

Campaign Analysis In Retail Marketing

BITS ZG628T: Dissertation

by

Pushkar Singh Bhauryal

2016HT12302

Dissertation work carried out at

Cognizant Bangalore



**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)**

Campaign Analysis In Retail Marketing

BITS ZG628T: Dissertation

by

Pushkar Singh Bhauryal

2016HT12302

Dissertation work carried out at

Cognizant Bangalore

Submitted in partial fulfillment of M.Tech. Software Systems degree programme

Under the Supervision of
Agniteja Reddy, Manager
Philips, Bangalore



**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE
PILANI (RAJASTHAN)**

CERTIFICATE

This is to certify that the Dissertation entitled **Campaign analysis in Retail Marketing** and submitted by **Pushkar Singh Bhauryal** having **ID-No. 2016HT12302** for the partial fulfillment of the requirements of M.Tech. **Software System** degree of BITS, embodies the bonafide work done by him/her under my supervision.



Signature of the Supervisor

Place : Bangalore

Agniteja Reddy, Manager
Philips, Bangalore

Date : 15/09/2019

Name, Designation & Organization & Location

ABSTRACT

The Indian retail industry is one of the fastest growing in the world. Retail industry in India is expected to grow to US\$ 1,200 billion by 2021 from US\$ 672 billion in 2017. India is the fifth largest preferred retail destination globally. The country is among the highest in the world in terms of per capita retail store availability. India's retail sector is experiencing exponential growth, with retail development taking place not just in major cities and metros, but also in Tier-II and Tier-III cities. India's online retail industry is one of the booming industries. Revenue generation from online sale is projected to reach \$60 Billion by 2020. India is expected to become the world's third largest consumer economy , reaching US\$ 400 billion in consumption by 2025.

Relying on retail analytics and hard data rather than guesswork enables you to make smarter decisions toward higher profits, better customer satisfaction, and having a more awesome store overall. [A survey by Alteryx and RetailWire](#) of nearly 350 retailers and brand manufacturers found that 81% of respondents say they gather shopper insights and 76% consider insights to be critical to their performance. According to the study, only 16% consider themselves experts when it comes to data harnessing, while 24% and 60%, respectively, describe themselves as “newbies” and “getting there.”

The purpose of this project is to make the Campaign for retail marketing is more effective using minimum budget. We will be discussing different issues faced by the organisation while campaigning / marketing and how to overcome these issues.

Acknowledgment

This project is supported by Cognizant. I thank my colleagues from cognizant who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper. I also thanks the team for providing me the sample data.

I would also like to show our gratitude to Agniteja Reddy for sharing their pearls of wisdom with me during the course of this project, and I thank 2 of my colleagues for reviewing this project and for their insights.

Pushkar Singh Bhauryal

Contents

Introduction	1
Data Analytics in Retail industry	2
Campaign Goals and KPI's	3
Getting Started with Data	5
Exploratory Data Analysis	7

List of Figures

Figure 1: Retail market size by year	1
Figure 2: Online Retail trends in India	1
Figure 3: Retail market Sector Composition	2
Figure 4: Use of data analytics in effective way by industry	2
Figure 5: Top 3 shoppers' insight by market	3
Figure 6: scatter matrix plot for customer purchase data	9
Figure 7: Scatter matrix for customer monthly data	10
Figure 8: Scatter matrix for customer socio economic data	11
Figure 9: customer socio economic data after cleaning and normalizing	11

Introduction

The Indian retail industry has emerged as one of the most dynamic and fast-paced industries due to the entry of several new players. Total consumption expenditure is expected to reach nearly US\$ 3,600 billion by 2020 from US\$ 1,824 billion in 2017. It accounts for over 10 per cent of the country's Gross Domestic Product (GDP) and around 8 per cent of the employment. India is the world's fifth-largest global destination in the retail space.

Retail market in India is projected to grow from an estimated US\$ 672 billion in 2017 to US\$ 1,200 billion in 2021. Online retail sales are forecasted to grow at the rate of 31 per cent year-on-year to reach US\$ 32.70 billion in 2018.

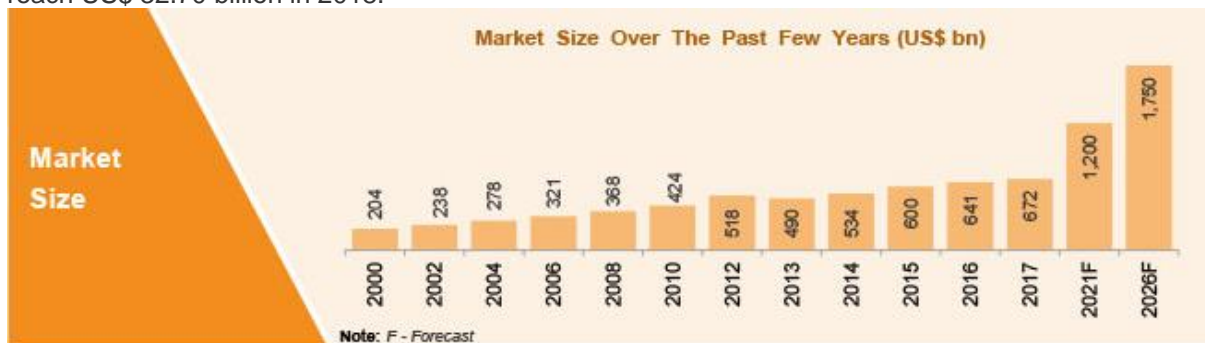


Figure 1: Retail market size by year

India is expected to become the world's fastest growing e-commerce market, driven by robust investment in the sector and rapid increase in the number of internet users. Various agencies have high expectations about growth of Indian e-commerce markets.

E-commerce is expanding steadily in the country. Customers have the ever increasing choice of products at the lowest rates. E-commerce is probably creating the biggest revolution in the retail industry, and this trend will continue in the years to come. India's e-commerce industry is forecasted to reach US\$ 53 billion by 2018. Retailers should leverage the digital retail channels (e-commerce), which would enable them to spend less money on real estate while reaching out to more customers in tier-2 and tier-3 cities.



Figure 2: Online Retail trends in India

It is projected that by 2021 traditional retail will hold a major share of 75 per cent, organised retail share will reach 18 per cent and e-commerce retail share will reach 7 per cent of the total retail market.

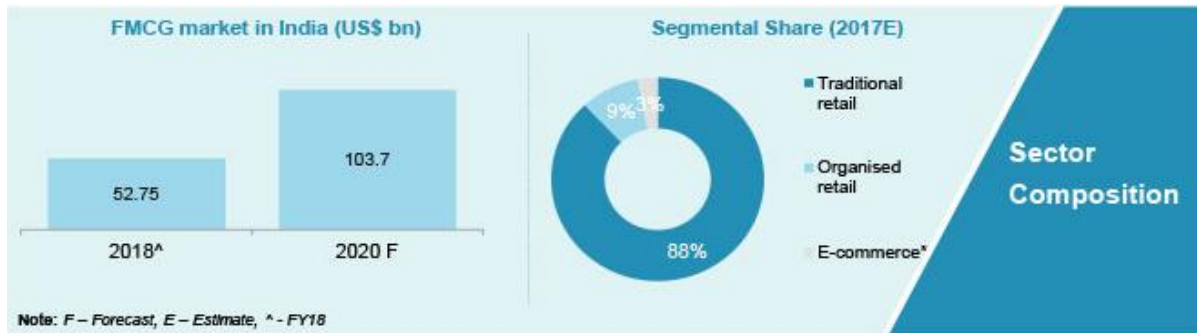


Figure 3: Retail market Sector Composition

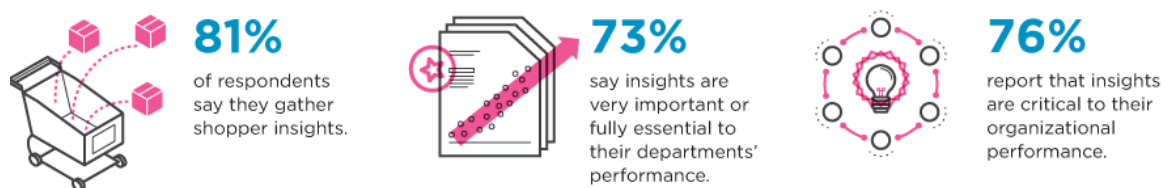
With the rising need for consumer goods in different sectors including consumer electronics and home appliances, many companies have invested in the Indian retail space in the past few months.

Data analytics In retail industry

In a price sensitive and highly competitive market one of the key to achieving success is to be data centric. Relying on retail analytics and hard data rather than guesswork enable you to make smarter decision towards higher profit, better customer satisfaction and having a better competitor overall.

A survey by Alteryx and RetailWire of nearly 350 retailers and brand manufacturers found that 81% of respondents say they gather shopper insights and 76% consider insights to be critical to their performance. Although many merchants are collecting data, most aren't using it effectively. According to the study, only 16% consider themselves experts when it comes to data harnessing, while 24% and 60%, respectively, describe themselves as "newbies" and "getting there."

The retail industry recognise the importance of insight, but it fail short on driving retail decision based on data



BUT MANY COMPANIES DON'T KNOW HOW TO HARNESS THE DATA THEY COLLECT.

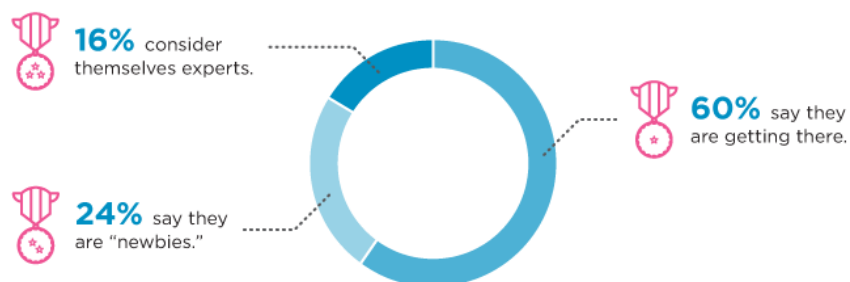
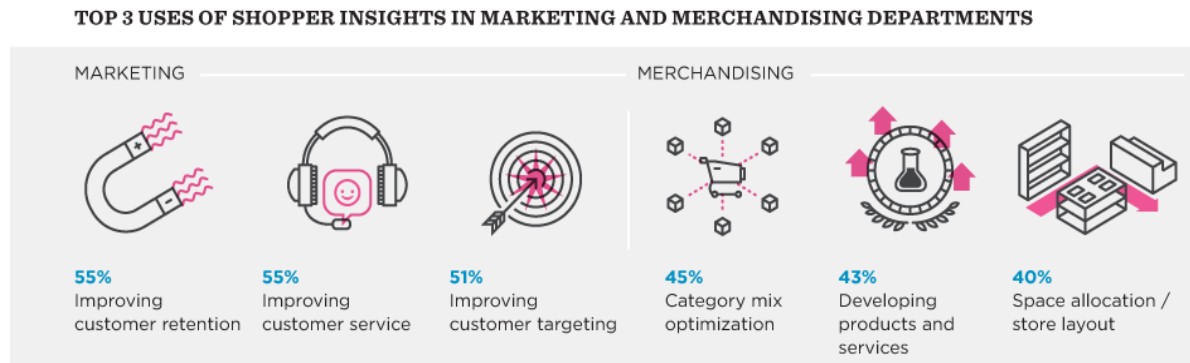


Figure 4: Use of data analytics in effective way by industry

According to survey, Even today marketing departments using shopper insight more extensively than merchandising, and the difference is clearly visible.



This survey also explain some of the missed opportunity regarding getting the most from the data and insight from that.

MISSED OPPORTUNITIES



Figure 5: Top 3 shoppers' insight by market

In fact, only 22 percent of marketers are confident that they have insight into the ROI of their marketing activities, according to a survey by the [American Marketing Association](#). More than twice that number, or 48 percent, even admitted that they are not confident in their marketing ROI.

This lack of confidence in marketing ROI is the byproduct of an over complicated approach to [marketing attribution](#). For retailers in particular, the key to overcoming this challenge is to use clear retail marketing marketing analytics to track and prove the impact marketing has on revenue and market share.

Campaign Goal and KPI's

Here we'll look at the best marketing KPIs and metrics for supporting retail marketing goals that are tied to revenue.

Gaol - Increase marketing revenue and marketing - influenced sales

KPI - Marketing influenced sales

The first KPI we will be considering is the number of sales influenced by marketing campaigns and activities. This will help us to track witch campaign and channel customer interact with before buying the product from us.

KPI - Transaction value and quantity

As a retail marketer, you are just not looking for more customer, But looking for each customer to purchase more items and spend more money. This is why tracking the average value and quantity of item purchased per transaction is important.

Goal - Customer acquisition cost reduction

KPI- Number of touchpoint

The fewer paid touchpoints customers require before making a purchase, the less money you are likely to have to spend on each customer.

KPI- Spending efficiency

Matric like cost per click(CPC) , return on investment(ROI) can help us to predict the spending of our next marketing campaign.

Goal- Increase conversion efficiency

Only about 23% of businesses say they are satisfied with their conversion rate according to Econstancy.

KPI-The most important KPI for tracking increased conversion efficiency is the conversion rate, Or the number of customers who bought the product,out of all who we reached using campaign.

KPI Shopping cart abandoned rate

This step is one of the most critical step track the efficiency of the campaign. This is the percentage of shoppers who add a product in the cart but don't end up purchasing anything.

Goal - Increase customer loyalty

KPI - Average customer lifetime value

Customer lifetime value is a prediction of the total net profit a company makes from a customer over the entire future of their relationship.

KPI - repeat buyer

Track the percentage of customers who return to purchase from your company again. This is the strongest metric for measuring customer loyalty in terms of willingness to purchase again.Having loyal customers is a huge benefit to retail marketers because when a customer purchases again, there are no new customer acquisition costs, the path the purchase is shorter.

Getting Started with Data

In this project we will create a machine learning model to decrease the customer acquisition cost. We have sample data for 5000 customers from previous campaign. We will go step by step to investigate data to make a model that will maximise the efficiency of the campaign and reduce the cost of acquisition of the customer.

Data Set - Explanation

Table 1- Customer profile data

Single dataset combining customer data for different subject areas including demographics for 5000 customers.

1. Customer Id (Unique Id)
2. Age
3. Job- Type of job
4. Marital Status- (Categorical: 'admin', 'blue-collar' etc.)
5. Ethnicity-- Marital Status (Categorical: 'single', 'married' etc.)
6. Education- Education (Categorical: 'High School', 'Illiterate' etc.)
7. Previous default flag - Has credit in default (Categorical: 'Yes', 'No')
8. Loan available flag - Has Personal Loan (Categorical: 'Yes', 'No')
9. House owned flag-Owns a house (Categorical: 'Yes', 'No')
10. Employment Type of Employment (Categorical: 'Unemployed', 'Retired' etc.)
11. Credit rating-Credit Score (Categorical: 'High', 'Critical' etc.)
12. Life stage Code -Situation of Customer based on family, age and marital status
13. Income USD - Income Range (Continuous)

Table 2 - Past Purchase Data

Single dataset summarizing sales data by item, date, campaign and customer for past one year for 5000 customers.

1. Customer Id
2. Loyalty Tier- Indicates Customer Loyalty Category (Categorical: 'Gold', 'Bronze' etc.)
3. Reward point earned- Reward Points Earned history (Numeric)
4. Average Basket Size - Number of items purchased in each transaction
5. Count of transaction- Number of transactions based on purchase history

Table 3 - Campaign Coverage Data

Single dataset summarizing characteristics and customer response for previous campaign. Details related with the last contact of the previous campaign

1. Customer ID
2. Mode of contact - Contact communication type (categorical: 'mail', 'in-store flyers')
3. Month last contacted - last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'dec')
4. Day of week last contacted - last contact day of the week (categorical: 'mon', 'tue',...)
5. Duration last contacted- duration, in seconds (numeric)
6. Duration last contact type - Call duration classification (categorical: Long, Medium, Short)
7. Num of contact campaign - number of contacts performed during this campaign and for this client
8. Num of contact campaign type -number of contacts during campaign classified (categorical: Often, Once, Average, Very Frequent)

9. Days elapsed last contact - number of days that passed by after the client was last contacted from a previous campaign
10. Num contact before campaign -number of contacts performed before this campaign and for this client
11. Num contact before campaign- type number of contacts before campaign classified(categorical: Often,Once,Average,Very Frequent)
12. Outcome last campaign -outcome of the previous marketing campaign (Categorical:'failure','nonexistent','success')
13. Buy decision flag-tells whether customer bought particular product (Binary:'Yes', 'No')
14. Target buy-whether customer purchased the product (Binary: 1,0)
15. Campaign start date-Start date for the campaign

Table 4 - Month level customer data

Transaction level data by item, date, campaign and customer at month level for past 12 months for 5000 customers

1. Customer Id
2. Month- Base unit to record sales (Categorical)
3. Revenue-Net Revenue based on past purchases (Numeric)
4. Net Profit-Net Profit for each customer (Numeric)
5. Net Margin-Indicates Profitability level of customer based on purchase history (Numeric)

Table 5 - Socio Economic data

Single dataset summarizing characteristics of overall economy performance and consumer spending confidence for past 3 months

1. Customer Id
2. Month last contacted- last contact month by which economic factors vary (categorical: 'jan'..)
3. Qtrly Emp variation rate-employment variation rate - quarterly indicator (numeric)
4. Monthly consumer price idx- consumer price index - monthly indicator (numeric)
5. Monthly consumer conf idx-consumer confidence index - monthly indicator (numeric)
6. US fed 3 month rate-Federal 3 month interest rate - daily indicator (numeric)
7. Quarterly num of emp-number of employees - quarterly indicator (numeric)

Data Exploration and Descriptive Analysis

One of the most critical step in the analytics project is data exploration or exploratory data analysis(EDA) and description statistical analysis. EDA will help us to find the relationship between multiple variables, find anomalies in the data set. Descriptive statics help us to get a basic understanding of the data, like data counts, max value, min value etc.

Exploring customer profile data

Lets check the basic stats of the data, and check if table has valid data or not, data has no missing values etc

Getting basic stats

```
customer_profile.describe()
```

	Customer_ID	age	Income_USD
count	5148.000000	5148.000000	5148.000000
mean	2574.500000	39.867133	94534.816239
std	1486.243923	10.560282	56004.541887
min	1.000000	17.000000	10003.000000
25%	1287.750000	32.000000	49125.000000
50%	2574.500000	38.000000	79281.500000
75%	3861.250000	47.000000	150483.250000
max	5148.000000	89.000000	199894.000000

Getting table info

```
display(customer_profile.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5148 entries, 0 to 5147
Data columns (total 13 columns):
Customer_ID      5148 non-null int64
age              5148 non-null int64
job              5148 non-null object
marital_status   5148 non-null object
education        5148 non-null object
Previous_Default_Flag  4066 non-null object
House_Owned_Flag  5148 non-null object
Loan_Availded_Flag  5032 non-null object
Income_USD       5148 non-null int64
Credit_rating    5148 non-null object
Life_Stage_Code  5148 non-null object
Ethnicity        5148 non-null object
Employment       5148 non-null object
dtypes: int64(3), object(10)
memory usage: 522.9+ KB
```

By looking on stats of customer profile data income of the customer is distributed in a large range. We will try to normalize the data while using our algorithms.

By looking in the info stats we can see that there are some null values in the previous default flag and loan available flag columns. We will also deal with null values in our data cleaning section.

Exploring customer purchase data

Getting table description

```
customer_purchase.describe()
```

	Customer_ID	Count_of_Transactions	Avg_Basket_Size	Reward_Points_Earned
count	5148.000000	5149.000000	5149.000000	5.149000e+03
mean	2574.500000	134.396582	207.289154	9.227399e+04
std	1486.243923	4821.030493	119.256848	3.310124e+06
min	1.000000	22.000000	33.405789	7.657000e+03
25%	1287.750000	51.000000	110.931340	1.653800e+04
50%	2574.500000	64.000000	184.594247	4.437900e+04
75%	3861.250000	91.000000	279.179362	7.337900e+04
max	5148.000000	346004.000000	775.884000	2.375594e+08

Getting table info

```
print(customer_purchase.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5149 entries, 0 to 5148
Data columns (total 5 columns):
Customer_ID          5148 non-null float64
Count_of_Transactions 5149 non-null int64
Loyalty_Tier         5148 non-null object
Avg_Basket_Size      5149 non-null float64
Reward_Points_Earned 5149 non-null int64
dtypes: float64(2), int64(2), object(1)
memory usage: 201.2+ KB
None
```

Creating scatter plot to check the relationship of numeric variable

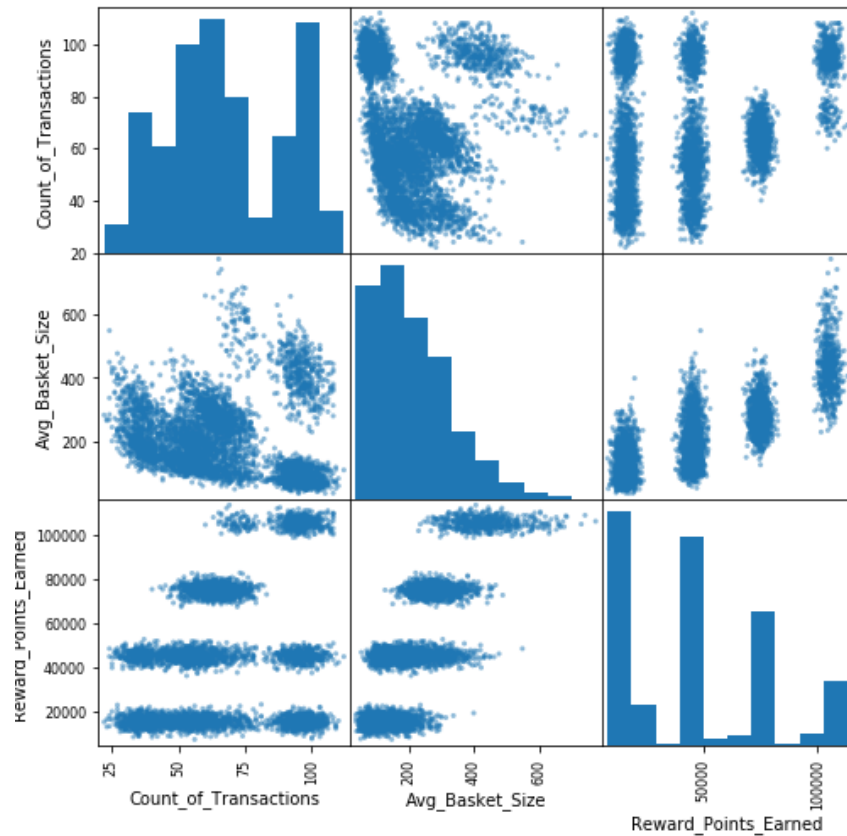


Figure 6: scatter matrix plot for customer purchase data

By looking into the chart and description table we can say that average basket size and reward points are highly skewed. We will deal with the data skewness in data preparation section.

There is one null value in the loyalty tier column. We will remove this row to clean our data.

By looking at the graph we can also see that there is a positive correlation between avg_basket_size and preward_point_earned.

Exploring customer monthly data

Getting table descriptive stats

	Customer_ID	Revenue	Cost	Net_Profit
0	1	798.046667	704.122425	93.924242
1	2	843.015000	741.555683	101.459317
2	3	1143.240000	1003.617667	139.622333
3	4	500.138333	457.216067	42.922267
4	5	638.154167	581.438650	56.715517

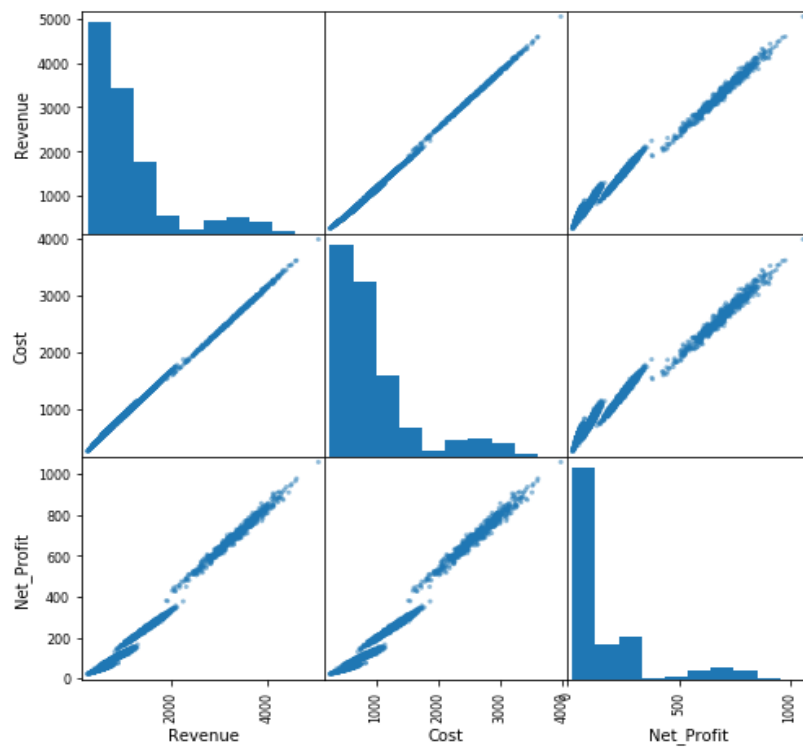


Figure 7- Scatter matrix for customer monthly data

By looking into the above graph we can see that most of the columns are skewed and highly correlated in nature. We will handle the skewness in the data preparation section.

Exploring Campaign coverage data

Getting table descriptive stats

```
campaign_coverage.describe()
```

	Customer_ID	duration_last_contact	num_of_contact_campaign	days_elapsed_last_contact	num_contact_before_campaigning	Target_buy
count	5148.000000	5148.000000	5148.000000	5148.000000	5148.000000	5148.000000
mean	2574.500000	254.616356	2.595183	962.542735	0.172300	0.113054
std	1486.243923	250.417132	2.742720	186.765117	0.497159	0.316689
min	1.000000	4.000000	1.000000	0.000000	0.000000	0.000000
25%	1287.750000	100.000000	1.000000	999.000000	0.000000	0.000000
50%	2574.500000	176.000000	2.000000	999.000000	0.000000	0.000000
75%	3861.250000	316.000000	3.000000	999.000000	0.000000	0.000000
max	5148.000000	3785.000000	34.000000	999.000000	6.000000	1.000000

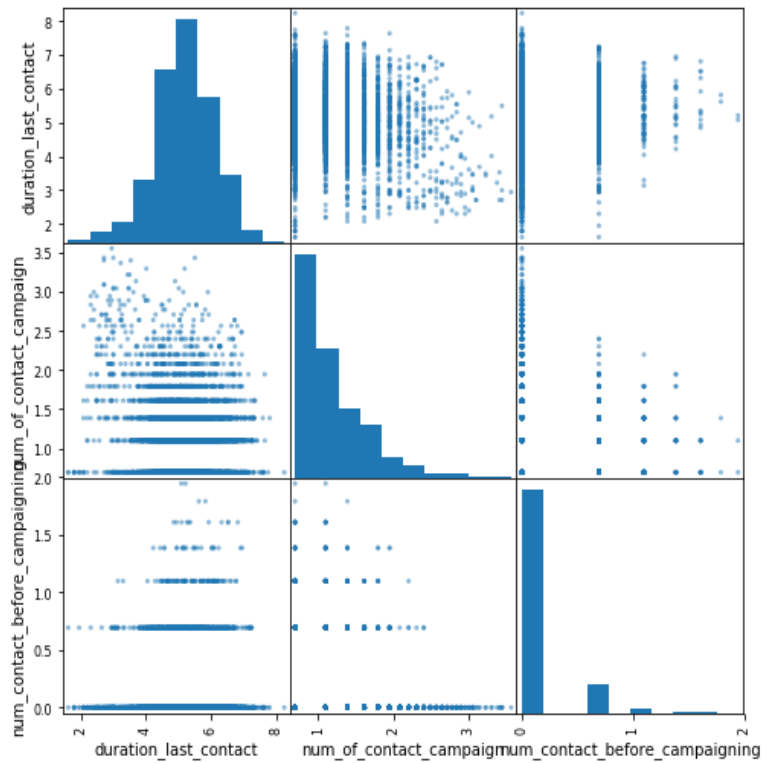


Figure 8- Scatter matrix for customer socio economic data.

By looking at the above plot and table we can see that almost all the columns data is skewed. Since column **days_elapsed_last_contact** has most of the data as 999 I removed that column from the data set. I also removed the column **days_elapsed_last_contact** and **day_of_week_last_contacted** as the data is already available in other dataframe.

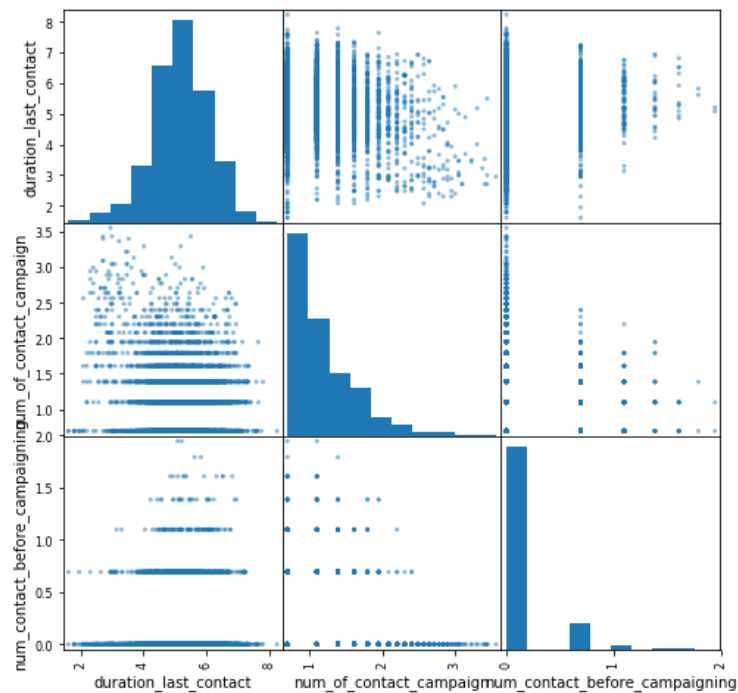


Figure 9- customer socio economic data after cleaning and normalizing

SCHEDULE AND RESULTS

SNo	Description of Work	Start Date	End Date	Status
1	System Study, Software Installation, defining problem statement, data collecting and problem-solving technique	20-Aug-2019	30-Aug-2019	Done
2	Data Cleaning, Exploratory data analysis, Basic descriptive statistical analysis, data preparation and visualization	1-Sep-2019	15-Sep-2019	Done
3	Data Transformation and machine learning algorithms Implementation	20-Sep-2019	5-Oct-2019	Ongoing
4	Selecting best model among the developed model, implementing parameter tuning to increase the efficiency of the model	6-Oct-2019	15-Oct-2019	Not-Started
5	Performance Testing and measurement testing of the model	16-Oct-2019	20-Oct-2019	Not-Started

References

<https://trackmaven.com/blog/retail-marketing-analytics-prove-sales/>

<https://www.ibef.org/industry/retail-india.aspx>