**Customer Personality Analysis –Clustering Algorithm**

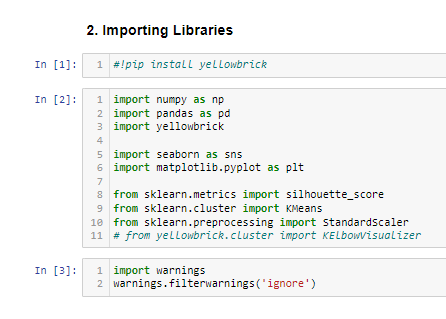
**Problem Statement:**

Customer buying behavior Analysis is a detailed analysis of a company’s ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers.

This analysis helps a business to modify its product based on its target customers from different types of customer segments. For example, instead of spending money to market a new product to every customer in the company’s database, a company can analyze which customer segment is most likely to buy the product and then market the product only in that segment.

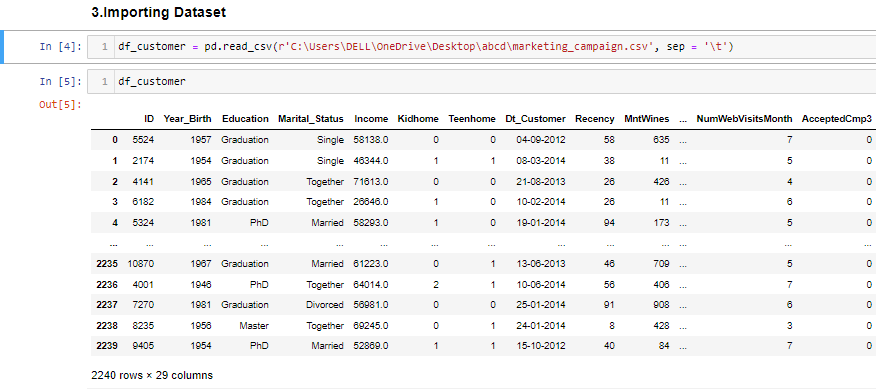
**1.Importing required Libraries:**

Importing required libraries to load and build model.



**2.Loading Dataset:**

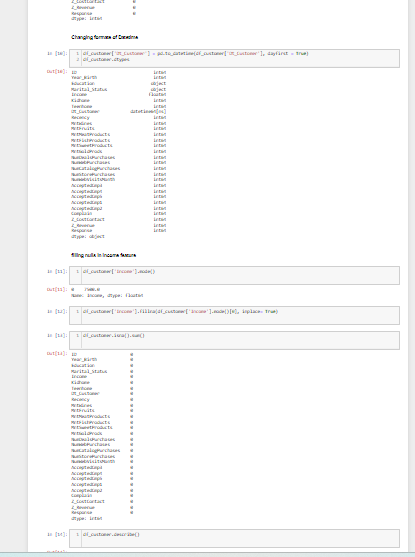
Using Pandas Lib loading marketing\_campaign.csv file.



### **3. Data Exploration and Cleaning:**

In this ,

* Checking top 10 rows with all columns.
* Find the shape and data types of dataset.
* Checking null values and filling with mode in the income column.

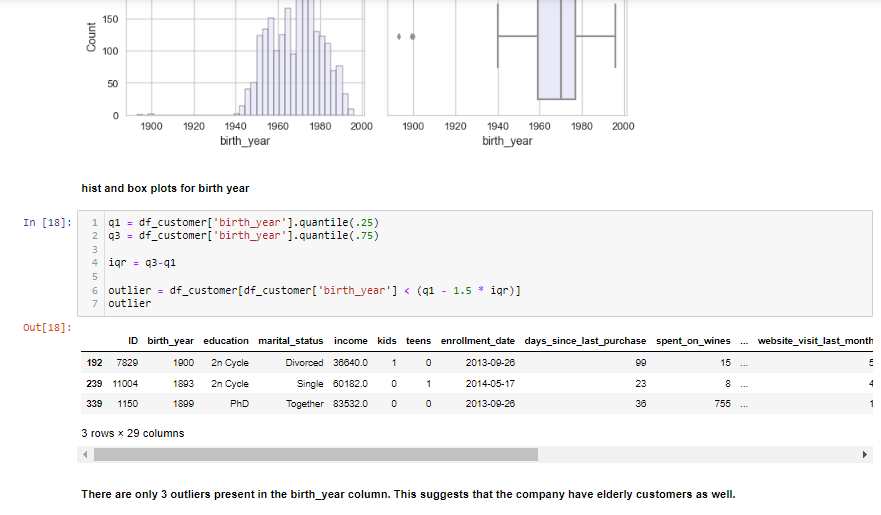


* Renaming the column for better understanding.



**4.** **Exploratory Data Analysis:**

Plotting histogram and box plots for every column in dataset for checking out layers in it.



#### Finding the total percentage of customers who made purchases across all six campaigns.

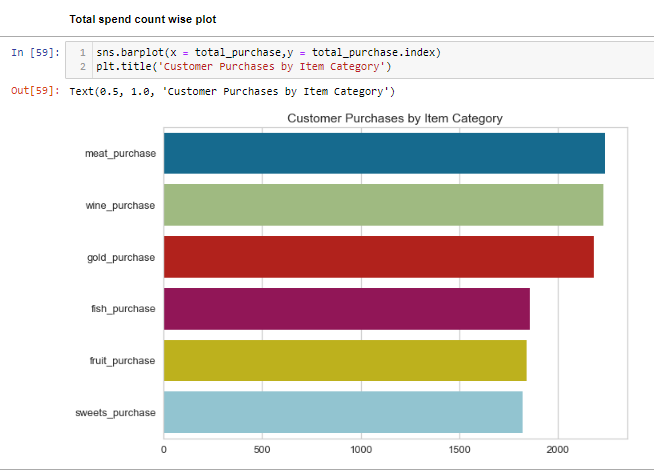


* Dropping the unaffected column for dimensionality reduction.
* Adding all wanted columns and create a new DF for counts of each column.



#### **Total spend count wise plot**

* Plotting a bar plot for all items purchases with count



* Plotting a bar plot for all items purchases with Sum

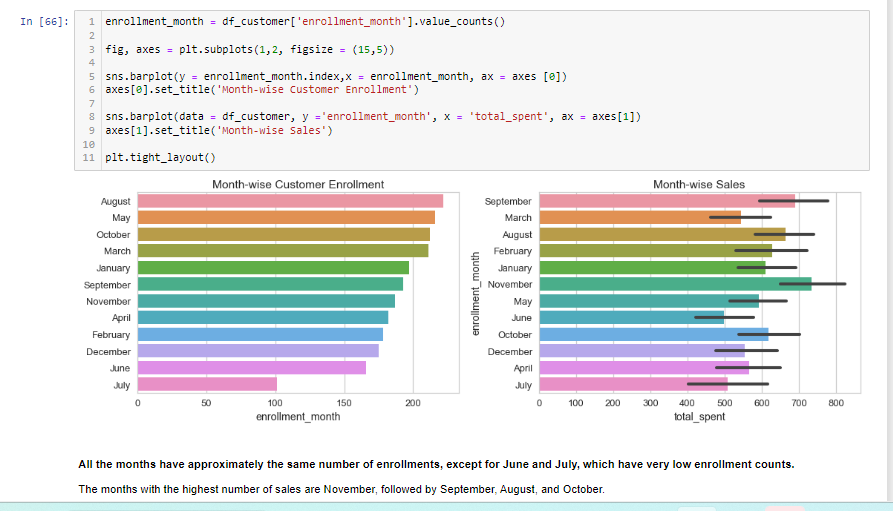


**5.** **Feature Transformation/Extraction**

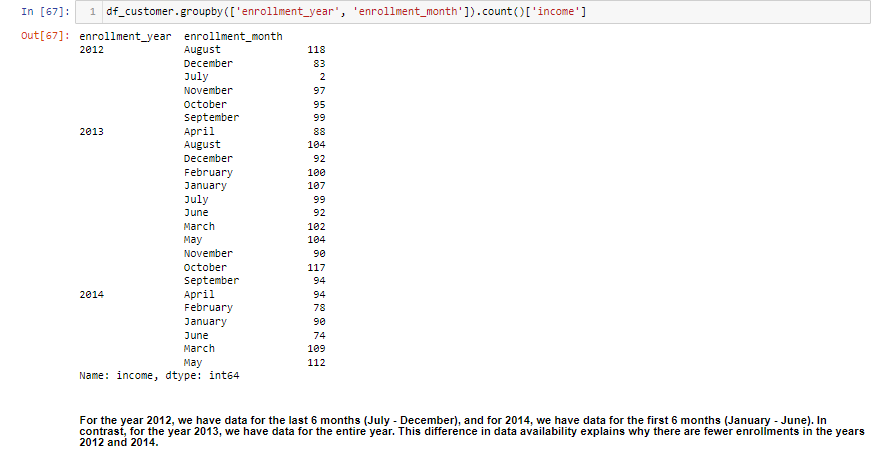
* Extracting the date-time column into separate month and year for better analysis.
* Creating a new children column by adding kids and teens columns.
* Combing all campaigns from (Campaign 0-6) and creating a new campaign purchase column.



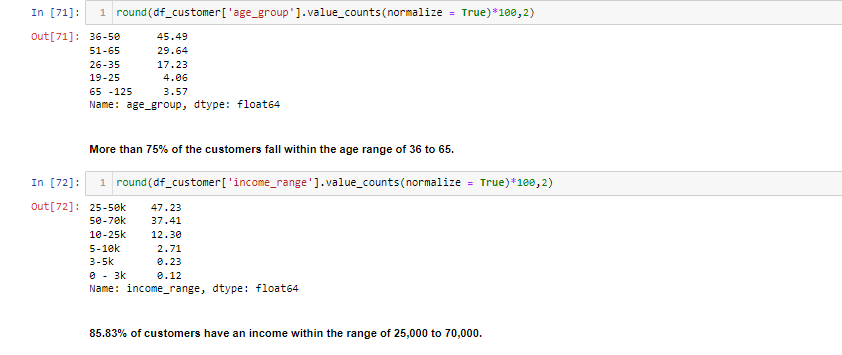
* Creating box plot for month wise total spent.



* Checking the earning of income, monthly and yearly wise.



* Finding the round of percentages, age averages and salary ranges.



#### Dropping the 'enrollment date' column, we have already created two new variables from it, which will be used in our analysis. Create a new Dataset name (after drop).

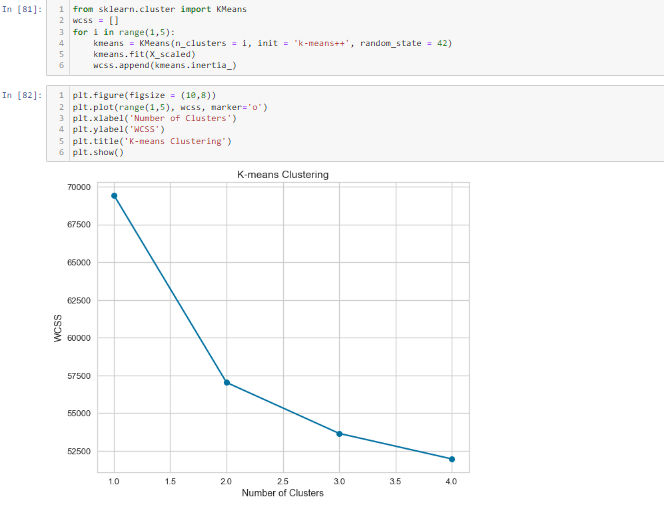
**6.** **Label Encoding and Standardization.**

* Assigning labels for each and every month (January to December).
* Applied standard scalar () on finalized dataset is our X.

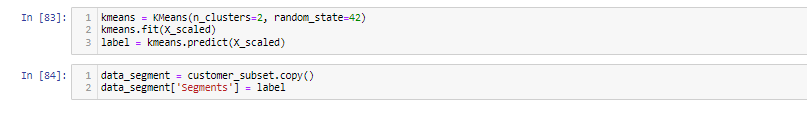


## **7.** **Model Building - K-Means**

* Finding the number of clusters and fit scaled-X in the clustering algorithm.
* Plotting a line chart to find number of clusters using Elbow method.



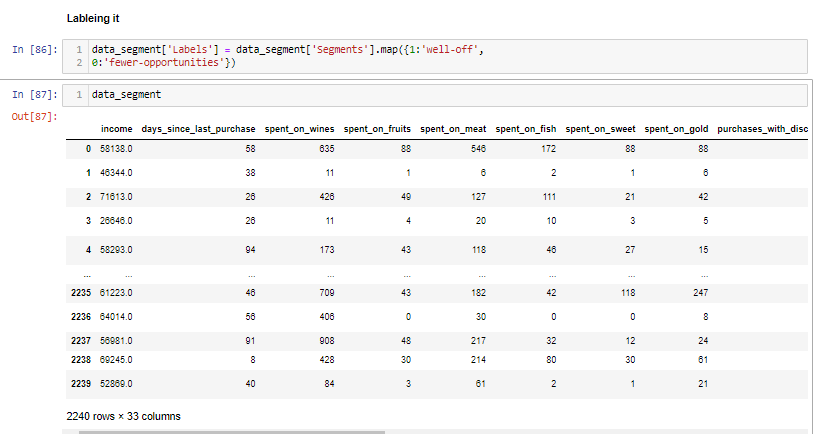
* Predicting the grouping of 2 clusters using K-Means assigned as labels.
* Adding these labels as a segment for our finalized data set.



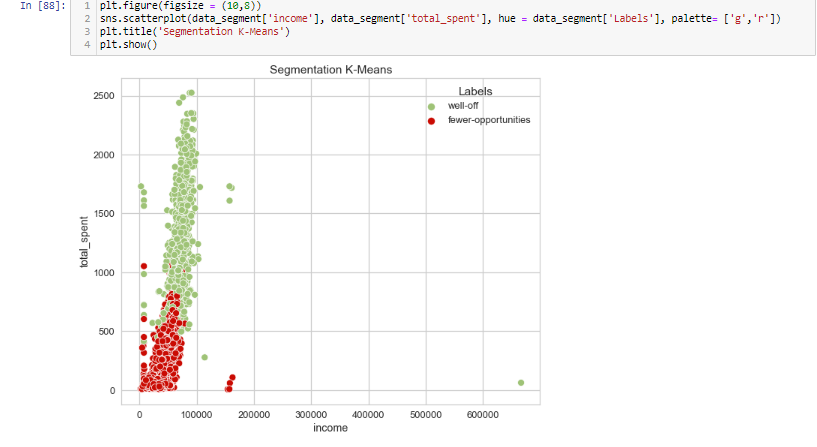
* Segment 0 (few opportunities) = Lower income, have more kids, higher deals purchase, higher web visits, low expense, low accepted campaign

### Segment 1 (well-off) = Higher income, less kids, high web, catalog, and store purchases, high expense, higher accepted campaign.

* Adding these segment labels (few opportunities and well-off) to our dataset.



* Plotting a scatter plot for label wise total spent with income.

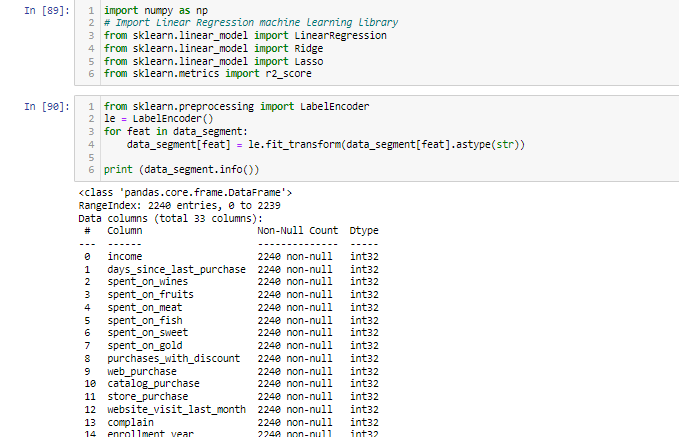


## **8.** **Conclusion:**

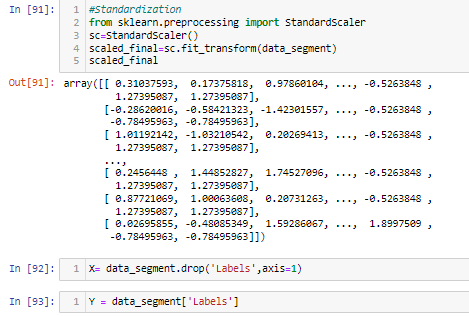
### The clustering is mostly based on income, expenses, number of purchases by it’s category, and total accepted campaign. Education level, marital status, and age did not affect the clustering. So, there are 2 segmentations which is the best number of clustering based on the model. Segment 0 which the customers had low income and low expenses. Segment 1 which the customers had high income and high expenses, which is better to focus on.

**9.** **Finding coefficients of each column for statistical analysis using LASSO Regression (Regularization):**

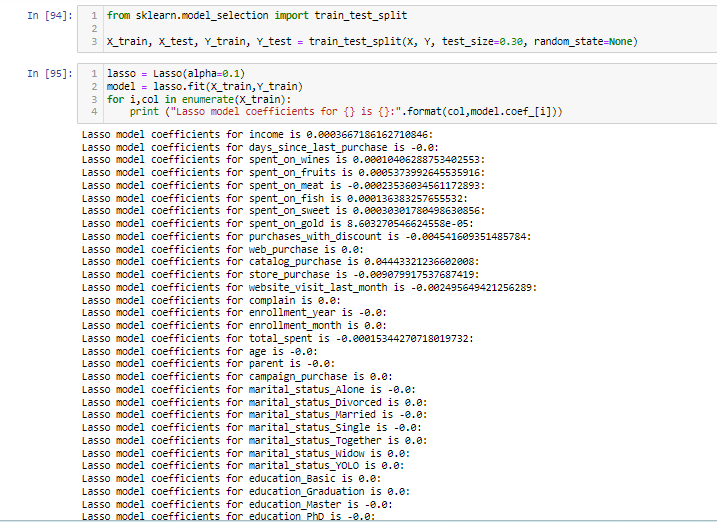
* Importing libraries for Lasso Regression
* Applying label encoding for dataset to converting strings to Numerics.



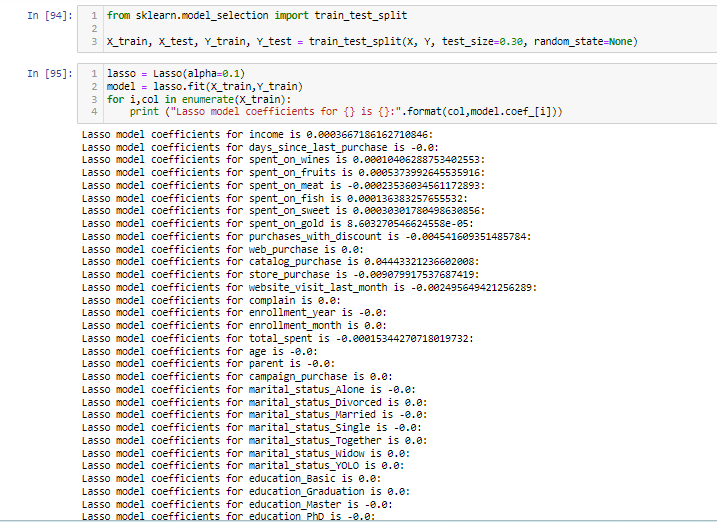
* Applied Standard scalar for scaling to values from (-1 to 1).
* Taking X and Y from the dataset.



* Splitting the X and Y for train test split.
* Build a Lasso Regression Model and fit the X\_train, Y\_train.
* Finding the coefficients of the each column in the model.



* Finding the accuracy score of the model.
* Appling Grid\_searchCV for finding best score at best\_params (99%).



## **10. Conclusion:**

* **This coefficient values are that the effect of every independent variable on target column (Dependent)**
* **In clustering we have seen the outputs graphically but in Lasso Regression is statistically we calculated how much effects on X to Y.**