PORTOFOLIO

DATA ANALYST

SALES STRATEGY INACCURACIES IN SUPERSTORES

Nahdiyah Purnama

About Me

Nahdiyah Purnama

Education

D3 Teknik Elektro Politeknik Negeri Jakarta

Experience

Jasa Marga - Internship

Kalbe Nutritionals X Rakamin - Virtual Internship (Project Based)

Contact

082118134474

Cilegon, Banten

E-Mail : nahdiyahpurnama@gmail.com

PROJECT BACKGROUND

Superstore developed an application system to collect transaction data at each branch, the data is expected to be utilized to enhance the company's sales performance.The inaccuracy of the sales strategy at each Superstore branch, which is to give a sizable discount program when the sales level is extremely high, is the issue that is frequently brought up as a complaint.This can be detrimental to the store because large discounts can reduce profitability. To be able to overcome these problems, an analysis is carried out which has the goals of increasing sales and profits, maximizing efficient resource allocation at each Superstore branch; and identifying mistakes or unsuccessful sales strategy decisions in order to make changes and prevent such difficulties in the future.

Tools

- Google Colab
- Looker Studio
- Python

Data Set

Field Name	Description
Order_ID	Unique Order ID for each Customer
Customer_ID	Unique ID to identify each Customer
Postal_Code	Postal Code of every Customer
Product_ID	Unique ID of the Product
Sales	Sales of the Product
Quantity	Quantity of the Product
Discount	Discount provided
Profit	Profit/Loss incurred
Category	Category of the product ordered
Sub-Category	Sub-Category of the product ordered

Field Name	Description
Product_Name	Name of the Product
Order_Date	Order Date of the product
Ship_Date	Shipping Date of the Product
Ship_Mode	Shipping Mode specified by the Customer
Customer_Name	Name of the Customer
Segment	The segment where the Customer belongs
City	City of residence of of the Customer
State	City of residence of of the Customer
Region/Country	Region where the Customer belong/ Country of residence of the Customer

Data Cleaning

Data null

```
df.isnull().sum()
```

```
Order_ID          0
Customer_ID       0
Postal_Code       0
Product_ID        0
Sales             0
Quantity          0
Discount          0
Profit            0
Category          0
Sub-Category      0
Product_Name      0
Order_Date        0
Ship_Date         0
Ship_Mode         0
Customer_Name     0
Segment           0
Country/Region    0
City              0
State             0
Region            0
dtype: int64
```

Data Duplicate

```
[ ] df_filtered.duplicated().sum()
```

```
1
```

```
[ ] df = df_filtered.drop_duplicates()
```

```
[ ] df.duplicated().sum()
```

```
0
```

Converting Format Time

Before

Order_Date	Ship_Date	Sh
11/8/2019	11/11/2019	
11/8/2019	11/11/2019	
6/12/2019	6/16/2019	
10/11/2018	10/18/2018	
10/11/2018	10/18/2018	

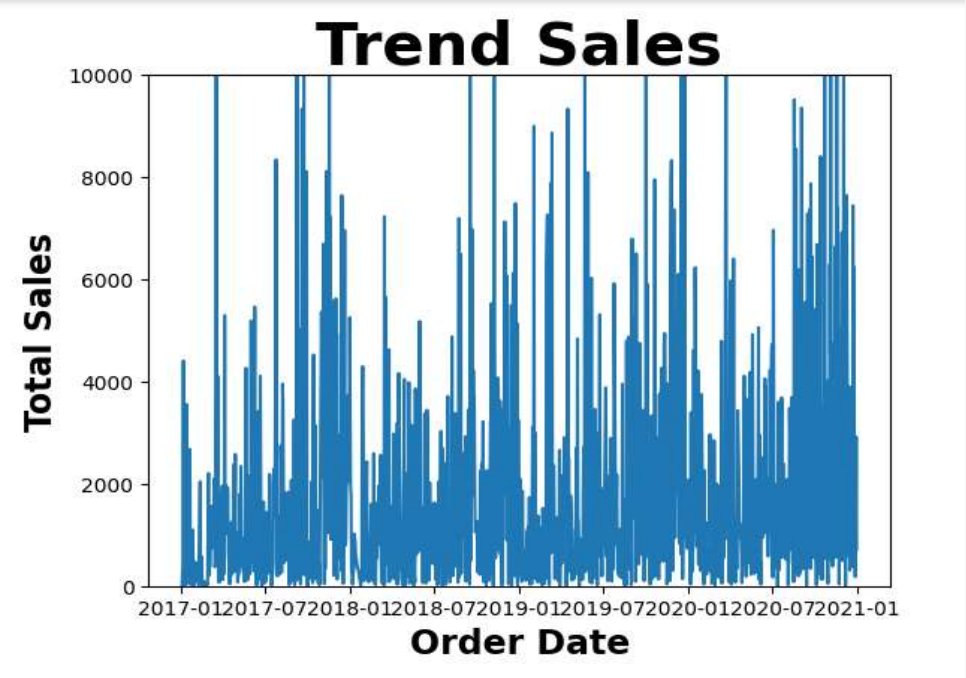
After

Order_Date	Ship_Date	S
2019-11-08	2019-11-11	
2019-11-08	2019-11-11	
2019-06-12	2019-06-16	
2018-10-11	2018-10-18	
2018-10-11	2018-10-18	

E D A

(EXPLORATORY DATA ANALYSIS)

Trend Sales

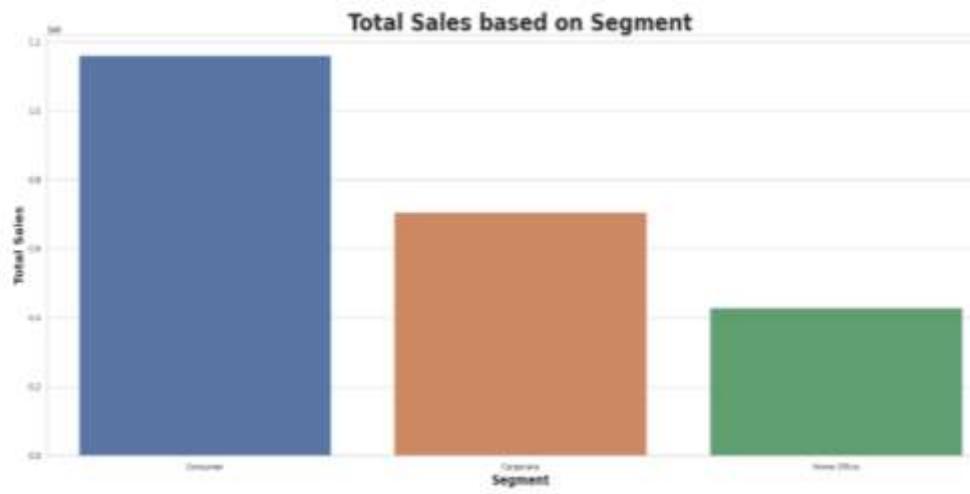


From January 2017 through January 20221, total sales likely to vary. Due to the huge year-end discount, sales increased significantly at the end of each year. The largest sales happened in November 2020, with 118,447 units sold. The majority of sales are less than 5000 units. There are signs of outliers in the period around 2017-02.

Trend Sales based on Category Products

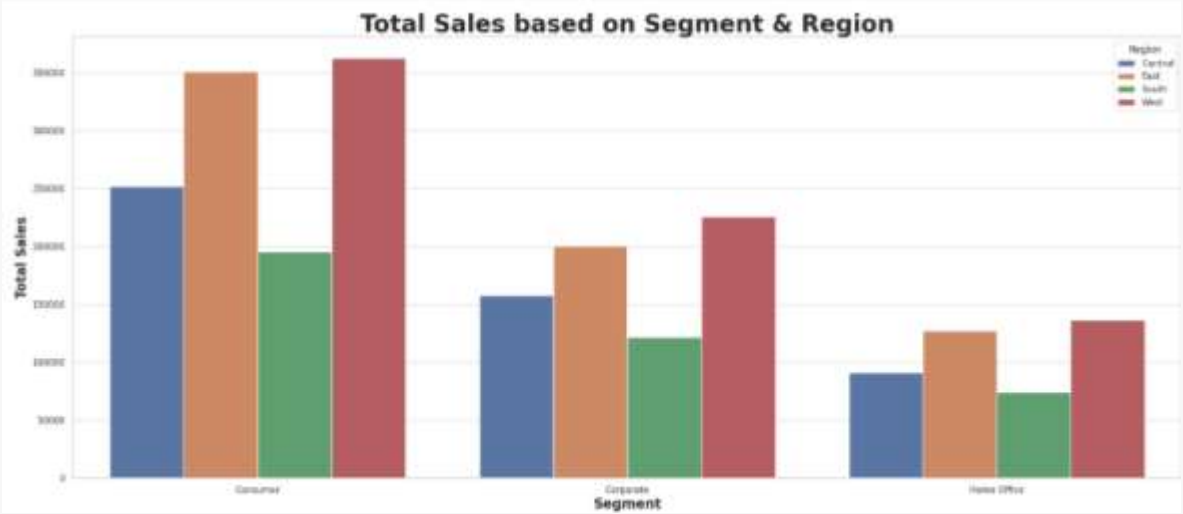
Technology sector dominate sales every year. Furniture is the lowest sales product. There are indications of outliers in some periods due to technology products.

Total Sales based on Segment



The total sales of the Consumer segment are the most dominating,t, which is around 50% of all total sales, while total sales of Home Office are the lowest at around 18.7% and total sales of the Corporate segment are around 30.7% .

Total Sales based on Segment and Region

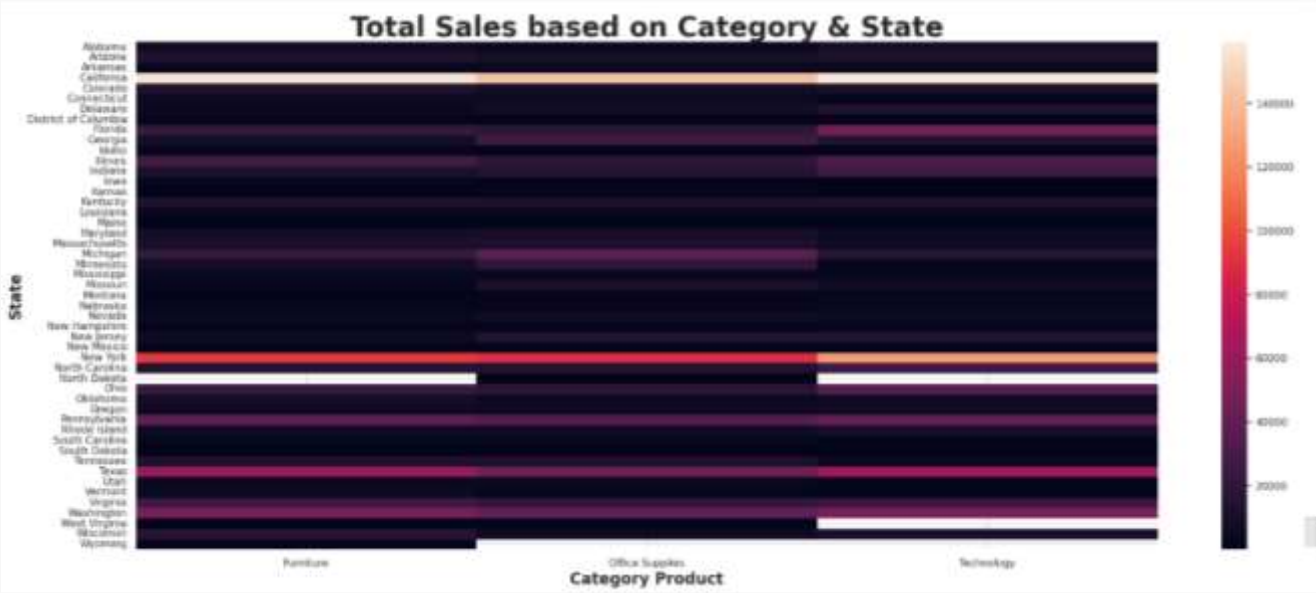


Sales patterns based on segments in various regions are relatively the same. Sales are dominated in the west and east regions.

Total Sales and Total Profit based on Segment

All total sales and total profit patterns are almost the same. The highest total sales and total profit came from the corporate segment while the lowest total sales and total profit are from the consumer segment, which suffered a considerable loss.

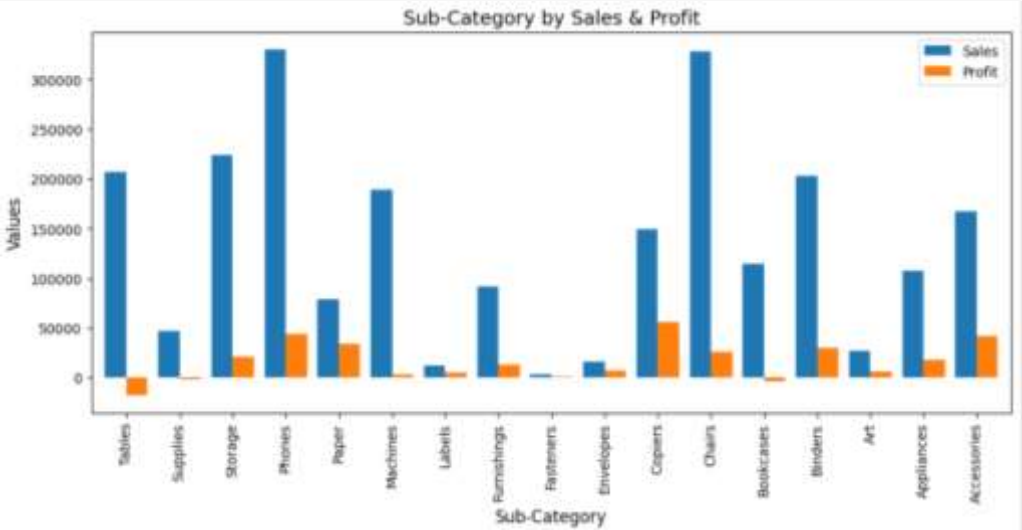
Total Sales based on Category and State



The highest sales were in California & North Dakota with more than 140,000 sales.

In the North Dakota region, office supplies were the smallest of the other two products. Average sales at each superstore branch ranged under 4000.

Sub-Category by Sales and Profit



Sales of subcategory items that table ranks fourth in sales but incurred the most losses, in addition to tables, book cases and supplies are also subcategory products that suffer sales losses. Copiers are the most profitable subcategory product. Phones are high-profit and high-sales products.

HOW TO IMPROVE ?

If the sub-categories products of supplies, bookcases, and tables are removed, the profit margin will increase to 16%.

DATA MODELLING

Data Modelling menggunakan Decision Tree Classifier

```
Decision Tree - Akurasi Training: 0.9995678478824547
Decision Tree - Akurasi Testing: 0.6997840172786177
Classification Report - Decision Tree:
              precision    recall  f1-score   support

 Accessories       0.75      0.76      0.76       194
  Appliances       0.44      0.39      0.42       117
         Art       0.65      0.64      0.64       199
     Binders       0.72      0.78      0.75       381
      Chairs       0.86      0.82      0.84       154
     Copiers       0.59      0.59      0.59        17
   Envelopes       0.29      0.25      0.27        63
   Fasteners       0.30      0.31      0.31        54
 Furnishings       0.89      0.92      0.90       239
      Labels       0.46      0.44      0.45        91
    Machines       0.62      0.55      0.58        29
       Paper       0.67      0.67      0.67       343
      Phones       0.80      0.80      0.80       222
     Storage       0.74      0.69      0.72       212

    accuracy                           0.70      2315
   macro avg       0.63      0.62      0.62      2315
weighted avg       0.70      0.70      0.70      2315
```

Sub-Category target modeling using the Decision Tree Classifier obtained a training accuracy value of 99% and testing accuracy of 69%.

Data Modelling menggunakan Gradient Boosting

```
Gradient Boosting - Akurasi Training: 0.8252664938058196
Gradient Boosting - Akurasi Testing: 0.7239740820734342
Classification Report - Gradient Boosting:
              precision    recall  f1-score   support

 Accessories       0.71      0.78      0.74       194
  Appliances       0.51      0.34      0.41       117
         Art       0.67      0.74      0.70       199
     Binders       0.77      0.78      0.77       381
      Chairs       0.89      0.82      0.86       154
     Copiers       0.71      0.59      0.65        17
   Envelopes       0.45      0.08      0.14        63
   Fasteners       0.52      0.30      0.38        54
 Furnishings       0.89      0.94      0.91       239
      Labels       0.66      0.49      0.57        91
    Machines       0.81      0.59      0.68        29
       Paper       0.61      0.80      0.69       343
      Phones       0.75      0.73      0.74       222
     Storage       0.75      0.75      0.75       212

    accuracy                           0.72      2315
   macro avg       0.69      0.62      0.64      2315
weighted avg       0.72      0.72      0.71      2315
```

Sub-category target modelling using Gradient Boosting gets a training accuracy of 82% and and a testing accuracy of 72%.

Data Modelling menggunakan Random Forest Classifier

Accuracy : 68.7257019438445

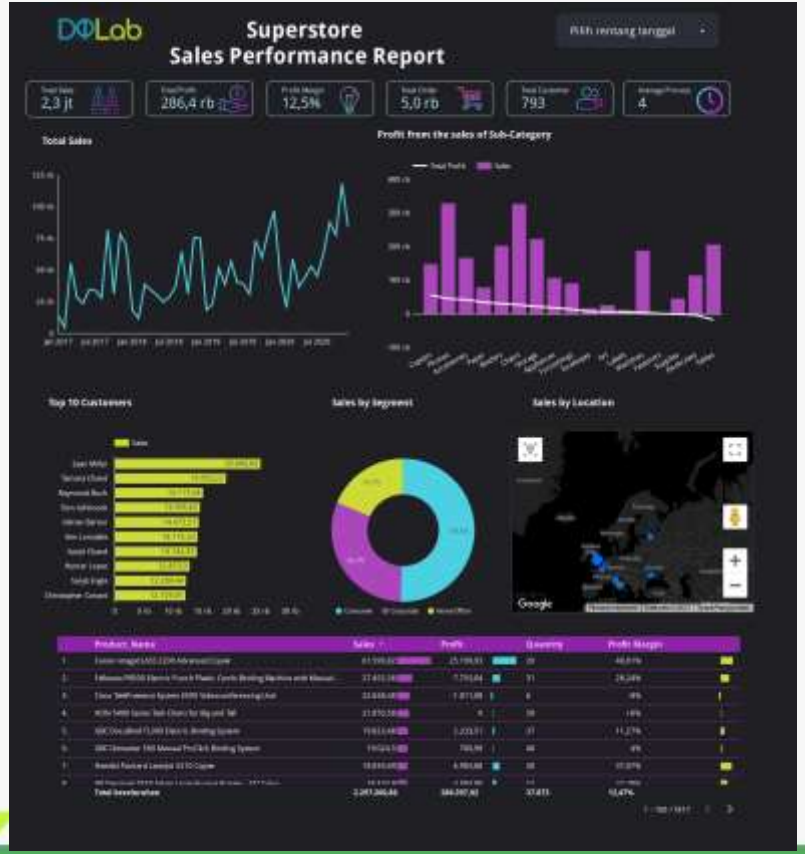
Report :

	precision	recall	f1-score	support
Accessories	0.71	0.79	0.75	194
Appliances	0.41	0.21	0.27	117
Art	0.58	0.68	0.63	199
Binders	0.73	0.80	0.76	381
Chairs	0.93	0.90	0.91	154
Copiers	0.60	0.35	0.44	17
Envelopes	0.10	0.05	0.06	63
Fasteners	0.28	0.13	0.18	54
Furnishings	0.93	0.95	0.94	239
Labels	0.31	0.24	0.27	91
Machines	0.74	0.48	0.58	29
Paper	0.58	0.69	0.63	343
Phones	0.77	0.76	0.76	222
Storage	0.70	0.73	0.71	212
accuracy			0.69	2315
macro avg	0.60	0.55	0.57	2315
weighted avg	0.67	0.69	0.67	2315

Sub-category target modelling using Random Forest Classifier gets a training accuracy 68%.

DASHBOARD SUPERSTORE

DASHBOARD



Link :
https://lookerstudio.google.com/reporting/e74a1f49-b328-49d5-8bd8-d67ba1d30cc5

KESIMPULAN

According to the findings of the EDA (Expaloratory Data Analysis), the Consumer category dominates the most, accounting for around 50% of total sales. Due to the huge year-end discount, sales increased significantly at the end of each year. Copiers are the most profitable subcategory product sales. Sales of sub-category products that suffered the losses come from tables, book cases, and supplies.

To overcome the loss of inaccurate sales strategies at each superstore branch, the sub-category products of supplies, bookcases, and tables are eliminated, the profit margin will increase to 16%. from 3 machine learning modelling data it can be concluded that modelling using Decision Tree Classifier is higher in accuracy than using Gradient Boosting and Random Forest Classifier, hence Decision Tree Classifier is utilized.

Link Google Colab EDA (EXPLORATORY DATA ANALYSIS) :

https://colab.research.google.com/drive/1VZXVhbCbrwx-ojybK7R17sGcmLLR8YaM?usp=sharing

Link Google Colab Data Modelling Machine Learning :

https://colab.research.google.com/drive/1jw3XTFQgaC0ANgbvWhNQqF6aw6y_fW0u?usp=sharing

Link Looker Studio Dashboard :

https://lookerstudio.google.com/reporting/e74a1f49-b328-49d5-8bd8-d67ba1d30cc5