ABSTRACT

In this dissertation , we study initially about analysis of data of sales of the company, **FLAMINGO EXPORTS**. Findings from these research can be used to determine prevalence and relationships among sales data, and to forecast events from current data and knowledge.

Contents

1.INTRODUCTION

The below data is the sales register data of “FLAMINGO EXPORTS” for the month of May-2023

The data contains the tax details and HSN code details of the product being saled.

Using the data it is analysed

- ✓ If there is any significant relationship between the product value before and after adding taxes
- ✓ If the taxes are distributed properly to the Central and states.

➢ **IMPORTING DATA.**

library(readr)

SD <- read_csv("E:/SD.csv")

View(SD)

	UQC	Total Quantity	Total Value	Taxable Value	tax percentage	CGST	SGST	Place Of Supply	hsn number	hsn value
1	NOS-NUMBERS	66	53500.69	45339.59	0.18	4080.55	4088	TAMILNADU	8302	1228.81
2	NOS-NUMBERS	722	75644.72	64105.66	0.12	5769.53	5780.9	KERALA	3917	1228.81
3	PACKS	42	7157.36	6065.54	0.12	545.91	555	MAHARASTRA	9603	540
4	NOS-NUMBERS	27	7990.18	6771.36	0.18	609.41	611	KARNATAKA	9603	1627.14
5	KLR-KILOLITRE	3	382	323.74	0.18	29.13	32.5	TAMILNADU	3506	161.02
6	NOS-NUMBERS	8	108.56	92	0.12	8.28	12	TAMILNADU	8308	436
7	NOS-NUMBERS	4	1280.02	1084.76	0.12	97.63	97.63	KARNATAKA	8460	1084.76
8	NOS-NUMBERS	23	21795.81	18471.05	0.18	1662.38	1600.36	MAHARASTRA	3209	1467.72
9	BAG-BAGS	10	1930.02	1635.62	0	147.2	140.3	TELUNGANA	32141000	455
10	NOS-NUMBERS	14	363.44	308	0.12	27.72	35	TELUNGANA	6805	1080
11	MTR-METERS	288.5	37892.59	32112.37	0.28	2890.11	2860	TAMILNADU	3917	240
12	NOS-NUMBERS	18	2069.7	1753.98	0.12	157.86	160	KERALA	3506	5185.16
13	PACKS	30	443.68	376	0.18	33.84	30.8	TAMILNADU	82029990	27.34
14	PACKS	5943	14238.24	12066.3	0.28	1085.97	1066	KARNATAKA	7318	8778.3

15	NOS-NUMBERS	42	4830.53	4093.65	0.12	368.44	390	TELUNGANA	8465	5400
16	NOS-NUMBERS	62	1958.8	1660	0	149.4	155	TAMILNADU	6802	559.24
17	NOS-NUMBERS	2	1612	1366.1	0.12	122.95	130	MAHARASTRA	32091010	900
18	KLR-KILOLITRE	14	1888	1600	0.28	144	167.12	TELUNGANA	29031990	773.43
19	KLR-KILOLITRE	0.1	89.99	76.27	0.18	6.86	12	KARNATAKA	3209	1050.05
20	NOS-NUMBERS	35	9711.1	8229.74	0.28	740.68	760.23	KERALA	8301	600
21	KGS-KILOGRAMS	45.67	1676.6	1420.84	0.28	127.88	135.6	KERALA	7317	580
22	NOS-NUMBERS	23	867.3	735	0.05	66.15	69.3	MAHARASTRA	6804	1673.73
23	OTH-OTHERS	1	507.4	430	0.28	38.7	45	TAMILNADU	7302	1085.95
24	SET-SETS	6	613.6	520	0.28	46.8	40.6	TELUNGANA	8302	612.38
25	SET-SETS	8	3689.52	3126.72	0.05	281.4	278.6	KARNATAKA	8301	25
26	NOS-NUMBERS	29	5201.44	4408	0.28	396.72	303.6	TELUNGANA	8202	235
27	KLR-KILOLITRE	20.25	40341.93	34188.07	0.28	3076.93	3090.3	TAMILNADU	3214	1990
28	SET-SETS	16	123842	104950.9	0.28	9445.58	9500.69	TELUNGANA	6910	620
29	NOS-NUMBERS	62	30926.35	26208.79	0	2358.78	2290.3	KARNATAKA	8481	4027.8
30	NOS-NUMBERS	3	495.6	420	0.28	37.8	36.55	KERALA	7306	180
31	KGS-KILOGRAMS	3	424.8	360	0.28	32.4	20.4	MAHARASTRA	3214	1645
32	BAG-BAGS	22	884.91	691.29	0.05	96.81	106.12	TELUNGANA	2523	5286
33	MLT-MILILITRE	20	3524.42	2986.78	0.12	268.82	245.69	TELUNGANA	3506	1456
34	NOS-NUMBERS	12	198.24	168	0.28	15.12	16.2	TAMILNADU	4811	1185
35	MLT-MILILITRE	7	123.9	105	0.12	9.45	15	KERALA	4811	10373.36
36	NOS-NUMBERS	54	1611.92	1366.04	0.18	122.94	135.64	MAHARASTRA	3920	1530
37	MTR-METERS	2	153.4	130	0.05	11.7	12.36	TELUNGANA	6802	2745
38	GMS-GRAMMES	0.25	106.2	90	0.05	8.1	13	TELUNGANA	3405	27.34
39	KGS-KILOGRAMS	1	300	254.24	0.18	22.88	35.12	KARNATAKA	3506	5646.09
40	PAC-PACKS	3	94.4	80	0.18	7.2	17.12	KARNATAKA	3917	1286.6

The above 40 enteries are the sales of the month May-2023

- **CREATING A DATAFRAME**

flami<-data.frame(SD$Invoice.Number,

SD$Total.Quantity,

SD$Total.Value,

SD$CGST,

SD$SGST,

SD$Taxable.Value,

SD$hsn.number,

SD$hsn.value)

flami

```
> flami
```

```
> flami
```

	SD.Invoice.Number	SD.Total.Quantity	SD.Total.Value	SD.CGST	SD.SGST	SD.Taxable.Value	SD.hsn.number	SD.hsn.value
1	457	66.00	53500.69	4080.55	4088.00	45339.59	8302	1228.81
2	458	722.00	75644.72	5769.53	5780.90	64105.66	3917	1228.81
3	459	42.00	7157.36	545.91	555.00	6065.54	9603	540.00
4	460	27.00	7990.18	609.41	611.00	6771.36	9603	1627.14
5	461	3.00	382.00	29.13	32.50	323.74	3506	161.02
6	462	8.00	108.56	8.28	12.00	92.00	8308	436.00
7	463	4.00	1280.02	97.63	97.63	1084.76	8460	1084.76
8	464	23.00	21795.81	1662.38	1600.36	18471.05	3209	1467.72
9	465	10.00	1930.02	147.20	140.30	1635.62	32141000	455.00
10	466	14.00	363.44	27.72	35.00	308.00	6805	1080.00
11	467	288.50	37892.59	2890.11	2860.00	32112.37	3917	240.00
12	468	18.00	2069.70	157.86	160.00	1753.98	3506	5185.16
13	469	30.00	443.68	33.84	30.80	376.00	82029990	27.34
14	470	5943.00	14238.24	1085.97	1066.00	12066.30	7318	8778.30
15	471	42.00	4830.53	368.44	390.00	4093.65	8465	5400.00
16	472	62.00	1958.80	149.40	155.00	1660.00	6802	559.24
17	473	2.00	1612.00	122.95	130.00	1366.10	32091010	900.00
18	474	14.00	1888.00	144.00	167.12	1600.00	29031990	773.43
19	475	0.10	89.99	6.86	12.00	76.27	3209	1050.05
20	476	35.00	9711.10	740.68	760.23	8229.74	8301	600.00
21	477	45.67	1676.60	127.88	135.60	1420.84	7317	580.00
22	478	23.00	867.30	66.15	69.30	735.00	6804	1673.73
23	479	1.00	507.40	38.70	45.00	430.00	7302	1085.95
24	480	6.00	613.60	46.80	40.60	520.00	8302	612.38
25	481	8.00	3689.52	281.40	278.60	3126.72	8301	25.00
26	482	29.00	5201.44	396.72	303.60	4408.00	8202	235.00
27	483	20.25	40341.93	3076.93	3090.30	34188.07	3214	1990.00
28	484	16.00	123842.01	9445.58	9500.69	104950.85	6910	620.00
29	485	62.00	30926.35	2358.78	2290.30	26208.79	8481	4027.80
30	486	3.00	495.60	37.80	36.55	420.00	7306	180.00
31	487	3.00	424.80	32.40	20.40	360.00	3214	1645.00
32	488	22.00	884.91	96.81	106.12	691.29	2523	5286.00
33	489	20.00	3524.42	268.82	245.69	2986.78	3506	1456.00
34	490	12.00	198.24	15.12	16.20	168.00	4811	1185.00
35	491	7.00	123.90	9.45	15.00	105.00	4811	10373.36
36	492	54.00	1611.92	122.94	135.64	1366.04	3920	1530.00
37	493	2.00	153.40	11.70	12.36	130.00	6802	2745.00
38	494	0.25	106.20	8.10	13.00	90.00	3405	27.34
39	495	1.00	300.00	22.88	35.12	254.24	3506	5646.09
40	496	3.00	94.40	7.20	17.12	80.00	3917	1286.60

```
>
```

__The above dataframe carries the 40 data of numeric values alone.__

- **ACCESSING**

```
> taxable<-SD$Taxable.Value
> total<-SD$Total.Value
> cgst<-SD$CGST
> sgst<-SD$SGST
> quantity<-SD$Total.Quantity
> invoiceno.<-SD$Invoice.Number
```

```
> hsnno.<-SD$hsn.number
> hsnamount<-SD$hsn.value
```

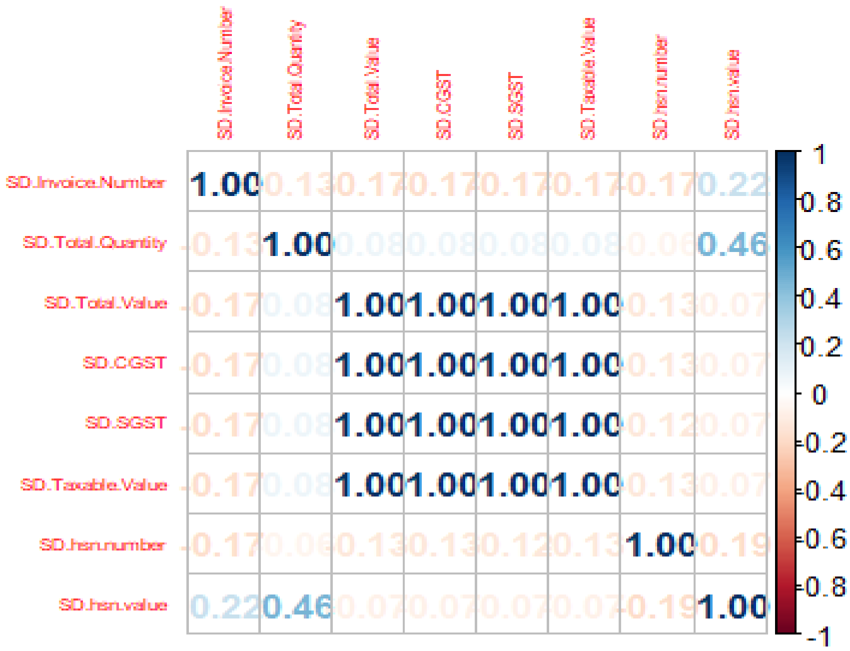
2.CORRELATION

2.1 CORRPLOT

library(corrplot)

m<-cor(flami)

corrplot(m,method="number",tl.cex=0.5)



INFERENCE

Among the columns of the dataset ,there exist a PERFECT POSITIVE CORRELATION between

- *total value and CGST*
- *total value and SGST*
- *total value and Taxable value*

- *CGST and SGST*
- *Taxable value and CGST*
- *Taxable value and SGST*

This implies that total value is a Dependent variable and CGST, SGST, Taxable value are the independent variables. Thus , a difference in the CGST,SGST ,taxable amount would change the total value of the sales.

2.1 .GRAPHICAL PLOT

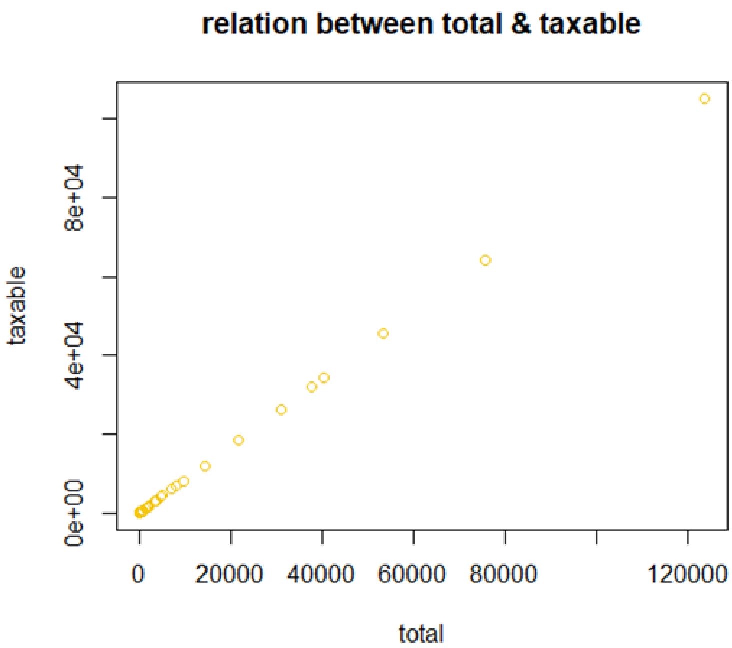
plot(total,taxable,col=7,main="realtion between total & taxable")

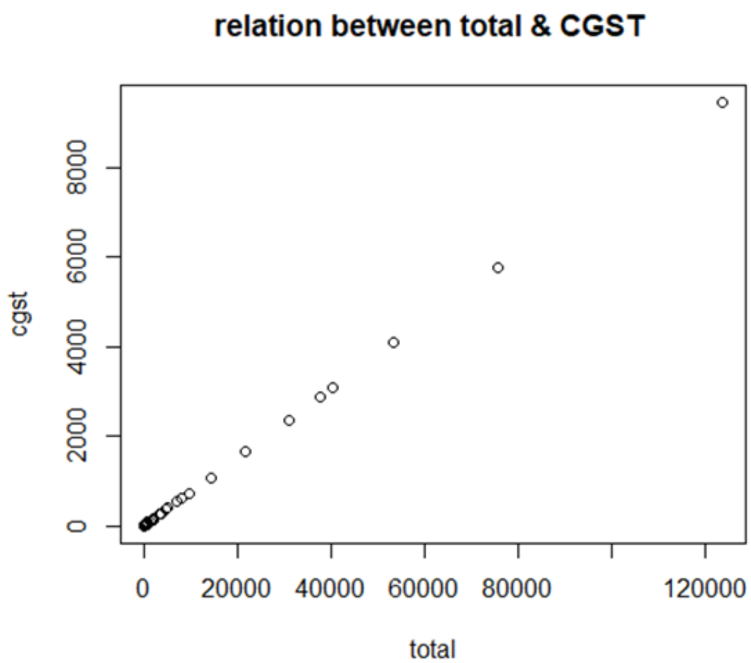
plot(total,cgst,col=5,main="realtion between total & CGST ")

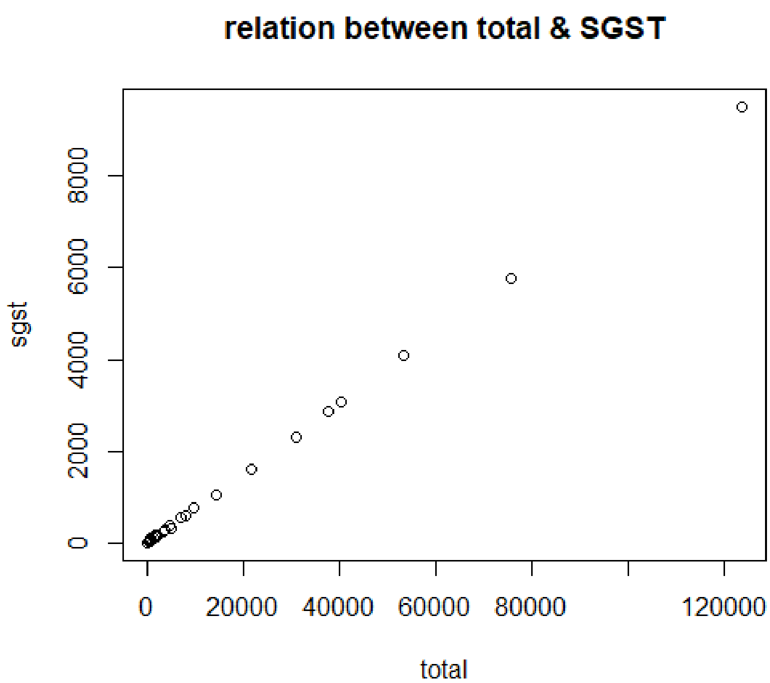
plot(total,sgst,col=9,main="relation between total & SGST")

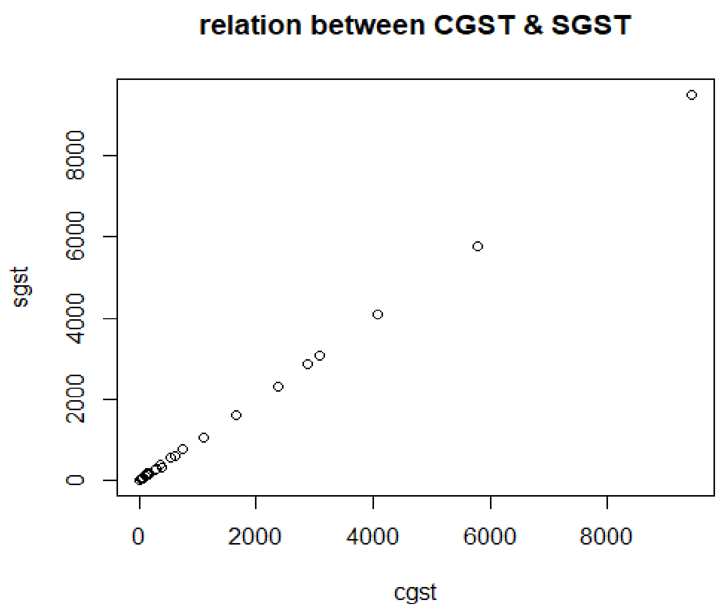
plot(cgst,sgst,col=2,main = "relation between CGST & SGST")

PRODUCTION




relation between total & CGST
cgst
0
2000
4000
6000
8000
0
20000
40000
60000
80000
120000
total


relation between total & SGST
sgst
0
2000
4000
6000
8000
0
20000
40000
60000
80000
120000
total



INFERENCE

As one value increses other values also increases which shows the total value,taxable value,CGST,SGST are perfectly positively correlated.

3.REGRESSION ANALYSIS

3.1 MULTIPLE REGRESSION

model12<-lm(total~taxable+cgst+sgst,flami=SD[,c("total","taxable","cgst","sgst")])

model12

```
> model12
```

```
Call:
lm(formula = total ~ taxable + cgst + sgst, flami = SD[, c("total",
  "taxable", "cgst", "sgst")])

Coefficients:
(Intercept)     taxable       cgst       sgst
 6.903e-12   1.000e+00   2.000e+00   6.858e-15
```

- **SUMMARY**

```
> summary(model12)

Call:
lm(formula = total ~ taxable + cgst + sgst, flami = SD[, c("total",
  "taxable", "cgst", "sgst")])
```

```
Residuals:
      Min        1Q    Median        3Q       Max 
-1.239e-11 -1.968e-12 -5.400e-13  5.780e-13  4.859e-11 

Coefficients:
             Estimate Std. Error   t value Pr(>|t|)    
(Intercept) 6.903e-12  1.563e-12 4.417e+00 8.77e-05 ***
taxable     1.000e+00  2.307e-14 4.334e+13  < 2e-16 ***
cgst        2.000e+00  2.671e-13 7.488e+12  < 2e-16 ***
sgst        6.858e-15  5.624e-14 1.220e-01    0.904    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.708e-12 on 36 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1 
F-statistic: 1.039e+32 on 3 and 36 DF,  p-value: < 2.2e-16
```

INFERENCE

Here total value is our dependent variable and the taxable,cgst,sgst are the predictors.
The null hypothesis is that there is no relationship between the dependent variable and the independent variable and the alternative hypothesis is that there is a relationship.
*Here the **f-statistic** is larger i.e, (1.039e+32) .which implies that the null hypothesis should be rejected.*
*Also ,the **p-value** is much lesser (i.e. 2.2e-16),which implies that the alternate hypothesis is accepted*

Hence ,there exist a cause and effect relationship between the variables

3.2 PREDICTION

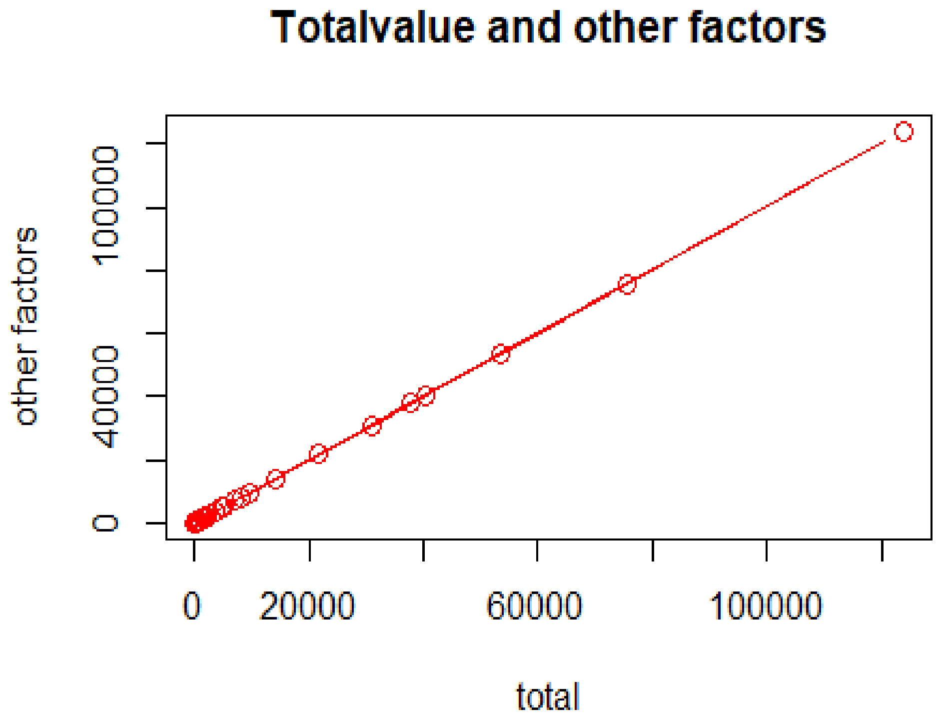
```
newflami<-data.frame(taxable=560,cgst=28,sgst=28)

newtotal<-predict(model12,newflami)

newtotal

> newtotal
  1 
616
```

INFERENCE

According to the gst sales data,total value is the sum of taxable,cgst,and Sgst.
i.e., total=taxable+CGST+SGST.
hence for we arrive at 616 as the resultant total value.

3.3 GRAPHICAL REPRESENTATION

```
plot(total,taxable+cgst+sgst,col="red",main="Totalvalue and other factors",
   abline(lm(taxable+cgst+sgst~total)),
   cex=1.3,pch=1,type="b",xlab="total",ylab="other factors")
plot
```



INFERENCE

From the least-squares regression line, we notice that the line perfectly flow through each of the points and that there is no residual between the point and the line. Hence there is a relationship between the values

4.PIECHART

- **Pie chart between quantity of sales and statetax(SGST):**

quantity1ofsale<-SD$SGST

place<-SD$Place.Of.Supply

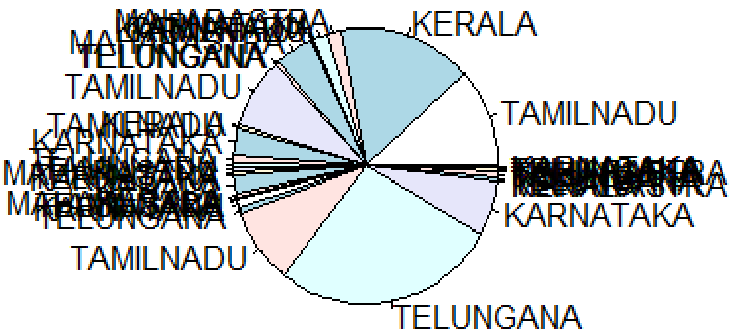
pie(quantityofsale,place)

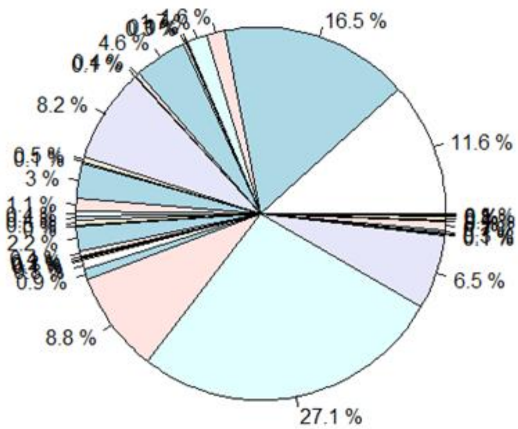
pie

piepercent<-paste(round(100*quantity1ofsale/sum(quantity1ofsale),1),"%")

pie(quantity1ofsale,labels=piepercent)

- **PIECHART BETWEEN QUANTITY AND SGST**





- **Pie chart between quantity of sales and central tax(CGST**):

quantity2ofsale<-SD$CGST

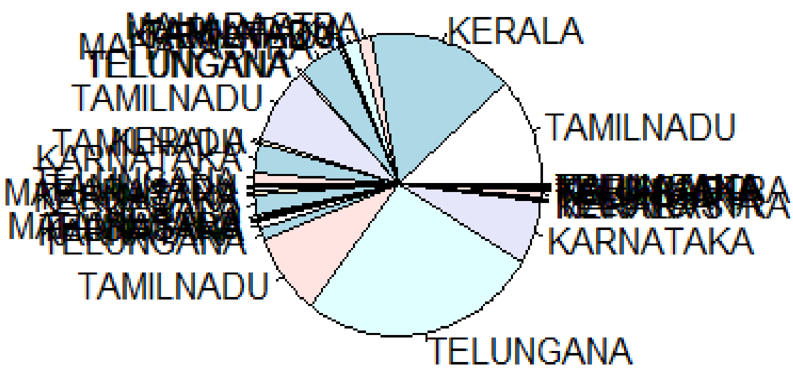
place<-SD$Place.Of.Supply

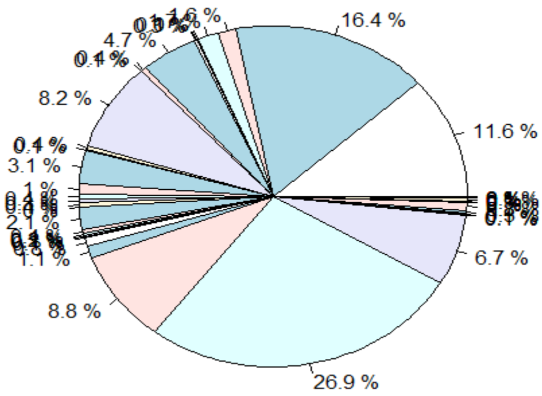
pie(quantity2ofsale,place)

piepercent<-paste(round(100*quantity2ofsale/sum(quantity2ofsale),1),"%")

pie(quantity2ofsale,labels=piepercent)

- **PIECHART BETWEEN QUANTITY AND CGST**





INFERENCE

The piechart between quantity and SGST depicts the quantity of sales among different states ,where telungana and kerala were the large consumers

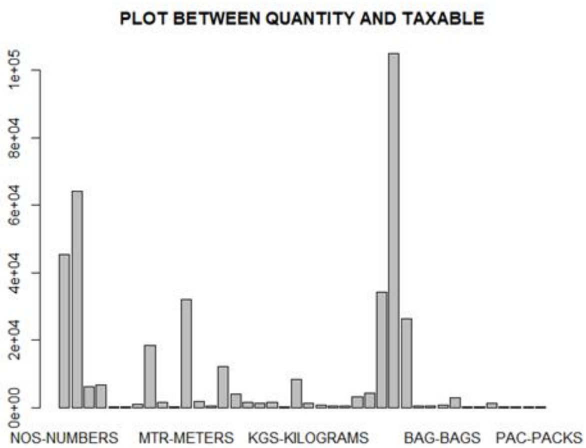
Comparitively,the pie chart between quantity and CGST was also spread in the same way,resulting in kerala and karnataka as the large consumers.

5.BAR CHART

x<-SD$UQC

y<-SD$Taxable.Value

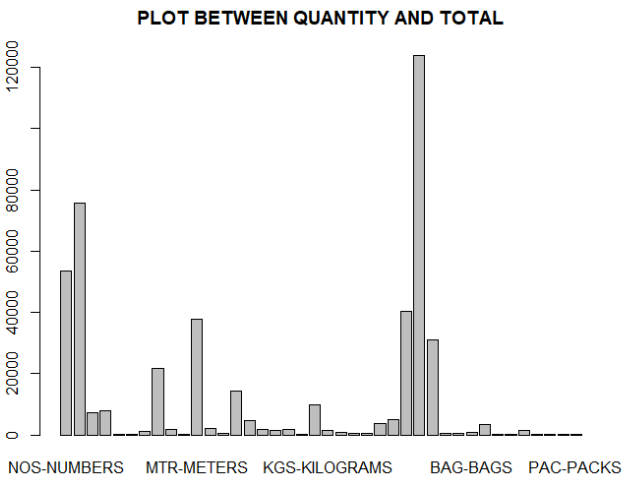
barplot(y,names.arg=x)



d<-SD$UQC

g<-SD$Total.Value

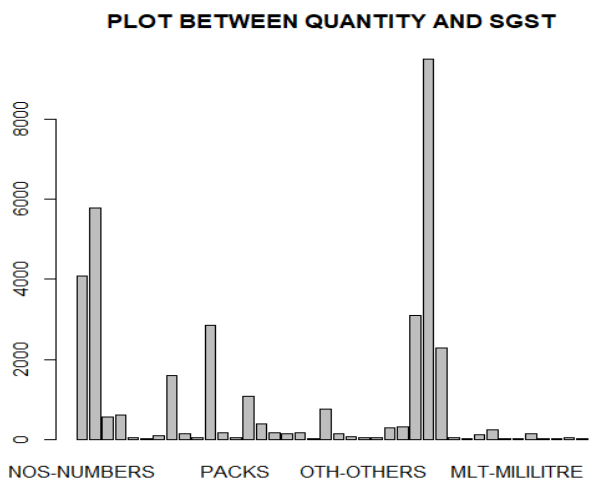
barplot(g,names.arg = d)



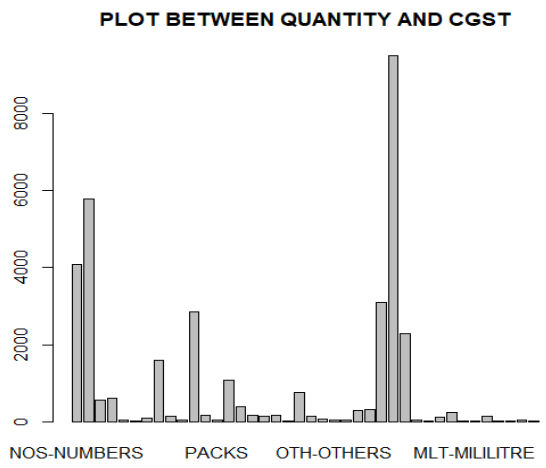
t<-SD$UQC

h<-SD$SGST

barplot(h,names.arg=t,main="PLOT BETWEEN QUANTITY AND SGST")



```
j<-SD$UQC

k<-SD$SGST

barplot(k,names.arg=j,main="PLOT BETWEEN QUANTITY AND CGST")
```



INFERENCE

From regression analysis we concluded that total value , taxable value,CGST,SGST have a linear relationship.

Hence the plots representation between the

'quantity and total value'

'quantity and taxable '

'quantity and CGST'

'quantity and SGST' are similar

6.PLOTTING

```
a<-c(SD$Taxable.Value)

b<-c(SD$Total.Value)

c<-c(SD$CGST)

d<-C(SD$SGST)
```

dataa<-data.frame(a,b,c,d)

dataa

```
> dataa
           a         b       c       d
1   45339.59  53500.69 4080.55    4088
2   64105.66  75644.72 5769.53  5780.9
3    6065.54   7157.36  545.91     555
4    6771.36   7990.18  609.41     611
5     323.74    382.00   29.13    32.5
6      92.00    108.56    8.28      12
7    1084.76   1280.02   97.63   97.63
8   18471.05  21795.81 1662.38 1600.36
9    1635.62   1930.02  147.20   140.3
10    308.00    363.44   27.72      35
11  32112.37  37892.59 2890.11    2860
12   1753.98   2069.70  157.86     160
13    376.00    443.68   33.84    30.8
14  12066.30  14238.24 1085.97    1066
15   4093.65   4830.53  368.44     390
16   1660.00   1958.80  149.40     155
17   1366.10   1612.00  122.95     130
18   1600.00   1888.00  144.00  167.12
19     76.27     89.99    6.86      12
20   8229.74   9711.10  740.68  760.23
21   1420.84   1676.60  127.88   135.6
22    735.00    867.30   66.15    69.3
23    430.00    507.40   38.70      45
24    520.00    613.60   46.80    40.6
25   3126.72   3689.52  281.40   278.6
26   4408.00   5201.44  396.72   303.6
27  34188.07  40341.93 3076.93  3090.3
28 104950.85 123842.01 9445.58 9500.69
29  26208.79  30926.35 2358.78  2290.3
30    420.00    495.60   37.80   36.55
31    360.00    424.80   32.40    20.4
32    691.29    884.91   96.81  106.12
33   2986.78   3524.42  268.82  245.69
34    168.00    198.24   15.12    16.2
35    105.00    123.90    9.45      15
36   1366.04   1611.92  122.94  135.64
37    130.00    153.40   11.70   12.36
38     90.00    106.20    8.10      13
39    254.24    300.00   22.88   35.12
40     80.00     94.40    7.20   17.12
```

plot(dataa$a,type="o",col="red",main="sales data",xlab="value before tax" ,

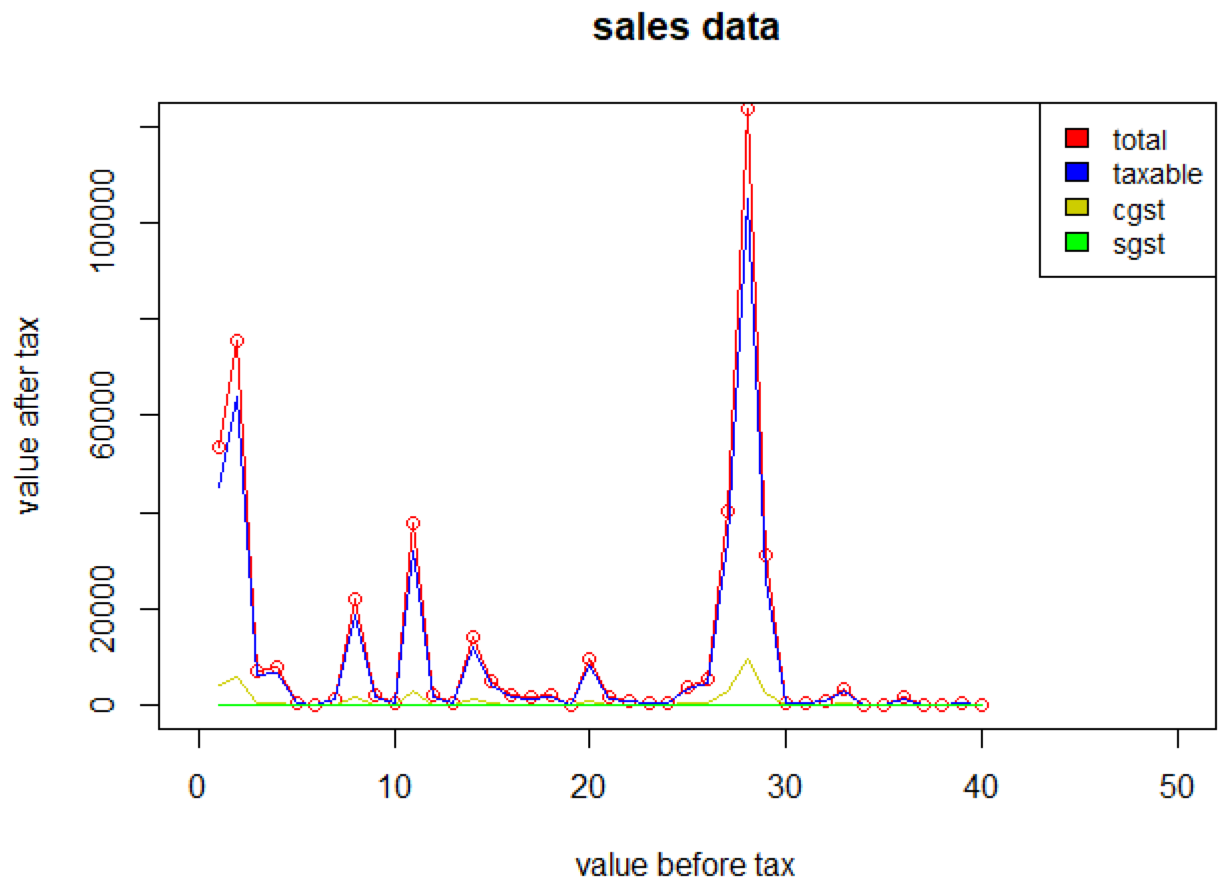
ylab = "value after tax",xlim=c(0,80),ylim = c(0,100))

lines(dataa$b,col="blue")

lines(dataa$c,col="brown")

lines(dataa$d,col="green")

legend("topright",c("taxable","cgst","sgst"),cex = 0.9,fill = c("blue","brown","green"))



INFERENCE

From the plot ,it is observed that total value(red) is high, and the taxable value (blue) follows the same wave since they are closely related. Also, the CGST and SGST follows the wave below.

7.ANOVA

7.1 .ONE WAY ANOVA

production=lm(SD$Total.Quantity~SD$Total.Value)

anova(production)

```
> anova(production)
Analysis of Variance Table
```

```
Response: SD$Total.Quantity
              Df   Sum Sq Mean Sq F value Pr(>F)
SD$Total.Value  1   220124  220124  0.2442  0.624
Residuals      38 34252670  901386
```

INFERENCE

The p value of the above data is high i.e,0.624>0.05.
Hence the null hypothesis is accepted such that there is no difference
Between the quantity and total value.

7.2.TWO WAY ANOVA.

```
fit=aov(SD$Total.Quantity~cgst+sgst)
anova(fit)
```

```
> anova(fit)
Analysis of Variance Table

Response: SD$Total.Quantity
          Df   Sum Sq Mean Sq F value Pr(>F)
cgst       1   219801  219801  0.2419 0.6257
sgst       1   638141  638141  0.7024 0.4074
Residuals 37 33614852  908510
```

INFERENCE

We see that there is no significant differences between quantity and both CGST & SGST
Since 0.6257 > 0.05
0.4074 > 0.05

8.STATISTICAL ANALYSIS BETWEEN CGST AND SGST:

- **CGST**

```
> mean(cgst)
[1] 878.7503

> median(cgst)
[1] 125.415

> var(cgst)
[1] 3523806

> range(cgst)
[1]    6.86 9445.58

> max(cgst)
[1] 9445.58
```

```
> min(cgst)
[1] 6.86

> sd(cgst)
[1] 1877.18

> IQR(cgst)
[1] 530.2025
```

- **SGST**

```
> mean(sgst)
[1] 877.2758

> median(sgst)
[1] 135.62

> var(sgst)
[1] 3541309

> range(sgst)
[1]   12.00 9500.69

> max(sgst)
[1] 9500.69

> min(sgst)
[1] 12

> sd(sgst)
[1] 1881.836

> IQR(sgst)
[1] 534.625
```

8.1 TO COUNT THE FREQUENCY OF THE DATA

table(SD$Total.Value)

```
> table(SD$Total.Value)

    89.99      94.4     106.2    108.56     123.9     153.4    198.24 
        1         1         1         1         1         1         1 
      300    363.44       382     424.8    443.68     495.6     507.4 
        1         1         1         1         1         1         1 
    613.6     867.3    884.91   1280.02   1611.92      1612    1676.6 
        1         1         1         1         1         1         1 
     1888   1930.02    1958.8    2069.7   3524.42   3689.52   4830.53 
        1         1         1         1         1         1         1 
  5201.44   7157.36   7990.18    9711.1  14238.24  21795.81  30926.35 
        1         1         1         1         1         1         1 
 37892.59  40341.93  53500.69  75644.72 123842.01 
        1         1         1         1         1 
```

```
table(SD$Taxable.Value)
```

```
> table(SD$Taxable.Value)

    76.27        80        90        92       105       130       168 
        1         1         1         1         1         1         1 
   254.24       308    323.74       360       376       420       430 
        1         1         1         1         1         1         1 
      520    691.29       735   1084.76   1366.04    1366.1   1420.84 
        1         1         1         1         1         1         1 
     1600   1635.62      1660   1753.98   2986.78   3126.72   4093.65 
        1         1         1         1         1         1         1 
```

```
     4408    6065.54    6771.36    8229.74    12066.3   18471.05   26208.79
        1          1          1          1          1          1          1
 32112.37   34188.07   45339.59   64105.66  104950.85
        1          1          1          1          1
```

INFERENCE

From the above two data sets we observe that each and every sale of the company is distinct.

8.2 TO FIND THE RANGES OF THE SALES

```
> range(SD$Total.Quantity)
[1]    0.1 5943.0

> range(SD$Total.Value)
[1]     89.99 123842.01

> range(SD$Taxable.Value)
[1]     76.27 104950.85

> range(SD$CGST)
[1]    6.86 9445.58

> range(SD$SGST)
[1]   12.00 9500.69

> range(SD$hsn.value)
[1]    25.00 10373.36
```

We obtain the maximum and minimum value of each of the dataset from the above range function.

9 . CONCLUSION

This repot aims to analyse the cost of a particular product and its associated tax relationship. Through this analaysation using various function of R,the structures of the may month sales data has been perfectly Showed and also it helped to forecast the data.

~*~