PA3_2

G01337418

Vikas chand Gunnam

The dataset chosen is supermarket sales obtained from Kaggle. The dataset is thoroughly analyzed and the following operations are done by the following columns below and division operation is done. The result is saved in the combination of first two names of the column data.

```
In [25]: df["UNQU"] = df["Unit price"] / df["Quantity"]
         df["TATO"] = df["Tax 5%"] / df["Total"]
         df["GRTO"] = df["gross margin percentage"] / df["Total"]
         df["COTO"] = df["cogs"] / df["Total"]
         df["GRGR"] = df["gross margin percentage"] / df["gross income"]

         df
```
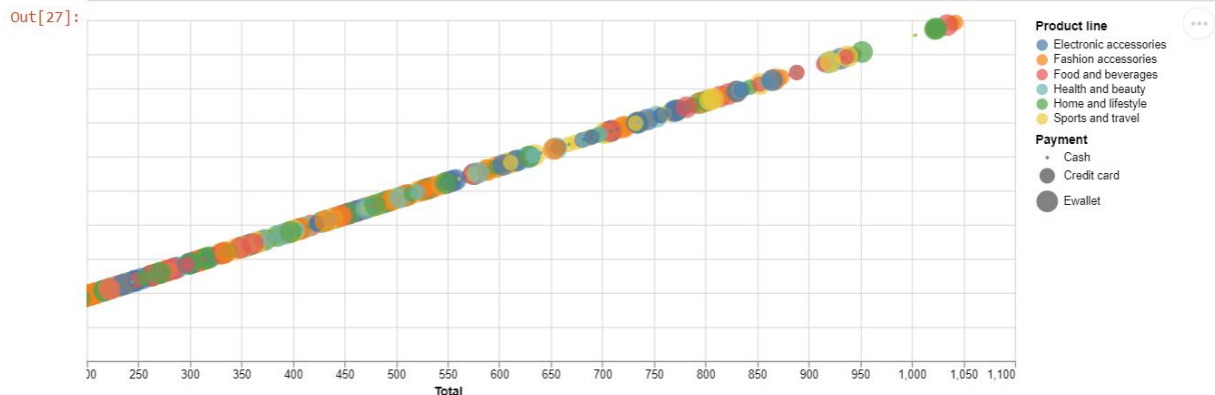
Later few graphs are plotted.

Tax 5% VS Total

The graph is plotted for tax of the product and total price of the products.The size of the points is representing which type of payment.
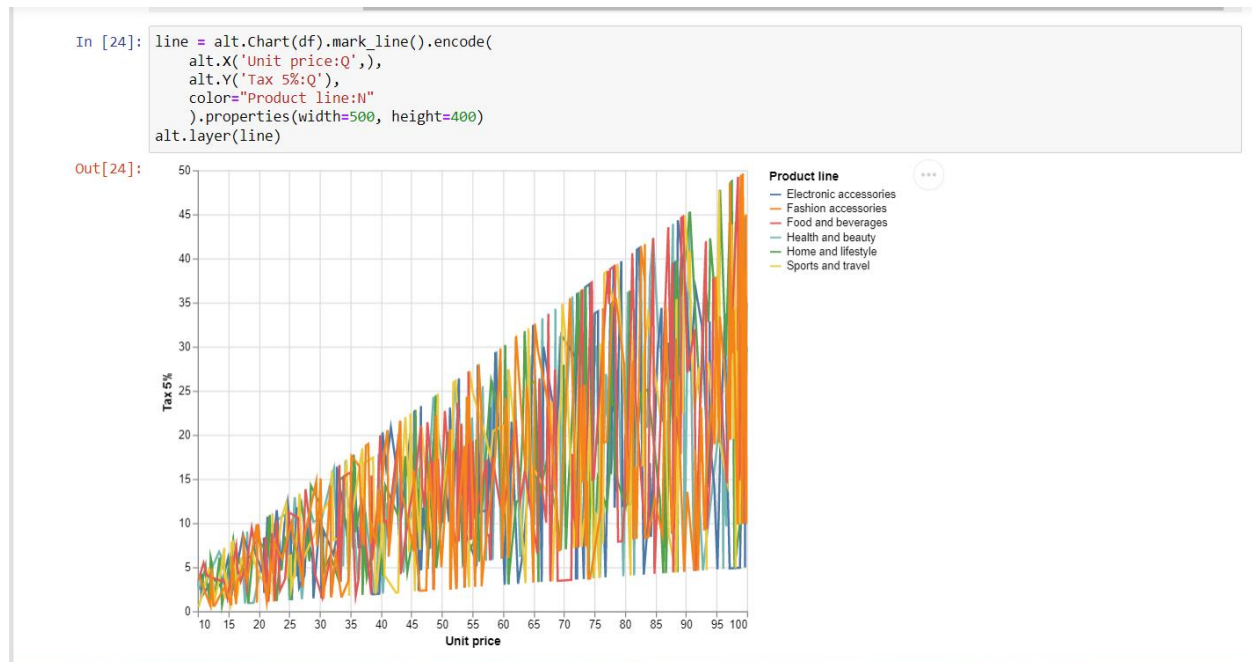
```
In [26]: import altair as alt
         from altair import datum
```

```
In [27]: alt.Chart(df).mark_circle().encode(
             x='Total:Q',
             y='Tax 5%:Q',
             size='Payment:N',
             color='Product line:N'
         ).properties(width=1000)
```

Out[27]:

Tax vs Unit price

The graph is plotted between tax and unit price of the product.It is clearly evident that the unit price is varied according to product line.
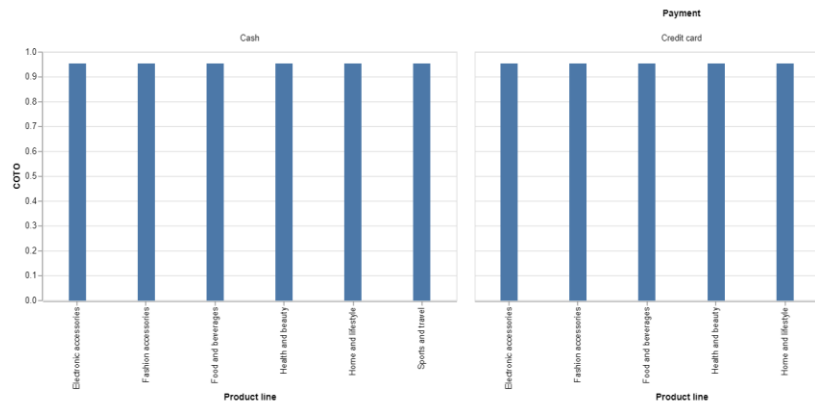
```
In [24]: line = alt.Chart(df).mark_line().encode(
             alt.X('Unit price:Q',),
             alt.Y('Tax 5%:Q'),
             color="Product line:N"
             ).properties(width=500, height=400)
         alt.layer(line)
```


COTO VS PRODUCT LINE(Payment mode)

The graph is plotted for coto for different type of payments made and their corresponding products.

Payment

Cash | Credit card

COTO

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Electronic accessories | Fashion accessories | Food and beverages | Health and beauty | Home and lifestyle | Sports and travel | Electronic accessories | Fashion accessories | Food and beverages | Health and beauty | Home and lifestyle
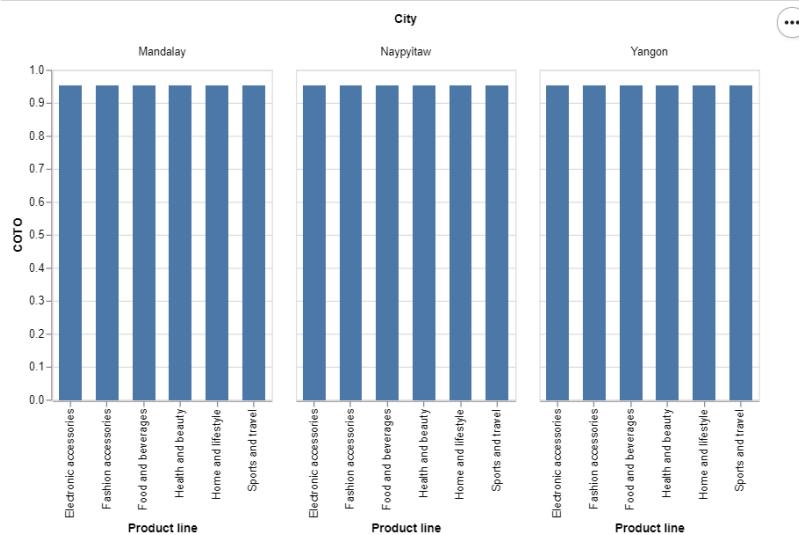
Product line

COTO VS PRODUCT LINE(City)

The graph is plotted for coto and cities and their corresponding products.

```
In [59]: alt.Chart(df).mark_bar(size=20).encode(
             alt.X("Product line:N"),
             alt.Y("COTO:Q"),
             column="City:N",
         ).properties(width=200)

Out[59]:
```

City

Mandalay | Naypyitaw | Yangon

COTO

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0

Electronic accessories | Fashion accessories | Food and beverages | Health and beauty | Home and lifestyle | Sports and travel | Electronic accessories | Fashion accessories | Food and beverages | Health and beauty | Home and lifestyle | Sports and travel | Electronic accessories | Fashion accessories | Food and beverages | Health and beauty | Home and lifestyle | Sports and travel

Product line

UNQU VS PRODUCT LINE(PAYMENT)

The graph is plotted for UNQU and payments and their corresponding products.

```
alt.Chart(df).mark_bar(size=20).encode(
    alt.X("Product line:N"),
    alt.Y("UNQU:Q"),
    column="Payment:N",
).properties(width=500)
```

Out[53]:
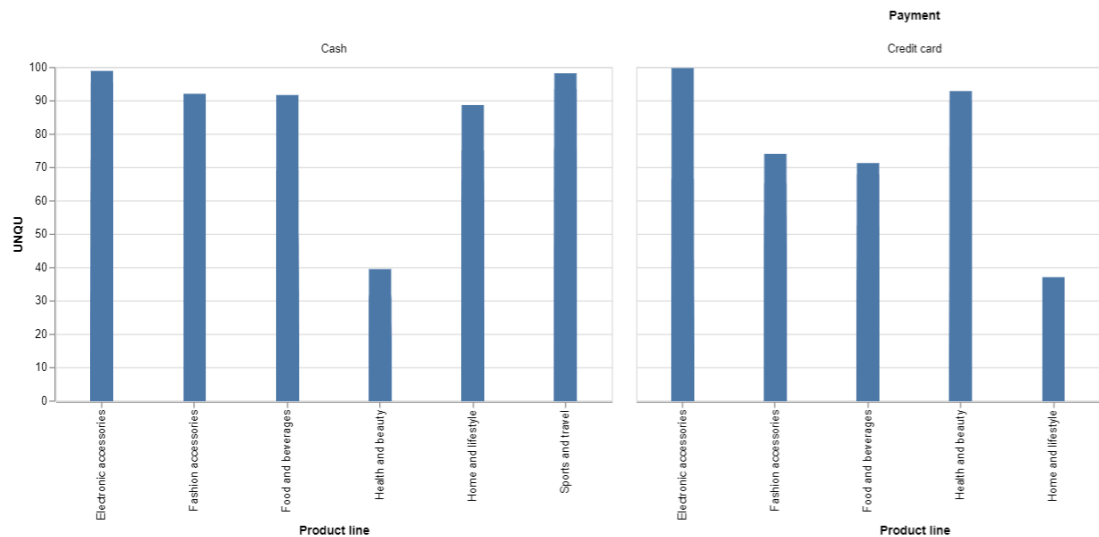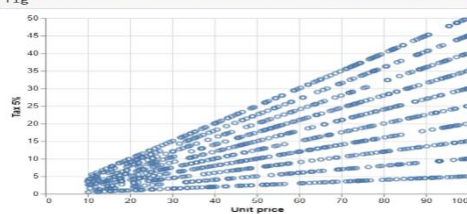

TAX VS UNIT PRICE

The regression line is plot ed between tax and unit price.

```
In [54]: fig = alt.Chart(df).mark_point().encode(x='Unit price',y='Tax 5%')
         fig
```

Out[54]:


```
In [55]: final_plot = fig + fig.transform_regression('Unit price','Tax 5%', method="linear").mark_line(color="#123c1", size=3)
         final_plot
```

Out[55]: