BUSINESS DATA MANAGEMENT

CAPSTONE PROJECT

MID-TERM SUBMISSION

"Optimizing Retail Performance:

An Analysis of a Fashion Shop"

NAME: VIVEK KUMAR

ROLL NO: 21F3000834

INDIAN INSTITUTE OF TECHNOLOGY, MADRAS

CHENNAI, TAMIL NADU – 600036

1. Executive Summary:

The Business Data Management Capstone Project is a project a part of the Diploma course of BS in Data Science and Applications offered by the IIT Madras. This capstone project explores the use of business data management and data science techniques to analyze the performance of a retail fashion shop. The project aims to identify the business problems and develop a solution or recommendations that address the business problem.

The project began with data collection and cleaning, including removing duplicate entries and incomplete entries. Most of the required data is collected in an unstructured format. I have to put the data in an MS Excel sheet in a structured way for further analysis. Then the cleaned data is used to create visualizations to understand the graphs, trends, and product performance.

The shop (Flying Colors of Fashion) was started on Oct 2020, three years ago. During the pandemic, the sale is almost zero because as we know that there is very low economic activity. After the pandemic, their business started to grow at an expected pace. The shop's sales and revenue have been normal or strong over the few past years, with the highest sales during the holiday season.

Further analysis will focus on identifying patterns in sales & revenue and stock analysis. This information provides a foundation for future analysis and is used to make data-driven recommendations.

2. Proof of Originality of the Data:

All the collected data are primary and provided by the owner of the shop named 'Flying Colours of Fashion'. Please find below some photos of the shop and a letter given from the shop to validate the originality of the data.

Figure 1 – Photo of the Shop with Stock

Figure 2 & 3 – Photos of the Shop Owner and the Customers

FLYING COLORS OF FASHION

CERTIFICATE OF AUTHENTICITY

TO WHOMSOEVER IT MAY CONCERN

This is to certify that I, Mr. [redacted] Kumar owner of the shop 'Flying Colors of Fashion' has provided the Business data for a few months to Mr. Vivek Kumar for academics purpose for the capstone project in BS in Data Science and Applications. The data provided by me is real up to my knowledge. This student promises us that he will use this data solely for academics and the use of this data outside of academics does not own any credibility.

So, I am certifying that I issue some of my business data to Mr. Vivek Kumar who is claiming himself as a student of the Indian Institute of Technology, Madras.

Date: 22nd Feb 2023

Place: Ranchi, Jharkhand

22.02.23

Business Owner Signature

Ground Floor, Circular Road, Ranchi

Phone: [redacted] E-Mail: [redacted]2468@gmail.com

Figure 4 – Authenticity letter

3. Metadata and Descriptive Statistics:

3.1. Stored and Cleaned data:

The raw data is collected and entries are done in MS Excel Sheets. There are three sheets where data is collected and cleaned for further use in analysis in MS Excel as follows with names and attributes(columns):

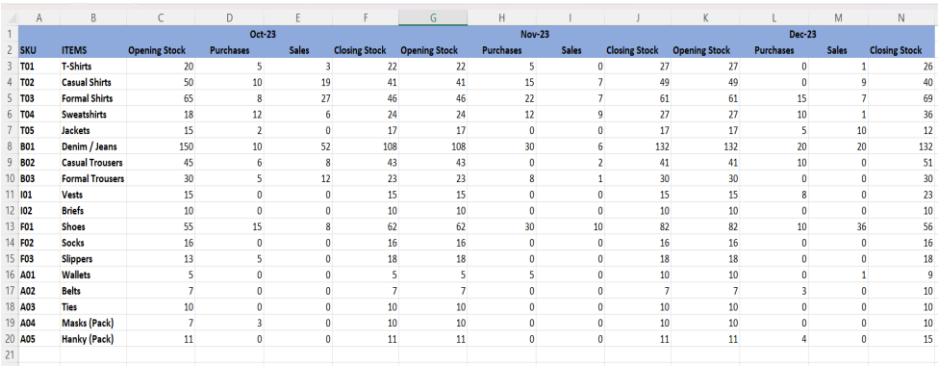
- **Items List**
  - Col A – Categories of items like Topwear, Bottomwear, Accessories, etc.
  - Col B – I have named short termed names of main categories with the proper number format. For ex – If a category is Topwear and has five items in this category, so I named those five items as T01, T02, ……., T05, and so on. I have named this column SKUs (Stock Keeping Units) as it helps eliminate long names and use them in analysis.
  - Col C – Items name as T-shirts, casual shirts, etc.
  - Col D – MRP (Maximum Retail Price) of the various items
  - Col E – Average Selling Price of the items
  - Col F – Average Cost Price of the items

	A	B	C	D	E	F
1	**CATEGORY**	**SKU**	**ITEMS**	**MRP**	**AVG SP**	**AVG CP**
2	Topwear	T01	T-Shirts	₹ 1,399.00	₹ 549.00	₹ 365.00
3	Topwear	T02	Casual Shirts	₹ 2,199.00	₹ 699.00	₹ 550.00
4	Topwear	T03	Formal Shirts	₹ 2,299.00	₹ 739.00	₹ 620.00
5	Topwear	T04	Sweatshirts	₹ 2,499.00	₹ 899.00	₹ 714.00
6	Topwear	T05	Jackets	₹ 4,999.00	₹ 1,379.00	₹ 890.00
7	Bottomwear	B01	Denim / Jeans	₹ 3,499.00	₹ 849.00	₹ 530.00
8	Bottomwear	B02	Casual Trousers	₹ 2,999.00	₹ 1,000.00	₹ 768.00
9	Bottomwear	B03	Formal Trousers	₹ 2,099.00	₹ 680.00	₹ 570.00
10	Innerwear	I01	Vests	₹ 250.00	₹ 199.00	₹ 140.00
11	Innerwear	I02	Briefs	₹ 399.00	₹ 219.00	₹ 132.00
12	Footwear	F01	Shoes	₹ 4,199.00	₹ 1,469.00	₹ 1,184.00
13	Footwear	F02	Socks	₹ 150.00	₹ 89.00	₹ 59.00
14	Footwear	F03	Slippers	₹ 475.00	₹ 310.00	₹ 206.00
15	Accessories	A01	Wallets	₹ 899.00	₹ 540.00	₹ 397.00
16	Accessories	A02	Belts	₹ 799.00	₹ 360.00	₹ 248.00
17	Accessories	A03	Ties	₹ 499.00	₹ 250.00	₹ 190.00
18	Accessories	A04	Masks (Pack)	₹ 249.00	₹ 120.00	₹ 85.00
19	Accessories	A05	Hanky (Pack)	₹ 199.00	₹ 80.00	₹ 55.00
20						

➢ Stock Inventory

- Col A – SKUs
- Col B – Items names
- Col C - F – Opening Stock, Purchases, Sales, and Closing Stock of October month respectively from columns C to F.

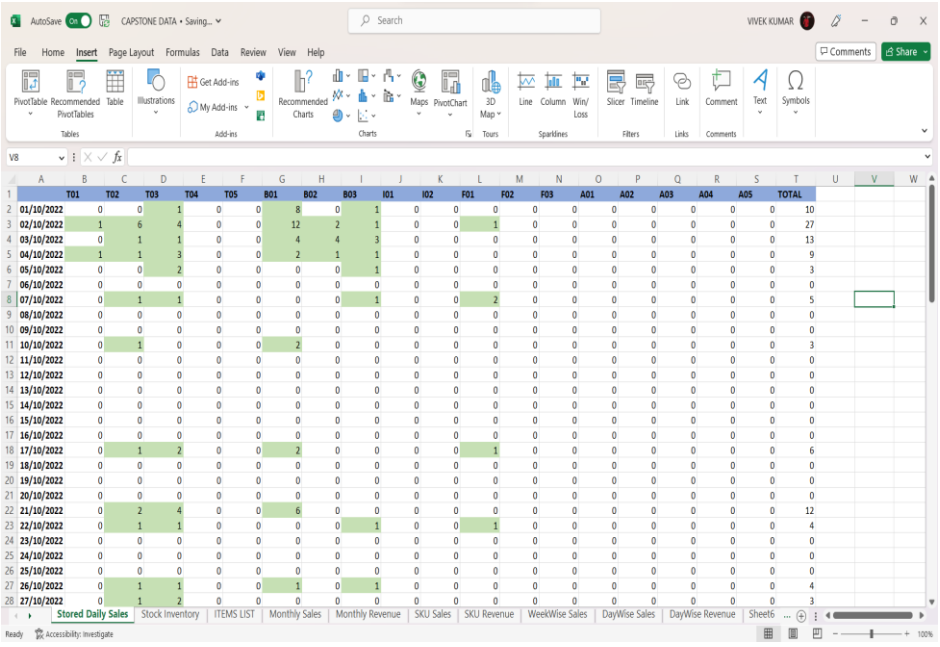
And further columns for Nov and Dec months same as Oct month.



		Oct-23				Nov-23				Dec-23			
SKU	ITEMS	Opening Stock	Purchases	Sales	Closing Stock	Opening Stock	Purchases	Sales	Closing Stock	Opening Stock	Purchases	Sales	Closing Stock
T01	T-Shirts	20	5	3	22	22	5	0	27	27	0	1	26
T02	Casual Shirts	50	10	19	41	41	15	7	49	49	0	9	40
T03	Formal Shirts	65	8	27	46	46	22	7	61	61	15	7	69
T04	Sweatshirts	18	12	6	24	24	12	9	27	27	10	1	36
T05	Jackets	15	2	0	17	17	0	0	17	17	5	10	12
B01	Denim / Jeans	150	10	52	108	108	30	6	132	132	20	20	132
B02	Casual Trousers	45	6	8	43	43	0	2	41	41	10	0	51
B03	Formal Trousers	30	5	12	23	23	8	1	30	30	0	0	30
I01	Vests	15	0	0	15	15	0	0	15	15	8	0	23
I02	Briefs	10	0	0	10	10	0	0	10	10	0	0	10
F01	Shoes	55	15	8	62	62	30	10	82	82	10	36	56
F02	Socks	16	0	0	16	16	0	0	16	16	0	0	16
F03	Slippers	13	5	0	18	18	0	0	18	18	0	0	18
A01	Wallets	5	0	0	5	5	5	0	10	10	0	1	9
A02	Belts	7	0	0	7	7	0	0	7	7	3	0	10
A03	Ties	10	0	0	10	10	0	0	10	10	0	0	10
A04	Masks (Pack)	7	3	0	10	10	0	0	10	10	0	0	10
A05	Hanky (Pack)	11	0	0	11	11	0	0	11	11	4	0	15

➢ Stored Daily Sales

In this sheet, I have recorded the number of sales of different items over three months i.e 1st Oct 2023 to 31st Dec 2023. I have put the date in the rows and different SKUs or items in different columns.



	T01	T02	T03	T04	T05	B01	B02	B03	I01	I02	F01	F02	F03	A01	A02	A03	A04	A05	TOTAL
01/10/2022	0	0	1	0	0	8	0	1	0	0	0	0	0	0	0	0	0	0	10
02/10/2022	1	6	4	0	0	12	2	1	0	0	1	0	0	0	0	0	0	0	27
03/10/2022	0	1	1	0	0	4	4	3	0	0	0	0	0	0	0	0	0	0	13
04/10/2022	1	1	3	0	0	2	1	1	0	0	0	0	0	0	0	0	0	0	9
05/10/2022	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3
06/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
07/10/2022	0	1	1	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	5
08/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
09/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10/10/2022	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3
11/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17/10/2022	0	1	2	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	6
18/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21/10/2022	0	2	4	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	12
22/10/2022	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	4
23/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25/10/2022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26/10/2022	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	4
27/10/2022	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

3.2. Metadata and descriptive statistics:

After recording entries or data points in the excel file, the data is sorted or filtered for the months of Oct 2023 to Dec 2023. For easier and further analysis, I have put the stored daily sales in 2-d in a new sheet named Daily Sales containing different dates for different SKUs or items with the number of sales. There are a total of 1657 data points as different dates in one column against different items (18 items).

Monthly Sales data:

Month	Total Sales	Range (Min-Max)
October	135	0 - 27
November	42	0 - 7
December	85	0 - 11

Monthly Revenue data:

Month	Total revenue (Rs)	Range (Min-Max) (Rs)
October	1,12,335.00	0 – 22,036.00
November	40,621.00	0 – 6,863.00
December	97,106.00	0 – 11,379.00

SKUs Volume and Revenue data:

SKU/Items	Total Sales	Sales Range	Total Revenue (Rs)	Revenue Range (Rs)
B01	78	0 – 13	66,222.00	0 – 11,037.00
F01	54	0 – 5	79,326.00	0 – 7,345.00
T03	41	0 – 4	30,299.00	0 – 2,956.00
T02	35	0 – 6	24,465.00	0 – 4,194.00
T04	16	0 – 6	14,384.00	0 – 5,394.00
B03	13	0 – 3	8,840.00	0 – 2,040.00
B02	10	0 – 4	10,000.00	0 – 4,000.00
T05	10	0 – 3	13,790.00	0 – 4,137.00

T01	4	0 – 1	2,196.00	0 – 549.00
A01	1	0 – 1	540.00	0 – 540.00

- There is zero amount of sales and revenue other than these items given in the above table.

4. Detailed Explanation of Analysis Process/Method:

The main tool used for this analysis is MS Excel. After entering and cleaning the raw data, various methods or tasks are to be done like imputing, sorting, filtering, etc. in excel. The various Excel features are used for creating graphs, pivot tables, and different trends to draw insights and conclusions from the data.

The following actions are taken for the analysis process:

- The entered data in the new sheet naming Daily Sales have three columns (Date, SKUs/Items, and Number of sales). The data is analyzed from 1st Oct 2023 to 31st Dec 2023.
- Further adding different columns or attributes from other excel sheets such as Day, Week No., Average Selling Price, Average Cost Price, Revenue, Total Cost Price, and Profit.
- Using different formulas and relations above added columns are calculated.
- In Stock Inventory data, relevant attributes or columns are calculated using different formulas of excel.
- Further, creating different pivot tables from daily sales data according to different parameters using filters, sorting, etc.
- After creating different tables according to different parameters using the pivot table method, various graphs and trends are created to analyze the data for better visualization and understanding.

5. Results and Findings (Mid-term):

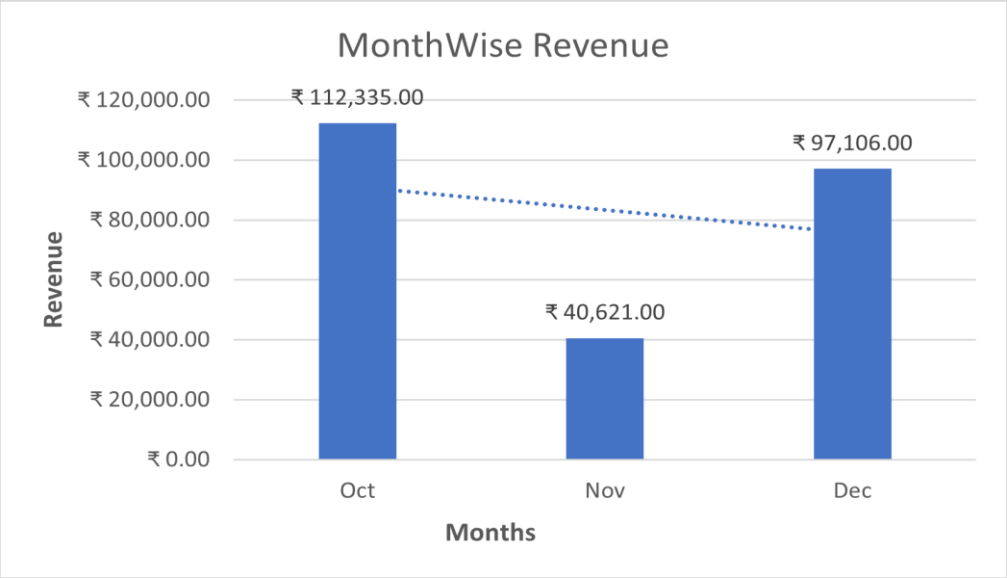
1. **Monthly Sales Analysis**:

   In this graph, we can see a trend of sales volume of first decreasing, then increasing back. The growth rate is linear. By analyzing the graph, we can say that there are maximum sales in the month of Oct 2023 due to the festive season. But, there is a

drop in the volume of sales in Nov 2023 and a further increase in volume by Dec 2023.

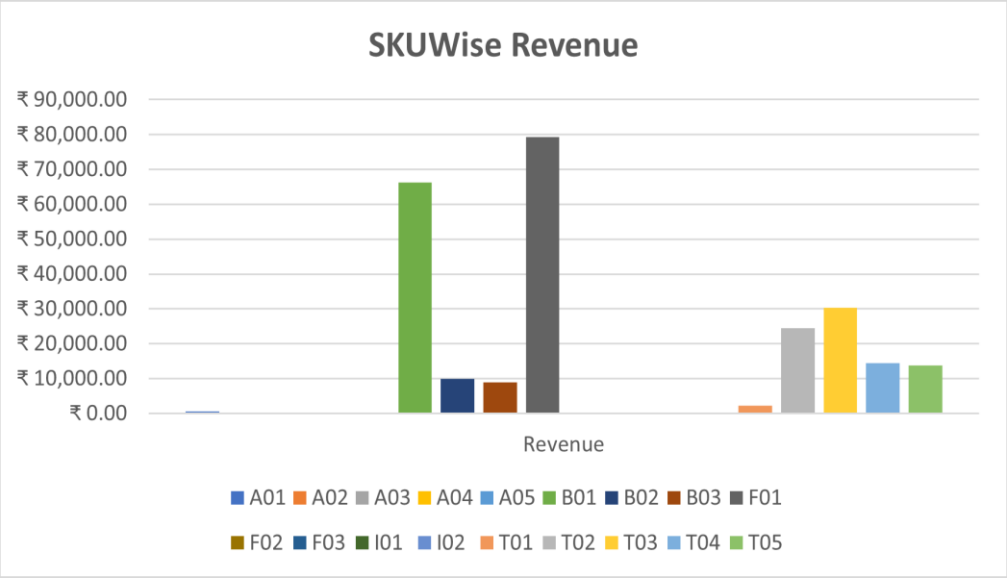
2. **Revenue Growth Analysis:**

In this graph, it is obvious that the volume of sales is directly proportional to the revenue. So, similar inference can be seen in this graph as the monthly sales graph. There is maximum revenue in the month of October and drop in revenue in the month of Nov 2023 and an increase in the month of December 2023.



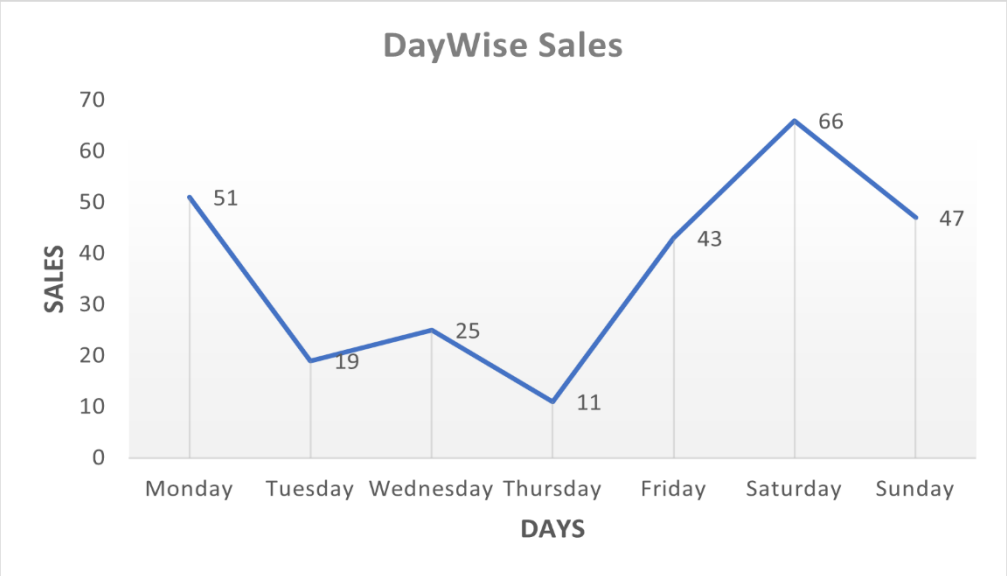
3. **Items-wise Revenue Contribution Analysis:**

From this revenue contribution graph, we can conclude that the most revenue is contributed by F01 followed by B01. Also, T03's revenue contribution growing very fast. The focus is more on these four items or SKUs that generate more of the total revenue.



4. **Day-wise Volume Analysis:**

In this graph of day-wise volume analysis, we can see that the maximum volume of sales is happening on the day of Saturday followed by Monday and Sunday. The lowest volume of sales of items is on Thursday.



❖ Way Forward:

- We can see the various detailed descriptive statistics such as maximum, minimum, and range of various items and months.
- We also see some graphs, plots, and some trends according to some filters.
- Further, a detailed analysis or interpretation regarding other data or objectives needs to be done that will be submitted in the final submission of the project.