**Insights Report**

BY:

Name: Kalluri Vishnu Vardhan Reddy

Email: vishnu15018@gmail.com

Mobile No: 9391586197

To 360digitmg

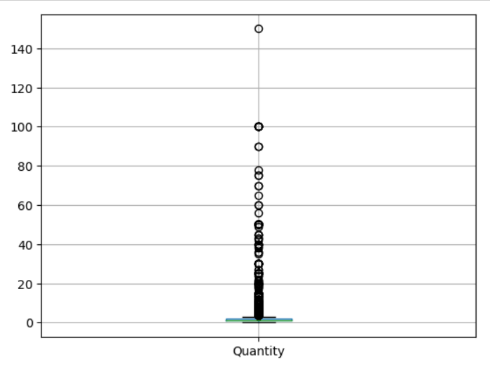
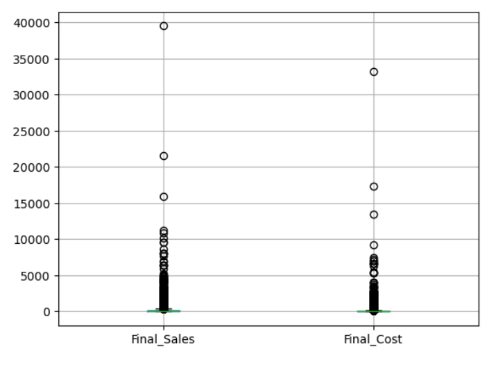
**Table of Contents:**

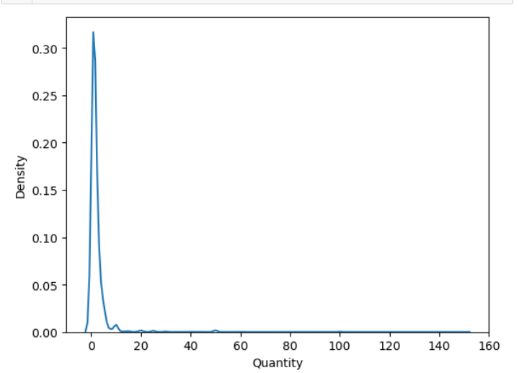
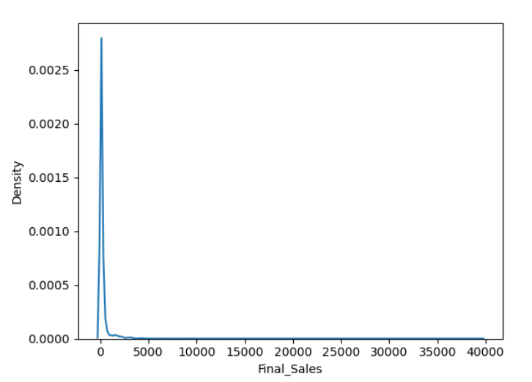
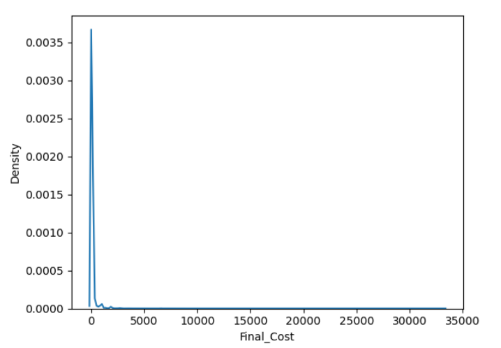
1. Insights after EDA
2. Data Preprocessing steps
3. EDA Insights after preprocessing
4. Insights Useful for business decision.

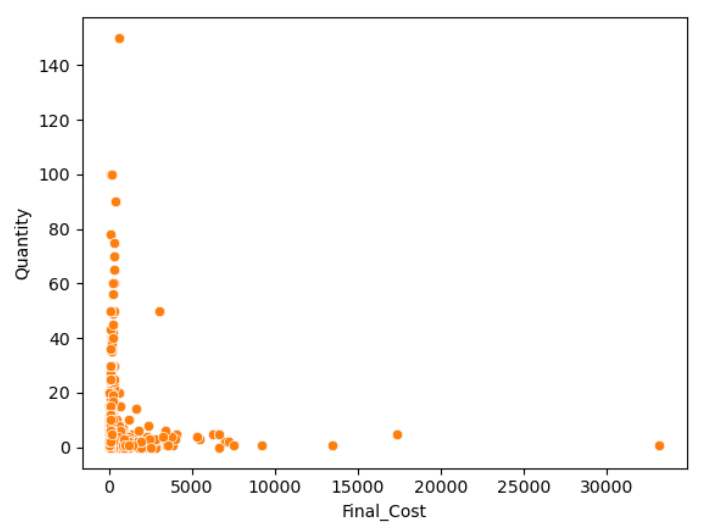
**1.Insights after EDA from the raw data**

**[Descriptive and Diagnostic analysis]**

1. Variance for the numerical columns is very high indicating the data is widely distributed [ note outliers may affect the variance]
2. Numerical columns are heavy tailed, high peak in visualization and asymmetric, indicate the presence of outliers and also few drugs have high sale values relative to others. [Analysed using the kurtosis and skewness]
3. Almost 83% of the transaction values are less than 300.
4. Reason for the outliers
5. outliers in this dataset are natural outliers.
6. These outliers occurs because sometimes some patient needs antibiotics and other costly drug to cure a specific disease and also when a patient bought more quantity of a drug it leads to outliers in sales







1. Some Costly drugs are present and some patients bought few drugs in high quantity

**Based On the Subact1 and Drug Name analysis**

1. Anti-infectives and Intravenous and other sterile solution are the top most performed drugs with respect to sales, but the demand with respect to quantity is highest in Intravenous and other sterile solutions
2. There may be viral, bacterial infections, abdominal related infections, heart related conditions, fever related disease or any specific medical condition more prevalent.

**Based on the sub cat and subcat1**

### Almost 90.8 sales belong to Injections, fluids electrolytes, tablets and capsules.

### In the Injections category Anti-infectives and cardiovascular & Haemopoietic condition related injections has highest sales

### For details about the infections and diseases we need to go through the patient disease data. [Those data are not in our scope]

### Given the high sales of "INJECTIONS," "IV FLUIDS, ELECTROLYTES, TPN," and "TABLETS & CAPSULES," it may be wise to ensure sufficient inventory levels for these subcategories to meet customer demand and avoid stockouts.

**VII. Based on specialisation and sub cat**

1. top5 specialisation are specialisation4, 7, 8,3 ,20. there may be a chance that among these specialisations any one must belong to treating the corona related cases. [ as the data is taken from 2022]
2. In Every specialisation the top3 categories are INJECTIONS, IV FLUIDS, ELECTROLYTES, TPN, TABLETS & CAPSULES, except for specialisation 20 the 3rd most demanded category is INHALER & CAPSULES.

VIII)  **Based on the department and specialization**

1. Almost 87% percent of drugs bought patients are from deparment1
2. Department1 has highest demand and sales compared to all other departments (almost 89% sales are from department1), and also among all three departments money was spent on dept1 is more compared to other departments.
3. Interestingly in department3 the top most demanded specialization is Specialization 2 rather than 4

**2. DATA PREPROCESSING STEPS**

**STEP1:** Removing the duplicate rows. Because it causes redundancy.

**STEP2:**

Type Casting of **Dateofbill** column from string to DATE type. **Because** to analyse the data month wise quarter and week wise effectively.

**Step3: Handling NULL VALUES**

It is necessary to fill the null values with an imputation strategy or with any value.

### 1. Formulation column has 650 null values

### 2. Drug Name has 1668 null values

### 3. Subcat has 1668 null values

### 4. Subcat1 has 1692 null values

### 5. If DrugName is null then Subcat and Subcat1 is also null.

## **Solution to treat NULL values:**

### 1. As all the null values are present in the categorical column, so filling it with imputed strategies like mode, backfill won’t give good results. Because when we impute the null values with mode or backfill it will provide the wrong sales results for the category that is mode.

### 2. So I am going to fill it with the ‘Unknown’ value, so that the sales related to the null category will be analysed separately.

### STEP4: Outlier Analysis

### Reason for the outliers

### 1.Outliers in this dataset are natural outliers.

### 2. These outliers occurs because sometimes some patient needs antibiotics and other costly drug to cure a specific disease and also when a patient bought more quantity of a drug it leads to high transaction value, these high values cause outliers in sales

### 3.We need to keep these outliers as well to perform sales analysis

### 4.As these outliers are natural, we need to keep those outliers.

### 5.We need to analyse these outliers separately, by creating a new column cost\_per\_unit and price \_category.

### FEATURE ENGINEERING:

### ADDING THE DERIVED COLUMNS:

### 1. cost\_per\_unit = dividing the final\_cost column by quantity for sales data and divide by return quantity for return data.

### 2.Price categories = based on the cost per unit price range, I want to distribute the data into three categories

### Low = 0 to 200

### medium = 201 to (mean + 3sigma) [z threshold)

### high >=medium

### NOTE: I choose the above limits such that the outliers lies in between medium and high category

## reason for creation of the cost\_per\_unit and price\_category columns

### i)To analyse the sales better, to find the market segments by separately analysing the outliers drugs that belongs to high and medium range.

### II)To gain more view into the sales segments of each department.

### 

### patients bought low-cost products in high quantities, compared to other medium and high-cost products

### ADDING PROFIT DERIVED COLUMN FOR PROFIT ANALYSIS

### profit column used for calculating profit that gained through each transaction // this is for profit analysis

### 3. EDA AFTER PREPROCESSING

### Based on the profit column:

### Department1 generating more profit than all other departments.

### but an interesting observation is from department 2 and 3, we are getting more profit percent 121 and 143 than the department 1 with 83percent with respect to cost spend.

### Marketing campaign for dep2 and dep3 related drugs in order to expand the sales and get more profit [ Note: we may take into the account that whether their location, need for expansion]

### In department3 has 929 profit percent in specialisation6 and 526 profit percent in specialisation 55, because a patient with ID 12018120000 bought high quantity of drugs

### Unknown drug is one of the highest sales and profit category in each top5 specialization

### Unknown drug subcategory has second most sales and highest demand with respect to quantity.

### Anaesthetics has 881 profit percent.

### Using the price category column:

### the low-cost drugs are in high demand compared to others.

### In every dept low price drugs has high demand compared to medium and high priced respectively.

### Getting Insights related to return:

### Total Return quantity: 4145.0

### The total cost of returned products: 190907.044

### Total no of returns: 1676

### There are no returns related to department 3. Almost all the returned drugs are related to the dep1, because the most patients bought drugs related to this Department

## e) Most of the drugs returned are related to Form1 and unknown from Specialisation 4 and 7

### f) By getting the feedback from these customers who returned more than 3 we can improve the service based on their feedback there by decreasing the bounce rate.

### Insights gained by analysing month wise

### December has the highest sales and highest demand with respect to quantity.

### But in august highest return quantity is recorded

### In April pharmacy gets the highest profit percent

### 

### INSIGHTS USEFUL FOR BUSINESS DECISIONS:

### BOUNCE RATE RELATED:

### Used this formula: No of return transactions / total No of transactions

### Bounce rate of the pharmacy: 11.809%

## Bounce rate of deparment1: 13.491743858236005%

## Bounce rate of department2: 0.06389776357827476%

## Bounce rate of department3: 0

**INSIGHTS:**

1. **There may be viral, bacterial infections, abdominal related infections, heart related conditions, fever related disease or any specific medical condition more prevalent, based on the top5 drugs in each condition**

### Marketing campaign for dep2 and dep3 related drugs in order to expand the department sales and get more profit [ Note: we may take into the account that whether their location, need for expansion]

1. **By getting the feedback from these customers who returned more than 3 we can improve the service based on their feedback there by decreasing the bounce rate.**
2. **Finally use the inventory management techniques by utilising the demand forecasting of top 5 or 10 drug conditions types. Use forecasting models for demand forecasting with time period month wise or week wise.**