# **Exploratory Data Analysis**

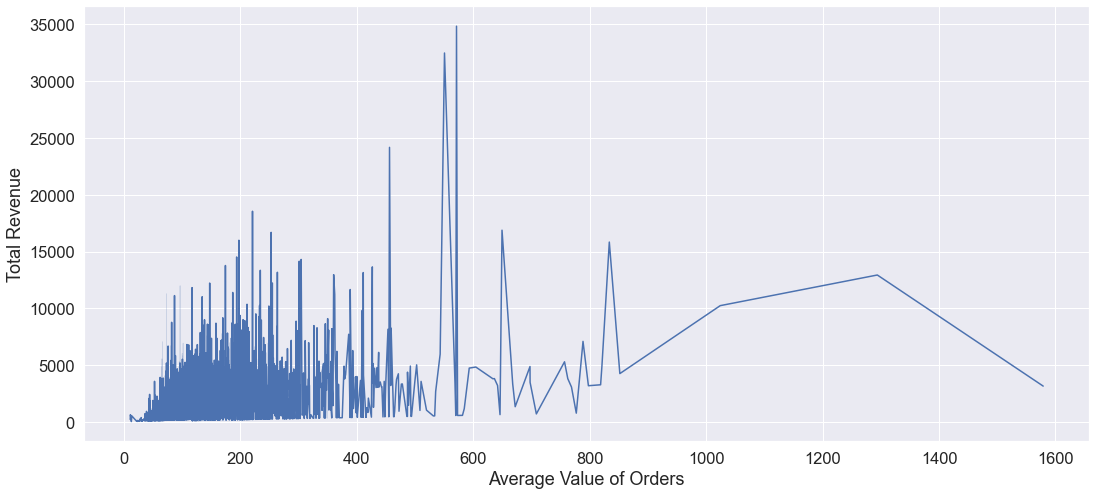
The Dataset shows detailed aggregate information of customers regarding their Orders and Revenue generated from those Orders. The count of Orders and Revenue has been categorized in different categories based on different frame of references.

## Orders v/s Revenue



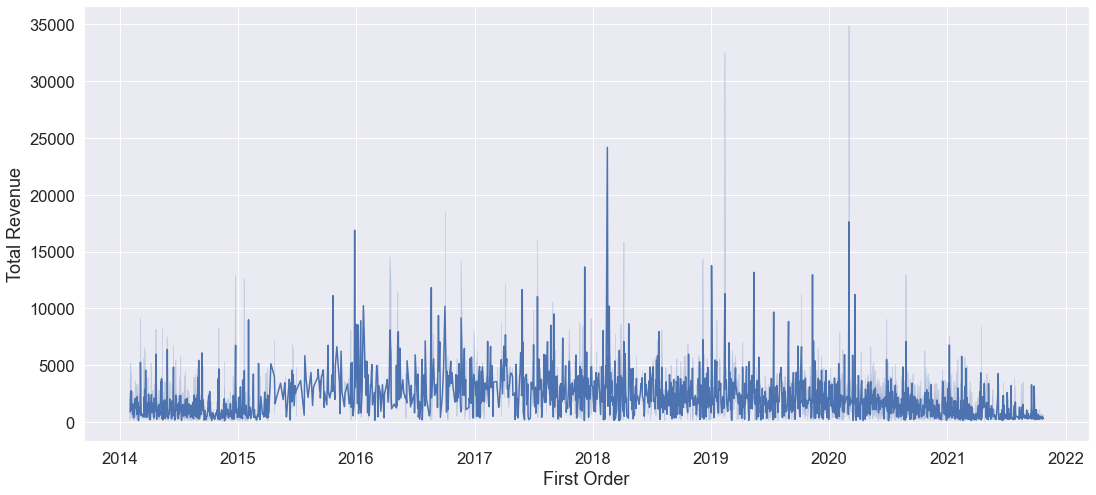
Results shows a **77% correlation** between Total Orders and Total Revenue Generated, which is quite predictable as more orders will yield more profit if the business is not working on a lossy model.

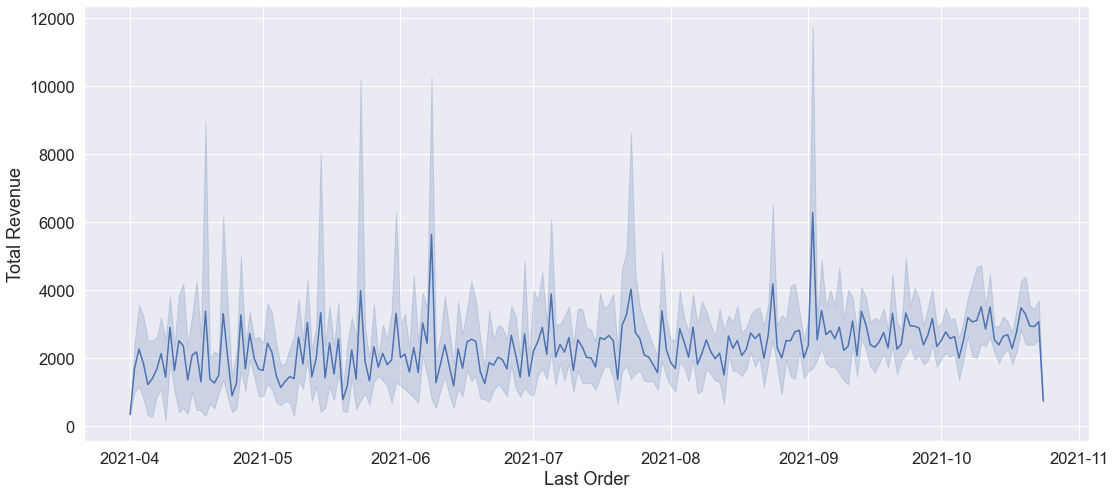
## Average Value of Orders v/s Revenue



Results show a **37% correlation** which shows that having a high Average order value does not correspond to a high overall revenue.

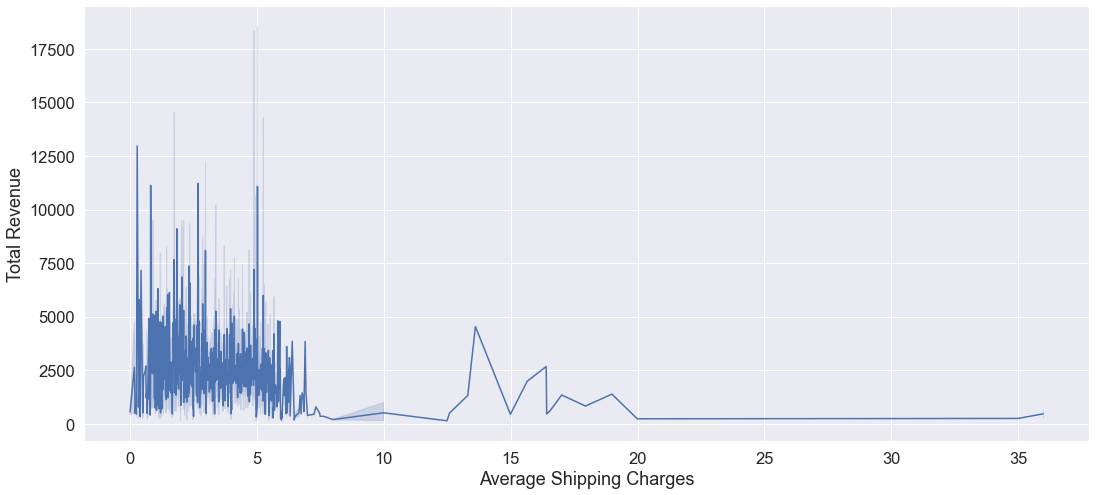
## Recency v/s Revenue

****

****

Recency is a critical feature used for customer segmentation, but in case of Revenue, it **does not correlate** **very promisingly**. For Last orders from customers, it shows that customer who placed orders recently had slightly more contribution towards higher Revenue, but in case of customers’ first orders, it doesn’t correlate.

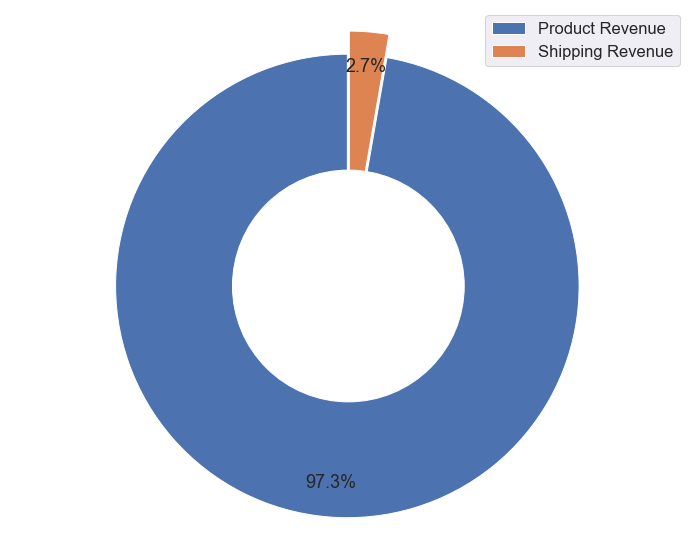
## Shipping/Carriage Revenue v/s Revenue



As we can see that customer with orders having higher shipping charges amounts to very low contribution in Revenue. Most of the orders placed by customers are with an Average Shipping Charges between **2** and **5** (**approx. 3300**).

## Breakings of Revenue

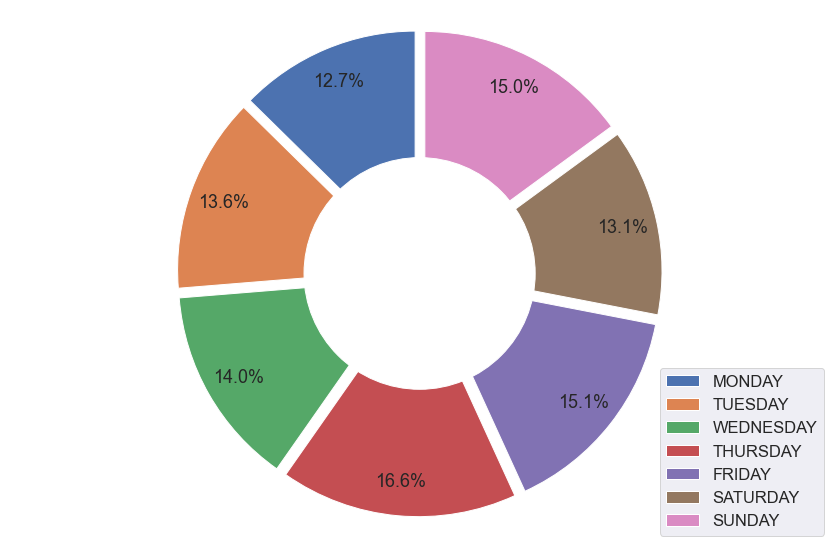
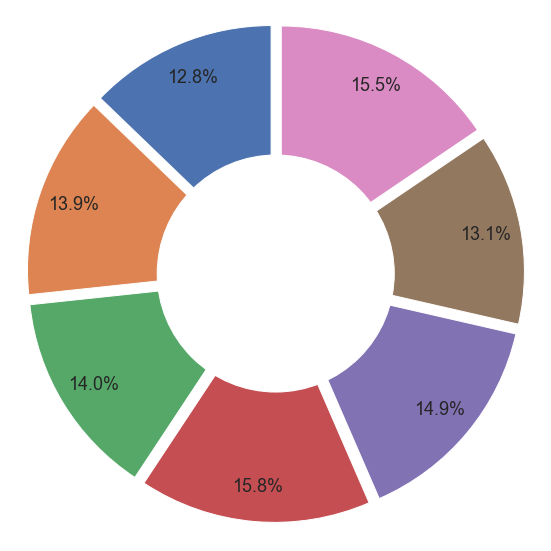
The Revenue is combined of two sub categories: 1). Product Revenue and 2). Shipping/Carriage Revenue. The Carriage Revenue is only **2.7%** of the Total Revenue.



## Weekday-Wise Orders and Revenue

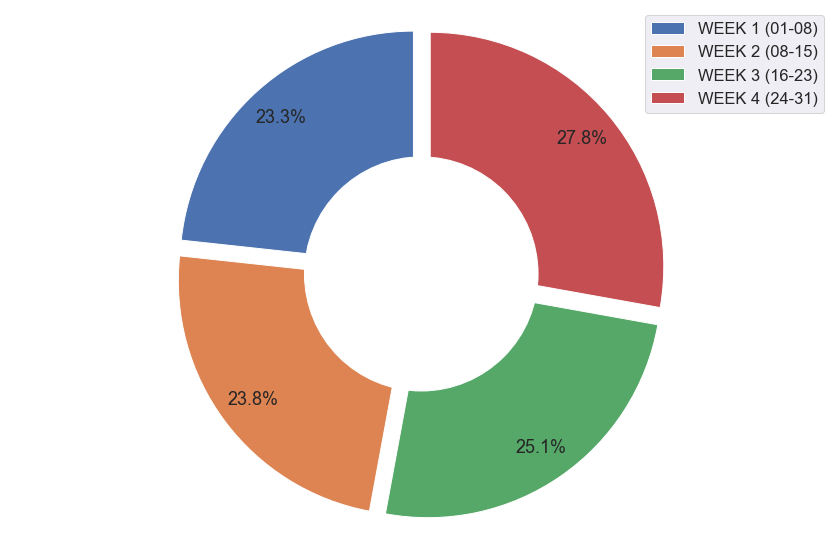
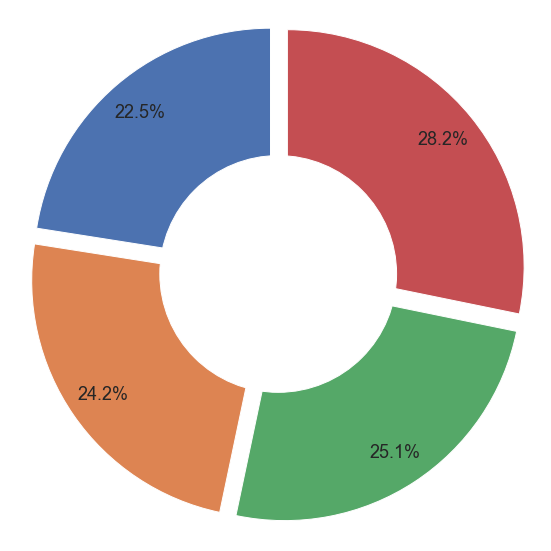
Revenue

Orders



The Data shows that the highest number of orders are placed on **Thursday** and hence most amount of revenue is generated on Thursday. Whereas lowest number of orders are placed on **Monday**, hence lowest revenue.

## Week-Wise Orders and Revenue

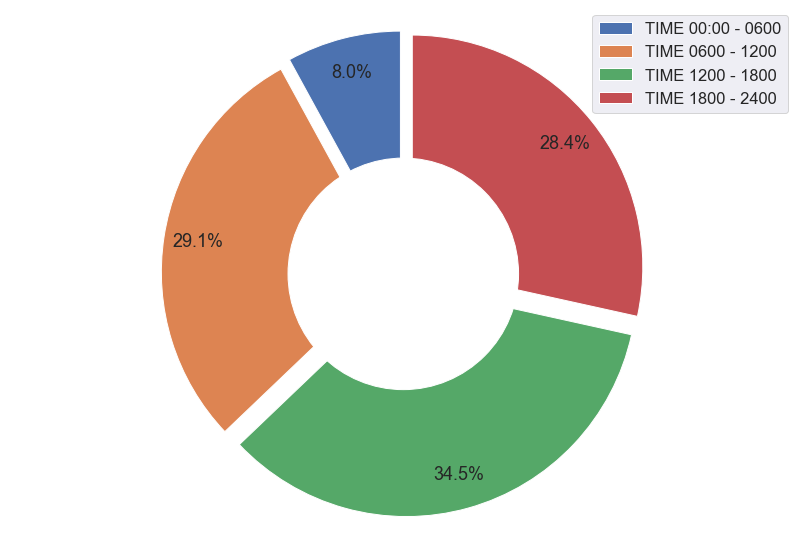
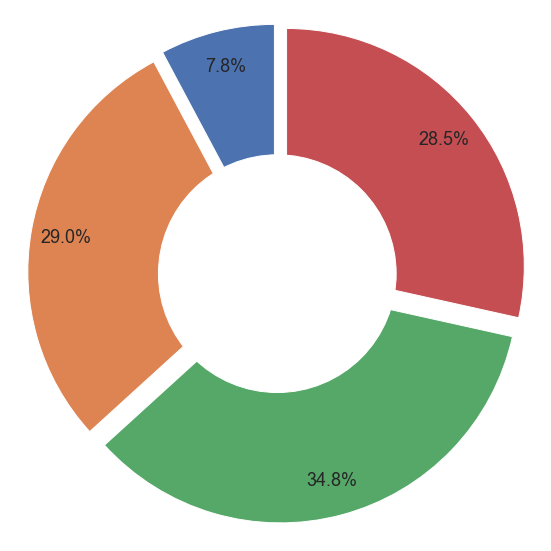


Revenue

Orders

The data shows that the **highest number** of orders are placed in **4th Week** of a month, hence the highest Revenue is also generated in 4th Week. Whereas, **lowest number** of orders are placed in **1st week** resulting in lowest Revenue.

## Time-Wise Orders and Revenue



Revenue

Orders

Time-Wise analysis shows that 3rd quarter of day from 12PM to 6PM is the prime time for highest number of Orders and hence most amount of Revenue.

# **RFM Segmentation**

**Aim:** Identify potential customer segmentation using RFM Model and provide some meaningful insights from each segment. Please find out the customers who are 'champions', 'Potential customers' and 'need attention'.

RFM stands for Recency, Frequency & Monetary. In case of this data, Recency Factor is given by the feature “**DAYSSINCELASTORDER**” (Number of days since Last Order), Frequency Factor is given by “**TOTAL\_ORDERS**” (Total Number orders placed by a customer) and Monetary Factor is given by “**REVENUE**” (Total Revenue generated by a customer).

Using these 3 factors, Customer are categorized into different classes or segments. These segments help to identify the nature of groups of customers and what type of strategies need to applied to cater the customers for better services and better business revenues & profits.

## Data Modelling Approach

Using the three features: **DAYSSINCELASTORDER , TOTAL\_ORDERS, REVENUE,** an RFM Dataset was created. This dataset included **CustomerID** and previously mentioned 3 Features.

RFM segmentation requires the division of datasets based on quantiles. Number of Quantiles can depend on the requirement of business/ request specified by business. In this case, business required 3 catogories:

1. Champions
2. Potential Customers
3. Need Attention

Therefore, the RFM Dataset was divided or cut into 3 quantiles for each feature at points **0.33** and **0.66**

Based on the cuts, each record/ Customer was assigned a score for each feature and stored in **R\_Score**, **F\_Score** and **M\_Score**.

All three scores were totaled and concatenated for sub groups in **RFM\_Group** and **RFM\_Score** respectively.

Using the qcut function of Dataframe, the dataset was labelled based on the categories provided and the labels were recorded in **Category** Feature.

**RFM Dataset**

ComputerID

DaysSinceLastOrder -> Recency

Total\_Orders -> Frequency

Revenue -> Monetary

**Creating Quantiles**

Points for cutting quantiles:

0.33 & 0.66

**Assigning Scores**

Based on cuts, scores were assigned

1-3 pts

**Total Score and Groups**

Adding and Concatenating the Scores

**Categorizing**

Based on Total Scores Customers were divided in 3 Categories:

1. Champions (3-5 pts)
2. Potential Customers (6-7 pts)
3. Need Attention (8-9 pts)

# **Strategies Suggestion based on Data Analysis and Segmentation**

1. There should be a type of premium access or Loyalty program to provide premium services to champions.
2. This may include having Free Shipping, Early Sale access, Exclusive Deals etc.
3. Using cross-selling for champion class customers to boost the sale of products with different categories.
4. There should also be some discount coupons and discount sales to encourage the Customers in “Potential Customers” and “Need Attention” category to purchase more items or more frequently.
5. Since Monday and Tuesday has the lowest number of Orders and Revenue, introducing the sales on Monday or Tuesday.
6. 1st Week of months has been observed for having lowest Orders and Revenue. Suggestion of having sales would for customers for other segments.
7. It has also been observed that people mostly order between 12PM and 6PM. Introducing the sales and discount offers between that time or between 6PM and 12AM can increase Revenue.