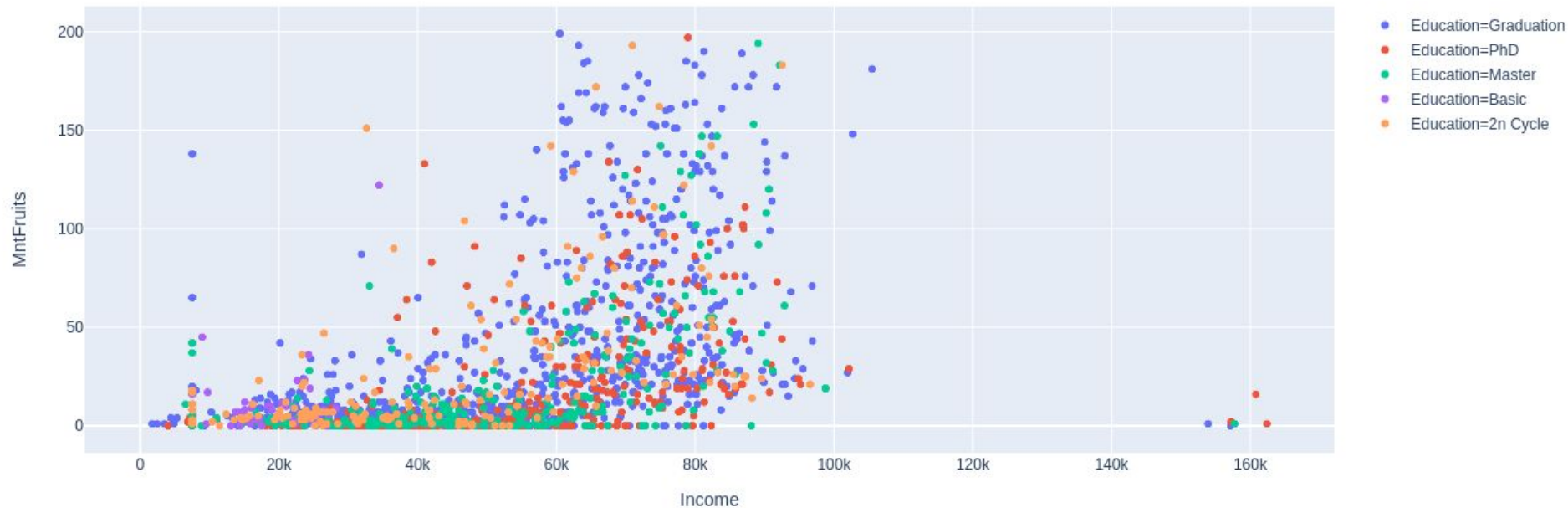
Purchase Analysis

Total Spend with Wine per Education



Purchase Analysis

Total Spend with Fruits per Education



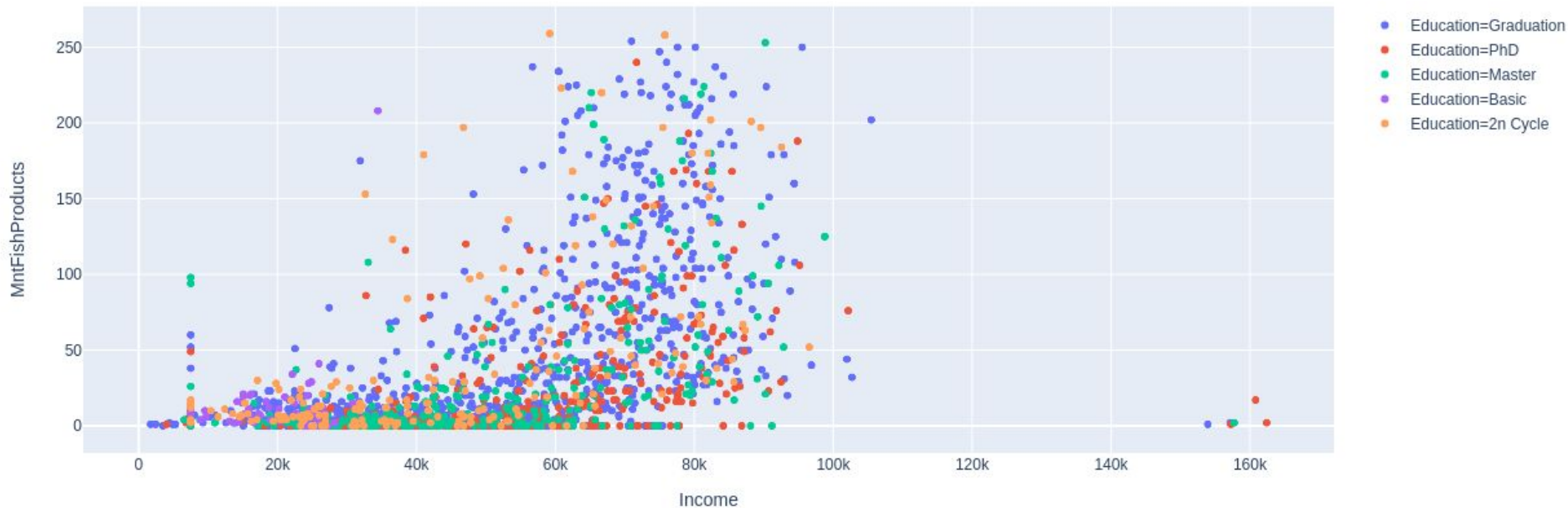
Purchase Analysis

Total Spend with Meat per Education



Purchase Analysis

Total Spend with Fish per Education



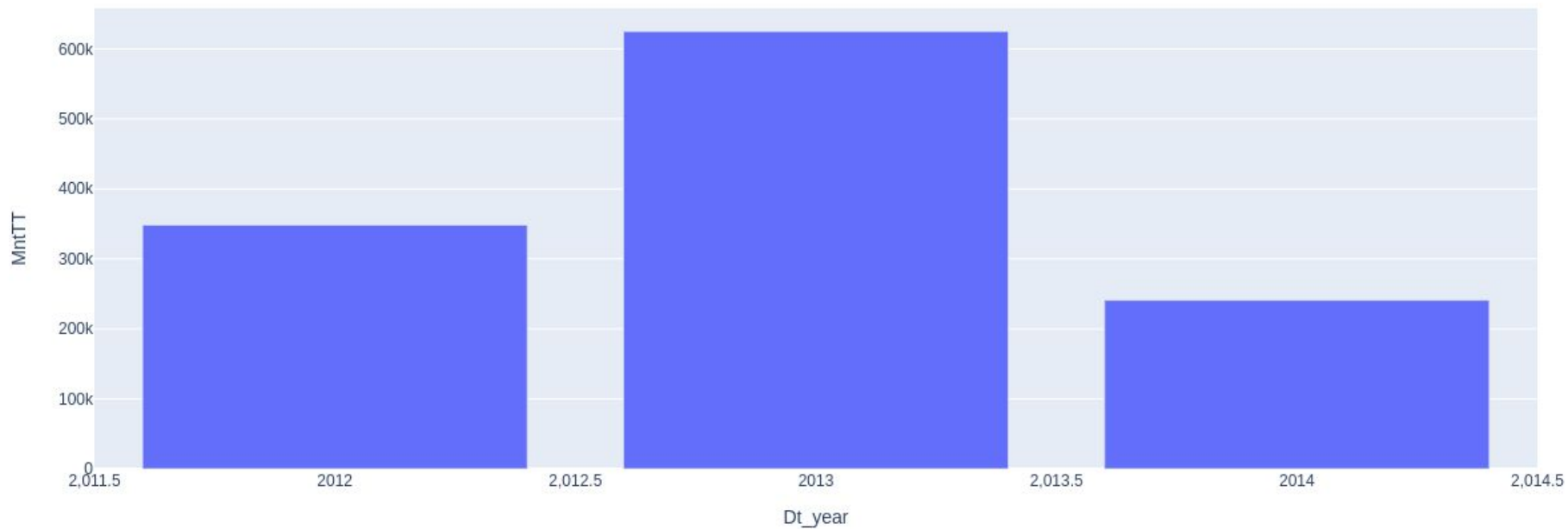
Purchase Analysis

Total Spend with Sweet Products per Education



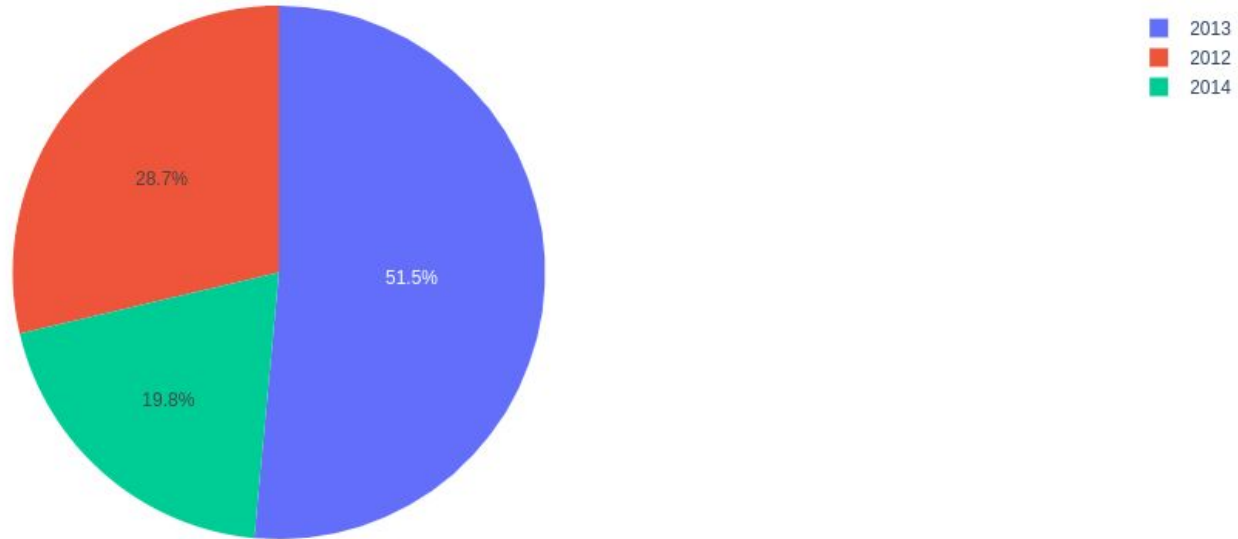
Frequency

Total Spend per Year



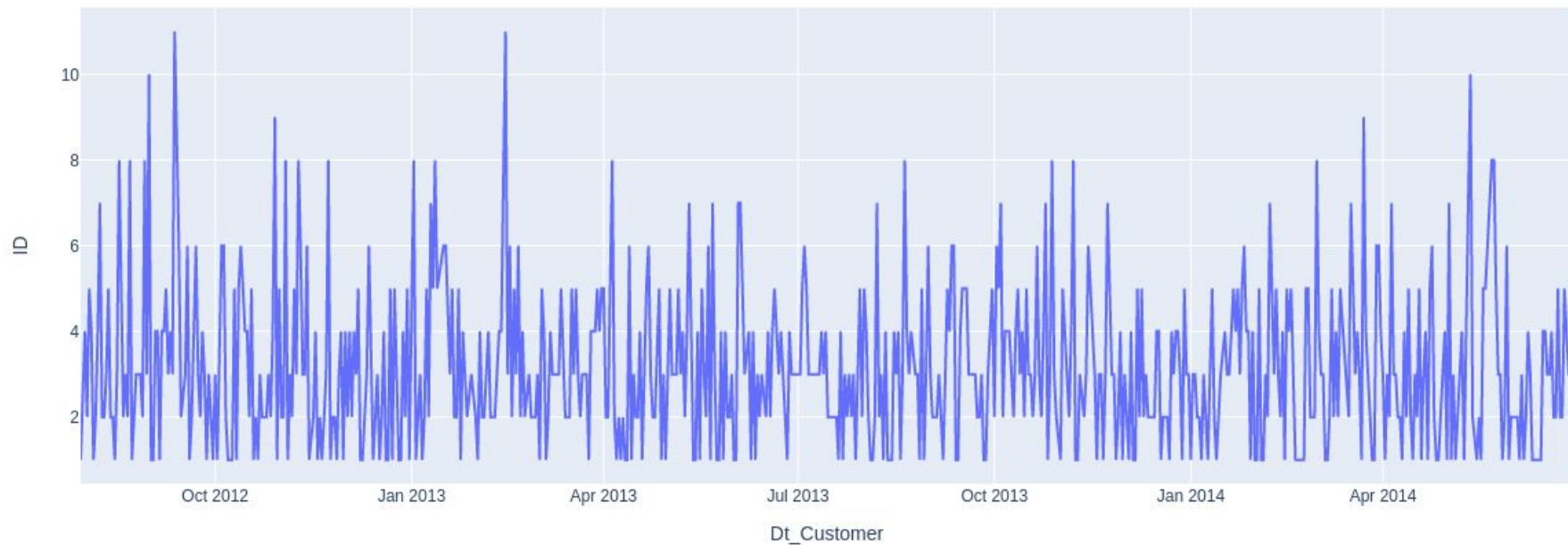
Frequency

Total Spend% per Year



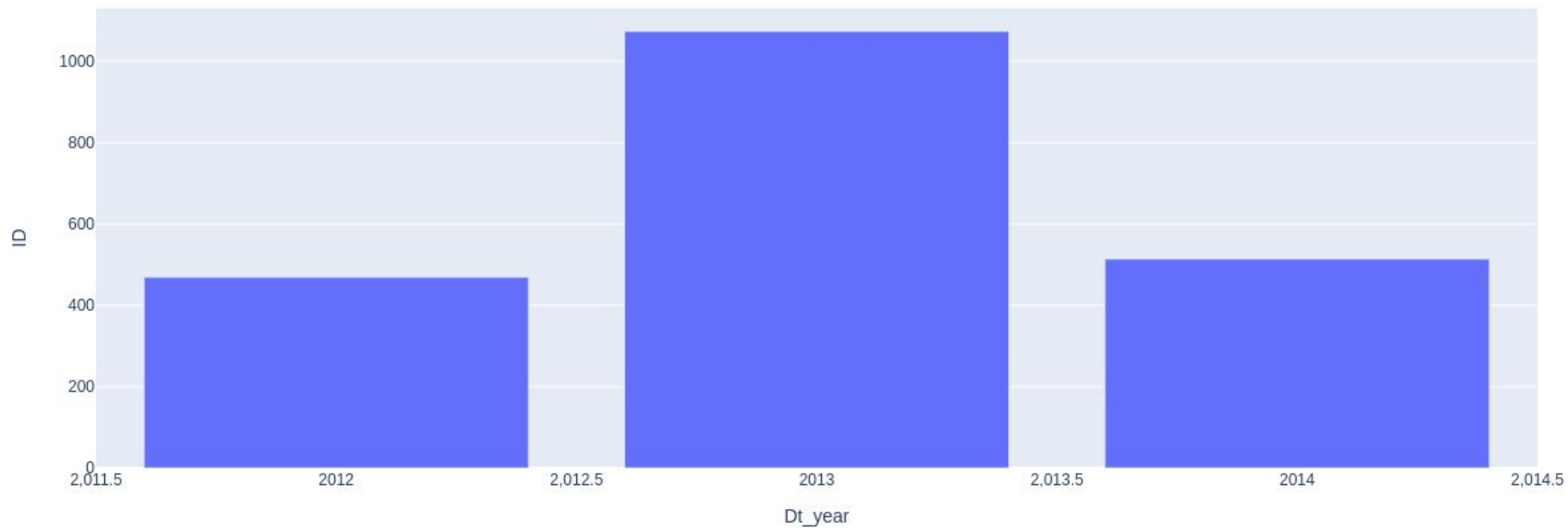
Frequency

Total of Clients per date



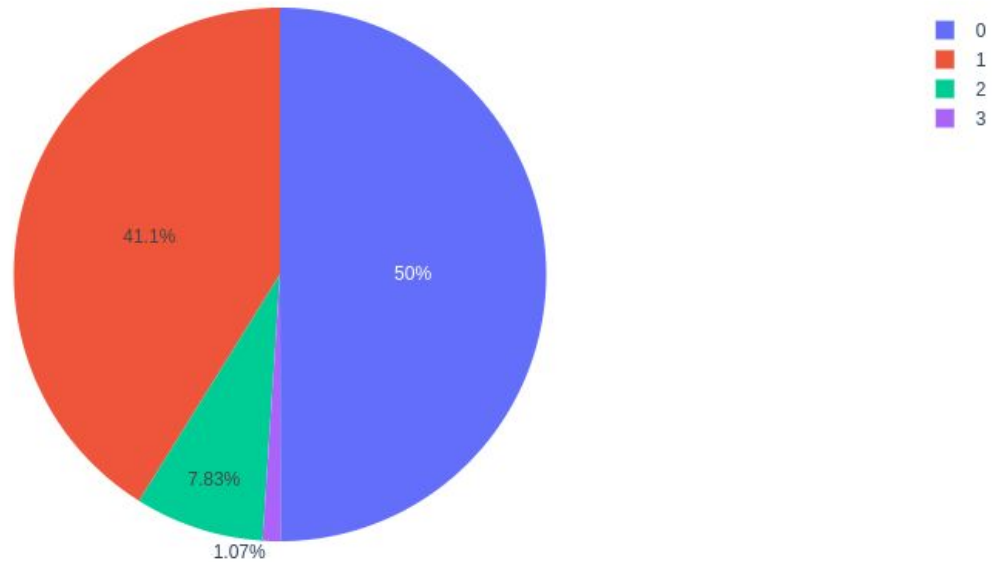
Frequency

Total of Customers per Year



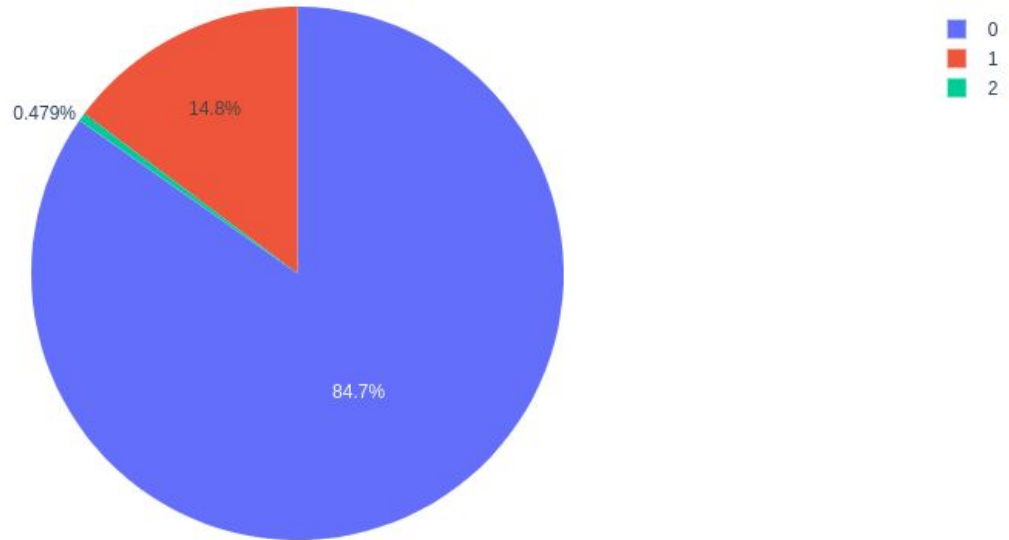
Dependents

Total Spend of Customers with Dependents%



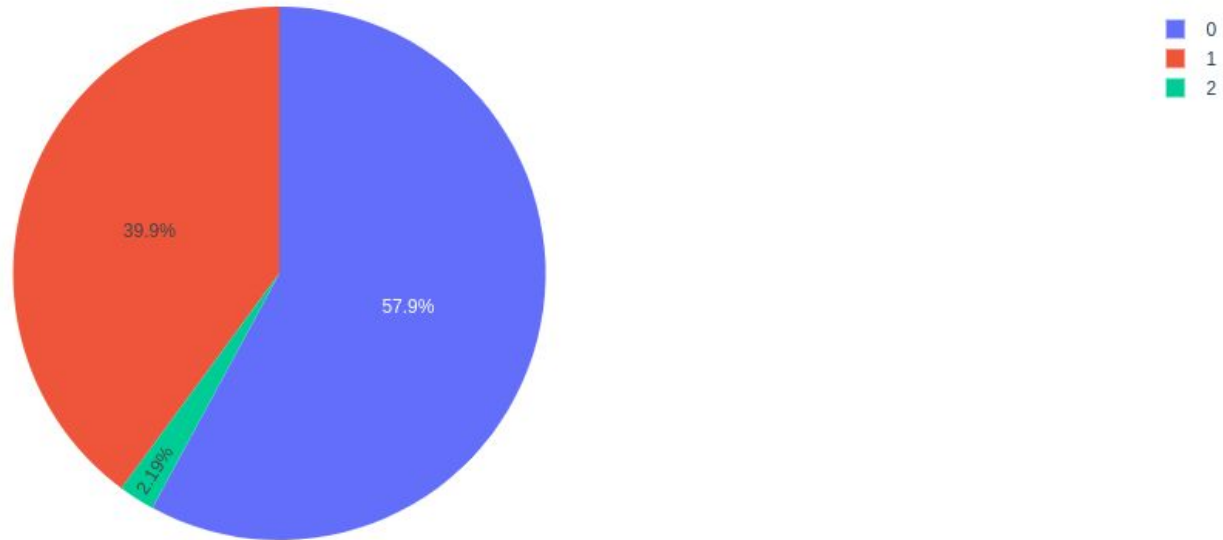
Dependents

Total Spend of Customers with Kids%



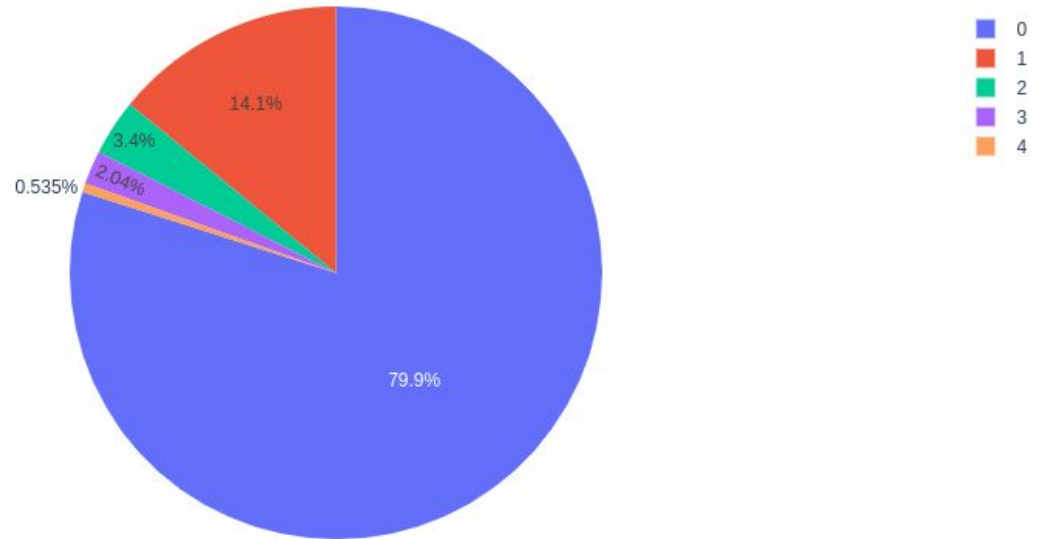
Dependents

Total Spend of Customers with Teens%



Campaigns

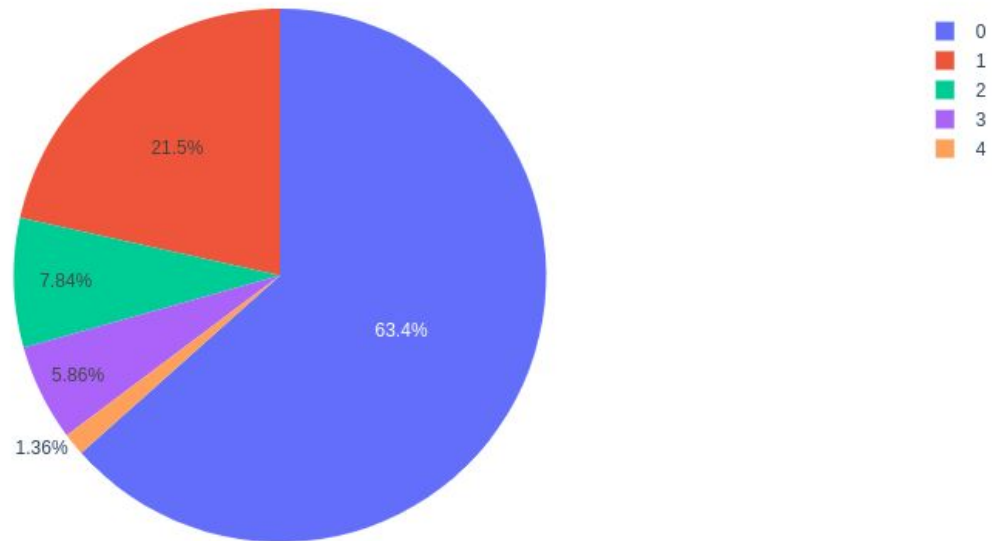
Total of Customers per Total of Campaigns Accepted



The overwhelming majority of customers buy without campaign (!)

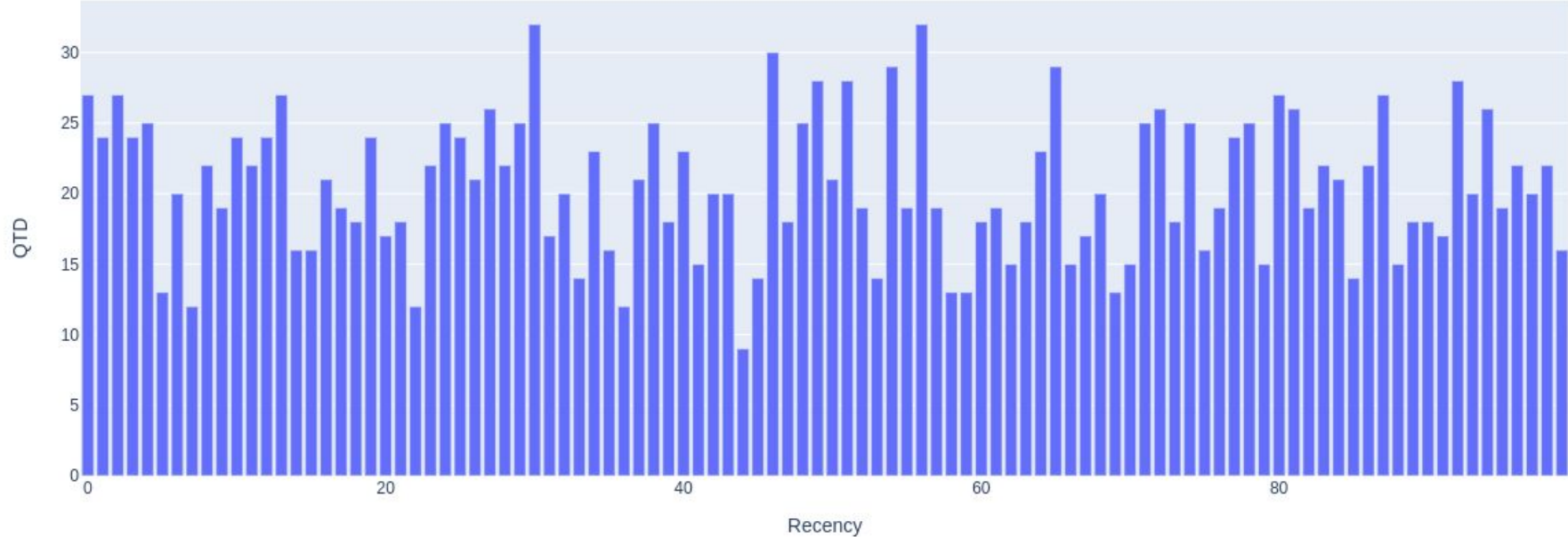
Campaigns

Total Spend per Number of Campaigns Accepted



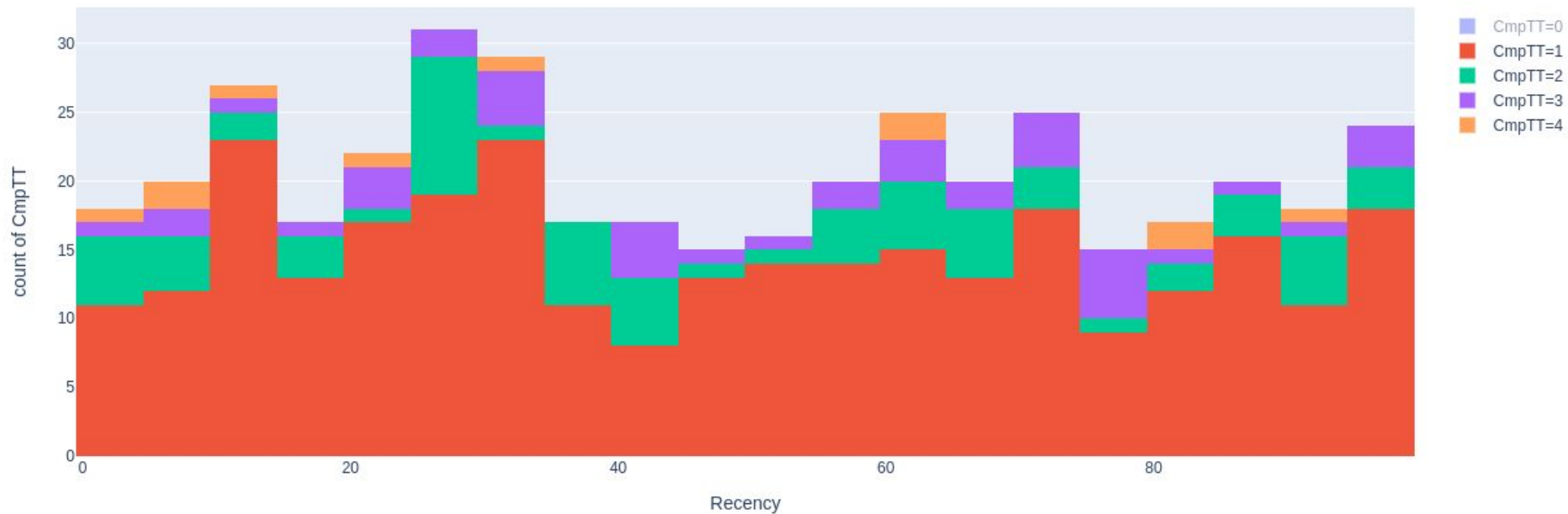
Recency

Total of Customers per Recency

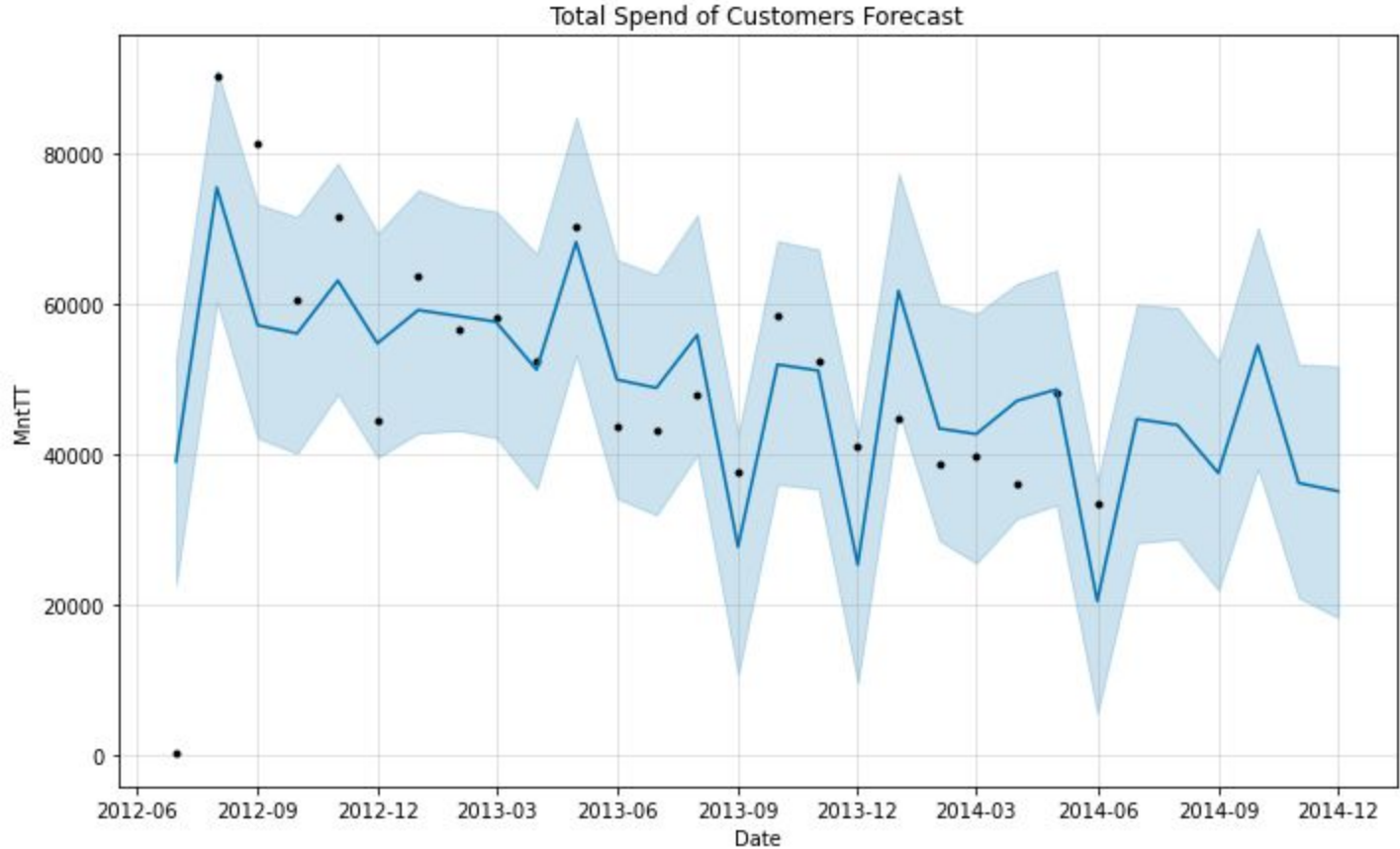


Recency

Total of Campaigns Accepted per Recency



Time Series Forecast



Six months forecast

| | ds | yhat |
|----|------------|--------------|
| 24 | 2014-07-01 | 44685.808586 |
| 25 | 2014-08-01 | 43874.656448 |
| 26 | 2014-09-01 | 37494.388518 |
| 27 | 2014-10-01 | 54475.231201 |
| 28 | 2014-11-01 | 36183.497223 |
| 29 | 2014-12-01 | 35083.189117 |

Classification

3 models were chosen for training, testing and validation:

Random Forest Classifier: It creates decision trees at random, where each tree will be used in the choice of the final result.

Logistic Regression: aims to produce, from a set of observations, a model that allows the prediction (probability) of values taken by a categorical variable.

XGBClassifier: It is an algorithm based on Decision Trees (Gradient Boosting). means that the algorithm uses the Gradient Descent algorithm to minimize the loss (loss) while new models are being added.

Classification

The models had the following results:

| Metric | Random Forest | Logistic Regression | XGBClassifier |
|-----------|---------------|---------------------|---------------|
| Recall | 73,33% | 61,66% | 36,66% |
| Precision | 37,60% | 25,69% | 62,85% |
| Accuracy | 80,04% | 70,85% | 88,56% |

Presentation by Alysson Guimarães

Contact me

[Linkedin](#)

[Github](#)