

Part II - (Analysis on Sales and Market Performance)

by (David Kipngeno Kiplangat)

Investigation Overview

Analysis task steps the analysis of this task took an approach as indicate below equivalently.

1.Data understanding. - This phase i took time to load the data print a sample of the lines and computing basic summary statistics to better understand the data.

2.Exploratory data analysis and visualization. This section covered the visualization of various variables and description of possible findings.

3.Explanatory data analysis. -the relationship in the data variables were identified, this was followed by attempt to explain the nature of the behaviour of the data.From this phase is where answering of most questions was done.

4.Conclusion and summary. This entailed the general findings from the data and the possible wisdom derivation from the data. An extensive report was produced providing the guidelines on the way to do the business and the best strategies to approach to better realize the companies goals and objectives

Analysis task steps the analysis of this task took an approach as indicate below equivalently.

1.Data understanding. - This phase i took time to load the data print a sample of the lines and computing basic summary atavistic to better understand the data.

2.Exploratory data analysis and visualization. -this section covered the visualization of various variables and description of possible findings.

3.Explanatory data analysis. -the relationship in the data variables were identified, this was followed by attempt to explain the nature of the behaviour of the data. From this phase is where answering of most questions was done.

4. Conclusion and summary. This entailed the general findings from the data and the possible wisdom derivation from the data. An extensive report was produced providing the guidelines on the way to do the business and the best strategies to approach to better realize the companies goals and objectives

Dataset Overview

The dataset employed for this task and analysis was originating from business setup environment. The Link to the data source is available : https://docs.google.com/spreadsheets/d/16F2awB6f0w-IWViGMgcEEgm5Xh4kigBkLGMSL_nbcnk/edit#gid=631415463 [.\(https://docs.google.com/spreadsheets/d/16F2awB6f0w-IWViGMgcEEgm5Xh4kigBkLGMSL_nbcnk/edit#gid=631415463\)](https://docs.google.com/spreadsheets/d/16F2awB6f0w-IWViGMgcEEgm5Xh4kigBkLGMSL_nbcnk/edit#gid=631415463) It is contains records of sales profits and items sold from different regions and segments and by different sales persons from a business environment.

```
In [1]: # import all packages and set plots to be embedded inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
# suppress warnings from final output
import warnings
warnings.simplefilter("ignore")
```

```
In [2]: # Load in the dataset into a pandas dataframe
# Loading the sales data from the csv file
data = pd.read_csv('SalesData.csv')
data.head()
```

Out[2]:

	CustomerID	CustomerName	BusinessSegment	Country	Region	State	City	PostalCode	Order ID	Order Date	...	Product Name	Price	Cost	
0	A33717C73120	Aaron Bergman	Consumer	United States	Central	Oklahoma	Oklahoma City	73120	CA-2013-140935	2020-11-11	...	Sauder Facets Collection Library, Sky Alder Fi...	142.8000	74.764398	20
1	A33717C73120	Aaron Bergman	Consumer	United States	Central	Oklahoma	Oklahoma City	73120	CA-2013-140935	2020-11-11	...	Samsung Convoy 3	76.4444	22.286997	20
2	A33717C76017	Aaron Bergman	Consumer	United States	Central	Texas	Arlington	76017	CA-2011-152905	2018-02-19	...	Akro Stacking Bins	7.1538	4.041695	20
3	A33717W98103	Aaron Bergman	Consumer	United States	West	Washington	Seattle	98103	CA-2011-156587	2018-03-07	...	Carina 42"Hx23 3/4"W Media Storage Unit	74.3636	41.543911	20
4	A33717W98103	Aaron Bergman	Consumer	United States	West	Washington	Seattle	98103	CA-2011-156587	2018-03-07	...	Newell 330	5.4545	3.099148	20

5 rows × 29 columns

Sales By Profit I was able to observe that the general sales as compared to profit had an intresting linear relationship. There was a significant positive relationship between the 2. sales made therefore can be said to directly influence the profit likely to meade. With this , it can be said that the sales of a product qualifies to be the predictor of that product as an increase of product sales leads to a corresponding increase of profit it generates.

Cost By Price Performance This was made to understand the behaviour of the business in terms of cost undertaken to produce the product and the reflexive amount the product is sold. it was discovered that the cost and price had a positive correlation between each other . A product that consume much resources to

produce was equivalent sold at a higher price. This helped to bring a better plough back profit which considerably was the business return on investment. I was as well noted that a certain 2 products had extremely high cost of production and were as well sold at a quite huge price. This is as can be clearly observed in the visualizations below.

Orders performance there was none uniform distribution for orders that were made. As it can be observed, there were many orders of 1 and 5. This means that most orders made could not exceed 5 and that this range was the range that was affordably possible to most customers. This suggests that, probably the class purchasing or rather making orders is different. There were significance small orders exceeding 10. That is to say few customers can actually afford this amount and are forced to only within their reach.

Most Sold Product Staples was the most sold product followed by every none sticky Binders. This was suggestive that the consumers of the product were mostly office workers. or it could mean that there was a great demand to office furnishing. With the season, the best to invest in would be that one of staples.

Sales Performance The sales behaviour for the data set appeared to have several of outlier. There was an exaggerate price of commodities. This can easily to explain whether the commodity also had majority sales and attracted most profits. The price of the commodity could have as well been influenced by the demand of the product. The time between ordering of the product and its sale would help to understand how the product traffic was. i.e, if the product takes less time then it translates to its high demand by customers.

Most Profitable Regions The regional profit performance was leaded by the region of west and east regions. This regions therefore are worth investing into in as far as profit is concerned.

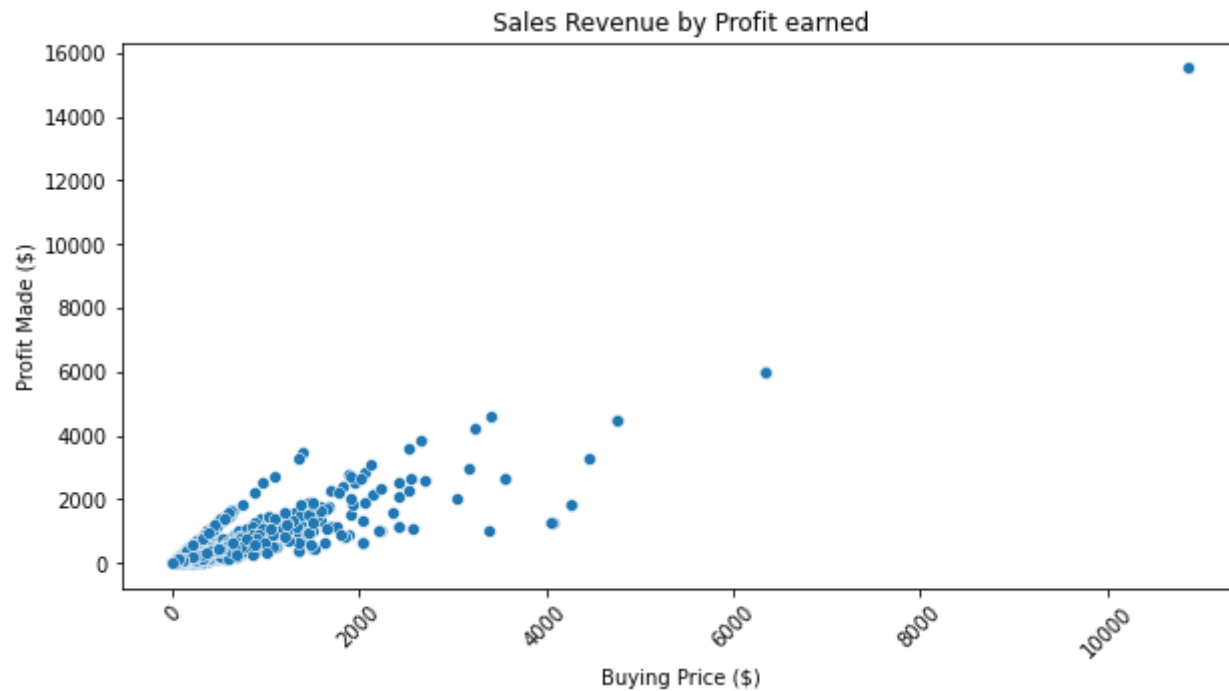
Most Profitable Business Segment the segment of Consumer brought in most profit followed by the segment of corporate and lastly was the segment of Home office.

Best Performing Year By Profit the year of 2021 was observed to be at pick of its performance in as far as profit income is considered. This could have be contributed by the marketing strategy employed by the sales persons which could have probably influenced the Purchase pattern of customers and thus attracted most profits.

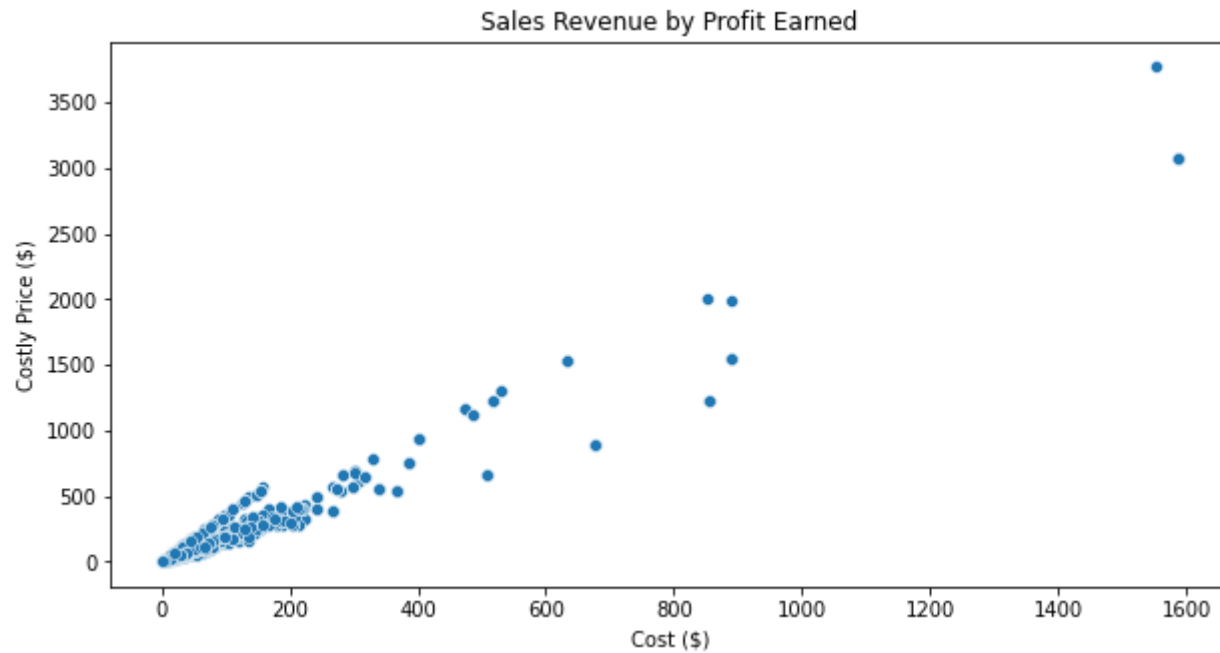
(Visualization)

This section showcases the visualization from which the findings and observations have been made in the above section.

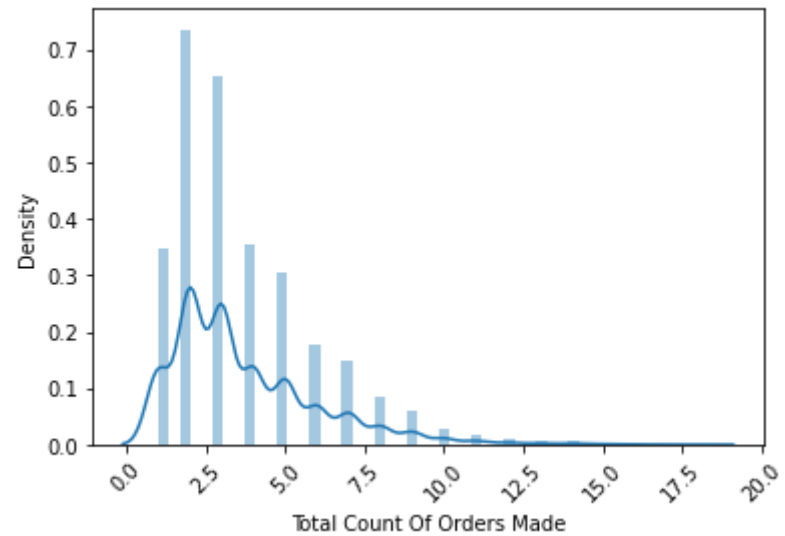
```
In [3]: # Sales made BY profit Earned
def plot_Profit_by_sales(data):
    # setting the figure size
    plt.figure(figsize=[10,5])
    # plotting the data
    sns.scatterplot(data=data,x='BuyingPrice',y='Profit')
    # labelling the figure
    plt.xlabel('Buying Price ($)')
    plt.ylabel('Profit Made ($)')
    plt.title('Sales Revenue by Profit earned')
    plt.xticks(rotation=45)
    plt.show()
plot_Profit_by_sales(data)
```



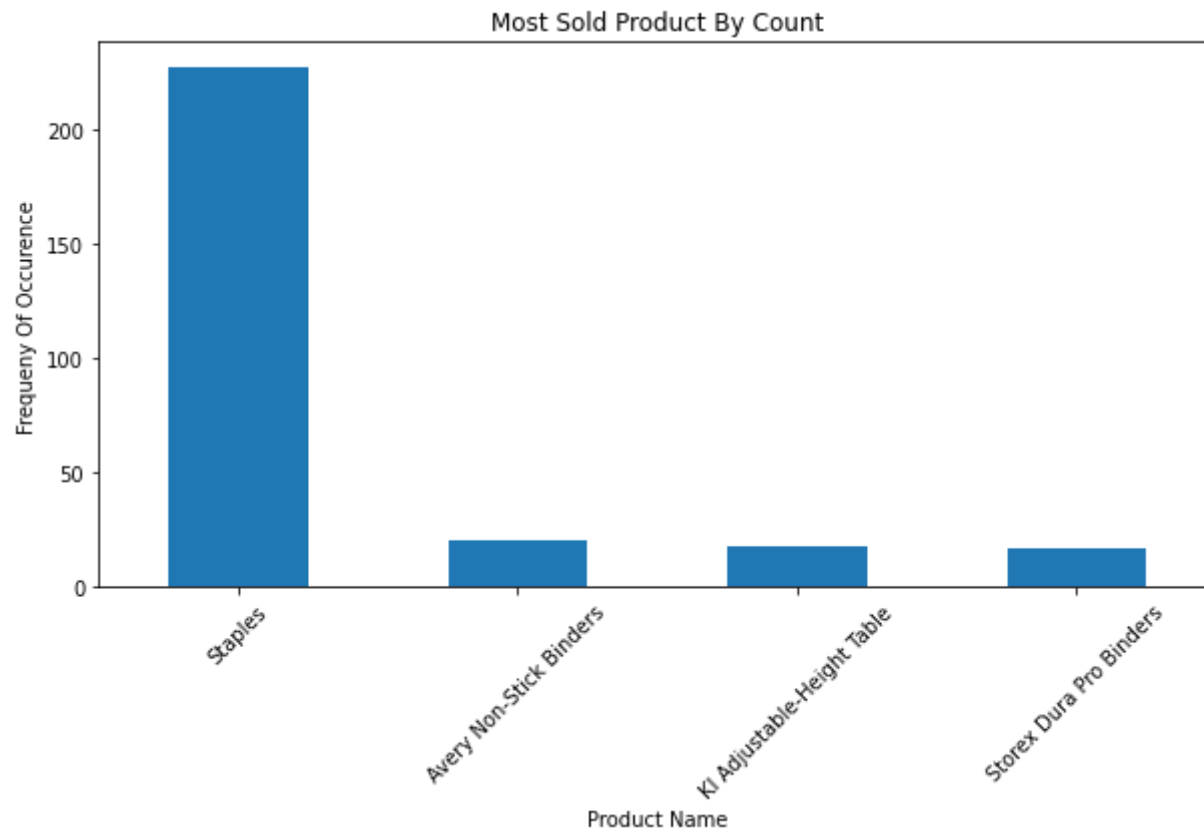
```
In [4]: # COST TO PRICE UNDERSTANDING AND PLOTTING.  
def plot_Profit_by_sales(data):  
    # setting the figure size  
    plt.figure(figsize=[10,5])  
    # plotting the data  
    sns.scatterplot(data=data,x='Cost',y='Price')  
    # labelling the figure  
    plt.xlabel('Cost ($)')  
    plt.ylabel('Costly Price ($)')  
    plt.title('Sales Revenue by Profit Earned')  
    plt.show()  
plot_Profit_by_sales(data)
```



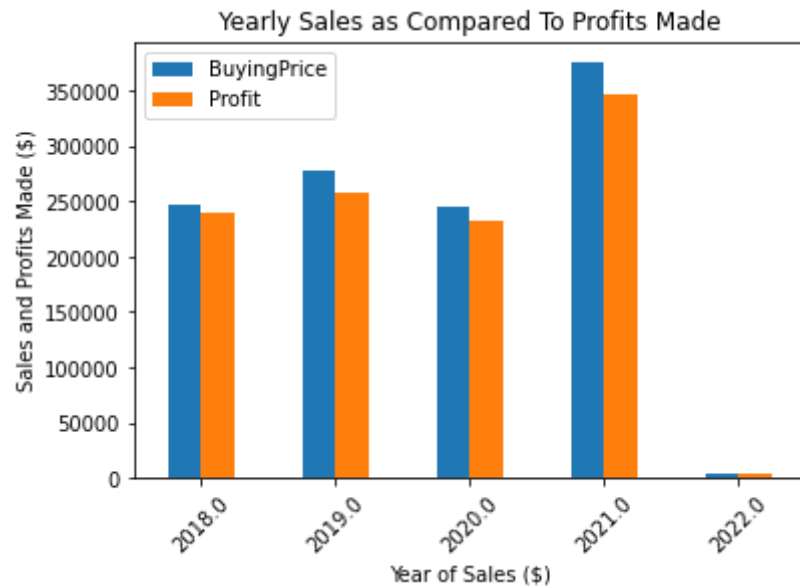
```
In [5]: # plotting a distribution of the orders sales.  
sns.distplot(data['OrderQty'])  
plt.xticks(rotation=45)  
plt.xlabel('Total Count Of Orders Made')  
plt.show()
```



```
In [6]: # plotting the most sold product.  
def plot_most_sold_product(data):  
    most_sold_product = data['Product Name'].value_counts().head(4)  
    plt.figure(figsize=[10,5])  
    most_sold_product.plot(kind='bar')  
    plt.xlabel('Product Name')  
    plt.ylabel('Frequency Of Occurrence')  
    plt.title("Most Sold Product By Count")  
    plt.xticks(rotation=45)  
    plt.show()  
# calling the function  
plot_most_sold_product(data)
```




```
In [7]: # creating a function
def yearly_sales(data):
    yearlySales = data.groupby('Year')[['BuyingPrice', 'Profit']].sum()
    yearlySales = pd.DataFrame(yearlySales)
    yearlySales.plot(kind = 'bar')
    plt.xlabel('Year of Sales ($)')
    plt.ylabel('Sales and Profits Made ($)')
    plt.title('Yearly Sales as Compared To Profits Made')
    plt.xticks(rotation=45)
    plt.show()
# calling the function
yearly_sales(data)
```



```
In [10]: # !jupyter nbconvert Part_II_david_sales_analysis.ipynb --to slides --post serve --no-input --no-prompt
```

```
In [ ]:
```

