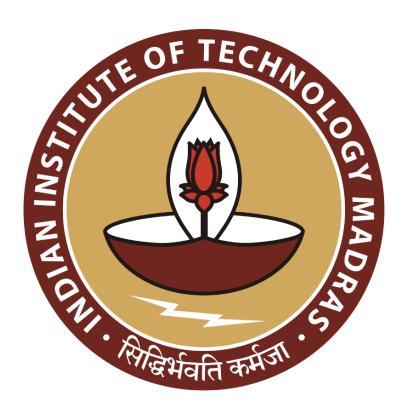
Data Analysis of Product & Services of Agrotech Enterprise

A Final submission report of the Business Data Management Capstone Project

Submitted by

Name: Saurao Upare Roll No: 21f1000355



IITM Online BS Degree Program

Indian Institute of Technology, Madras, Chennai Tamil Nadu, India, 600036

Contents

•	Executive S	Summary	4
---	-------------	---------	---

- Detailed explanation of Analysis process/method....5-10
- Results and Findings (with Graphs and other Pictorial Representation).....10-18
- Interpretation of the Result and Recommendations......18-20

Data Analysis of Product & Services of Agrotech Enterprise

Declaration Statement

I am working on a Project titled " Data Analysis of Product & Services of Agrotech

Enterprise". I extend my appreciation to "UPARE AGRO ENTERPRISE", for providing the

necessary resources that enabled me to conduct my project.

I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered through primary

sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the

data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the information of academic honesty and integrity, and I am

receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with

other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware

and prepared to accept disciplinary measures imposed by the relevant authority.

I agree that all the recommendations are business-specific and limited to this project exclusively,

and cannot be utilized for any other purpose with an IIT Madras tag. I understand that IIT

Madras does not endorse this.

Signature of Candidate: (**Digital Signature**)

Name: Saurao Madhukar Upare

Date: 15/09/2023

Executive Summary:

3

Upare Agro Enterprises started its adventure in 2013. From those times it has been a popular shop for the nearby farmers and plantation owners. At the time of the pandemic the shop faced a lot of losses but after that it started to flourish with the sale of different equipment and machineries. Whereas the sale of organic fertilizers was not significant so the data from the financial year april 2021- march 2022 and another year from april 2022-nov 2022 has been taken up by me for the analysis of the sales and other improvements which are analyzed by me and recommended to the owner.

The demand for all types of agricultural equipment, fertilizers, and seeds is at its highest during the rainy season; however, this need declines at the conclusion of the season and only slightly increases throughout the winter. And in the summer, this business is entirely dependent on small-scale entrepreneurs and people running plant nurseries, whereas homegrown plants, fruits, and vegetables are also sort of in high demand today, but people seem to avoid using organic manure and Gandul khat (manure made with the aid of earthworms and leftovers from home cooking) Neem khat, cocopit etc... because of the high cost. Due to time constraints, environmental awareness, soil health, and other factors, those concerns are gradually realizing their significance so I have given some sort of alternatives so the sale can be increased.

Analysis of both dataset of year 2021-2022 (called dataset1) and april 2022-nov 2022(called dataset2) is done month wise whereas how many unique items were sold by the enterprises across each month in that year. Pareto analysis is also being done and in which volume pareto and revenue pareto is being done by me. The contribution of machinery sale to the revenue is enormously high whereas the sale of the organic manure is significantly less in both the year. The data supplied were the basis for the comprehensive analysis in this case study. Some conclusions drawn from the data include that revenue and sales are increasing, which is positive, but the sale of organic manure is less and proper recommendation is given in the report and also for the machinery the sale revenue is good but the no of sale is not significant so for that a proper way to approach the customer is recommended by me.

Detailed explanation for analysis method:

Along with excel analysis (Pareto analysis, I had done the analysis on the google colab using the pandas dataframe and done an EDA on the both dataset.

There are two different voucher sale lists of the year(21-22) and (april22-nov22), and I saw two major issues in the sale and also when I discussed with the owner the following two major issues he pointed out to me. So I worked out on the following data and looked for the solution for the issues through EDA and excel analysis whose links are given below with proper description, what it contains:

The data sheet used for EDA and pandas profiling in the google colab is given below:

Data1(financial year 2021-22): ■ Primary data1 for BDM.xlsx

Data2(financial year 2022-23): Primary data2 for bdm

The link of colab for viewing the EDA done on data1 : •• Data1 of agro.ipynb

The link of colab for viewing the EDA done on data2 : •• Data2 of agro.ipynb

The link for the excel sheet where I have done all sort of calculations and data analysis:

Analysis of Primary data Upare Agro enterprises.xlsx

Sheet name

Data 1 : It is uncleaned data with null values and reductant values

Clean Data1 : It is a cleaned data with removed null values and rows and added

necessary column for analysis

Vol pareto1 : It is the Volume pareto analysis on data 1

Rev pareto1 : It is the Revenue pareto analysis done on data 1

 Trend1 for data1: It is a line chart drawn on Sales quantity and Revenue quantity for viewing which month has more sales and revenue generation.

 Trend2 for data1: It is the scatter chart for viewing the trend in different types of sales across each month in the financial year for data1

Data Analysis of Product & Services of Agrotech Enterprise

Data2 : It is uncleaned data with null values and reductant values of data2

•

• Cleaned Data2 : It is a cleaned data with removed null values and rows and added

• necessary column for analysis

Vol pareto2 : It is the volume pareto analysis of the data2
Rev pareto 2 : It is the revenue pareto analysis of the data2

• Trend2 for data2: It is the scatter chart for viewing the trend in different types of sales across each month in the financial year for data2.

Excel Analysis

The steps of cleaning data is below:

The spreadsheet illustrating the successive one and a half years' worth of monthly sales contains the List of Sales Vouchers, which consists of all voucher series, bill numbers, and the quantity of things sold. I am left with some of the most significant data elements that have been gathered in a single sheet and that I believe to be very relevant to my study after performing data cleaning, transformation, and column elimination.

Basic operations have been carried out, such as converting the supposedly numerical data type from text or a string of text to numbers, getting rid of all redundant and pointless columns, and adding a few useful columns to further simplify my own analysis. Same

I have eliminated from this sheet the prominent columns such as "Bill No," "Particulars," "TIN/GSTIN No," and "MRP." The Particulars, Bill No., and GSTIN No. are useless to me, so I removed them. However, I left the MRP column because some products have a zero MRP selling price because the product or the manure is produced locally and he is selling those products at market rate, so MRP was insignificant for me.The Month and Type columns are two new additions to my analysis(given below).

	1						
Date	Month	Туре	Item Details	Material Centre	Qty. Unit	Price	Amount
Apr 1, 2021	April	Tool	Carborater	Main Store	1.00 Pcs.	848.22	848.22
Apr 1, 2021	April	Tool	C35 FOUR ARMS SPRINKL	Main Store	10.00 Pcs.	118.64	1186.44
Apr 1, 2021	April	Tool	PACKING KIT	Main Store	1.00 Pcs.	127.12	127.12
Apr 1, 2021	April	Tool	Clutch Assembly	Main Store	2.00 Pcs.	223.21	446.42
Apr 2, 2021	April	Tool	Grass Weed Slasher	Main Store	1.00 Pcs.	2678.58	2678.58
Apr 2, 2021	April	Tool	Garden Hand Tools	Main Store	1.00 Pcs.	2542.38	2542.38
Apr 2, 2021	April	Tool	Grass Weed Slasher	Main Store	1.00 Pcs.	2321.42	2321.42
Apr 2, 2021	April	Tool	Harvesting Tools	Main Store	1.00 Pcs.	3750.00	3750.00
Apr 7, 2021	April	Tool	Spray Pump	Main Store	1.00 Pcs.	2142.86	2142.86
Apr 15, 2021	April	Tool	Battery Spary Pump 18 Lt	Main Store	1.00 Pcs.	2142.86	2142.86
Apr 20, 2021	April	Tool	Falcon Garden Tools Gras	Main Store	1.00 Pcs.	580.00	580.00
Apr 20, 2021	April	Tool	Spray Pump 5 Ltr	Main Store	1.00 Pcs.	900.00	900.00
Apr 20, 2021	April	Tool	Digger	Main Store	1.00 Pcs.	250.00	250.00
Apr 20, 2021	April	Tool	Golf King Hedge Shears	Main Store	1.00 Pcs.	600.00	600.00
Apr 20, 2021	April	Tool	Garden Gloves	Main Store	1.00 Pcs.	152.54	152.54
Apr 22, 2021	April	Tool	Mulching 25 Micron 1 Mt	Main Store	15.00 Roll	1677.97	25169.50

Fig 1. Cleaned data with additional columns of 'Month' and Type

Pareto Analysis:

According to Pareto analysis, 80% of a project's benefits can be obtained by doing 20% of the effort, or, in the opposite case, 80% of issues can be linked to 20% of the causes. A strong tool for decision-making and measuring quality is Pareto analysis.

Volume Pareto:

I've used the cleaned data and transformed it into a pivot table. Now I grabbed the item information and put it in the rows. I also took the quantity and made a column with the total number of units sold for the item in question. I then added up the quantity of sales for each item and added the cumulative percentage in the following column, and I discovered that dataset 1 contained a total of 252 items, with the 45 top-ranking items accounting for about 20% of the total revenue. Thus, for the provided dataset 1 of the years 21–22, the Volume Pareto is satisfactory.

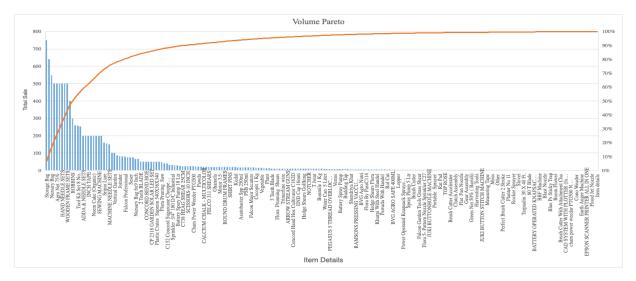


Fig 2. Volume Pareto (data1)

Similarly, I looked at the dataset 2 for the year 22's volume pareto and found that there were a total of 71 items and that the top item's total revenue, which added up to 80%, was generated by the top 10% of the items, which barely matched the volume pareto principle.

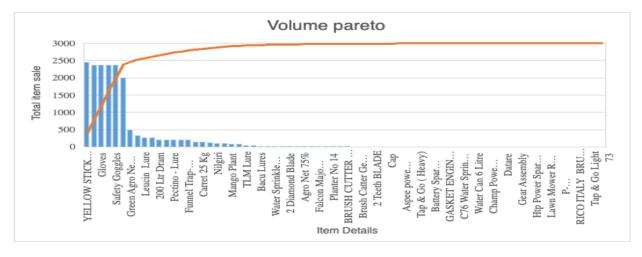


Fig 3. Volume Pareto(data2)

Revenue Pareto:

I've used dataset 1 that has been cleansed and transformed to a pivot table. I then sorted the item list in descending order, created a cumulative sum column, then created a percentage column, which demonstrates that the first 24 items barely satisfy the pareto principle by adding up the total revenue to 80% when the amount column was added to the values.

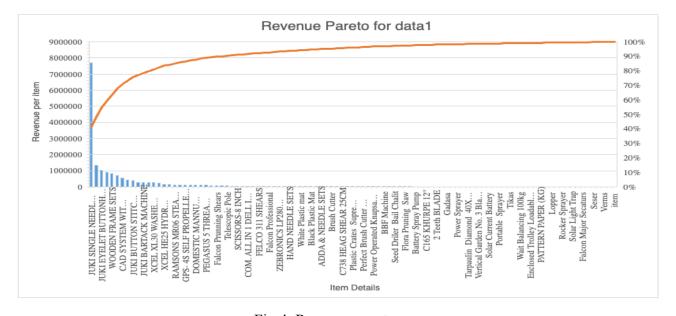


Fig 4. Revenue pareto

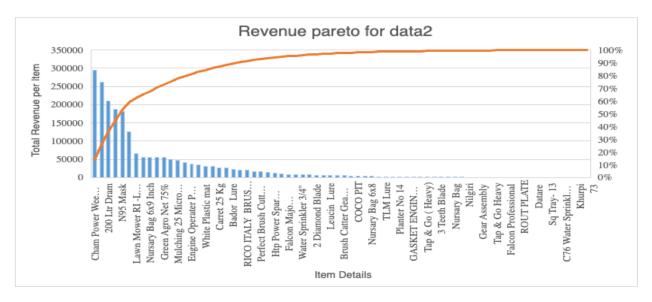


Fig 5. Revenue Pareto

I then applied the same procedures to dataset 2 for the year 22, and I discovered that there are 71 total products in this list, with the top 14 things accounting for 80% of the total revenue made in that year. Therefore, the 20% of the list's total goods add up to 80% of the sale's overall revenue. Pareto's concept so perfectly aligns with ideals in the above fig 5.

Month-wise sale of total items:

The total sale across each month tells us about the particular month of max sale where customers have farming time of the year. So first I made the pivot table with the values as the total volume of item sale vs the month on the x axis.

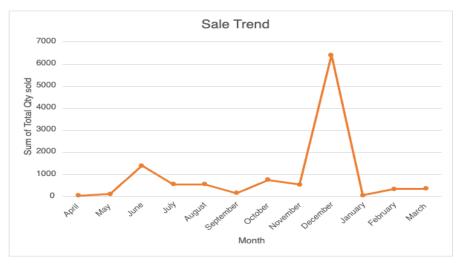


Fig6. Sale of total items sale for each month data1

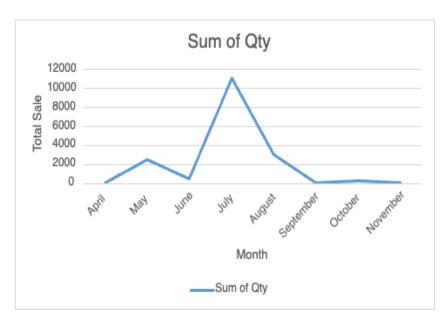


Fig 7.Sale of total items sale for each month data2

Month-wise revenue generation:

The month wise revenue generation is a plot of total revenue generated per month on y axis vs the month on the x axis. The pivot table was created and then the month column and the sum of amount is put on the values place and taken the total revenue from each month and plotted the trend of revenue generation across each month.

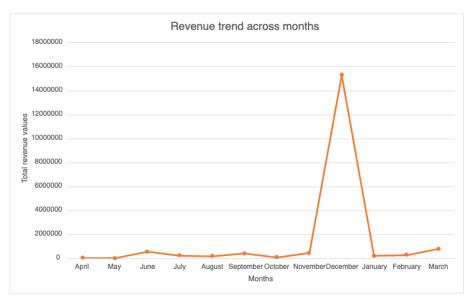
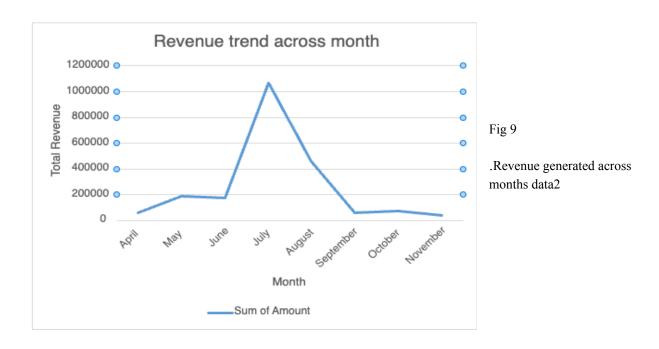
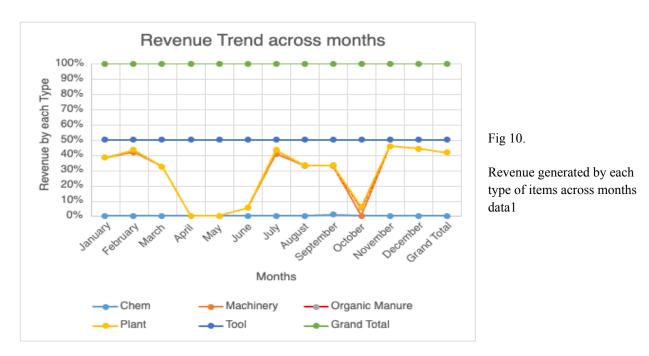


Fig 8. Revenue generated across months data1



Month-wise revenue generation by each Type of items:

The plot of total monthly revenue created by each type of item on the y axis vs the month on the x axis represents the month-wise revenue generation by each type of item. The month column and the sum of the values are placed on the pivot table after which the total income from each month is obtained, and the trend of revenue production by each item is plotted.



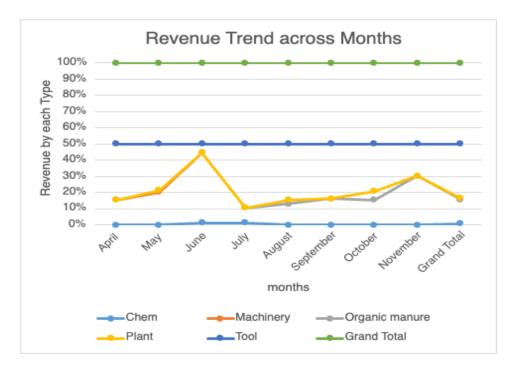


Fig 11.Revenue generated by each type of items across months data2

EDA through google colab:

The Exploratory Data Analysis is done on both the dataset1 and dataset2. The need for the EDA is that we get a very easy and detailed analysis in a very short code and the steps involved in doing so is below:

The steps for the EDA and other plots used for the analysis:

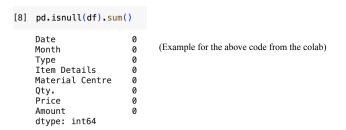
- → First of all I imported all the necessary modules required by me for the analysis
- → Then I uploaded the .xlsx file of both dataset in the google colab and read the excel file with the code of pd.read excel(--.xlsx) and called that as df file.

```
import numpy as np
import pandas as pd
from ydata_profiling import ProfileReport
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_excel('/content/Primary data1 for BDM.xlsx')

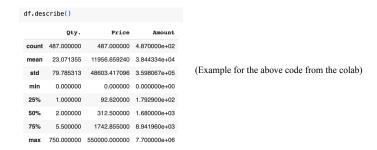
df.shape
(487, 9)
```

→ I checked if there are any null values in the dataset and removed the 'Unit' column from the dataset so that the unnecessary thing is removed.



→ Getting the shape of file, info() and describe() which gives the knowledge of the file and describe() gives the mean median mode of the dataset.

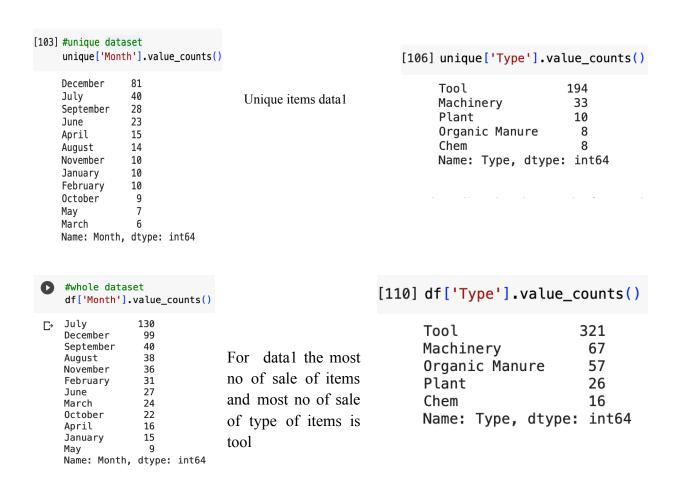
 \rightarrow



→ Now for the duplicated columns I use the code df.duplicated(subset='Item Details') for the duplicate items sold for the whole year and then taking the inverse of the code to give the unique items sold by the enterprise.

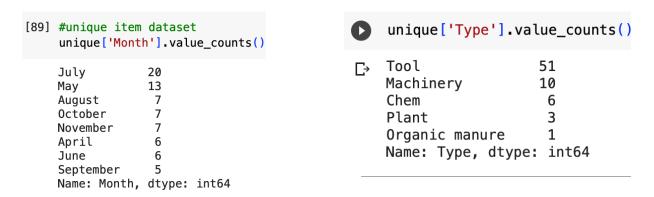


- → Now again for the duplicated columns I use the code df.duplicated(subset='Month', 'Type', 'Item Details') for the duplicate items sold for the whole year and then taking the inverse of the code to give the unique items sold across each month and type by the enterprise for that year.
- → Now taking the value counts across each month for the month-wise distribution of the unique items of both dataset.
- → Whereas the same thing can be done for the whole dataset df and taking the total items sold across each month. The analysis is done on both dataset



FOR DATA2:

Now for the dataset2 the same unique items count and type of items are calculated which is given below.



This above picture of code tells us that the data with unique items sale is most in July and Tool is the most sold item in this year For the data2 the total sale and most item sold is,

```
#whole dataset taken for analysis
                                         [96] df['Type'].value_counts()
df['Month'].value_counts()
                                              Tool
                                                                    80
             25
July
                                              Machinery
                                                                    17
October
             18
                                              Chem
                                                                     7
May
             16
November
             14
                                               Plant
                                                                     6
             12
June
                                              Organic manure
                                                                     1
August
             11
                                              Name: Type, dtype: int64
September
              9
              6
April
Name: Month, dtype: int64
```

In the month of July the most items are sold whereas Tools where the items sold mostly in this year and organic manure being the least.

For DATA1

→ So now for the total no of sales across each month i have plotted the bar plot using matplotlib given below and the type of items sold also plotted across months

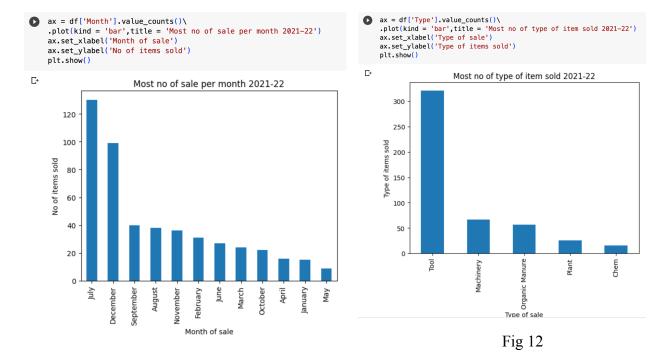


Fig 11.

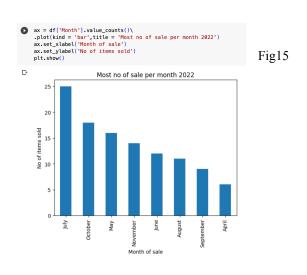
The above bar plot tells us about the total sale of items and the other bar plot at the right side tells us that the different types of items sold across each month, which is July month for highest sale of items in this year and Tools being the most type of item sold across the year.

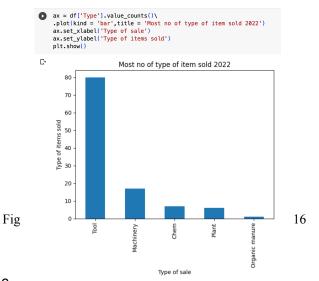


This bar plot to the left, tells us about the unique items sold across every month and December month being the max sold items which is around 81. And the most no of item sold was Tool and Chem being the least.

For DATA2

Plot of different item sold and the type of items sold across year



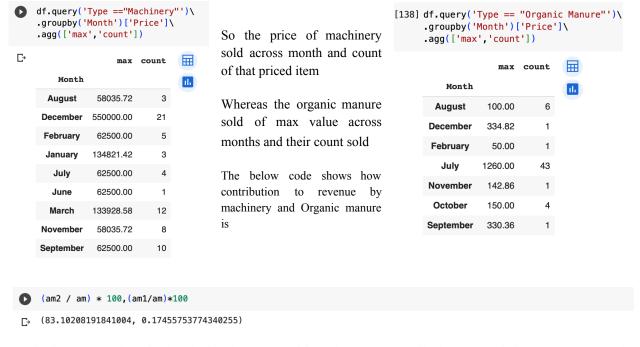


→ Now we have seen the whole dataset1 and dataset2 for most sales and most types of item sold but now I have done the analysis for the organic manure and machinery sold each year. The below fig shows the respective thing:

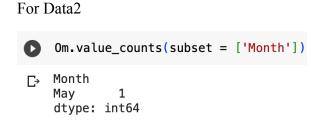
For Data1



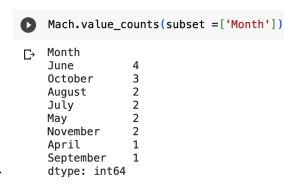
→ This code show how the organic manure sale is distributed across each month whereas same for the Machinery item sold across every month, this show that how the organic manure sale is sale among the months and whereas the sale of machinery is also very less but u can see that the max amount of price(below code) for each machinery is too high that it significantly contributes to the total revenue but the organic manure being less sold and the price is comparatively much less and have less contribution towards total revenue.



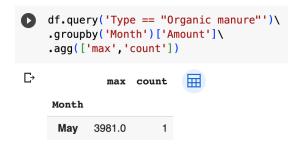
So the above percent tells us that the sale of Machinery is 83% of the total revenue generated by the company which enormous as compared to Organic Manure sale and which is 0.17 % of the total revenue generated by the company in that financial year



This shows that the organic manure sold in this time period of sale is only 1 that is in the month of May



→ Whereas the machinery sold has significantly reduced to last year and the most sold is in the month of June.



→ The max price of Organic manure and machinery is shown here and the Machinery has a very huge price and the sale count is less.



```
(am2 / am) * 100,(am1/am)*100

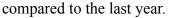
(29.518888974675967, 0.18825657383940403)
```

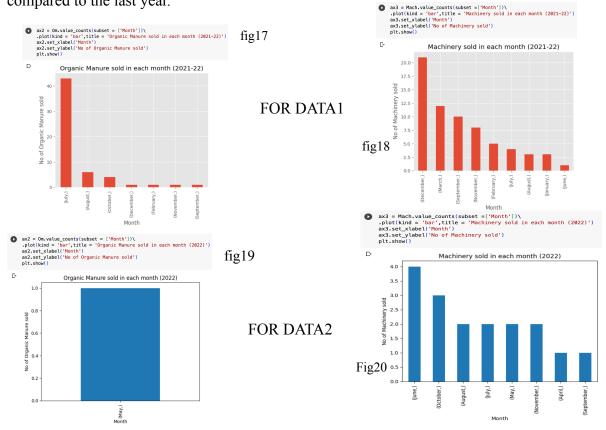
So the above percent tells us that the sale of Machinery is 29% of the total revenue generated by the company which enormous as compared to Organic Manure sale and which is 0.18 % of the total revenue generated by the company in that financial year

→ This percent of sales of organic manure and machinery shows how significantly less is sale of organic manure and the count of machinery is less but has a huge revenue contribution due to that pricing.

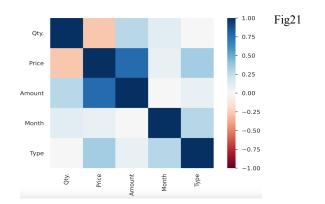
Interpretation of Results and Recommendation for the problem statements:

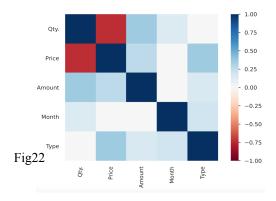
The above analysis showed that the organic manure has a less demand in the enterprise huge in the month of July for year(2021-22) and no demand or a single demand in the next year 2022 and the item machineries have also a less demand wrt to others items such as tools, but due to the high pricing of the machineries the contribution to the total revenue by machineries is enormous. In the month of December the Sale Machinery is highest in the year(2021-2022) with a total revenue contribution of approx 83 percent to the total revenue generation. The next year the contribution to the total revenue by the machineries is 23 percent which is significantly less





Correlation between different features of the dataset1 and dataset2:





There is a positive correlation between the feature column Price and Type is significant, similarly for the column Month and Type for the dataset1 and dataset2.

There is a negative correlation between the feature column Price and Qty for the dataset1 and 2 whereas for the data2 the negative correlation is high than that of data1.

4. Recommendations:

For Decrement in the sale of Organic manure

- ★ The enterprise should do proper marketing of the product in the towns nearby and city. The sale of manure says that the name of the product is not properly spread across the town.
- ★ The owner should make social media pages for the interaction with large masses and the display of products should be kept forward and make home deliveries of the small packets of manure for the one who plants in the house backyard or home nurseries(terrace farming)
- ★ There is a significant sale of other equipments in the shop so should give pamphlets to spread the name of the product among the regular customers
- ★ There are various exhibitions and environmental drives happening in the town, significantly advertising and spreading the word in the awareness camps such as reliability, product shelf life, environment friendliness, less storage management.
- ★ When the product is sold with the product he can give the customer the lure combo pack with organic manure(as the lure will attract the insects and germs toward it and less harm is done to the crops)
- ★ Among the sale of organic manure the sale of water supply equipment will also increase, so making a proper offer to the customer so that side by side both sales will flourish.(The organic manure requires more moist environment)

For Subsidized Agricultural Machineries leading to decrement of sale:

★ The Machineries have a high pricing and the number of items sold in this category is sale so the subsidy thing should be helped by the owner as the main problem is the total payment of the amount despite having a 30% to 80% subsidy in the agricultural equipments

So he can make a proper account of customers who are trusted and a regular customer of the shop and give him an offer that he would contribute to the purchase from his side but there he would certainly return the money after he gets the subsidized amount in his account in the bank. This would take word of mouth to spread and as the offer spreads he can scale sale and also add some 2% to 3% commission on the offer given to the farmers.

Also he can give offers for maintenance and services of the machineries in considerable amount to the farmers and farm delivery services for the one who purchased his equipments.