



Ecommerce Analytics

PROJECT

Nitish Singh | Data Science 360 | 17/09/2023

Project Introduction

The client is one of the leading online market place in India and wants help in measuring, managing and analysing performance of business.

Client has asked to provide data driven insights about business and understand customer, seller behaviors, product behavior and channel behavior etc.

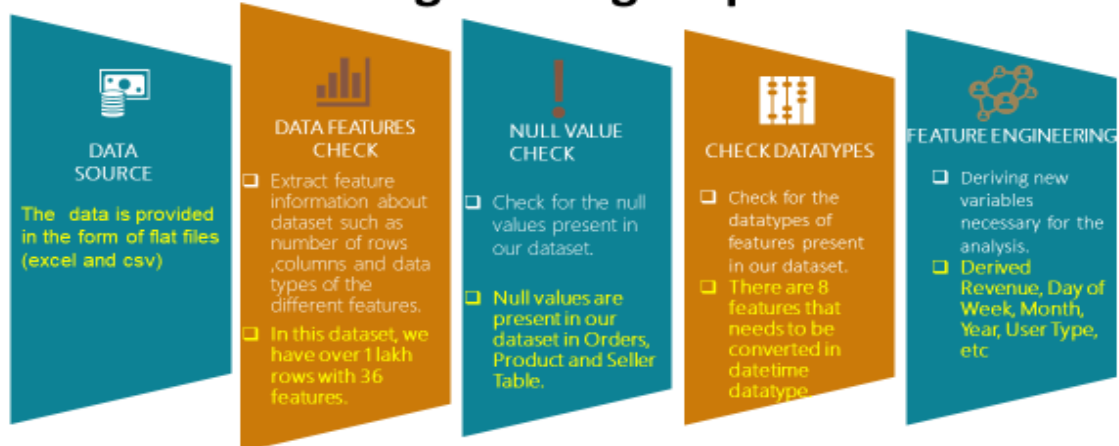
The below are few sample business questions to be addressed as part of this analysis.

- a) Perform Detailed exploratory analysis
- b) Performing Customers/sellers Segmentation
- c) Cross-Selling (Which products are selling together)
- d) Payment Behaviour
- e) Customer satisfaction towards category & product

Available Dataset

The sales dataset is a collection of anonymized data about orders placed from September 2016 to October 2018. It contains a wide range of information about each order, including the order date, product details, payment and shipping information, customer and seller IDs, and customer reviews. The dataset also includes information about the sellers who list their products on the website, as well as data on customer behavior and demographics. The dataset is designed to help analysts and researchers better understand the e-commerce landscape in India and identify opportunities for growth and optimization.

Meta Data Check/ Data Cleaning / Feature Engineering Steps



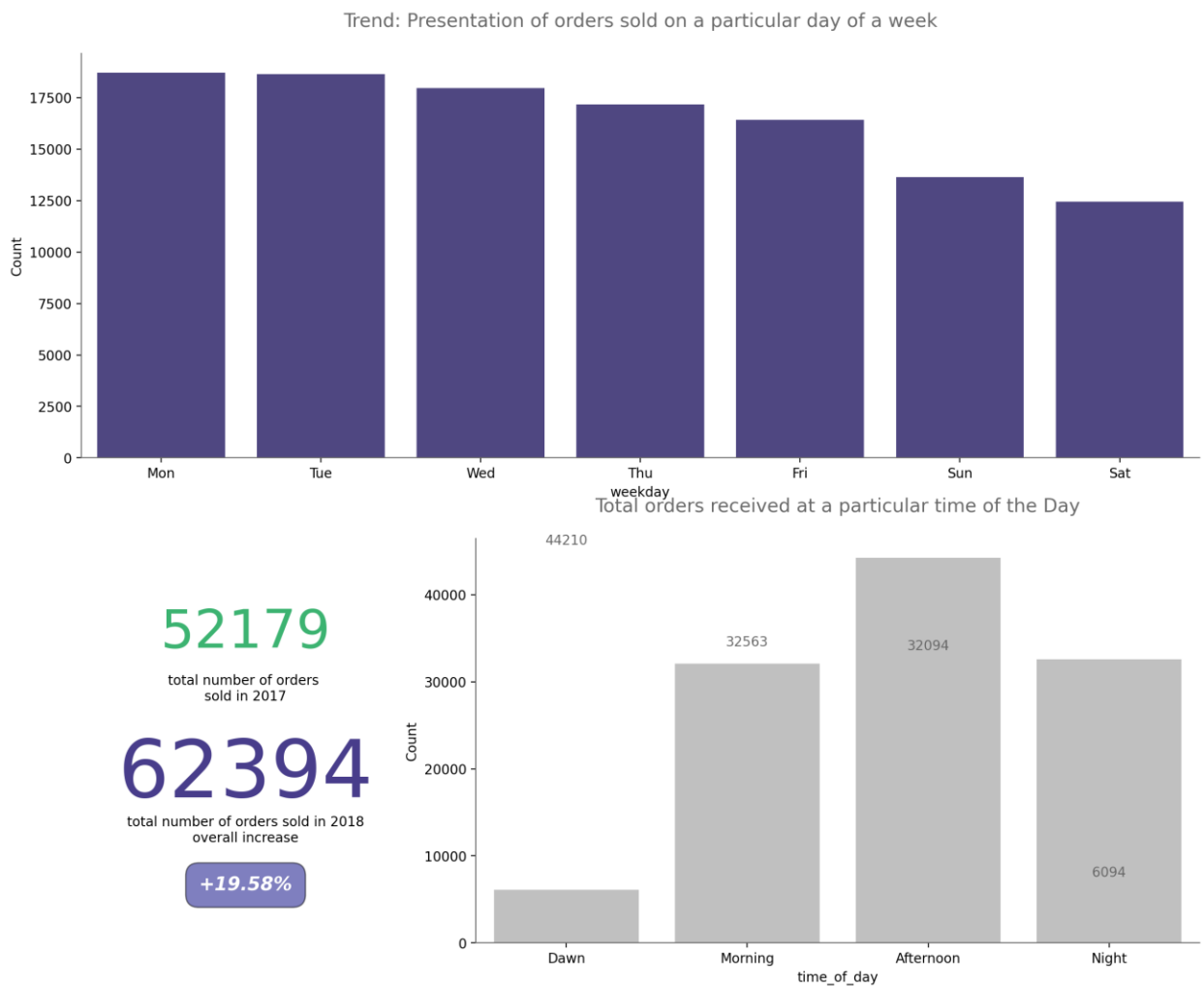
HIGH LEVEL MATRICES

Some high level matrices that are explained in this analysis are Total Revenue, Total quantity, Total products, Total categories, Total sellers, Total locations, Total channels, Total payment methods, Total Customers

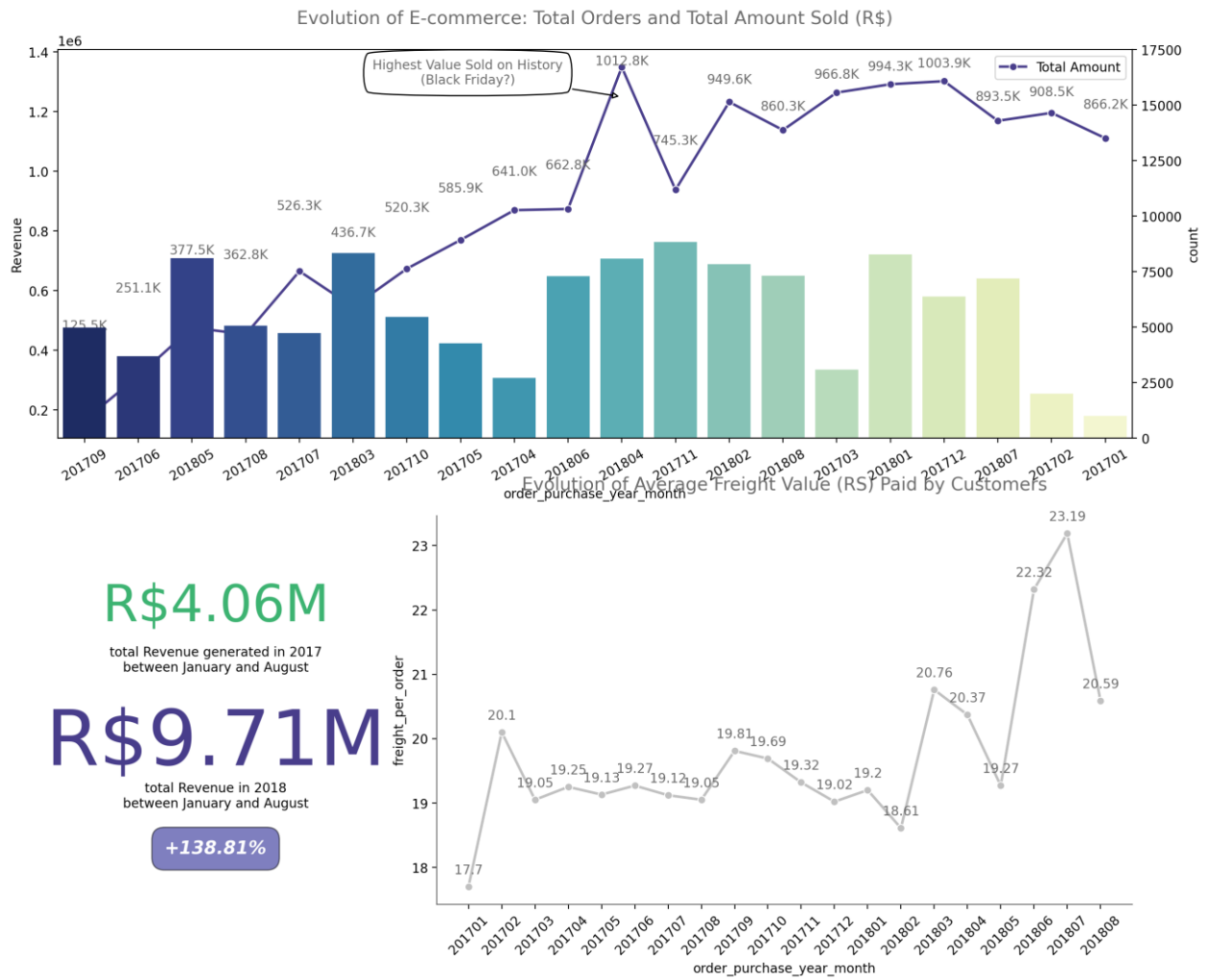
| | | | |
|--------------------------|------------------|----------------|----------------|
| Total Revenue (Millions) | Payment Methods | Total Products | |
| \$17.87M | 4 | 31.78k | |
| | | | |
| Total Locations | Total Categories | Total Sellers | Total Quantity |
| 20 | 71 | 2976 | 136.9k |

ORDERS

We can see some seasonality patterns here. Most of the orders were received on **Weekdays** and on Weekends the orders are less. The chances are that customers plans to get their product delivered by the weekend when they are available at home so that's why the orders are high on weekdays instead of weekends. **Monday's** are most the preferred day for customers and they tend to buy more at **afternoons**. Almost same number of orders were received at Morning and Night time.

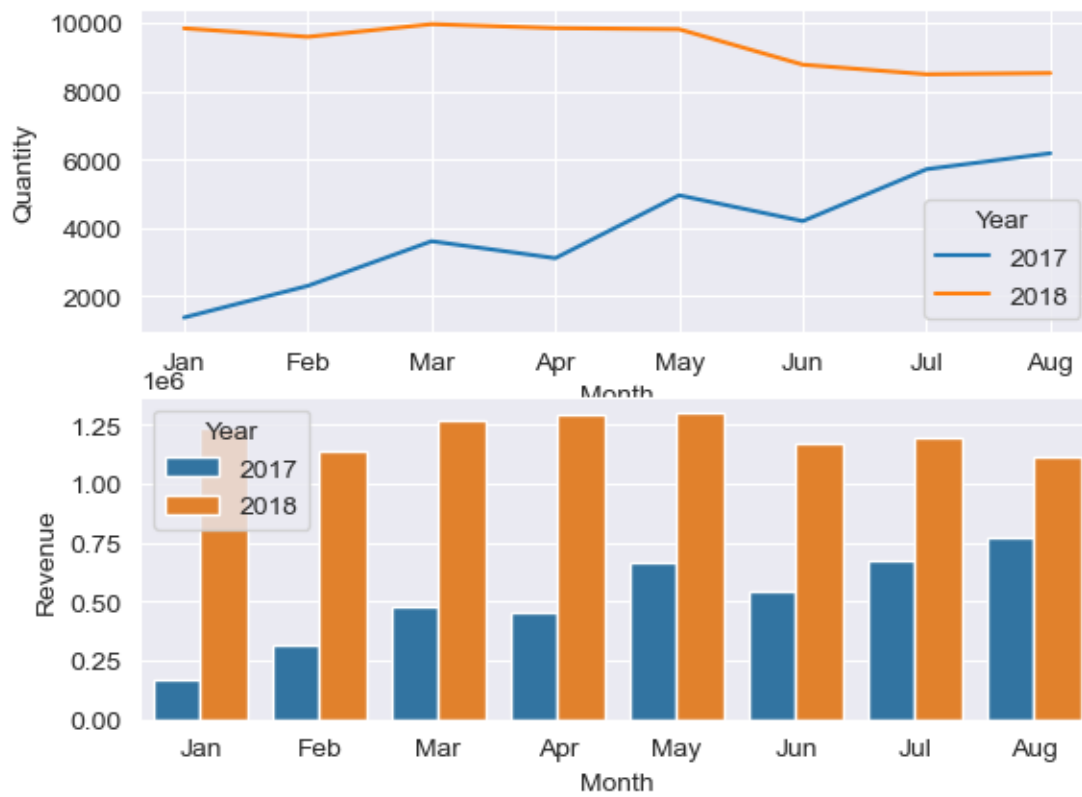


A **20% jump** in the **number of orders** received in the year 2018 is seen compared to year 2017. Similar trend but with a very huge increase has been observed in the case of revenue as well. The **total revenue increased by 138%** compared to the year 2017.



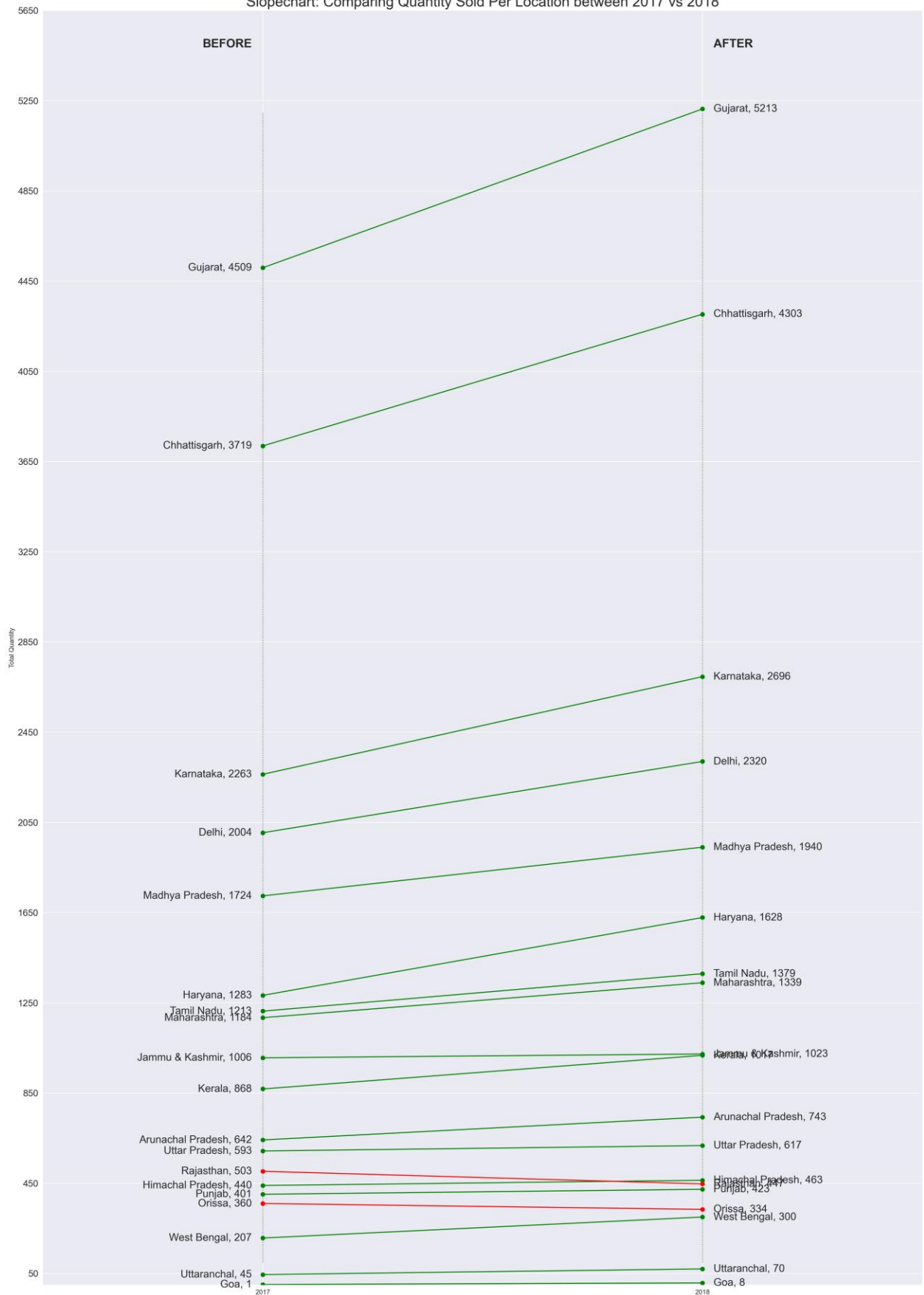
Year-Wise Comparison

Total Orders and Total Revenue: If we compare the year-wise trend of Revenue and Orders we will see a rising trend in the year 2017 but a steady trend in the year 2018.



Location: Company has seen an increase in the demand from all the states except the two (Rajasthan, and Orissa)

Slopechart: Comparing Quantity Sold Per Location between 2017 vs 2018



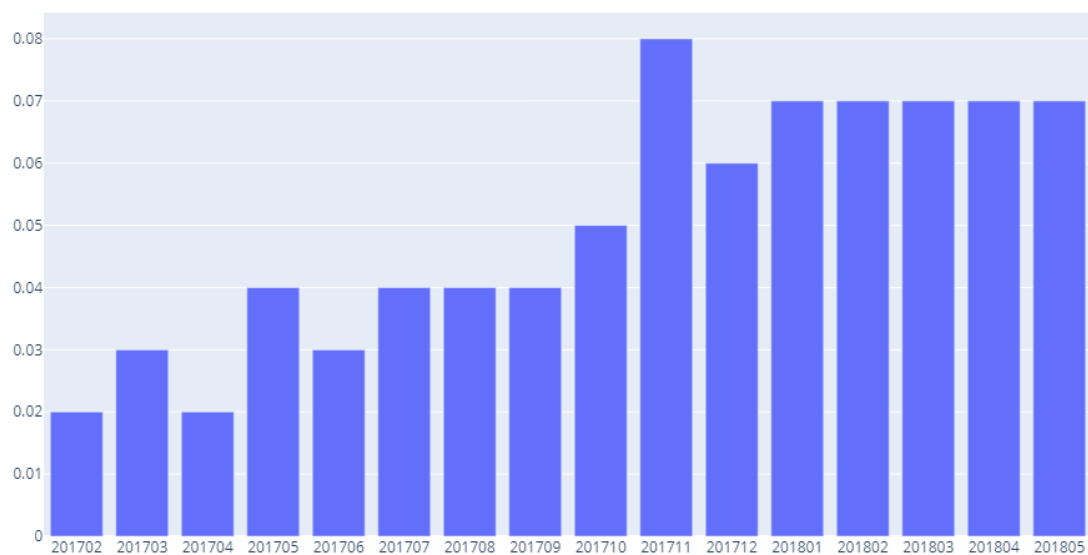
CUSTOMERS

The total number of customers are **92795** out of which total number of **repeat customers** are **1674** and new customers are 91121. **New Customer Ratio** is **98.2%**

Monthly Trend: New Customer Ratio

A steady increase from February 2017 to September 2017 then a jump in November 2017.

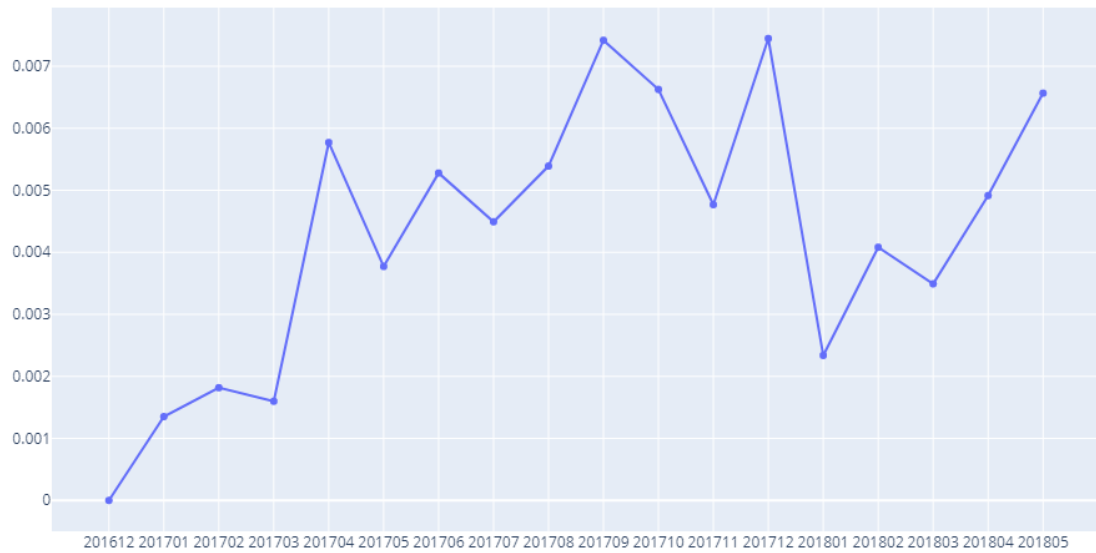
New Customer Ratio



Customer Retention Rate

We have observed an overall **steady increase** in the customer retention rate in the year 2017 (started at 0.01% and reached 0.07% at the end of the year). In the month of November, 2017 the retention rate had dipped. The reason might be because many other competitors too have been offering discounts on their platform as it was the Black Friday Month.

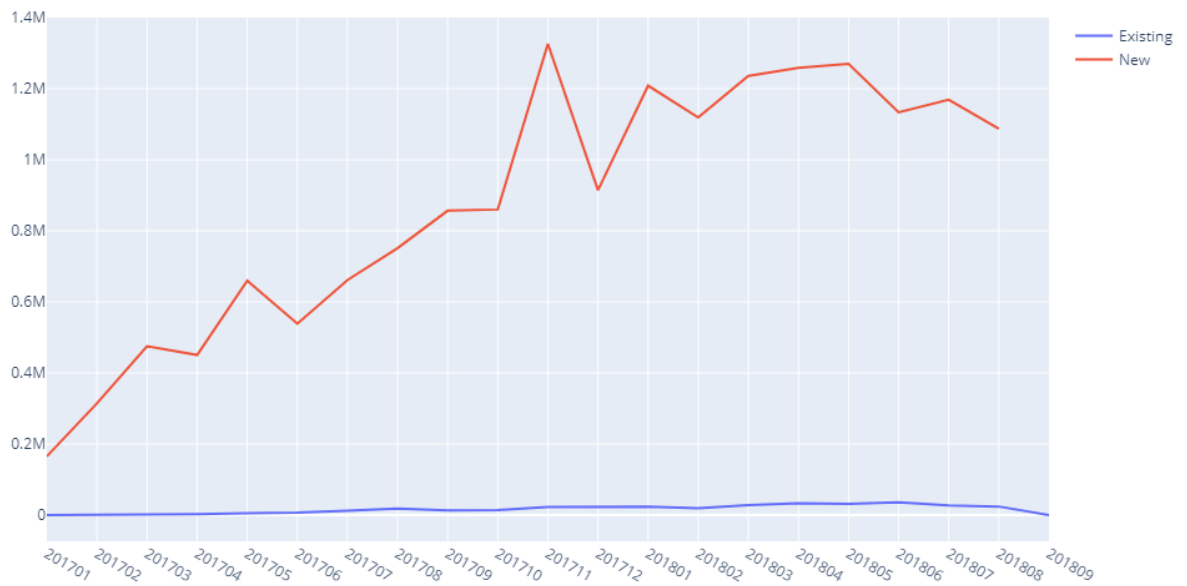
Monthly Retention Rate



Revenue from New Vs Existing Customer

Since the number of new customers is very high compared to the existing ones the same is reflected in the revenue comparison. Most of the revenue has come from the new customers (**17.45 Millions**) compared to existing customers (**0.34 Millions**)

New vs Existing



RATINGS AND REVIEWS

Overall average customer rating for the products is 4.

Rating Clusters



1) Number of Orders by Each Rating group –

Observation –

- high rating high orders low rating low orders → ratings positive relation with no. of orders
- 75% of the total orders received 4 or above rating

2) Number of Sellers by Each Rating group

80% of the sellers have a rating of 4 and above

Observations -

- **Majority of orders (92%)** are fulfilled by sellers who have average rating of 4 or 5. So we can say that Orders and Ratings have a positive correlation. The sellers who are not performing well are not receiving much order.
- **Suggestion:** Company should offer more benefits to these sellers since their contribution is more in terms of quantity as well quality.

3) Number of Categories by Each Rating group

Out of 71 Categories listed on the platform 44 have an average rating of 4 and above. Also there are no category who have average rating between 1 to 2.

Location-wise rating

In terms of popularity, in **70%** of the cities company is **very popular** if we go by average rating score where as in **23%** cities the company has an **average popularity**.

Overall **7%** cities are where company is **not popular** or we can say customers are not happy.

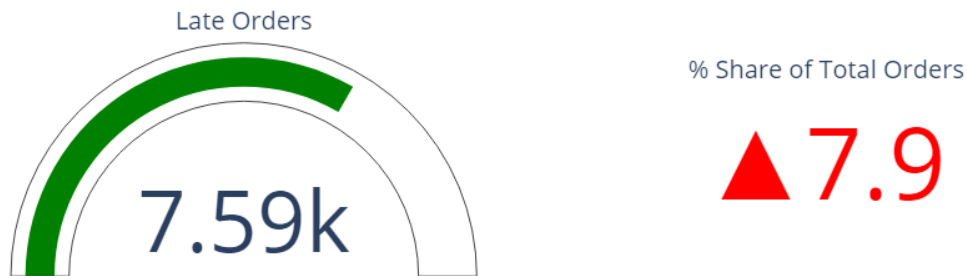
| rating_bucket | Total_City | percent |
|---------------|------------|---------|
| 1 to 2 | 128 | 3.14 |
| 2 to 3 | 169 | 4.15 |
| 3 to 4 | 938 | 23.04 |
| 4 to 5 | 2837 | 69.67 |

Ratings are reflection of customers' happiness. Company should explore the reasons why the rating score is low. Is it because of the quality, price, delays or some other reasons?

Company can make different marketing plans based on the ratings to generate more revenue

DELAYED ORDERS

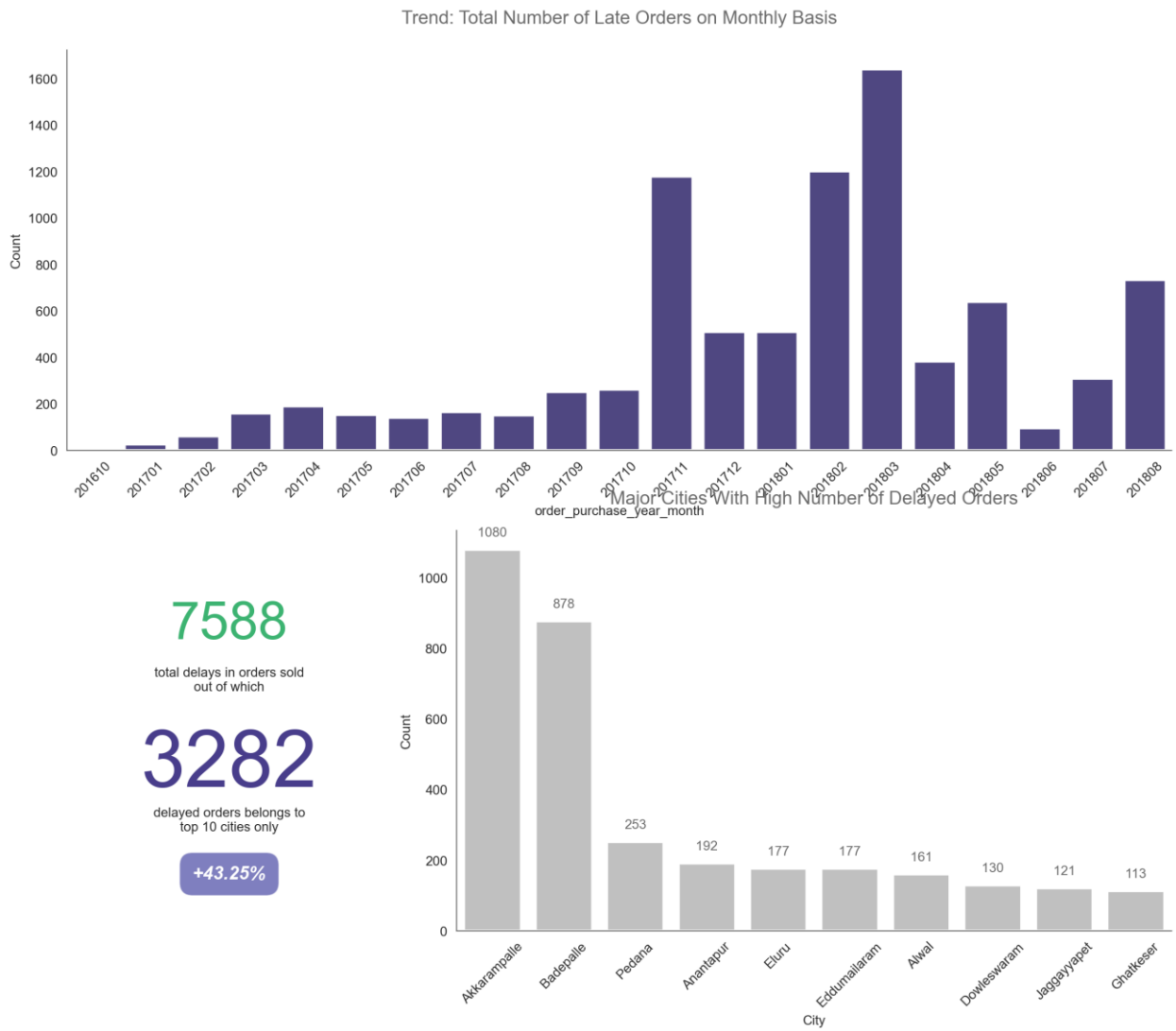
Total number of orders which were delivered late are **7827**. If we look at the Percentage share of Delayed order out of total orders it is **7.87%**



Major delays were found in case of **10 cities** – Akkarampalle, Badepalle, Pedana, Anantapur, Eluru, Eddumailaram, Alwal, Dowleswaram, Jaggayyapet and Ghatkeser.

These cities accounts for 37% of total delay. Important observation here is that **all these cities belongs to one state (Andhra Pradesh)**. Even if we extend our analysis to top 15 cities still it's Andhra Pradesh where all major delays were happened.

Suggestion: Company should do research for reasons of delay, any issues in logistics, any social scuffles, sudden rise in demand, weather calamities, disease outbreak. Although Andhra Pradesh is the State from where the demand is highest so that's the reason delayed orders are also major part of this state but still company needs to control these delays.



Monthly trend:

Sudden rise in delay figure in the month of November, 2017 (reason – black Friday). After a fall in the next two month again rise in delay has been observed in the month of February and March 2018 (reason – Holi).

So overall in November, December, January, February, March, August major delay has happened. These are the months when important festivals in India are celebrated (January – New Year, Makar Sankranti, Lohri, February – Valentine’s Day, March – Holi, August – Raksha bandhan, November – Black Friday, December – Christmas, New Year).

So we can see one of the reason of delay is due to increase in the demand. Company should plan in advance for the additional logistics services to reduce such delays during the time of high demand.

CANCELLED ORDERS

The total number of canceled orders is **554** due to which company lost **Rs131.8k** which is **0.74%** of it's total revenue generated.

Total Amount Lost Due to Cancellations



Total Revenue Loss (%)

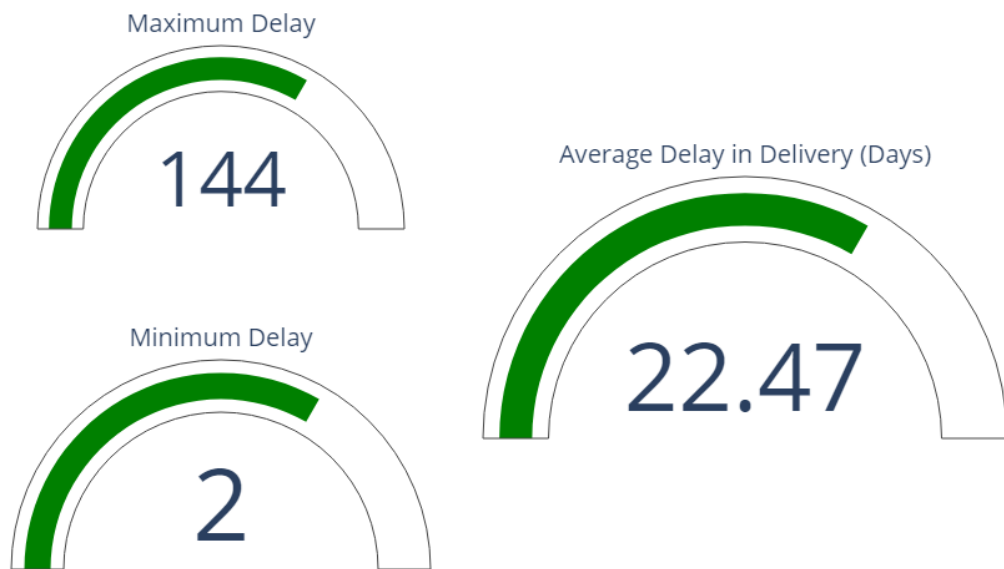


Total Number of Orders Cancelled



Is it because the estimate delivery time is too high?

The average estimated delivery time is **23 days** (approx.) so this might be one of the reasons for the cancellation.



Below is a plot showing the Categories which have high cancellation orders.



FREIGHTS

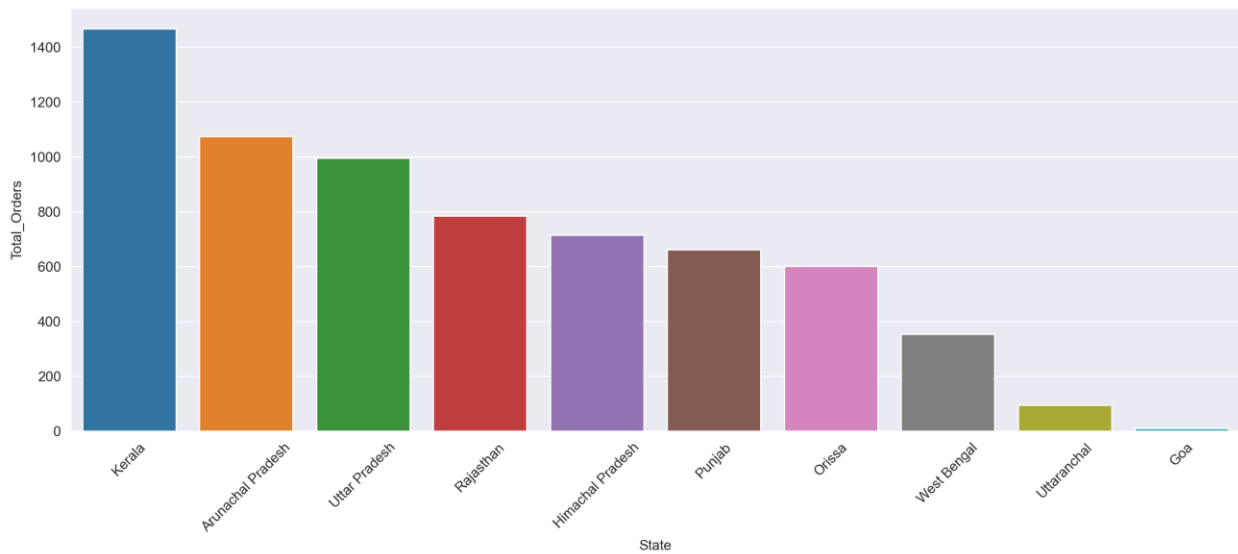
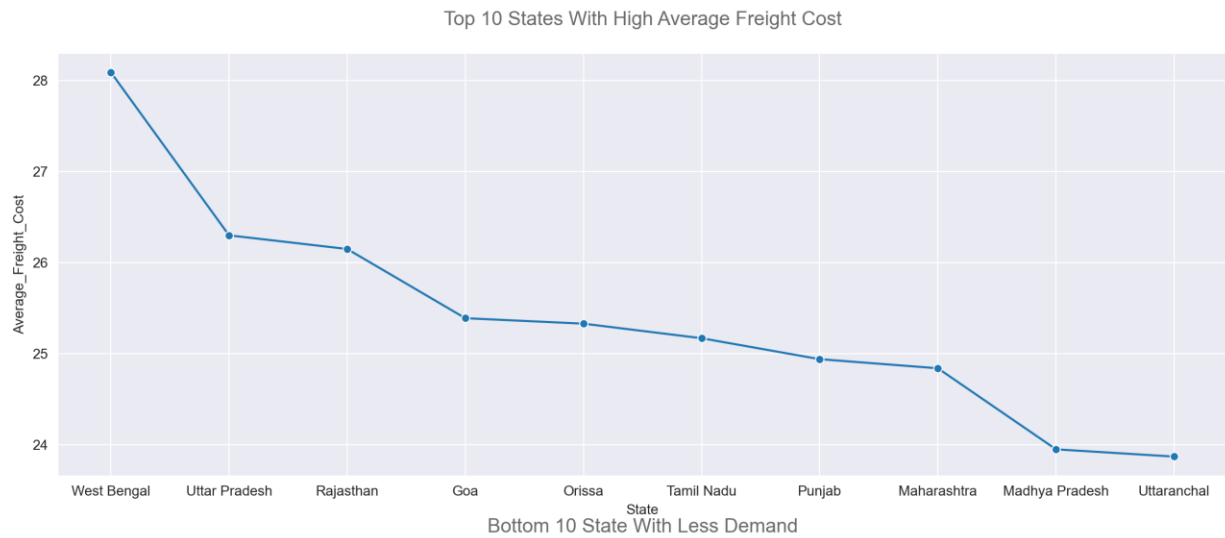
Freight Costs as a Percent of Revenue

Total Freight Costs to Revenue is the starting point when tracking and analyzing your Freight performance from a financial standpoint. Tracking this metric overtime tells you how much of your revenue goes towards your freight shipping and changes in this metric should be closely monitored.

This company's **Total Freight Costs to Revenue** figure is standing out at **12.83%**

Check if the cities with highest freight cost are among the cities where demand are low?

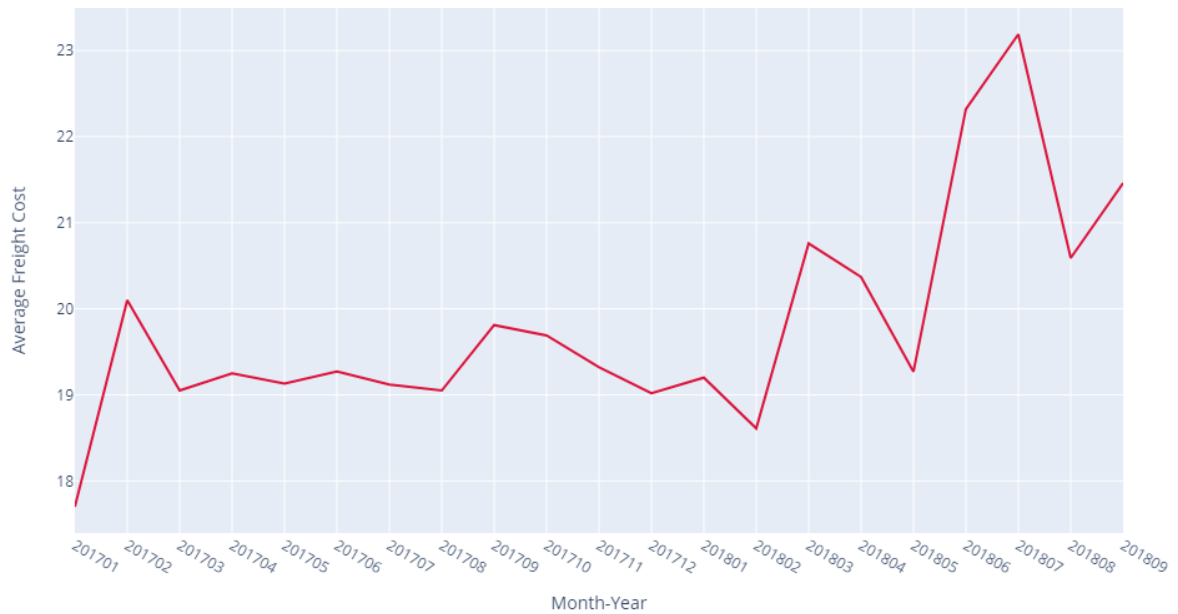
Yes, there is relation between the demand and freight cost. If we see below graph out of 10 states where demand is low 7 states have high average freight cost. The reason for high freight cost can be because of the less demand from these states.



Monthly Trend

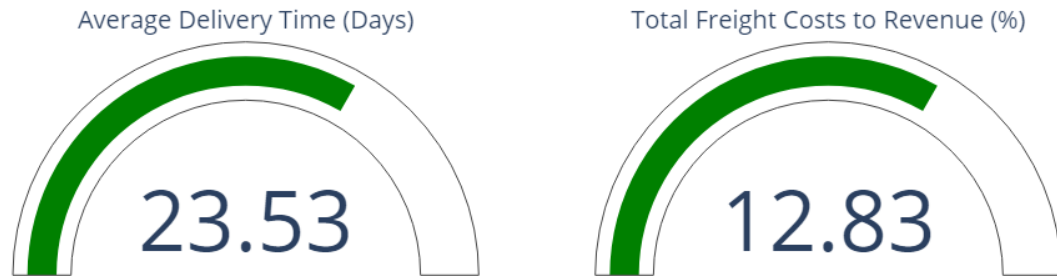
Freight cost has shown an increasing trend since 2017. The highest increase was recorded in July, 2018.

Trend: Monthly Average Freight Cost



Average Delivery Time

The average delivery time of the order is **23.53 Days** which is quite high. Company should try to see the reasons for such high number. This would help in increasing the retention rate as well.



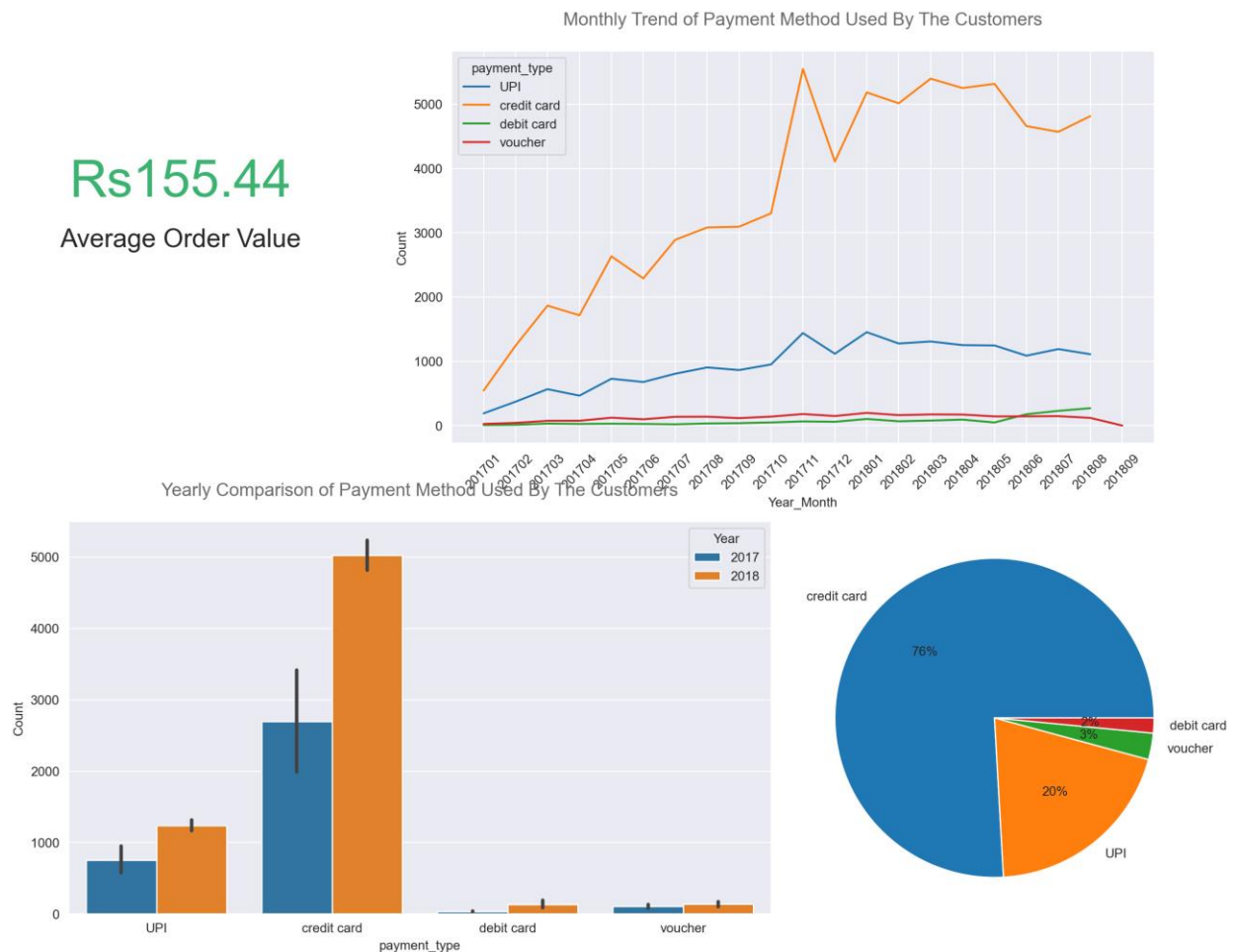
PAYMENTS

The average order value is Rs155.44

Payment methods:

In terms of **transaction counts**, Credit card is the most preferred payment method but debit card is not the favorite choice as it is the least used medium of transaction.

Rs155.44
Average Order Value



Further analysis can be done on customer behavior in using credit card, such as the selected tenure, which product categories are usually purchased with a credit card, etc.

As for credit card, we can see that there was a significant increase from 2017 to 2018. On the other hand, debit card was barely used a medium for payment and the same trend continued in year 2018.

It could be due to promotions/collaborations with certain credit cards as well as a not so attractive promotional methods used for debit card. Further analysis can be done by confirming with other departments, for example marketing or Business Development regarding this.