Superstore Data Visualization

By Pabba Abhishek

Presented on 09/06/21

Situation presented

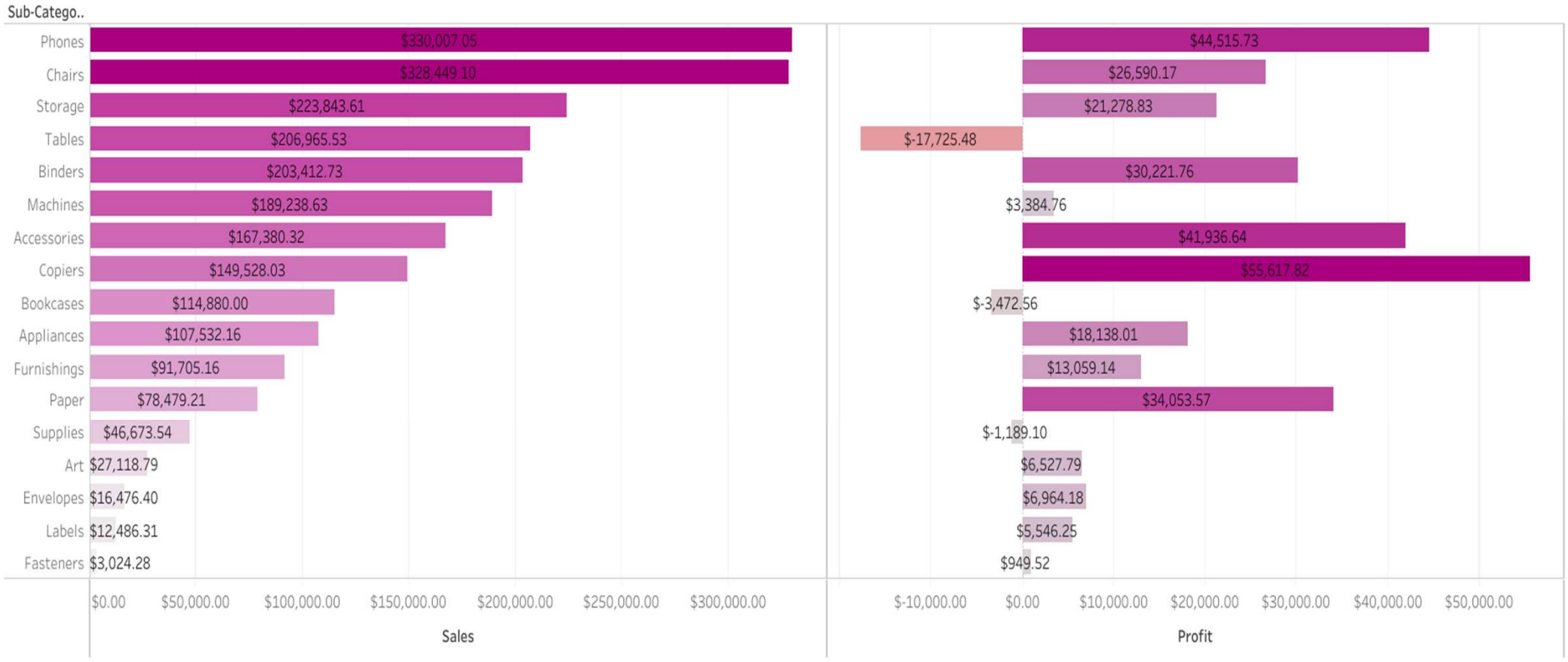
As a business manager of a superstore retail chain, I'm analyzing the data about sales and profits of the company. The following is the data gathered to analyse.

Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	Sales	Quantity	Discount	Profit
Standard Class	Consumer	United States	Seattle	Washington	98115	West	Furniture	Tables	1747.25	5	0	629.01
Standard Class	Corporate	United States	Seattle	Washington	98103	West	Furniture	Tables	3393.68	8	0	610.8624
First Class	Home Office	United States	Harrisonburg	Virginia	22801	South	Furniture	Tables	2244.48	7	0	493.7856
Second Class	Home Office	United States	Seattle	Washington	98105	West	Furniture	Tables	1913.4	9	0	401.814
Second Class	Home Office	United States	Seattle	Washington	98115	West	Furniture	Tables	2348.82	9	0	399.2994
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Standard Class	Consumer	United States	Jackson	Mississippi	39212	South	Furniture	Tables	2430.08	8	0	388.8128

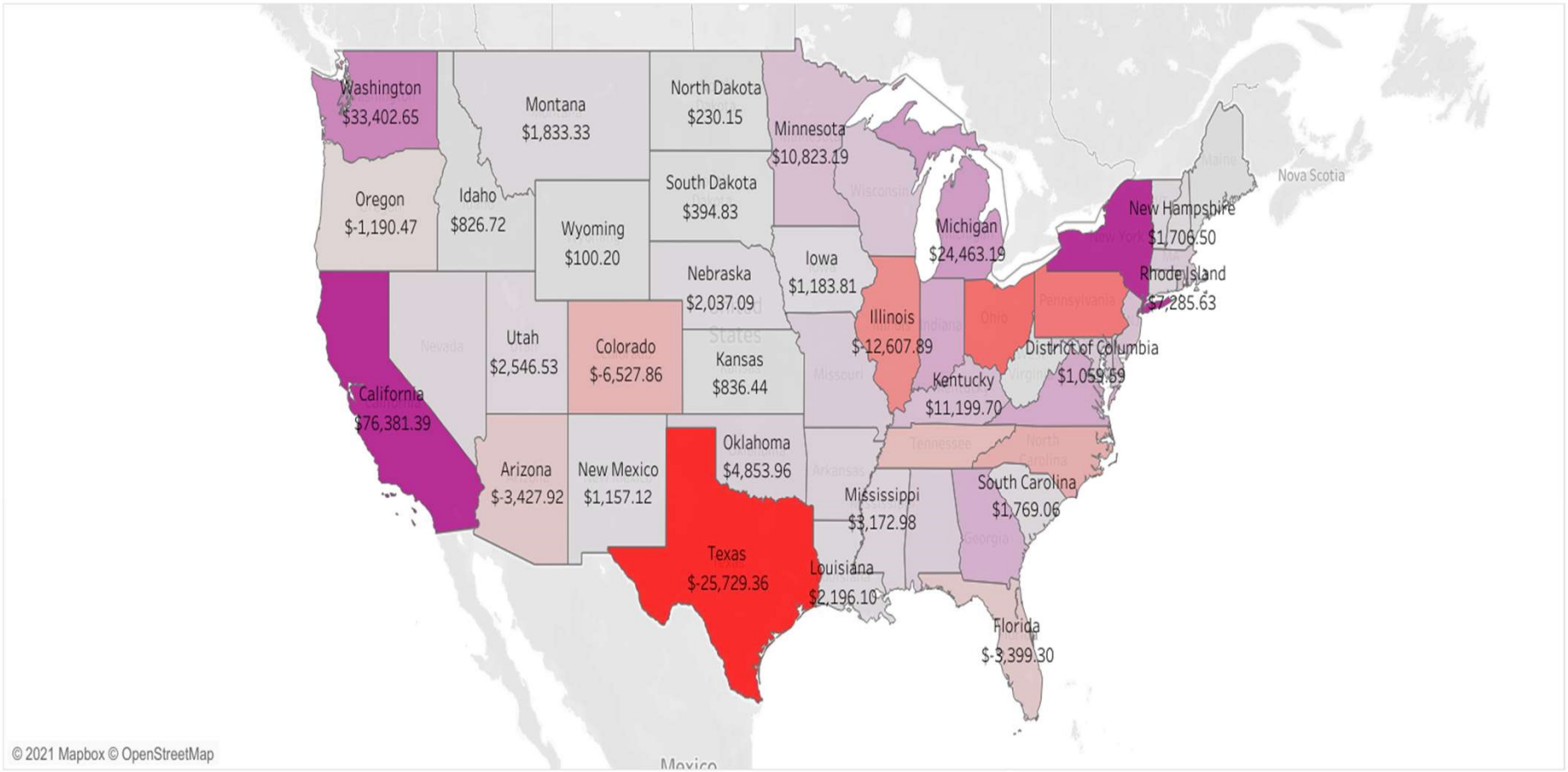
Goal of the project

1. Finding the best and least performing areas.
2. Visually presenting the findings and letting the shareholders see them.
3. Recommend solutions to increase profits in those areas.
4. Forecast the growth with the following suggestions.

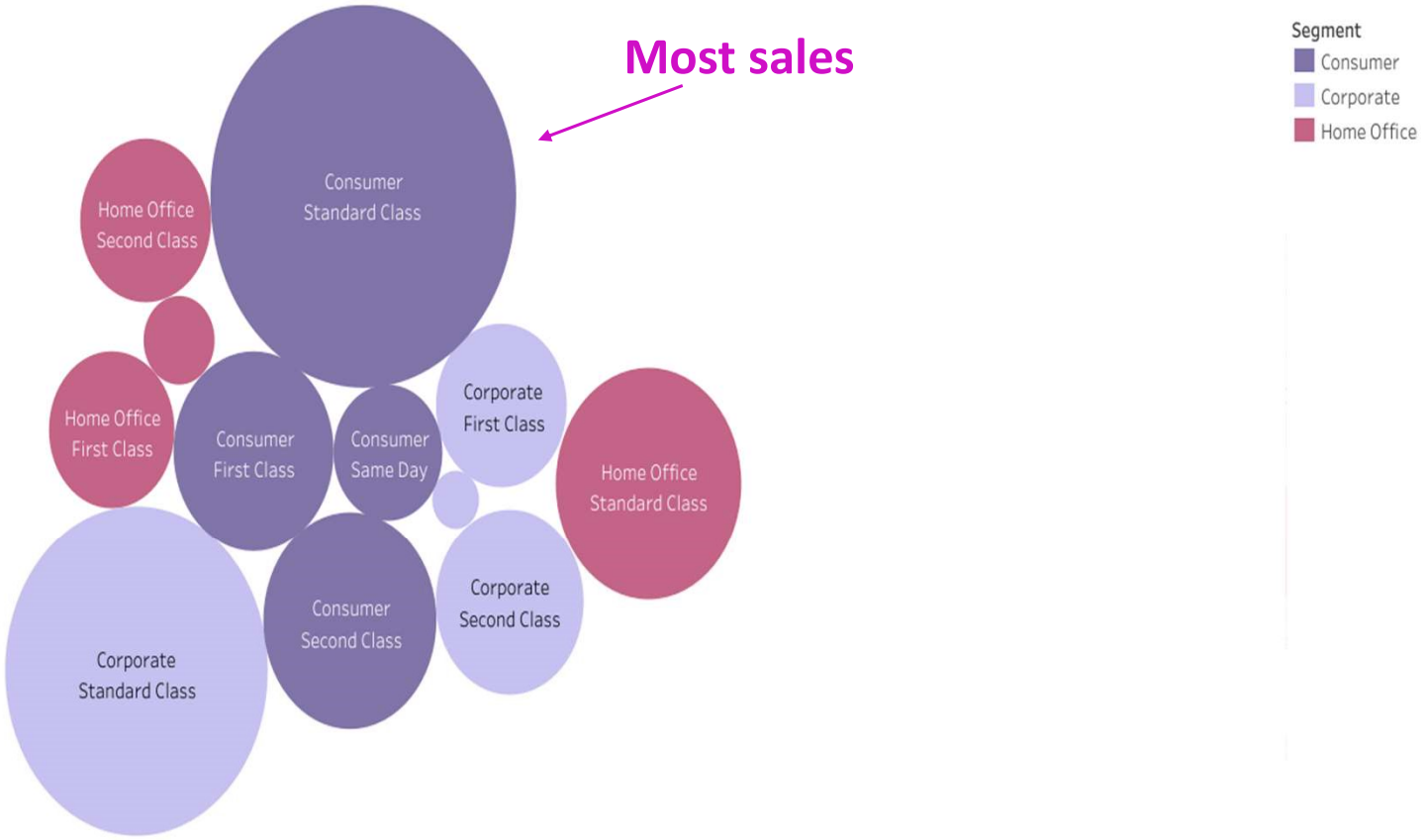
Profit and sales for each sub-category



State wise Profit



Segment and ship mode relation with sales



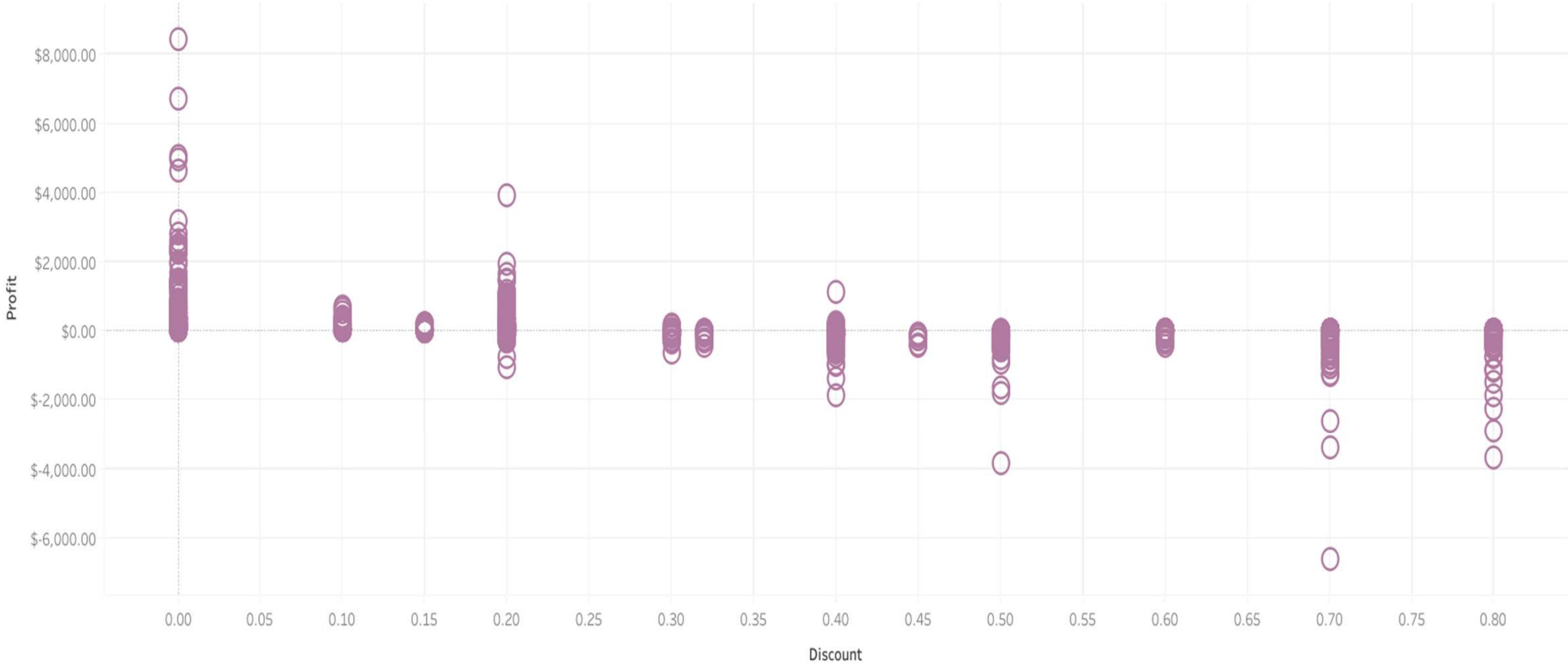
Things to be noted

- Most sold Sub-category is Phones
- Most profitable Sub-category:
  - ✓ Copiers
  - ✓ Phones
  - ✓ Accessories
- Best performing states:
  - ✓ California,
  - ✓ New York &
  - ✓ Washington
- Most sales from Standard and Consumer class.
- Least performing Sub-categories:
  - X Tables
  - X Supplies
  - X Bookcases
- Least performing states:
  - X Texas,
  - X Illinois,
  - X Pennsylvania & Ohio

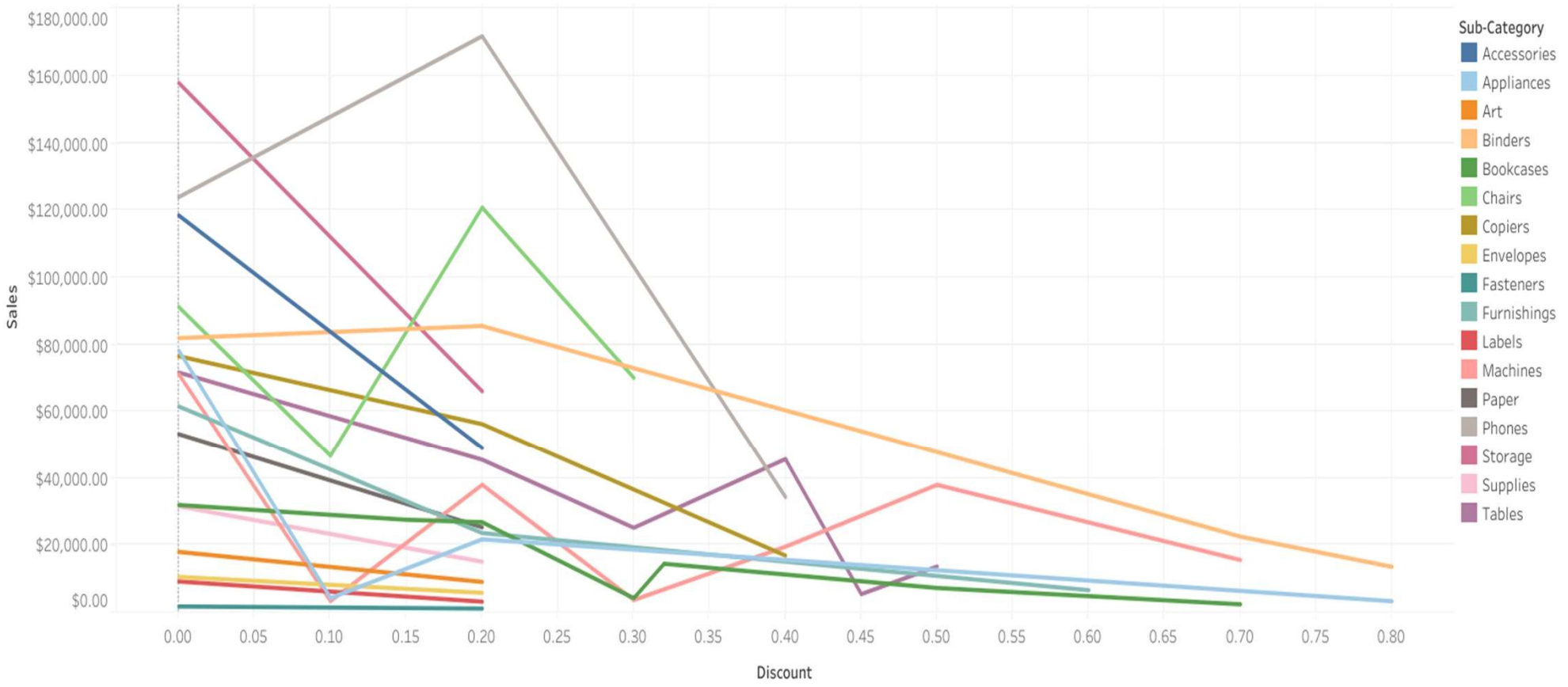
Things should be done further

1. Best performing and least performing state outlets to be contacted.
2. Further information regarding shipment mode and segment mode is to be gathered.
3. Relation between quantity, profits, discount and sales need to be further clarified.
4. Time period of the sales should be gathered to understand the periodic patterns in sales if any present.

Higher the discount higher the losses



Sales are independent of discount



Using ML to predict profits for 0 discount

As noticed earlier, huge discounts are leading to huge losses and discounts are nowhere promoting the sales.

Using a machine learning model, the profits are forecasted if for the same sales there was a zero discount

In [9]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.15,random_state=1940)
```

In [10]:
```
from sklearn.ensemble import RandomForestRegressor
regr = RandomForestRegressor(random_state=0)
regr.fit(x, y)
```

Out[10]:
```
RandomForestRegressor(random_state=0)
```

In [11]:
```
regr.score(x_test,y_test)
```

Out[11]:
```
0.99112172237077
```

In [12]:
```
forecast=pd.read_csv(r"C:\Users\Pabba Abhishek\Desktop\GRIP\SampleSuperstore-ML-Forecast.csv")
```

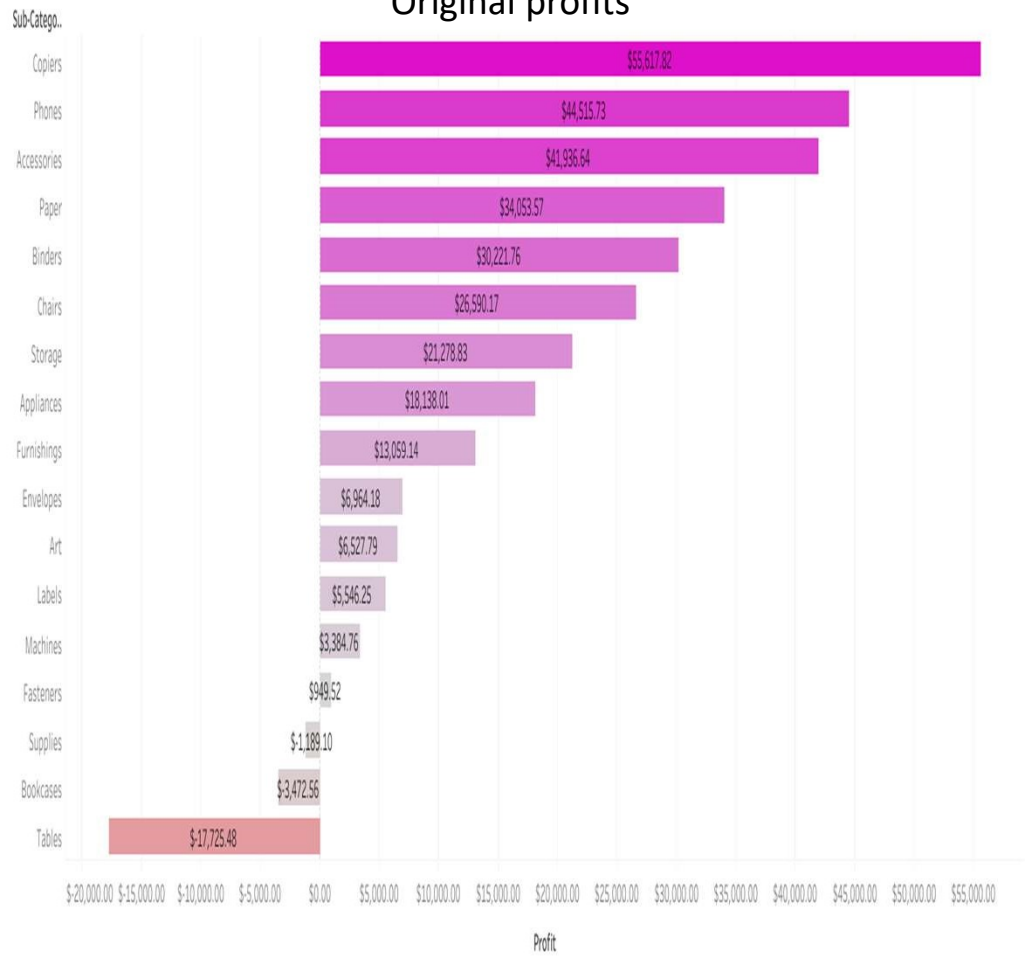
In [13]:
```
forecast
```

Out[13]:

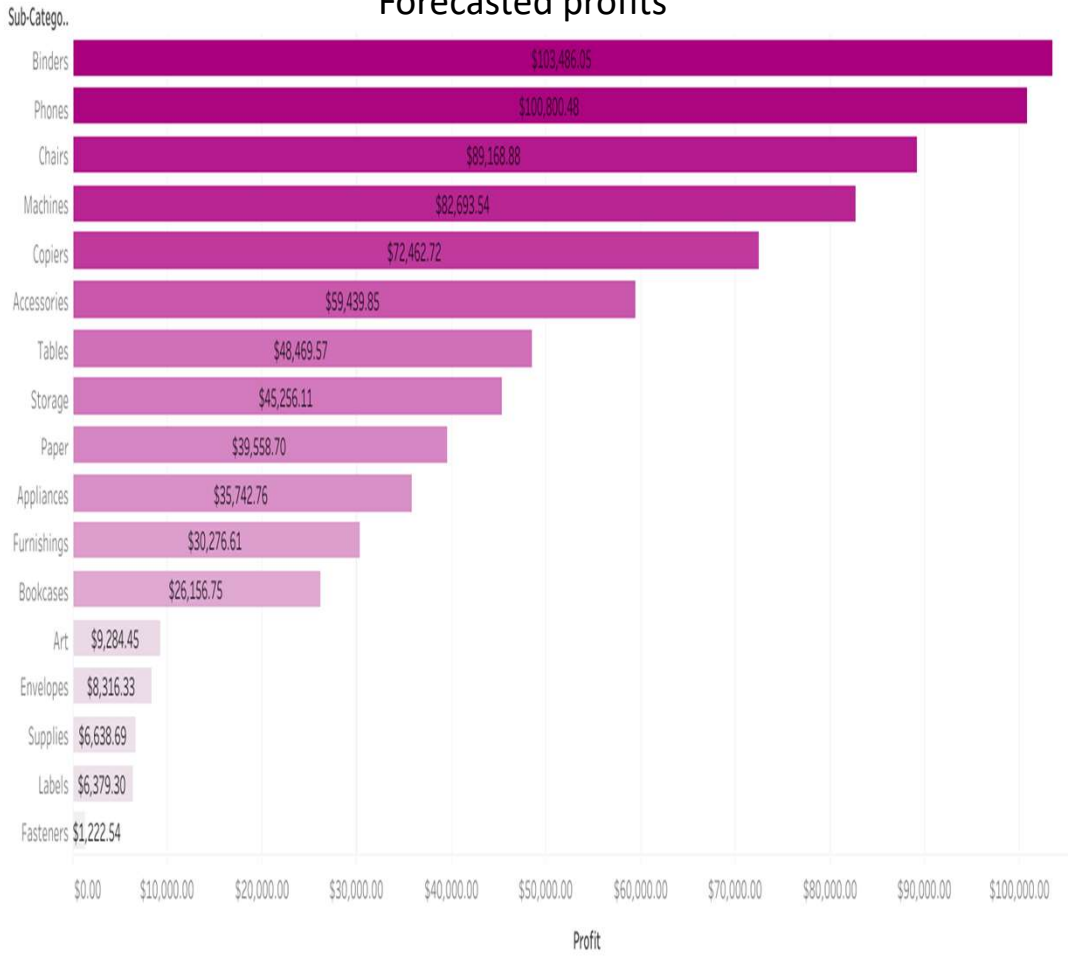
	Ship Mode	Segment	Postal Code	Region	Sub-Category	Sales	Quantity	Discount
0	0	1	60610	1	2	143.128	2	0
1	0	1	75701	1	2	2.688	3	0
2	0	2	76106	1	2	24.588	3	0
3	0	2	77506	1	2	1.624	2	0
4	1	1	79605	1	2	1.392	2	0

Original profits vs Forecasted profits at 0 discount

Original profits

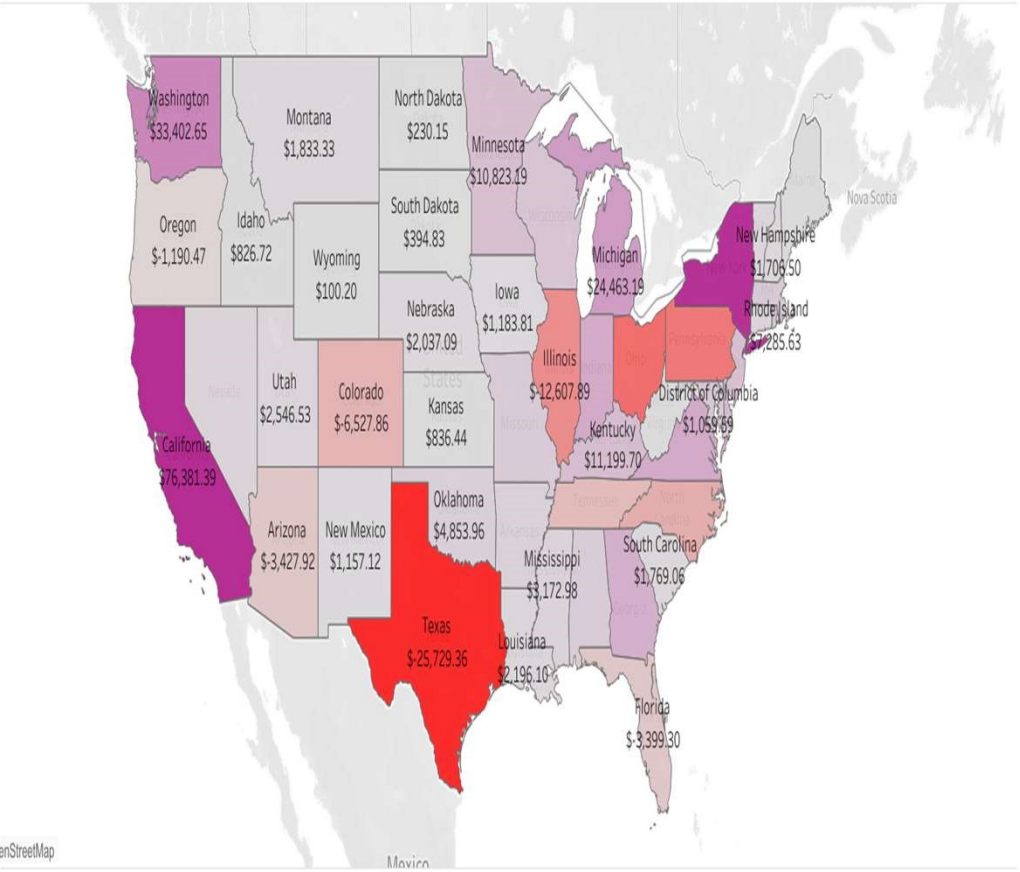


Forecasted profits

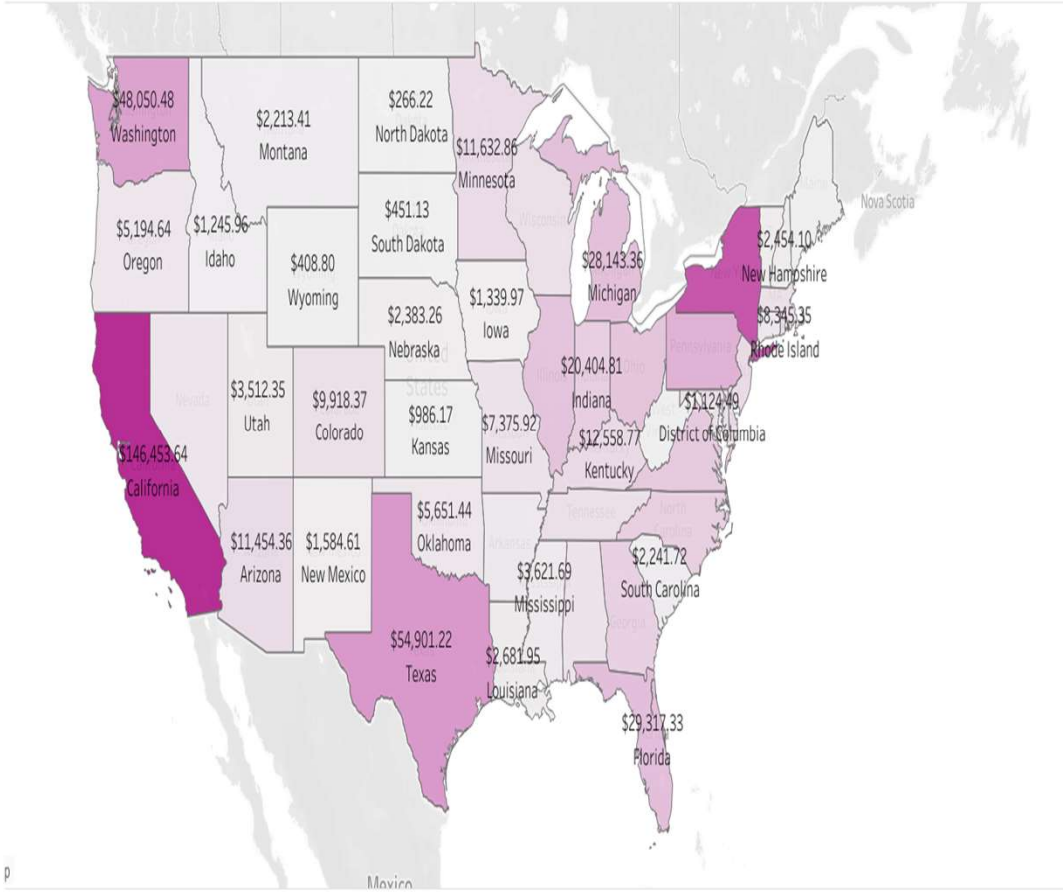


State wise improvement in profits

Original profits



Forecasted profits



Suggestions to increase profits