Dang Khoa Le
ID: 103844421

Assessment 2: Major Assignment Part A

FIN10002: Financial Statistics
Semester 1, 2023

Dang Khoa Le
ID: 103844421

Table of Contents

Introduction:

At foremost, the Super Store Data contain an enormous size of data with 9994 recorded transactions, which will thereby be randomly filtered to a dataset of 200 orders, varies by 9 different variables, including Order Date, Ship Mode, Segment, Region, Category, Quantity, Sales, Discount, and Profit. The ultimate aim of this report is to analyse different variables on their unique features, their relativity amongst other sectors and how they affect the orders suggestively.

In order to accomplish this task, numerous tools such as Excel, Jupyter Notebook Python, Pivot Table, Descriptive Statistics and graphic analysis with the reasonable methods are required. The initial step is to load the data into Jupyter Notebook, using Python to process the data in Tables and Graphs, followed by analysing the data using various corresponding techniques in Excel.

The project scope is to investigate and determine the relationships between variables and how they affect the identity of orders in tally.

The processes used in this report will include data clarification and preparation, data analysis using Pivot Table techniques, Graphic analysis and a methodical description. Pivot Table is utilised to summarise and visualise the data to gain insights into the relationships between the different entities. Meanwhile graphical illustration will be deployed to present the data in a visually appealing way.

In summary, the investigation are delved into processing data from the sample to acquire the relationship across different entries, therefore, suggestively provide an inform advice for business to have better economic decision.

Part A:

1. Select a random sample

In order to obtain 200 Random transactions from the 9994 Super Store Data on Excel, these steps were implemented:

- Data → Data Analysis → Random Number Generation,
- In the Random Number Generation window, type in:
  - Number of Variables: 1
  - Number of Random Number: 9994
  - Distribution: Uniform
  - Between: 1 and 9994
  - Select the Output Range at the box next to the last right column (Profit)
- Name the column header to be Random,
- Afterwards, go back to Home → Sort → Custom Sort → Sort by – Random – Smallest to Largest (or opposite of your choice),
- Then, select the first 201 rows (or the first 200 order), which is the required Random Sample.

When the Sample is ready, it is optional to make several copies of the Sample for later usages in different tasks.

2. Descriptive statistics

Below are the brief descriptions, summary tables and graphic illustrations accordingly to all 9 variables with data obtained from the Random 200 transactions generated before.

A. Order Date

Month	Monthly Order
Jan	7
Feb	7
Mar	8
Apr	13
May	14
Jun	20
Jul	16
Aug	11
Sep	20
Oct	26
Nov	24
Dec	34
Grand Total	**200**
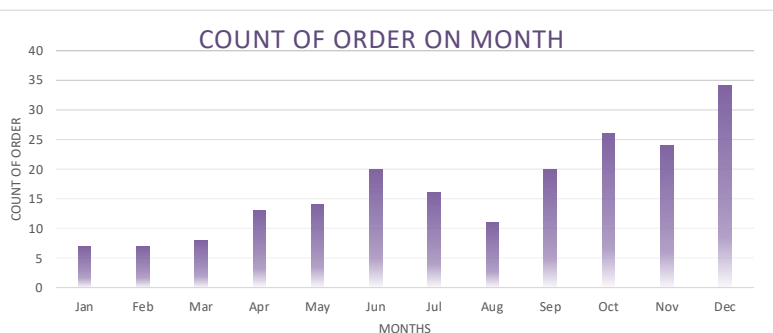
From visual observation of the Pivot Table and the Graph simultaneously, the data depicts the gradual differences in the number of monthly transactions varied from January to December. The data reveals that there is an increasing trend in the number of orders from January to March, which then significantly increases from April to June. Noticeably, the months of June and July show a peak in the number of orders, followed by a slight dip in August. September shows a sudden increase in the count of orders, which further rises and reaches its peak in October and November. Finally, December shows a considerable increase in the Count of Orders, with the highest number of transactions in the entire year.

B. Ship Mode

Row Labels	Count of Ship Mode	Relative Frequency
First Class	38	19.00%
Same Day	8	4.00%
Second Class	40	20.00%
Standard Class	114	57.00%
Grand Total	**200**	**100.00%**
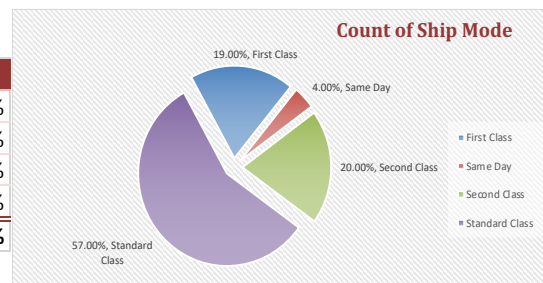
From the Pivot Table and the Graphical demonstration,

the data emphasizes characterized differences in the number of 4 Ship Modes, in consist of First Class, Second Class, Standard Class and Same Day. The highest number of orders (114) were delivered using the Standard Class, followed by Second Class (40), First Class (38), and Same Day (8). This insinuates customers high preference on the more economical and slower shipping options, like the First, Second Class and especially the Standard Class. Additionally, the remarkable low count of orders shipped using Same Day indicates its unpopularity amongst other methods.

C. Segment

Row Labels	Count of Segment	Segment Frequency
Consumer	104	52.00%
Corporate	64	32.00%
Home Office	32	16.00%
Grand Total	200	100.00%
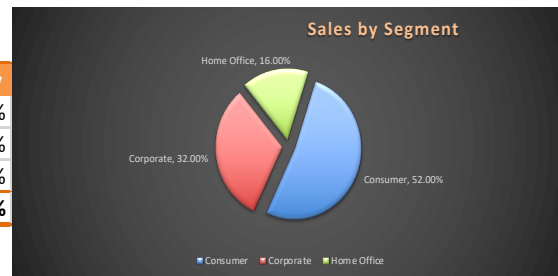
The data notations show the distribution of transactions across different Segments, including Consumer, Corporate and Home Office. Consumer segment has the highest number of orders with a count of 104, followed by Corporate counterpart with 64 orders and Home Office with 32 orders. This indicates that the majority of purchases are coming from the Consumer segment, while the other two segments have comparatively fewer orders. It is important to understand this difference in order to make informed decisions for the business, such as targeted marketing or adjusting Sales strategies based on the segment that generates the most revenue.

D. Region

Row Labels	Count of Region	Region Frequency
Central	50	25.00%
East	54	27.00%
South	32	16.00%
West	64	32.00%
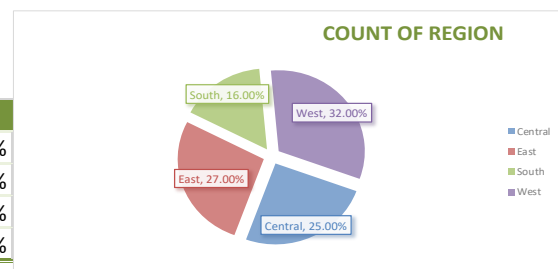
The given data shows the count of orders placed from four different Regions - Central, East, South, and West. It can be observed that the West region has the highest count of transactions with 64, followed by the East region with 54. The Central Region has 50 orders, while the South-side has the lowest count of 32 orders. This indicates that the West and East regions are more active and have more orders compared to the other regions. The data suggests that there may be differences in customer behaviour or demand in each region, which analysing data can enable understanding and identifying potential opportunities or challenges.

E. Category

Row Labels	Count of Category	Frequency of Category
Furniture	44	22.00%
Office Suppli	121	60.50%
Technology	35	17.50%
Grand Total	200	100.00%
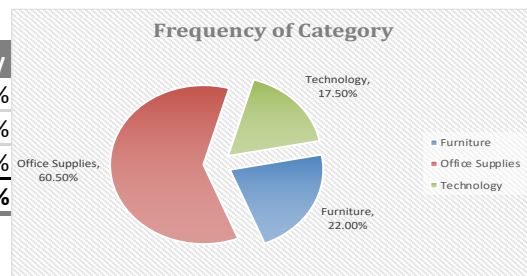
The table above reveal the Count of Category, which presents the number of orders in each category. We can observe that Office Supplies has the highest number of orders (121), followed by Furniture (44) and Technology with the lowest order count (35). This suggests that Office Supplies are more frequently ordered compared to the other categories. The data implies that there may be a significant low and high demand in each of these category customer types.

F. Sales

Sales Amount	Count of Sales	Bins	Frequency
<100	131	50	102
100-300	28	100	29
300-500	18	200	16
500-700	9	300	12
700-900	4	400	14
900-1100	4	500	4
1100-1300	2	600	4
1300-1500	1	700	5
1900-2100	1	3000	14
2500-2700	1		
2700-2900	1		
Grand Total	200		
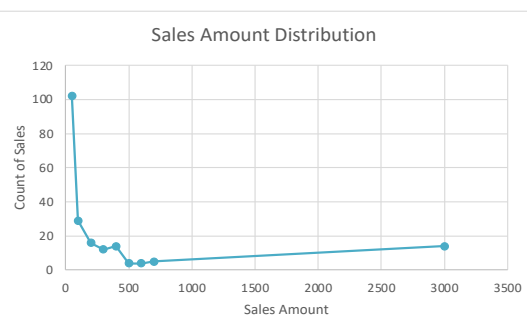
The data representations above depict the distribution of Sales amounts based on the count of times a specific amount occurred. The data indicates that the majority of sales were less than 100, with 131 occurrences. Only a small

percentage of sales fall into the higher value ranges, with only 1 occurrence each for sales greater than 1900. The distribution of sales is heavily skewed towards lower sales amounts, indicating that the majority of transactions placed in the less costly purchases. The frequency decreases as the sales amount increases, indicating that higher sales amounts were less common.

G. Quantity

Quantity Amount	Sum of Quantity
1	14
2	90
3	132
4	96
5	140
6	96
7	77
8	64
9	54
11	33
14	14
Grand Total	810

The data reveals the Quantity amounts and the relative Count of Quantity that products sold at each quantity level. The most common quantity size to be sold is 5, with a total of 140 products sold at that quantity. The table shows that most customers tend to buy products in the size of 3 to 6, with a total of 468 products sold at those quantity levels. quantities portion of 1 and 2 are also relatively common, with a total of 104 products sold at those quantity levels. The remaining quantity size (7 to 14) have fewer products sold, with only 290 products sold in total. The average quantity portion to be sold is approximately 4.12.

H. Discount

Discount Amount	Count of Discount
0%	97
10%	3
20%	82
30%	1
45%	1
60%	3
70%	4
80%	9
Grand Total	200

These data sets show the Count of Discount and the Discount Amount for each order. The majority of orders (97 out of 200) did not have any discount applied. The most common discount rate applied was 20% (82 out of 200), followed by 0% (no discount) and 80% (9 out of 200). The remaining choices of discount were less common, with only 1 order each for 30% and 45% discounts, and a few orders with higher discounts of 60% and 70%. Overall, the data shows that discounts are not applied very frequently or likely to be at a low rate.

I. Profit
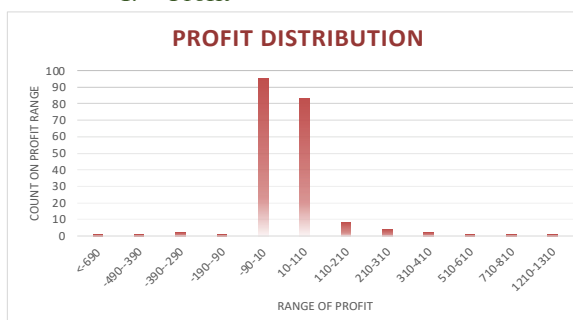
| Profit Amount	Count of Profit
<-690	1
-490--390	1
-390--290	2
-190--90	1
-90-10	95
10-110	83
110-210	8
210-310	4
310-410	2
510-610	1
710-810	1
1210-1310	1

Based on the given data source, the majority of the profits fall between the range of -90 to 10, as proved by the highest count of profit falling within that range. There are also noticeable Counts of Profit within the range of 10-110 and 110-210. On the other hand, there are very few Profits falling outside of the range of -390 to 610, indicating a relatively narrow range of profit amounts. Additionally, there are a few extreme values of profit, such as the expense (minus profits) of more than -690 or profits greater than 810, but these are rare occurrences. In overall, the distribution of profits seems to be skewed towards the lower end of the range, with the majority of profits falling within a relatively narrow range.

3. Dashboard

A. Monthly Sales by Region

Sum of Sales	Regional Sales			
Month	Central	East	South	West
Jan	$310.15	$15.17	$545.94	$24.85
Feb	$543.54	$151.04	$0.00	$0.00
Mar	$585.98	$577.84	$42.68	$140.26
Apr	$71.12	$31.10	$3,971.33	$304.19
May	$176.48	$522.36	$66.68	$803.09
Jun	$99.74	$783.98	$675.12	$1,416.94
Jul	$705.21	$655.20	$287.36	$730.80
Aug	$36.88	$117.14	$298.50	$129.75
Sep	$391.99	$3,776.25	$47.56	$1,964.40
Oct	$131.86	$93.04	$1,538.25	$1,186.93
Nov	$1,922.66	$2,200.40	$92.85	$3,876.79
Dec	$2,203.08	$1,735.86	$417.56	$3,820.21
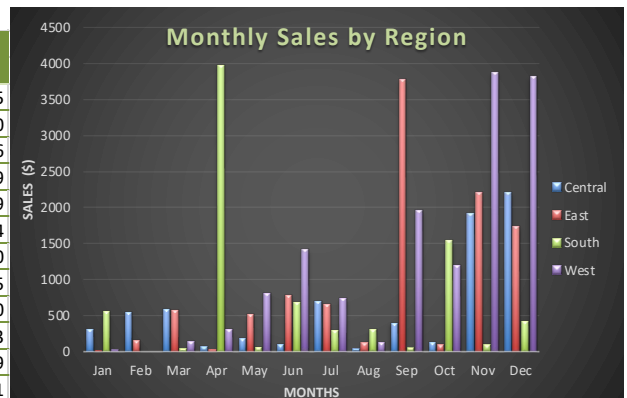
The analysis of the Table and Graphic notations shows that the highest sales occurred in December, with a total sales figure of $11,176.71, which is more than double the sales in the next highest month, November, with a total sales figure of $7,092.70. The Central region has the highest sales in December, with a total sales figure of $2,203.08, followed by the West region with a total sales figure of $3,820.21.

It is also noticeably that there were no sales in the South side in February. Additionally, the South region had the lowest sales in July, with a total sales figure of $287.36. The East region had the lowest sales in January, with a total sales figure of $15.17.

This Graphic investigation can be used to identify the best and worst performing regions by month and to identify sales trends over time. Alongside with the Graph, the Pivot Table can also be used to compare sales figures across different regions and to identify areas for improvement. This analysis is ideal for businesses to make methodical decisions on where and when to focus their marketing efforts and to optimize sales strategy.

B. Regional Profit by Category

Sum of Profit	Region Profit			
Category	Central	East	South	West
Furniture	-$219.22	$1,099.28	-$107.82	$342.70
Office Supplie	-$265.27	$744.54	$1,627.84	$697.48
Technology	$342.82	$654.09	$270.77	$617.76

This data sets illustrate the Regional profit by Category. It presents a comparison of the total profit in each region from three categories: Furniture, Office Supplies, and Technology.
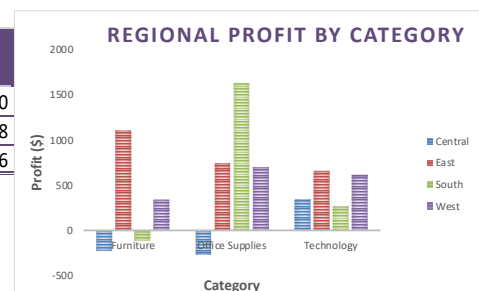
· The South region has the highest profit from Office Supplies, whereas the East region has the highest profit from Furniture.
· The Technology category has contributed positively to the profit of all four regions, and the Central region has the highest profit from this category.
· The Furniture category has resulted in losses for the Central and South regions, whereas the Office Supplies category has resulted in a loss only for the Central region.

This table acknowledge businesses to understand which product categories are most profitable in each region, and enable profitable decisions based on this information.

C. Monthly Average Discount by Region

Average of Discount	Regional Discount			
Month	Central	East	South	West
Jan	30%	20%	0%	0%
Feb	33%	10%	0%	0%
Mar	33%	0%	0%	10%
Apr	0%	0%	10%	5%
May	7%	35%	0%	3%
Jun	40%	8%	20%	3%
Jul	27%	8%	10%	13%
Aug	50%	0%	10%	0%
Sep	23%	7%	20%	8%
Oct	27%	10%	33%	11%
Nov	28%	8%	15%	18%
Dec	20%	12%	12%	22%

**Monthly Average Discount by Region**
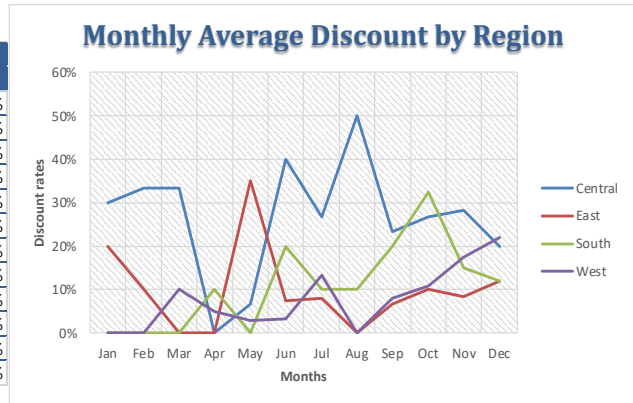
The above Table and Graph shows the Average Regional Discounts offered in different Months by four different regions. From data observation, the discounts offered by each region vary significantly from month to month.

· In January, Central had the highest average discount of 30%, whereas the East had an average discount of 20%. In February, Central had the highest average discount again, but this time at 33%, and the East had an average discount of 10%.
· In March, the East and West regions offered no discounts, while the Central region had an average discount of 33%, and the South region had an average discount of 10%.
· In April, the South region had the highest average discount of 10%, while the other regions had no discounts. In May, the East side had the highest average discount of 35%, while the other regions had lower discounts ranging from 3% to 7%.
· In June, the Central region had the highest average discount of 40%, while the other regions had discounts ranging from 3% to 20%. In July, the Central region had the highest average discount of 27%, while the other regions had discounts ranging from 8% to 13%.
· In August, the Central region had the highest average discount of 50%, while the other regions had no discounts. In September, the South side had the highest average discount of 20%, while the other regions had discounts ranging from 7% to 23%.
· In October, the South region had the highest average discount of 33%, while the other regions had discounts ranging from 10% to 27%. In November, the West zone had the highest average discount of 18%, while the other regions had discounts ranging from 8% to 28%.
· Finally, in December, the West region had the highest average discount of 22%, while the other regions had discounts ranging from 12% to 20%.

Overall, the discounts offered by each region vary significantly by month, with some regions offering higher discounts than others, depending on the month, which can somewhat insinuate the distribution of discount by occasional trends or by the relationship with customers in different area.

D. Regional Shipment Quantities

Average of Quantity	Regional Quantity			
Ship Mode	Central	East	South	West
First Class	5.0	4.2	3.8	3.7
Same Day	2.0	3.0	4.0	2.3
Second Class	4.0	3.6	5.2	4.4
Standard Class	4.2	4.6	3.3	4.1

REGIONAL SHIPMENT QUANTITIES
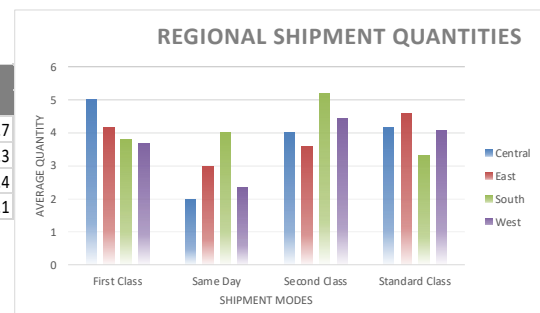
From observation, the Average Quantity of products shipped varies across the different regions and shipping modes. The Central region has the highest average quantity for all shipping modes except for Same Day, where the South region has the highest average quantity. The East region generally has lower average quantities than the Central and West regions. The South region has the lowest average quantity for First Class and Second Class shipments, but the highest for Same Day shipments. The West region has the highest average quantity for Standard Class delivery.

Overall, the fluctuation across each ship modes are minor whereas regional differences in all four areas doesn't seem to have any significant differential level. Thus, this information can be useful for understanding the shipping trends and preferences in different regions for better decisions making related to inventory management and delivery strategies of the business.

E.   Sales by Segment

Segments	Sum of Sales
Consumer	20110.5775
Corporate	16677.861
Home Office	3431.684

The Pie chart and Table shows the difference in the Sales by Segment ratio. It summarizes the total sales amount of each segment including Consumer, Corporate and Home Office.

From the table, we can observe that the Consumer segment has the highest sales amount with a total of 20110.5775, followed by Corporate with a sales amount of 16677.861 and Home Office with a sales amount of 3431.684. Based on this information, we can conclude that the Consumer counterpart is the most profitable segment for the company. Hence, this analysis can be used to perform a proper decision on future business strategies related to the segment selling points.
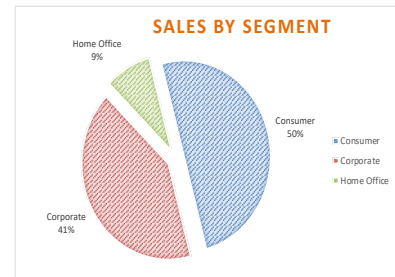
Conclusion

In conclusion, the analysis of the Super Store data set, filtered into a sample of 200 orders has provided useful insights into the different variables that affect sales, profit, and other factors that impact business operations. Through the use of various analytic tools such as pivot tables, descriptive statistics and graphical representations on Excel and Jupyter Notebook, we were able to understand the distribution of the data and the relationship between different variables.

The random sample of 200 transactions provided a lack of accuracy representation from the larger sample, which was evident from the confidence intervals calculated. However, it is indisputable that the sample are still able to keep an acceptable correspondence to the Super Store data, being able to fully provide all key variables, and displaying the relationship across entities, which also mentionable that the smaller sized sample allow better analytic accuracy with minimal effort. The descriptive statistics and graphical representations of each variable support identifying trends, patterns, and outliers. The dashboard with five pivot tables and graphs allowed us to analyse the relationships across different variables and provided us with valuable insights into the segments and regions that contributed the most to sales and profits.

However, it is noticeable that the analysis was based on a random sample of 200 transactions from the original dataset of 9994 orders. The results may not be fully representative of the larger dataset, and further analysis on a larger sample size may be required to confirm the findings. In addition, it is also noticeable that random and systematic errors from human can be undertaken within the assessment, including applying unappropriated method or graph, using wrong Excel formula, Python code and presenting an untechnical description, with lack of spelling or proper demonstration.

Overall, the assessment has demonstrated the usefulness of various analytical tools, including Pivot Tables, Descriptive Statistics and corresponding graphs. These tools can be utilized to gain insights into various factors and variables that impact business activities and aid in better decision making.