PROJECT REPORT

Name:	**Vinayak Soni**
Email ID:	**vinaysoni2310@gmail.com**
Phone no:	**8369944151**

Problem Statement:

Students need to analyze supermarket sales across different branches and provide insight to understand the customer better. They need to predict the gross income according to the cost of goods sold and the rating given by the customer.

Statistical Informations:

Shape – 1000 X 17

1. For Numerical columns:

	Unit price	Quantity	Tax 5%	Total	cogs	gross marg	gross inco	Rating
count	949.000000	949.000000	949.000000	949.000000	949.000000	9.490000e+02	949.000000	949.000000
mean	55.582055	5.487882	15.292389	321.140165	305.847777	4.761905e+00	15.292389	6.981454
std	26.439731	2.932991	11.682345	245.329247	233.646902	8.886467e-16	11.682345	1.723918
min	10.080000	1.000000	0.508500	10.678500	10.170000	4.761905e+00	0.508500	4.000000
25%	32.900000	3.000000	5.832000	122.472000	116.640000	4.761905e+00	5.832000	5.500000
50%	54.920000	5.000000	12.036000	252.756000	240.720000	4.761905e+00	12.036000	7.000000
75%	77.720000	8.000000	22.428000	470.988000	448.560000	4.761905e+00	22.428000	8.500000
max	99.960000	10.000000	49.650000	1042.650000	993.000000	4.761905e+00	49.650000	10.000000

2. For Categroical Columns:

	Branch	City	Gender	Gender	Product lin	Date	Payment
count	949	949	949	949	949	949	949
unique	3	3	2	2	6	54	3
top	A	Yangon	Member	Male	Fashion ac	#########	Cash
freq	322	322	480	475	172	392	330

Pre-processing:

ELITE TECHNO® GROUPS
Inspiring Young Generation

1. Coverted date time object into Date time format

df['Time'] =pd.to_datetime(df['Time'], format = '%H:%M’)

2. And from that extracted Hours for EDA and made a new column of Hour

df["Hour"] = df['Time'].dt.hour

3. Dropped all the nan values as ratio of nan value is very less

df.dropna(axis=0,inplace=True)

4. Used Standard scalar for scaling of Data as some of the models like linear Regression works on distcance matrix.

```
scaler = StandardScaler().fit(X)
rescaled_X_train = scaler.transform(X)
```

Exploratory Data Analysis (EDA):

1. From the insights we find that the number of branches are almost equal, and each branch corresponds to one city.

2. There's a noticeable difference in Sales as Branch C which also belongs to Naypyitaw City does the highest sales. But still the sales is almost similar in all branches.

3. Gross income generated by females is slightly more than males as wells the gorss income generated by C branch is comparatively more than other 2 branches.

4. The rating given is also similar among Male and Females across different city and branch.

5. Payment mode of Credit card is slightly less used than other 2 modes.

6. Cash payment is preferred method of payment in Naypyitaw city whereas Ewallet is mostly used in Yangon city.

7. The highest buying product by females is Fashion and least bought is Health products, whereas there's no significant difference in buying of males.

8. In Yangon city, the lowest selling product is of Health and Highest selling product is of Home and Sports Category.

9. In Naypyitaw city, the lowest selling product is of Home and Sports and Highest selling product is of Food and Fashion Category.

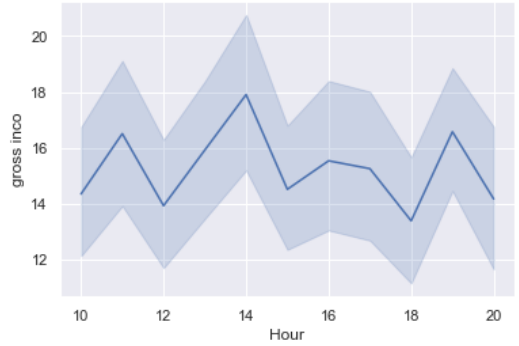
10. In Mandalay city, the lowest selling product is of Health, Home, Food and Highest selling product is of Fashion and Sports Category.

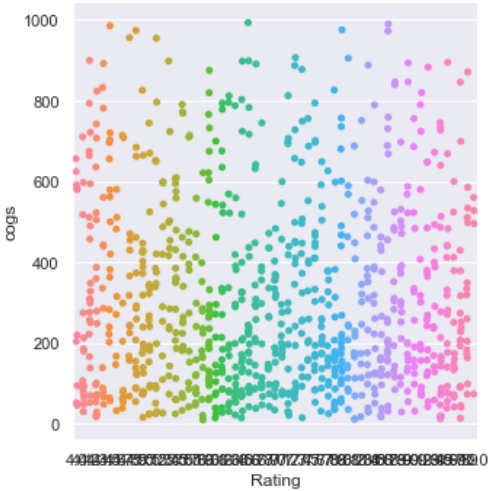
11. Sales is highest at 14 hour i.e 2pm.

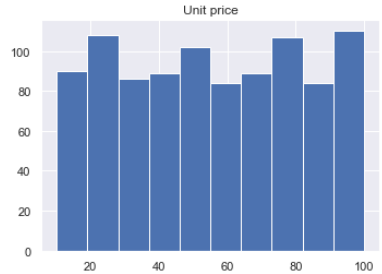
12. Majority of customers Total lies between 10-200, so as gross income and cogs. The majority of the products has the unit price of 25 or 95. Ratings given by customers is almost similar whereas majority of ratings received are of 4.5 and 9.6.

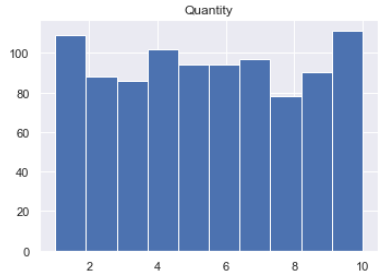
Data Visualization:

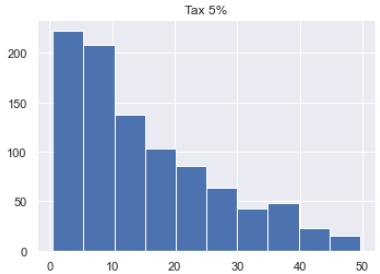
ELITE TECHNO® GROUPS
Inspiring Young Generation

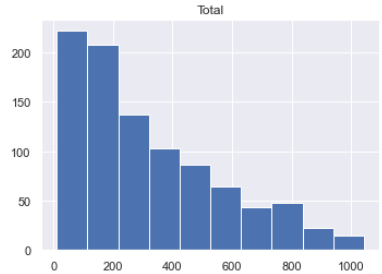


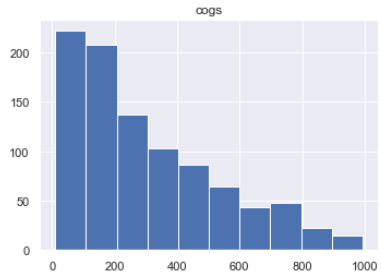


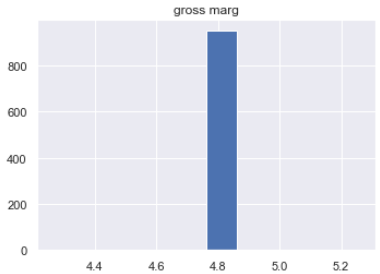


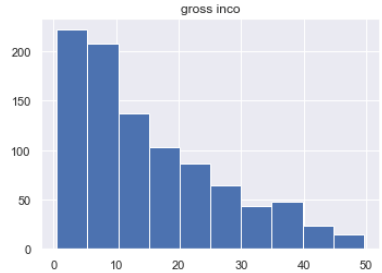


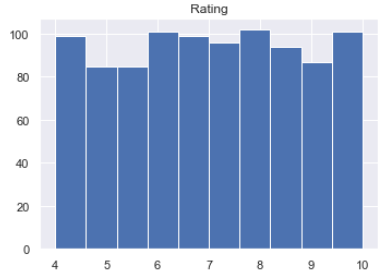


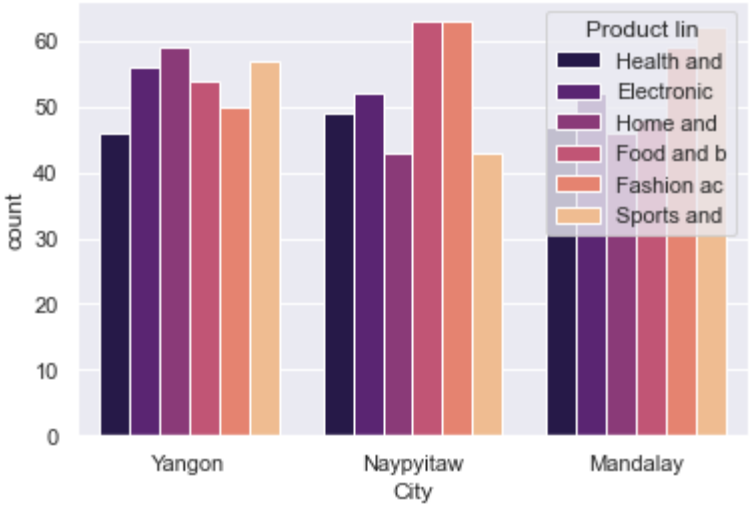


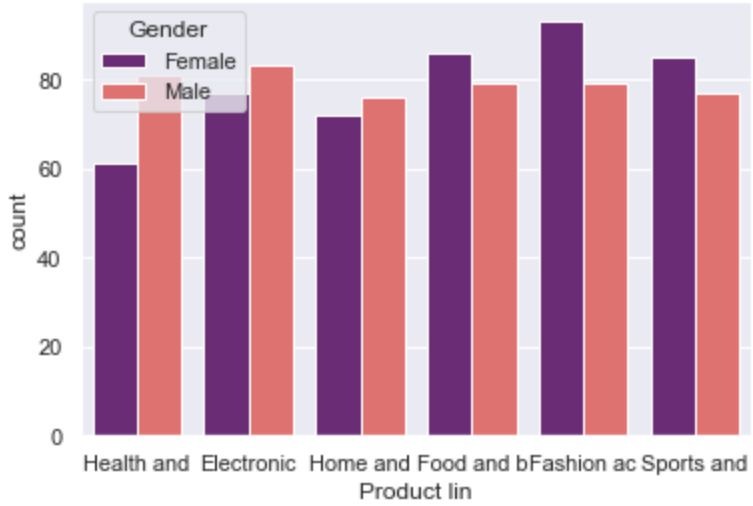


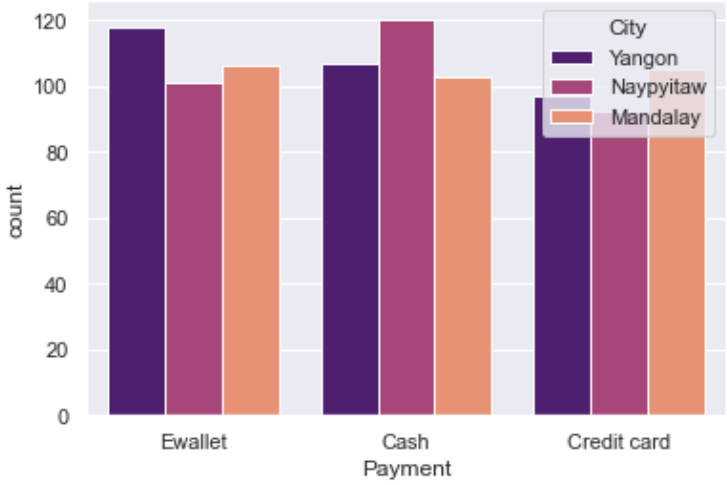


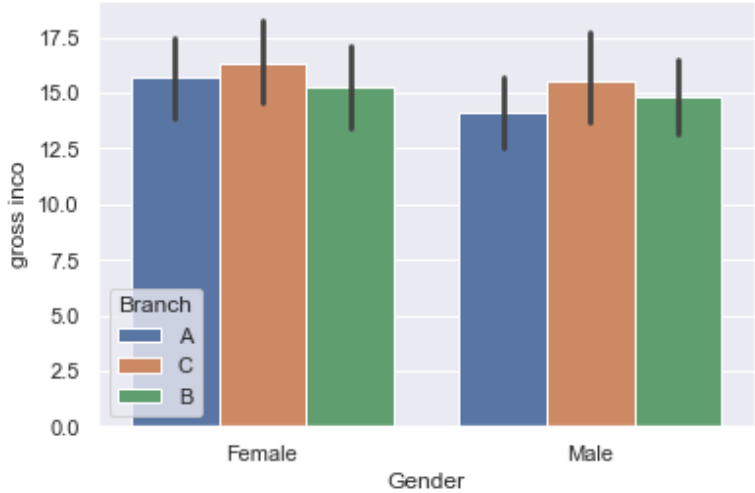


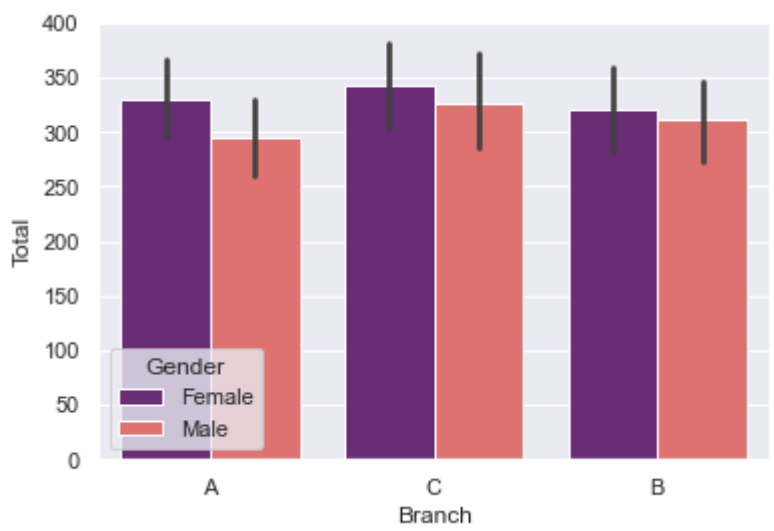


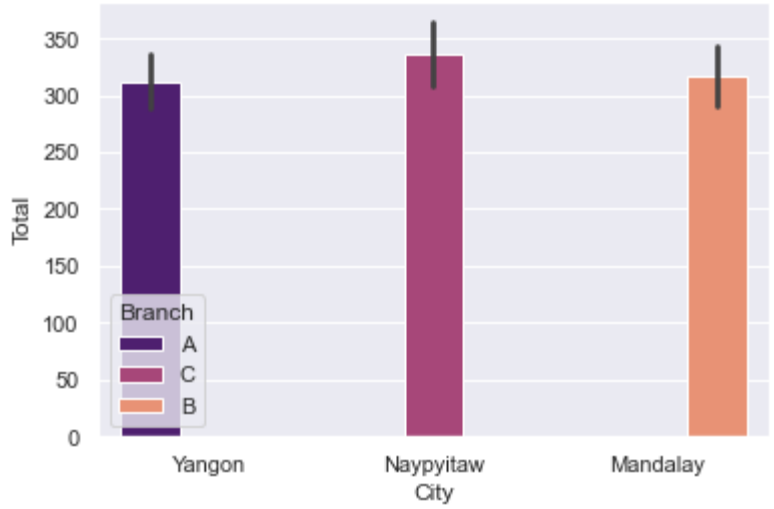

Product lin
Health and
Electronic
Home and
Food and b
Fashion ac
Sports and
count
0
10
20
30
40
50
60
Yangon
Naypyitaw
Mandalay
City

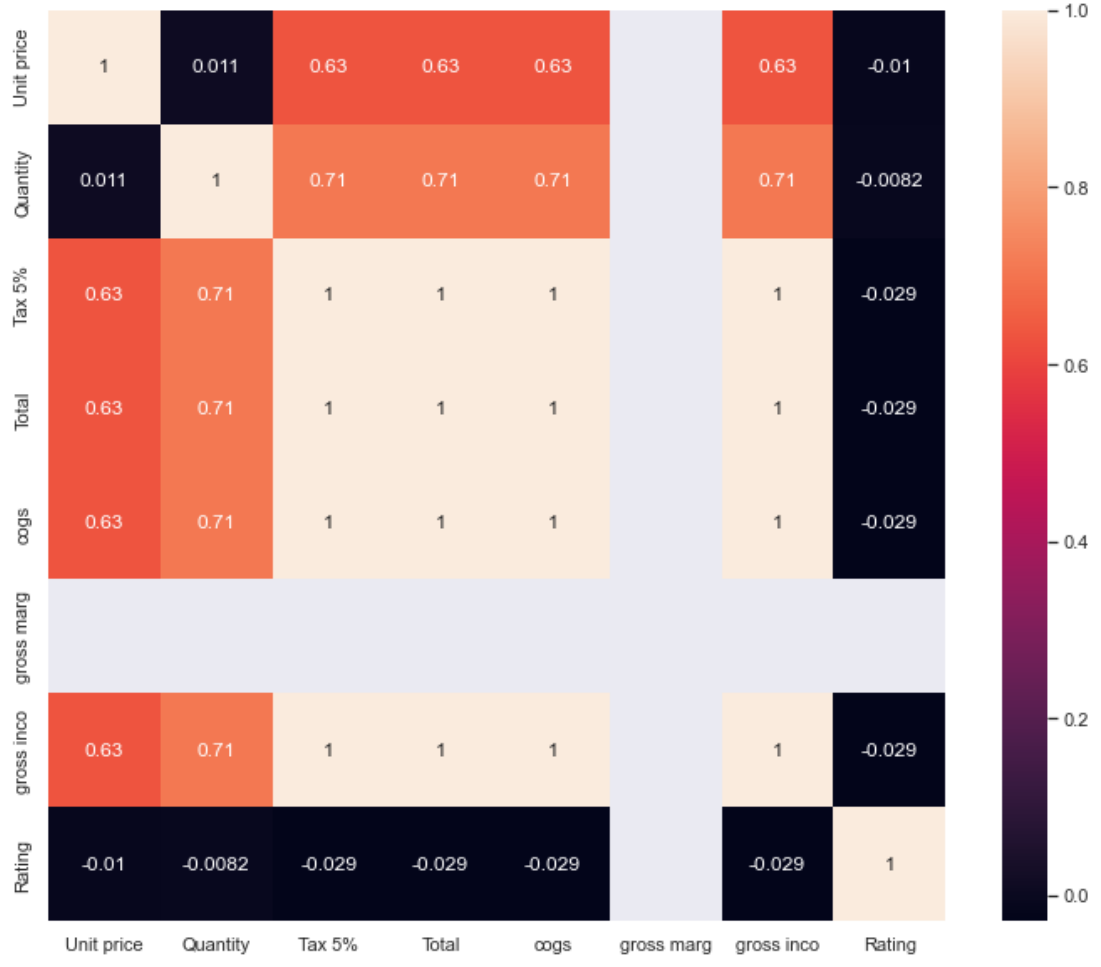

Gender
Female
Male
count
0
20
40
60
80
Health and
Electronic
Home and
Food and b
Fashion ac
Sports and
Product lin

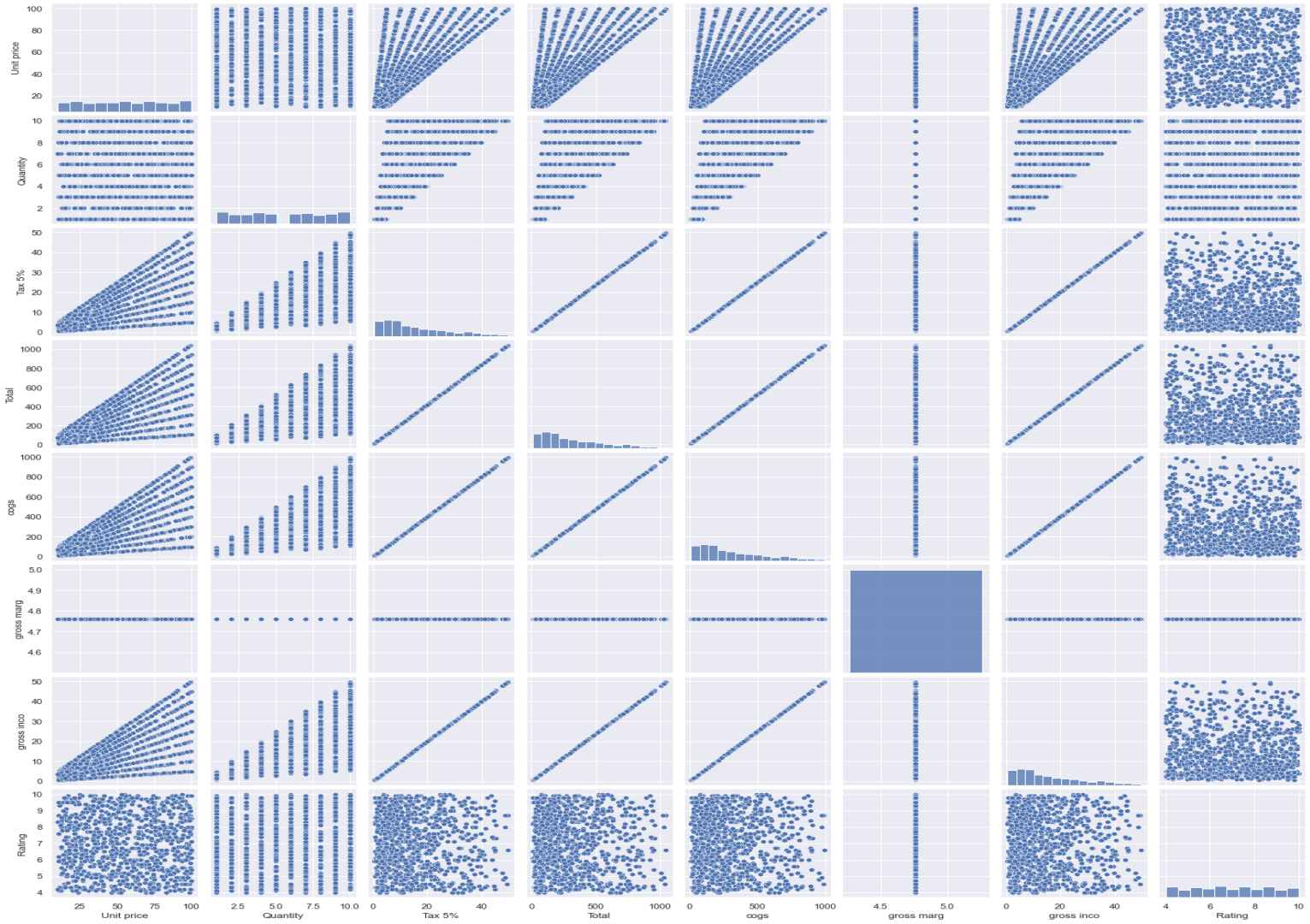

City
Yangon
Naypyitaw
Mandalay
count
0
20
40
60
80
100
120
Ewallet
Cash
Credit card
Payment


Branch
A
C
B
gross inco
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Female
Male
Gender


Total
Gender
Female
Male
A
C
B
Branch


Total
Branch
A
C
B
Yangon
Naypyitaw
Mandalay
City


Unit price
Quantity
Tax 5%
Total
cogs
gross marg
gross inco
Rating
1 0.011 0.63 0.63 0.63 0.63 -0.01
0.011 1 0.71 0.71 0.71 0.71 -0.0082
0.63 0.71 1 1 1 1 -0.029
0.63 0.71 1 1 1 1 -0.029
0.63 0.71 1 1 1 1 -0.029
0.63 0.71 1 1 1 1 -0.029
-0.01 -0.0082 -0.029 -0.029 -0.029 -0.029 1
1.0
0.8
0.6
0.4
0.2
0.0


Unit price
Quantity
Tax 5%
Total
cogs
gross marg
gross inco
Rating
Unit price
Quantity
Tax 5%
Total
cogs
gross marg
gross inco
Rating

Accuracies of each model:

ELITE TECHNO® GROUPS
Inspiring Young Generation

Linear Regression Model

MSE: 5.46164214817704e-29 RMSE: 7.390292381345301e-15

Decision Tree Model

MSE: 0.009595282894736797 RMSE: 0.09795551487658465

Random Forest Model

MSE: 0.003053991242631348 RMSE: 0.055262928284984573

Inference:

1.Popular payment method used by customers:- **Cash Payment**

2. Does gross income affect the ratings that the customers provide? :- **No, it doesn’t.**

3. Which branch is the most profitable? :**- Branch C**

4. Is there any relationship between Gender and Gross income? **:- No, but from females its slightly higher.**

5. Is there any time trend in gross income? **:- Yes**

6. What is the spending pattern of females and males and in which category do they spend a lot?
:- Females on Fashion products spends high, but for males no specific difference.

7.How many products are bought by customers? :**-Generally 9-10 products**

8. Which city should be chosen for expansion and which products should it focus on?
:- Naypyitaw City as the sales from this city is highest and some of the products like Home and Sports lack sales so we improve the sales from these products and generate higher revenue.
9. Which hour of the day is the busiest? :- **14 pm is the busiest.**

Conclusion:

As we can see from the chart, some of the products does very well but certain products are lacking sales and dues to this gross income generated is also. If we focus on the products which are not giving much sales in their respective cities and find a way to increase the sales of these products such as Home and Sports product in Naypyitaw city, health products in Yangon city we can increase the gross income drastically.