

Name : Eshwar Anugandula

Roll No: RK20KTB41

Reg. no:12019502

Course: INT 353

**E- Commerce Shipping Data**

**Introduction:**

This dataset is about the products sent by shipping mode. It consists about the information of different means of transport of which product has been delivered, Cost of product, Rating of customer, discount offered, weight of the product, It completely gives information about the details of products have been shipped.

**Reason for choosing this data set:**

An international e-commerce company based wants to discover key insights from their customer database. They want to use some of the most advanced machine learning techniques to study their customers. The company sells electronic products. Now a days everybody is experiencing the door delivery of the products. We can analyse whether the product received with safety, time taken for delivery, cost of shipping, It is easier to know which mode is best to order a particular product. There are 12 columns and 11000 rows which helps me analyse, visualise and interpret the data set in effective manner.

**Data set column description:**

|  |  |  |
| --- | --- | --- |
| 1 | ID | ID Number of Customers. |
| 2 | Warehouse block | The Company have big Warehouse which is divided in to block such as A,B,C,D,E. |
| 3 | Mode of Shipment | The Company Ships the products in multiple way such as Ship, Flight and Road. |
| 4 | Customer care calls | The number of calls made from enquiry for enquiry of the shipment. |
| 5 | Customer rating | The company has rated from every customer. 1 is the lowest (Worst), 5 is the highest (Best). |
| 6 | Cost of the Product | Cost of the Product in US Dollars. |
| 7 | Prior purchases | The Number of Prior Purchase. |
| 8 | Product importance | The company has categorized the product in the various parameter such as low, medium, high. |
| 9 | Gender | Male and Female |
| 10 | Discount offered | Discount offered on that specific product. |
| 11 | Weight in grams | It is the weight in grams. |
| 12 | Reached on time | It is the target variable, where 1 Indicates that the product has NOT reached on time and 0 indicates it has reached on time. |

**Insights from the data set:**

Importing necessary imports from library

Loading data

Finding no of rows and columns

Description of data

Information of columns

Checking null values

Data cleaning by dropping unwanted columns

heatmap of the data for checking the correlation between the features and target column

**Checking value counts of categorical columns**

Exploring relation of categorical columns with reached on time or not

looking at the warehouse column and what are the categories present in it

making a count plot of warehouse column and see the effect of Reached on time or not on the warehouse column.

looking at the gender column and what are the categories present in it

making a count plot of gender column and see the effect of Reached on time or not on the warehouse column.

making a count plot of mode of shipment column and see the effect of Reached on time or not on the warehouse column.

looking at the product importance column and what are the categories present in it

making a count plot of product importance column and see the effect of Reached on time or not on the warehouse column.

looking at the customer ratings column and what are the categories present in it

making a count plot of prior purchases column and see the effect of Reached on time or not on the warehouse column.

making a dist. plot of cost of the product column

making a dist. plot of discount offered column

## Which mode of shipment carries most weights?

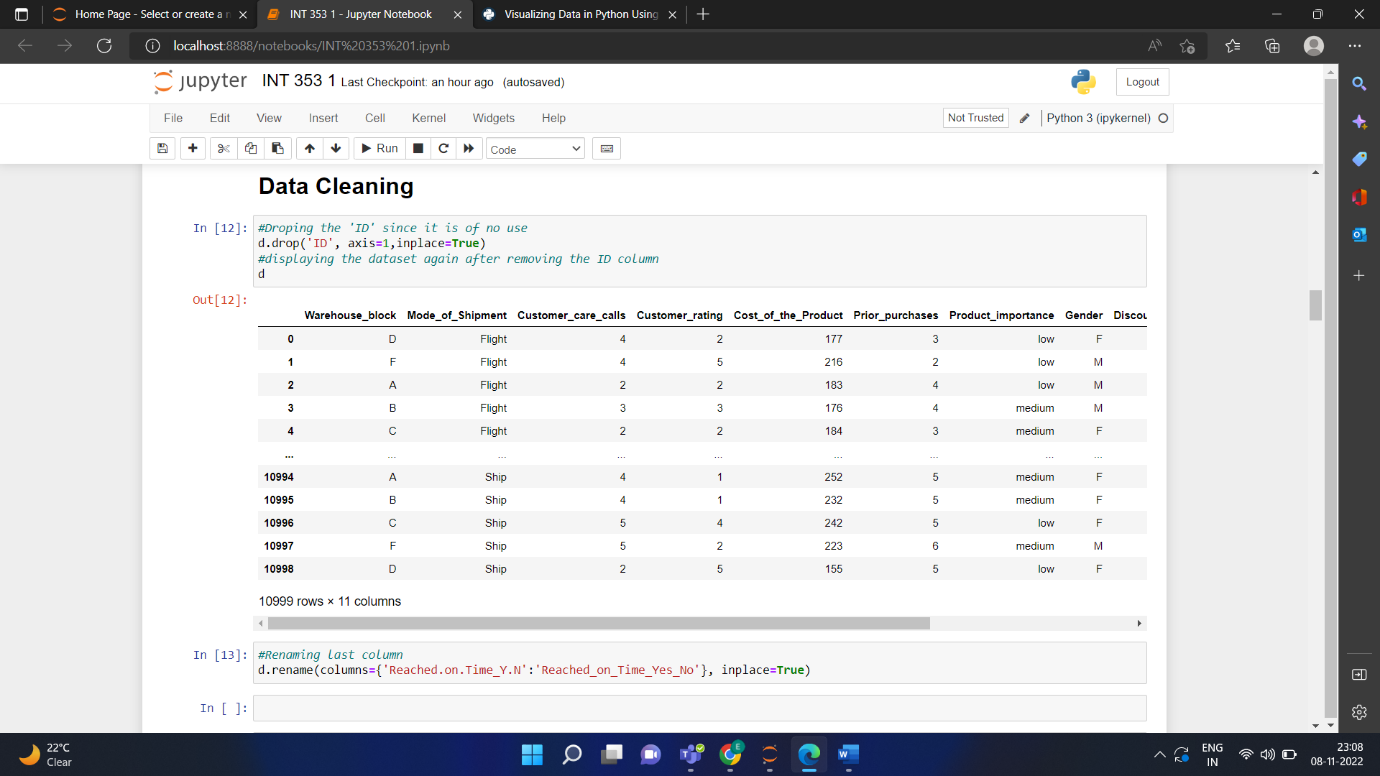
### Effect of Warehouse on Cost of Product

### Does Mode of Shipment effect Cost of Product?

### the relation between cost of the product and the discount offered and the relation with whether or not the product will reach on time.

conclusion

Data Cleaning:



A screenshot of a computer

Description automatically generated

Univariate Analysis:

Categorical features:

Icon

Description automatically generated

From the above bar graph we can see that most of the shippings are done by “Ship” that is in water ways Chart, bar chart

Description automatically generated

The above bar graph shows in percentage. About 68% of the shipments are done by ship and 16% of shipments are done by both roads and flights

Chart, pie chart

Description automatically generated

In this pie chart we can clearly depict that most of the shipments are delivered to block “F” (33.33%) and rest of the blocks has equal weightage.



In this bar graph it is clear that count of female’s shipments are slightly higher than male’s.  
  
  
  
A picture containing logo

Description automatically generated

In this bar graph we can conclude that most of the products are reached on time but the count of the products which are not reached on time is also high.

Chart, pie chart

Description automatically generated

In this pie chart it is clear that most of the products ordered where low cost, and medium cost products. Only 8.62% of the high cost products were shipped.

### Numerical features

Chart, histogram

Description automatically generated

The histogram clearly shows that most of the products ordered were in between the range of $250 to $255 where as least no products ordered were around $100 to $110

Chart

Description automatically generated

This is the density plot for cost of the products where can see that density is high at $250 to $300.

Chart

Description automatically generated

The no of customer care call received are 4 with highest frequency.

Icon

Description automatically generated with medium confidence

This is the density plot for customer care calls

Chart, bar chart

Description automatically generated

Highest prior purchases are 7 and lowest prior purchases are 3.

Chart, pie chart

Description automatically generated

From the above pie chart we can conclude that rating of “3” is given by most of the customers.

Box plot for the “Cost of the product”

Chart

Description automatically generated

count 10999.000000

mean 210.196836

std 48.063272

min 96.000000

25% 169.000000

50% 214.000000

75% 251.000000

max 310.000000

Name: Cost\_of\_the\_Product, dtype: float64

Boxplot for weight in grams

Chart

Description automatically generated

count 10999.000000

mean 3634.016729

std 1635.377251

min 1001.000000

25% 1839.500000

50% 4149.000000

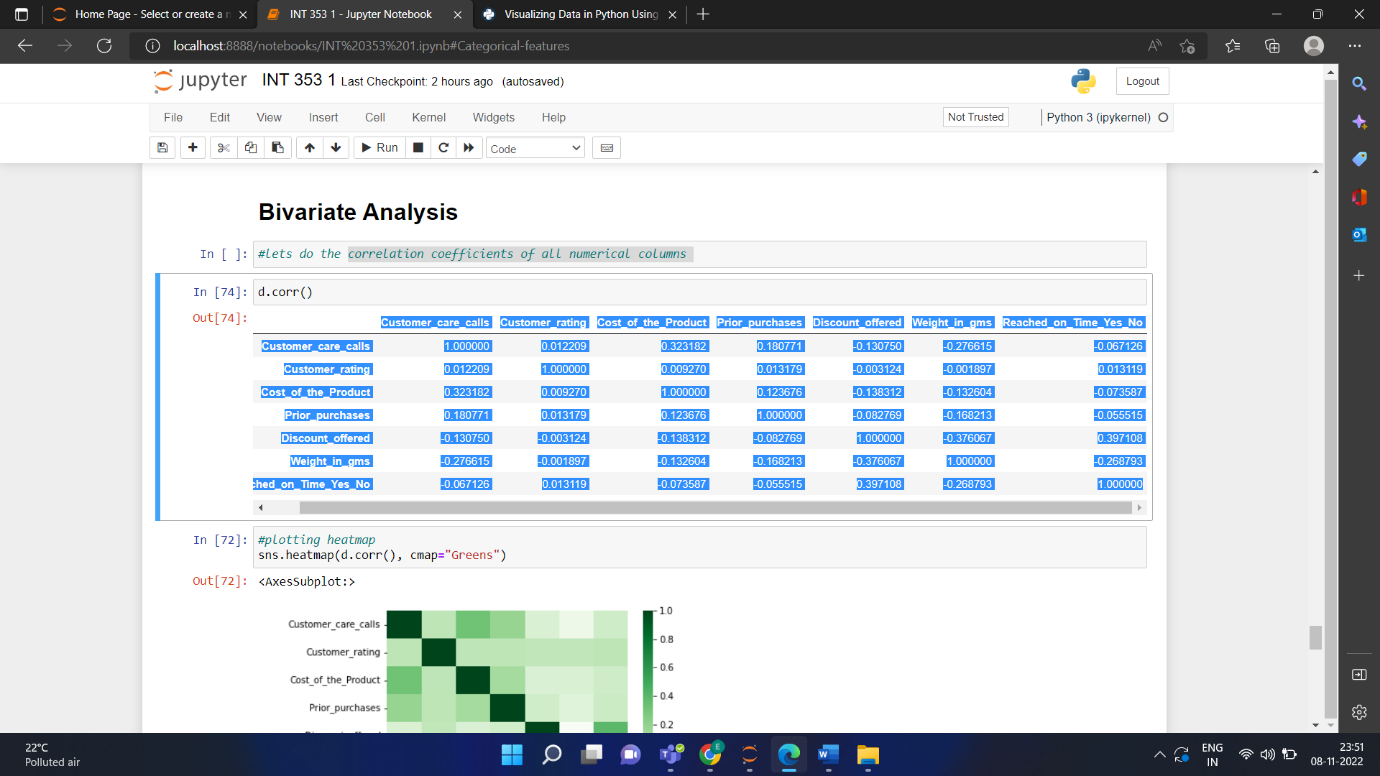
75% 5050.000000

max 7846.000000

Name: Weight\_in\_gms, dtype: float64

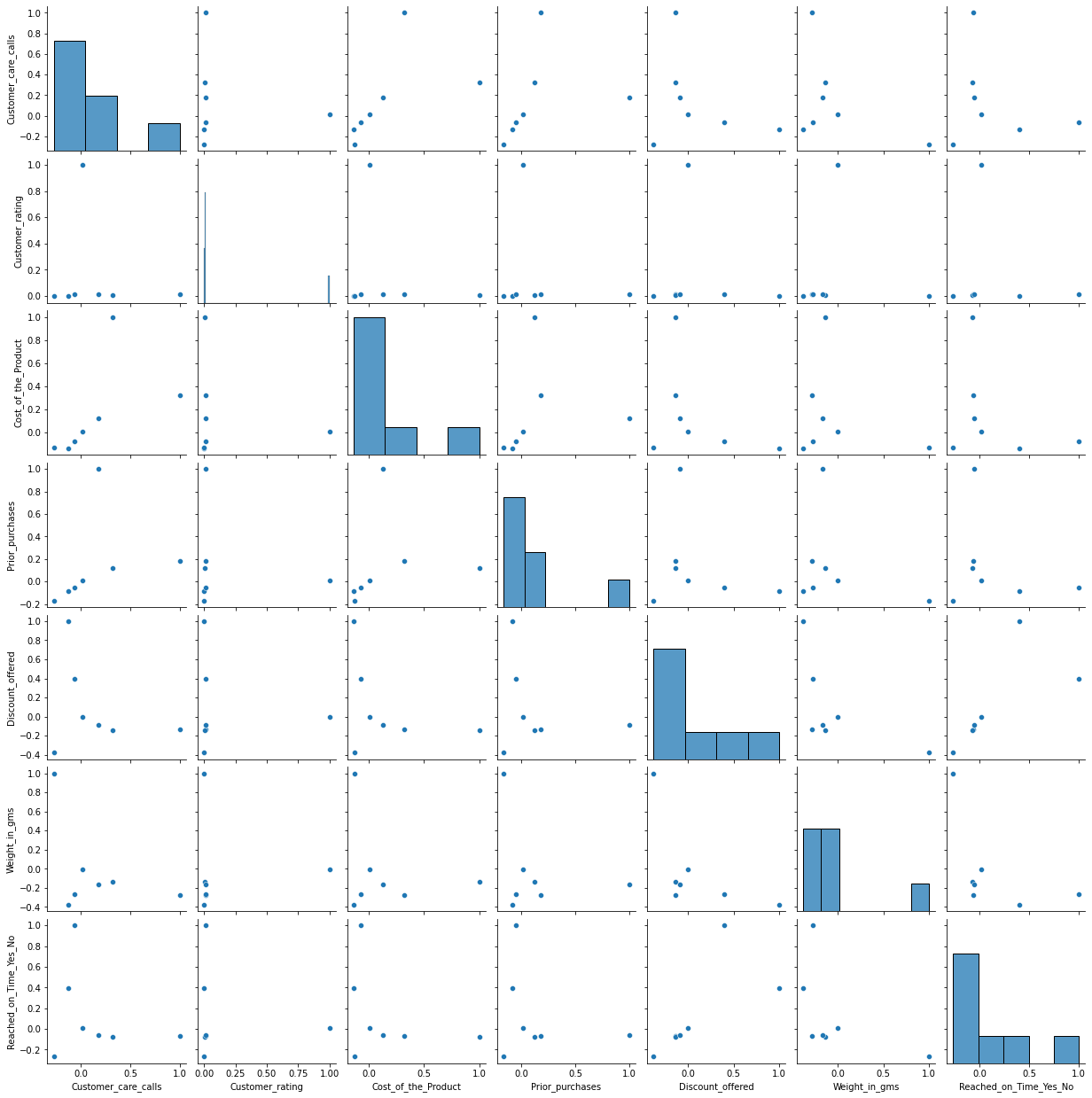
# Bivariate Analysis

correlation coefficients of all numerical columns:



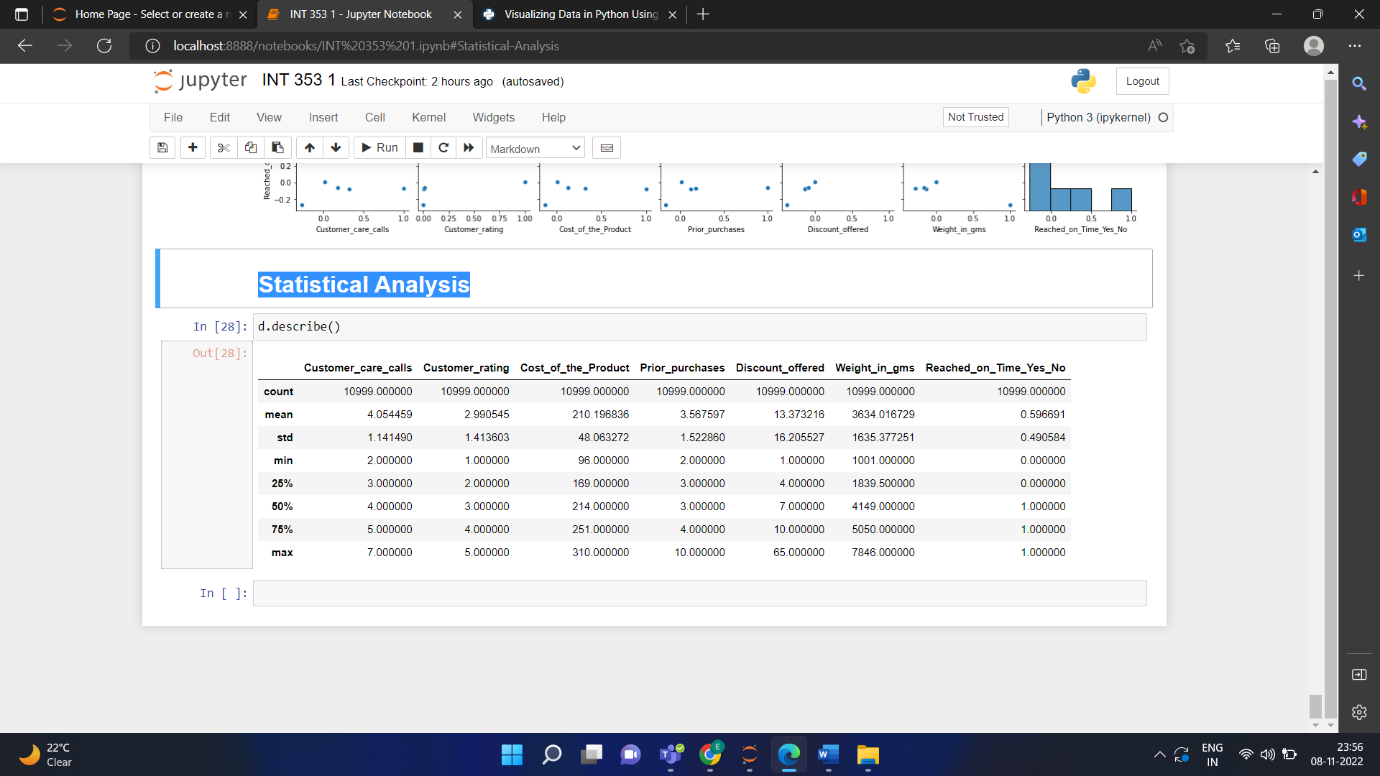
Heatmap for corelation coefficients:Square

Description automatically generated with medium confidence



This is pair plot for ear and every column of the data set. A very few graphs shows linear regression.

Statistical Analysis:



Multivariate Analysis:

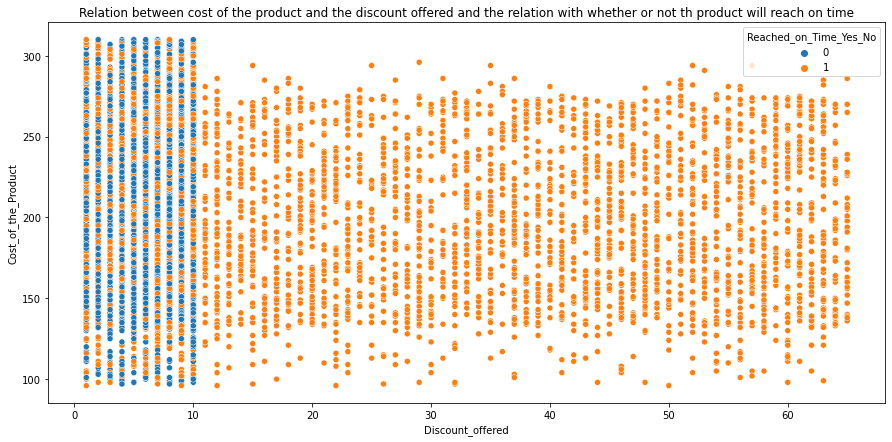
Correlation matrix for the columns in the data set

A picture containing background pattern

Description automatically generated

We can clearly depict that find that in most of the cases the correlation is almost equal to zero or negative. There is no cases where it showed positive correlation.

Scatter plot to see the relation between cost of the product and the discount offered and the relation whether the product reached on time



From the above graph we can conclude that discount offered between 0%-10% were only reached on time and rest of them were not reached time.

Profile plot:

Chart

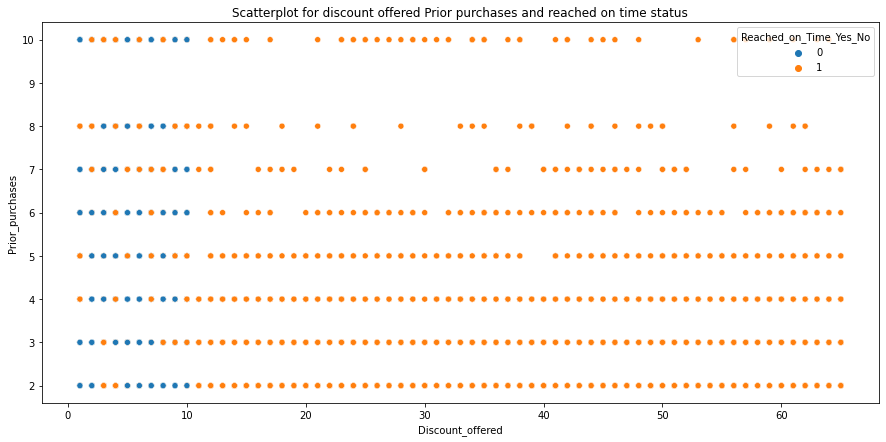
Description automatically generated

profile plot showing between customer care calls, customer calls along with Reached on time or not.

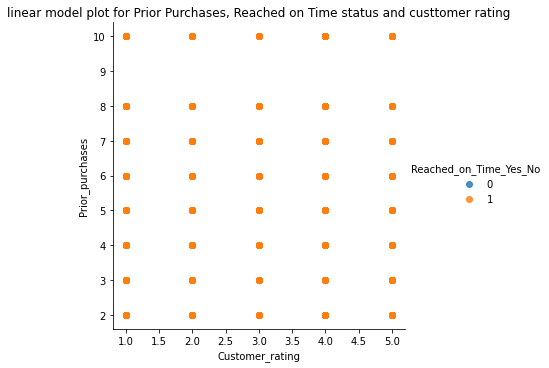


From the above scatterplot it is clear that discount offered between 0%-10% were only reached on time and rest of them were not reached time.

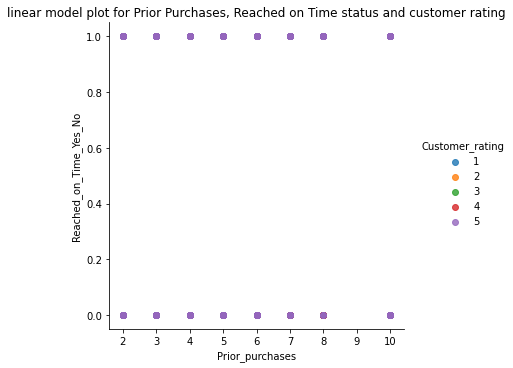
It is clear that the discount offer between 0 &10 where most of the products reached on time has customer rating as 4.0



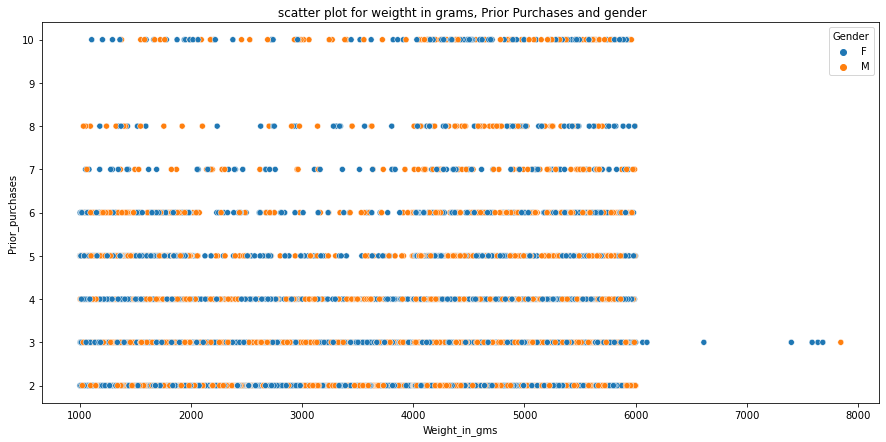
It is clear that there were zero number of prior purchases where prior purchases count is equla to 9.



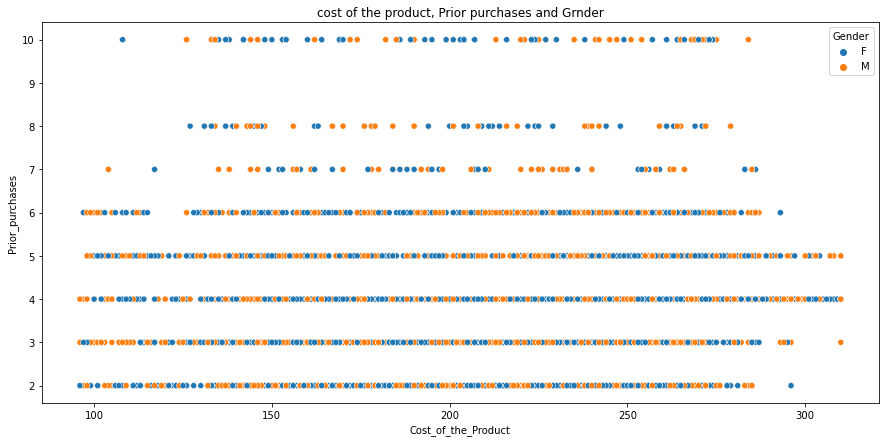
All the products which are were reached late irrespective of Prior purchases and customer rating.



Customer rating is 5 either it is reached on time or not reached.



scatter plot for weight in grams, Prior Purchases and gender



Scatterplot for cost of the product, Prior purchases and Gender.

Prediction Analysis:

The rating is low due to most of the products did not reached on time. So in order to increasing rating products must be reached on time. Prior purchases playing a crucial role most the people whose chosed shipping were actually the onces who ordered the products before. There are very few people who ordered for first time.

Conclusion:

Most of the shipping is done by “Ship” that is in water ways. About 68% of the shipments are done by ship and 16% of shipments are done by both roads and flights. The Count of female’s shipments are slightly higher than male’s shipments.  
The count of the products which are not reached on time is also high. Most of the products ordered where low cost, and medium cost products. Only 8.62% of the high-cost products were shipped. Most of the products ordered were in between the range of $250 to $255 where as least no products ordered were around $100 to $110.The no of customer care call received are 4 with highest frequency. Highest prior purchases are 7 and lowest prior purchases are 3. Rating of “3” is given by most of the customers. A very few graphs showed linear regression.