Medical Inventory Optimization using Customer Data for Reducing Bounce Rate

Yashraj Mishra[1], Ankita Jaiswal[2], Dr. Goldi Soni [3]

[1]Student , Amity University Chhattisgarh
[2]Student , Amity University Chhattisgarh
[3]Assistant Professor, Amity University Chhattisgarh

Abstract- *Healthcare establishments all over the world need to handle their medical inventories effectively. Patient care may suffer because of overstocking, stockouts, and higher expenses brought on by inaccurate medication demand predictions. This research proposes an advanced machine learning model-based data-driven approach for medical inventory management optimization. The suggested solution seeks to decrease bounce rates by 30%, increase customer happiness, decrease wasteful inventory costs, and improve product availability. The solution's effectiveness is shown by outcome analysis, where the Gradient Boost model performs better than the others. This method has a very low MAPE (Mean Absolute Percentage Error) score (less than 5 percent). Hospital scan work together and embrace best practices more readily because of the solution's scalability, utility, and capacity to be used across several facilities and specialties. The initiative seeks to offer insightful information on supplier selection for higher operational performance, better customer service, and cost efficiency. In this project there are two methodologies used: Python programming libraries and Structured Query Language programming which is used for data preprocessing.*

Keywords: **Python Programming, SQL (Structured Query Language), Exploratory Data Analysis, Machine Learning, Data Preprocessing, Data Visualization, Transportation Cost Reduction**

1.Introduction

This research study's goal is to minimize the transportation cost in the supply chain and logistics management system using operational research, Python programming and PostgreSQL Server. This study and development approach's Python programming portion is built on the mind map that contains a certain part as shown in [Fig. 1]. This mind map for Python programming serves as a useful tool for comprehending and effectively implementing Python code by providing a visual representation of the essential ideas and elements of the language.

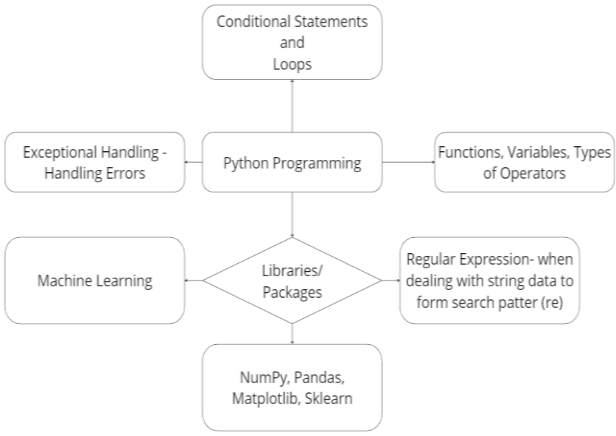

Fig – 1: Python Programming mind map

For manufacturing businesses to operate effectively, product transportation is essential. Reducing transportation expenses is crucial for sand production enterprises to preserve their profitability and competitive edge. This research study focuses on maximizing the amount of sand transported utilizing various logistics to minimize transportation costs from various warehouses to various demand destinations. A solver-based strategy that makes use of mathematical modeling and optimization approaches is used to accomplish this goal.

Effective medical inventory management is essential to delivering high-quality patient care in healthcare facilities across the globe. Medical supplies and necessary pharmaceuticals must always be available, but keeping the right amount of inventory on hand while reducing waste and expenses can be difficult. To solve this problem, healthcare facilities are using more and more data-driven strategies, utilizing cutting-edge machine learning methods to maximize medical inventory management. Moreover, CRISP-ML (Q)- Cross Industry Standard Process of Machine Learning with Quality Assurance. points say:

1.1. **Business Problem:** Bounce rate is increasing significantly leading to patient dissatisfaction.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

1.2. **Business Objective:** Minimize Bounce Rate
1.3. **Business Constraint:** Minimize Inventory Cost
1.4. **Business Success Criteria:** Reduce bounce rate by at least 30%.
1.5. **Economic Success Criteria:** Increase revenue by at least 20 lac INR by reducing bounce rate.

2. Methodologies and Strategies

In [Fig. 2] is an architecture that defines workflow of methodologies.

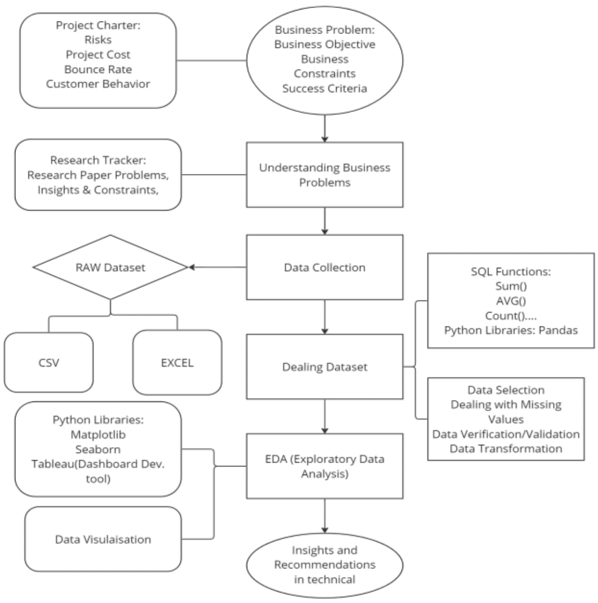

Fig – 2 : Architecture

2.1. Data Source and Collection:

Dataset is collected as well as extracted from the different sources and Client {leading Pharma} provided the dataset. Now, this dataset contains (df.shape) 14218 rows and 14 columns.

Using pre-existing data that has been gathered by other researchers, organizations, or sources is known as secondary data collecting. This method works well when gathering primary data is impractical or when researchers want to supplement their conclusions with more data. A popular technique for obtaining secondary data is to do a literature review, which is looking through books, articles, and other publications to compile pertinent material and pinpoint areas that still need research.

Official statistics, which are generated and disseminated by governments and organizations, are useful tools for academics to examine trends and patterns. Examples of these statistics include census data, economic indicators, and health statistics.

2.2. Data Dictionary:

There are in total 14 columns defining types of sales, patient ID, Specialization of Doctors, Department, Date of Bill, Quantity, Return Quantity, what is Final Cost and Final Sales of Drugs, MRP of returned Drug, Formulations and Subcategories. Now, the following table [Table 1] shows the detailed variable name and variable description.

Table - 1, Variable Dictionaries

Variable Name	Variable Description
Typeofsales	Type of sale of the drug. Either the drug is sold or returned.
Patient_ID	ID of a patient
Specialisation	Name of Specialisation (eg. Cardiology)
Dept	Pharmacy, the formulation is related with.
Dateofbill	Date of purchase of medicine
Quantity	Quantity of the drug
ReturnQuantity	Quantity of drug returned by patient to the pharmacy
Final_Cost	Final Cost of the drug (Quantity included)
Final_Sales	Final sales of drug
RtnMRP	MRP of returned drug (Quantity included)
Formulation	Type of formulation
DrugName	Generic name of the drug
SubCat	Subcategory (Type) to the category of drugs.
SubCat1	Subcategory (condition) to the category of drugs

2.3. Data Types:

Following [Fig. 3] After preprocessing dataset we get the following information like non-null values, Count and Data type.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14218 entries, 0 to 14217
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Typeofsales     14218 non-null  object
 1   Patient_ID      14218 non-null  int64
 2   Specialisation  14218 non-null  object
 3   Dept            14218 non-null  object
 4   Dateofbill      14218 non-null  datetime64[ns]
 5   Quantity        14218 non-null  float64
 6   ReturnQuantity  14218 non-null  float64
 7   Final_Cost      14218 non-null  float64
 8   Final_Sales     14218 non-null  float64
 9   RtnMRP          14218 non-null  float64
 10  Formulation     14218 non-null  object
 11  DrugName        14218 non-null  object
 12  SubCat          14218 non-null  object
 13  SubCat1         14218 non-null  object
dtypes: datetime64[ns](1), float64(5), int64(1), object(7)
memory usage: 1.5+ MB
```

Fig - 3 : Dictionaries data types.

2.4. Data Preprocessing:

In data processing this research paper uses two strategies: Python Preprocessing as well as Structured Query Language (PostgreSQL) preprocessing.

- **Data Cleaning:**

Data cleaning steps include dealing with missing values, handling duplicate data, and finally removing outliers.

The significance of missing values data in its row and unrefined form often presents a challenge that every data scientist must face a missing value. Missing values can occur for various reasons errors in data collection, system failure, survey nonresponses and more ignoring them is not option they can introduce bias distort correlation and impact the performance of overall machine learning models. There are in total 3 steps to deal with missing values see first, deletion of rows and deletion of the columns, second, is the method of imputation, it involves estimating missing values based on existing data. Techniques included mean, median, mode imputations and more advanced method like regression imputations and third, advanced imputation techniques, used for more sophisticated imputation consider method like K nearest neighbor where missing values are predicted based on their proximity to similar instances for time series data technique like forward fill or backward fill imputation can be effective. Dealing duplicate data is like a distorted mirror reflecting false pattern it can artificially inflate the bottoms of certain features leading to mist model training duplicate entries can also slow down computation and result in an inefficient allocation of resources.

Outliers hold the power to distort statistical analysis and machine learning models leading to inaccurate prediction and biased results. Outliers are data points that significantly derivate from normalization. They can have unusually high or low values that don't align with the overall pattern of the data set. Outliers can stem from various sources including measurement errors, data entry mistakes or genuine rare events, there are three basic methods to remove outliers first, is Z score and the second, is inter-quartile range and third, is outlier removal using percentile method. Following [Fig. 4] shows the inter-quartile range that this research paper has used:

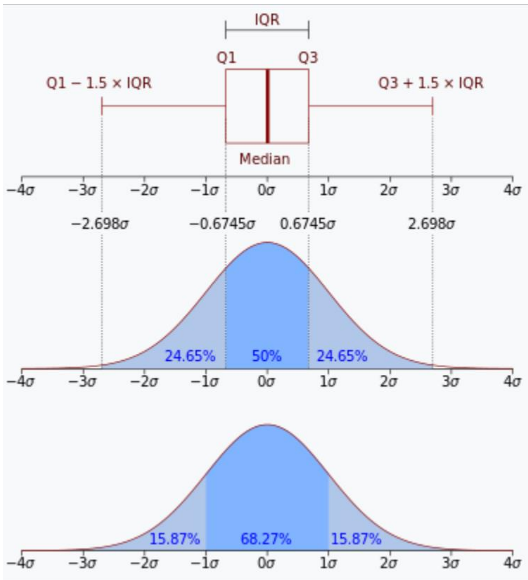

Fig – 4 : Box Plot + Interquartile Range, Source Wikipedia

Beyond this approaches there is one more approach: -

2.5. Solver Methodology:

One effective optimization technique that's frequently employed in operations research is the Solver method. To solve difficult problems, it makes use of mathematical models and algorithms. The transportation problem in this research study will be solved using the Solver approach by expressing it as a Linear Programming (LP) or Mixed-Integer Linear Programming (MILP) model.

The process flow in [Fig. 5] entails determining the issue, developing, and putting into practice a model, assessing its efficacy, reviewing test findings, and then putting the solution to deal with the identified problem into practice.

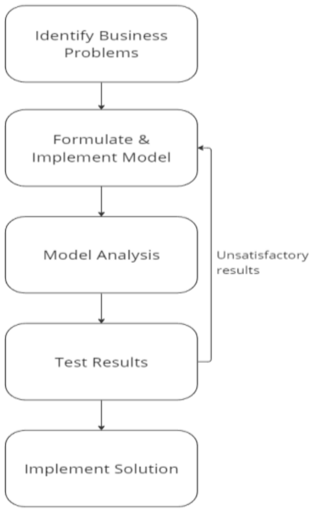

Fig - 5 : Visual model of problem-solving method.

- **Model Formulation:**
  Creating the model is the first stage in the Solver process. This entails putting the transportation problem into a mathematical model. The goal of the transportation problem is to minimize transportation costs by figuring out the best way to distribute items from various warehouses (sources) to various destinations (demand points). The model will consider variables like each warehouse's capacity, the demand at each site, and the expenses of each route's transportation. In this instance, a sand production company's goal is to reduce its overall transportation expenses.

- **Decision Variables:**
  Determine the variables that need to be ascertained, often known as decision variables, and express them using algebraic symbols. Now let's define the transportation model's decision variables. Assume that there are 'n' demand points and 'm' warehouses. X (i, j) is a matrix of choice variables that we can make, with 'i' stands for the warehouse and 'j' for the demand point. The amount of sand moved from warehouse 'i' to destination 'j' is represented by the decision variable X (i, j).

- **Objective Function:**
  The problem's objective can be mathematically represented in terms of decision variables. The model's goal is to reduce the overall cost of transportation.

  The total of the transportation expenses for each pair of warehouses and demand locations can be used to create the objective function [2].

  Minimize Z = C (i, j) * X (i, j)

  Here, C (i, j) represents the transportation cost per unit of medical supplies from warehouse i to destination j and X (i, j) represents the quantity of sand transported from warehouse i to destination j.

- **Constraints:**
  Constraints built into the model will guarantee that the total amount of sand supplied from each warehouse does not beyond its capacity and that the entire demand at each destination is satisfied. Non-negativity restrictions on the decision variables are another type of constraint that may exist. Due to the constraints, it clears what will be section we have to focus in data analysis.

3. Code Processing Observation Analysis:

After EDA (A method for examining and condensing data collections is called exploratory data analysis, or EDA. Prior to statistical modeling or machine learning, this is an important stage.). There are following observation:

3.1. Four Business Moment Decision:

There are three business moment decisions, here in the research paper to find out the outliers and other visualizations focused on co-relation matrix.

- **FIRST MOMENT BUSINESS DECISION:** Includes Measure of central Tendency (Mean, Median, Mode),
- **SECOND MOMENT BUSINESS DECISION:** Includes Variance, Standard Deviation, Range,
- **THIRD BUSINESS DECISION:** Includes Skewness, and
- **FOURTH MOMENT BUSINESS DECISION:** Includes Kurtosis.

Following [Table 2] and [Table 3] contains the four-business moment decision result before processing and after pre-processing respectively.

Table – 2 : Before Removing Outliers Four Moment Business Decision: Readings Unprocessed Data

Column_Name	**First Moment Business Decision**			**Second Moment Business Decision**			**Third Moment Business Decision**	**Fourth Moment Business Decision**
	Mean	Median	Mode	Variance	Standard Deviation	Range	Skewness	Kurtosis
Qunatity	2.23	1	1	26.34	5.13	150	11.34	183.09
ReturnQuantity	0.29	0	0	2.7	1.64	50	17.17	341.5
Final_Cost	124.82	53.65	49.352	216007.9	464.77	33138	34.5	2064.98
Final_Sales	234.04	83.44	0	450560.4	671.24	39490	21	980.93
RtnMRP	29.13	0	0	33218.35	182.26	8014	15.8	415.82

Table – 3 : Removing Outliers Four Moment Business Decision: Readings Processed Data

Column_Name	**First Moment Business Decision**			**Second Moment Business Decision**			**Third Moment Business Decision**	**Fourth Moment Business Decision**
	Mean	Median	Mode	Variance	Standard Deviation	Range	Skewness	Kurtosis
Qunatity	1.96	1	1	6.34	2.51	18		19.48
ReturnQuantity	0.22	0	0	0.57	0.75	5	4.33	20.956
Final_Cost	109.583	53.65	49.352	42251.4	205.55	1479	5.13	27.969
Final_Sales	209.96	86.424	0	155041	393.752	2248	3.71	14.68
RtnMRP	20.045	0	0	5974.73	77.296	578	5.31	31.62

From Table 2 and Table 3, the result indicates that the unclean data exhibits higher mean variance standard deviation, range skewness and kurtosis values compared to the clean data. Cleaning this data set has resulted in more stable and normalized distribution with reduced variability and potential basis and statistically best making it more reliable for business making decisions.

- After removing outliers from uncleaned/ unprocessed data the graph of numerical columns will be:

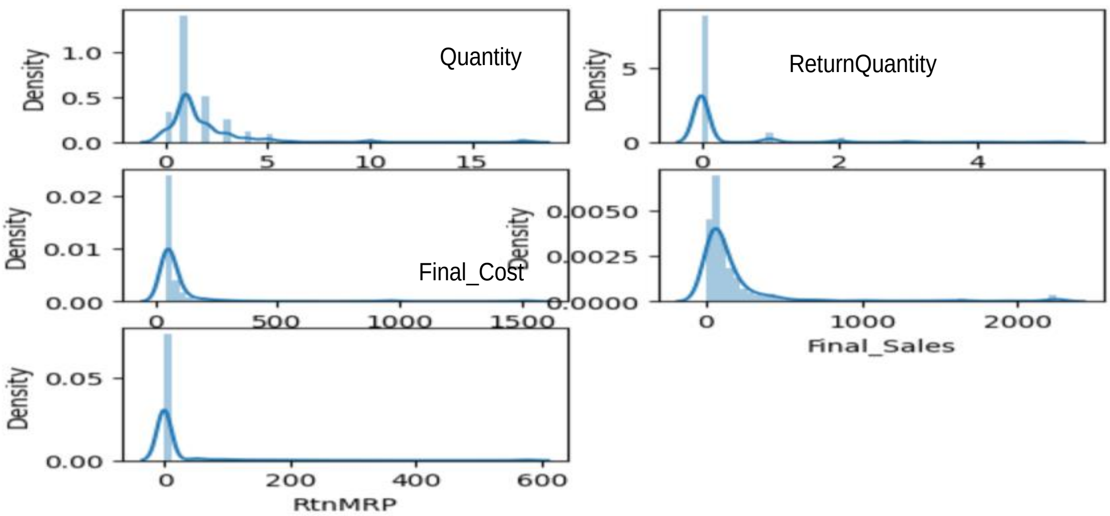

Fig - 6 : Columns graph with removed outliers

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

Here above graph [Fig. 6], states that, how the numerical column behaves after the removal of outliers, for the testing purpose all the values are plotted in the inter quartile range (IQR) which is the range between the first quartile i.e. 25th percentile and the third quartile i.e. 75th percentile [11]. Data points outside a certain range of the IQR are considered as outliers. So, the above table of four momentum graphs can be followed by outlier removal process. This graph was generated by seaborn and matplotlib plotting process using distplot and subplot functions. There are still outliers in some of the numerical columns because they are mostly treated as the highest value which is important for data wrangling/exploring.

3.2. Data Visualization:

- Plotting of the two important categorical data formulation and subcat1 (subcategory 1) of the total outcomes using Heatmap (a picture or chart that shows the variations in temperature or infrared radiation that were observed over a certain area or over a certain amount of time). Following [Fig. 7] shows the Heatmap,

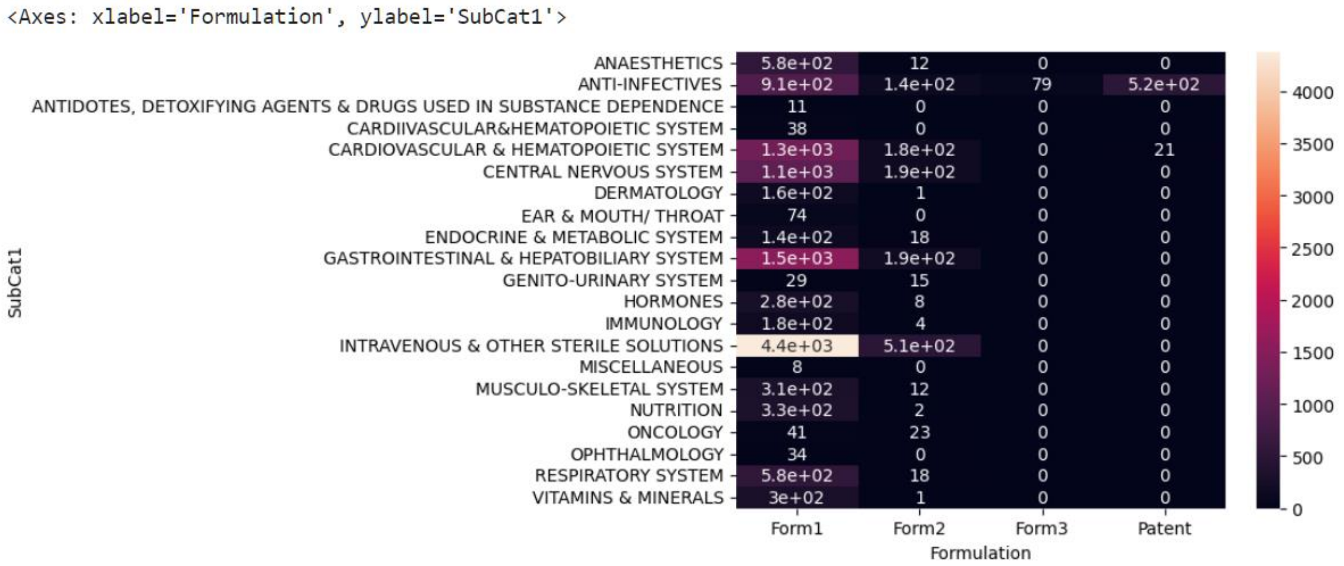

Fig – 7 : Heatmap (Subcat1 based Formulation)

- Plotting two correlated (df.corr()) numerical data, Final_Sales and Final_Cost based on the Patient ID. Following visuals [Fig. 8] shows the numerical insights using scatter plots: -

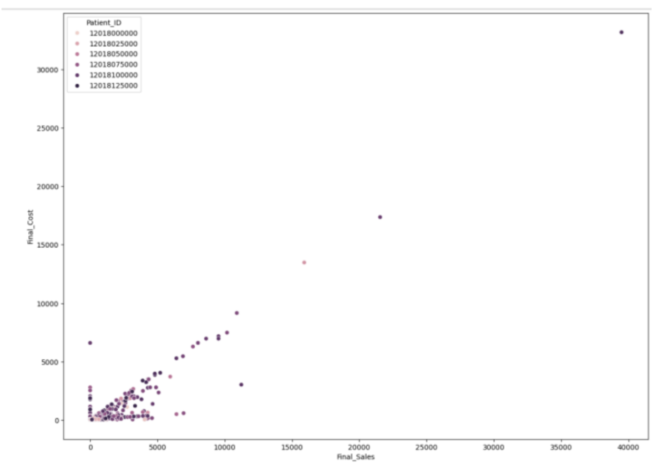

Fig – 8 : Numerical Insights (Final_Cost, Final_Sales based on PatientID).

- Let's plot the Most Subcategory Item that was visited by patients using pie plots, this visual [Fig. 9] includes the overall bounce rate that patient wants the drugs but due to unoptimized inventory they refuse to buy. As you can see from the visuals 'INJECTIONS' were highly wanted items by patients with 57.45 %.

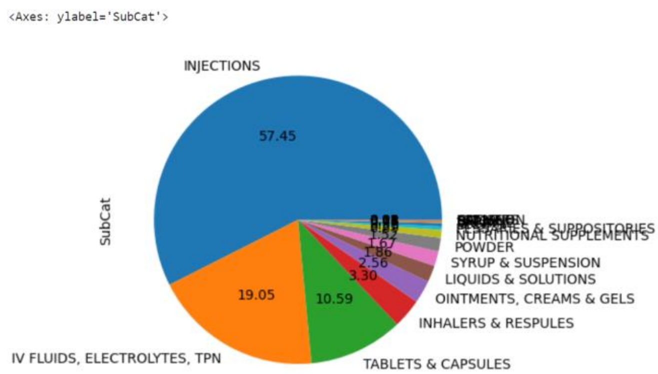

Fig – 9: Numerical Insights (Final_Cost, Final_Sales based on PatientID)

- Let's now see another constraint i.e. 'Formulation', In particular research we found that 'Form1' shows the highest type with 86.3%. The following visual [Fig. 10] shows the Formulation chart.

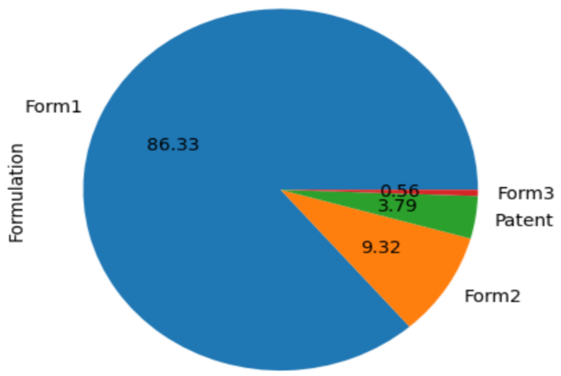

Fig – 10 : Formulation Chart Analysis.

- Also, we need to plot a pie chart for the most requested Department. In the department research shows [Fig. 11] 'Department1' maximum of 87.49%.

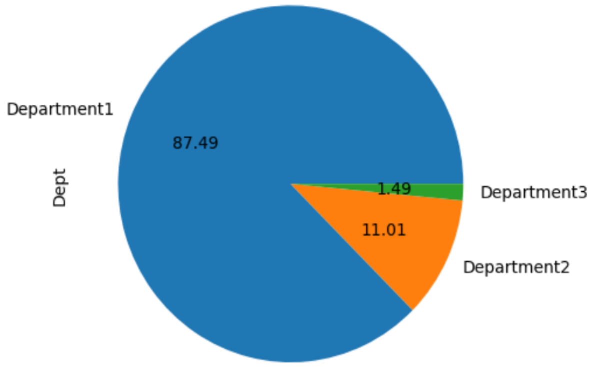

Fig – 11: Department Chart Analysis.

- When focusing on the Type of Sales there are around 2000 products that are returned.

- Now, the most important plot is Numerical Categorical Plot Analysis: In the following research we include four plots where different specialization is compared to amount of Quantity, Return Quantity, Final Sales and Return MRP. In the following visuals [Fig. 12, 13, 13, 15] shows the numerical-categorical plots.

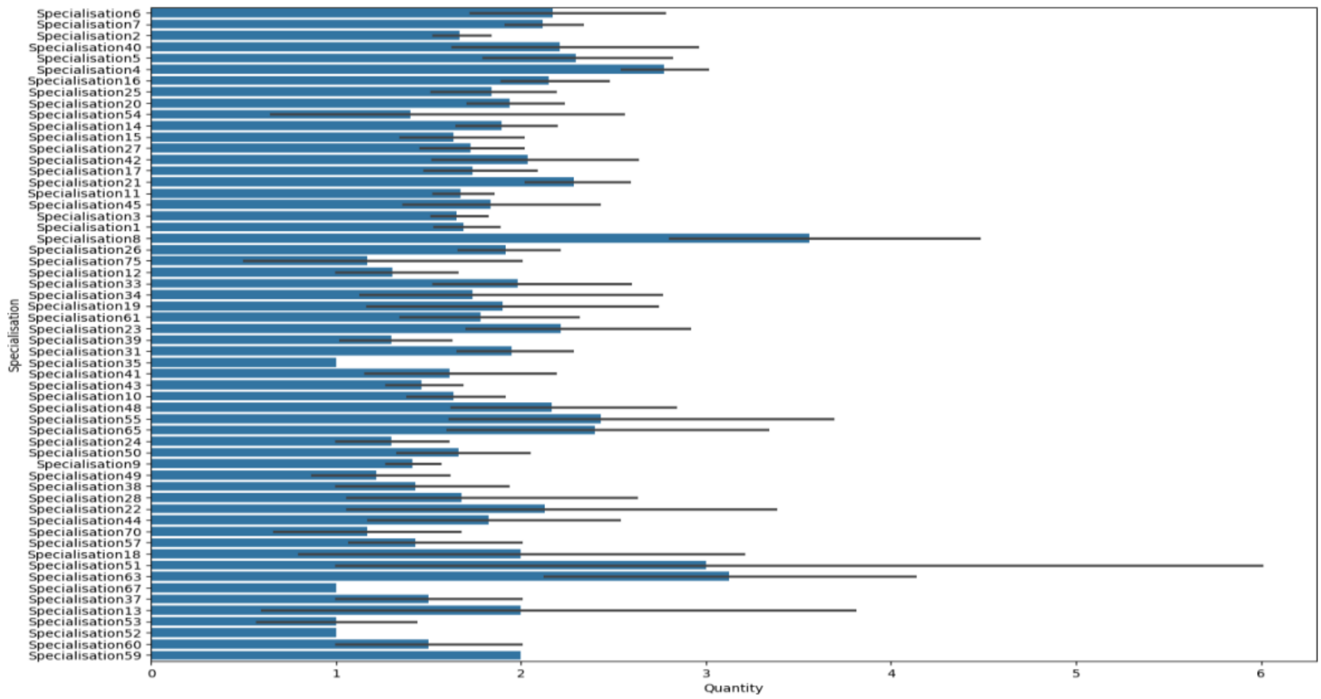

Fig – 12 : Specialization vs Quantity.

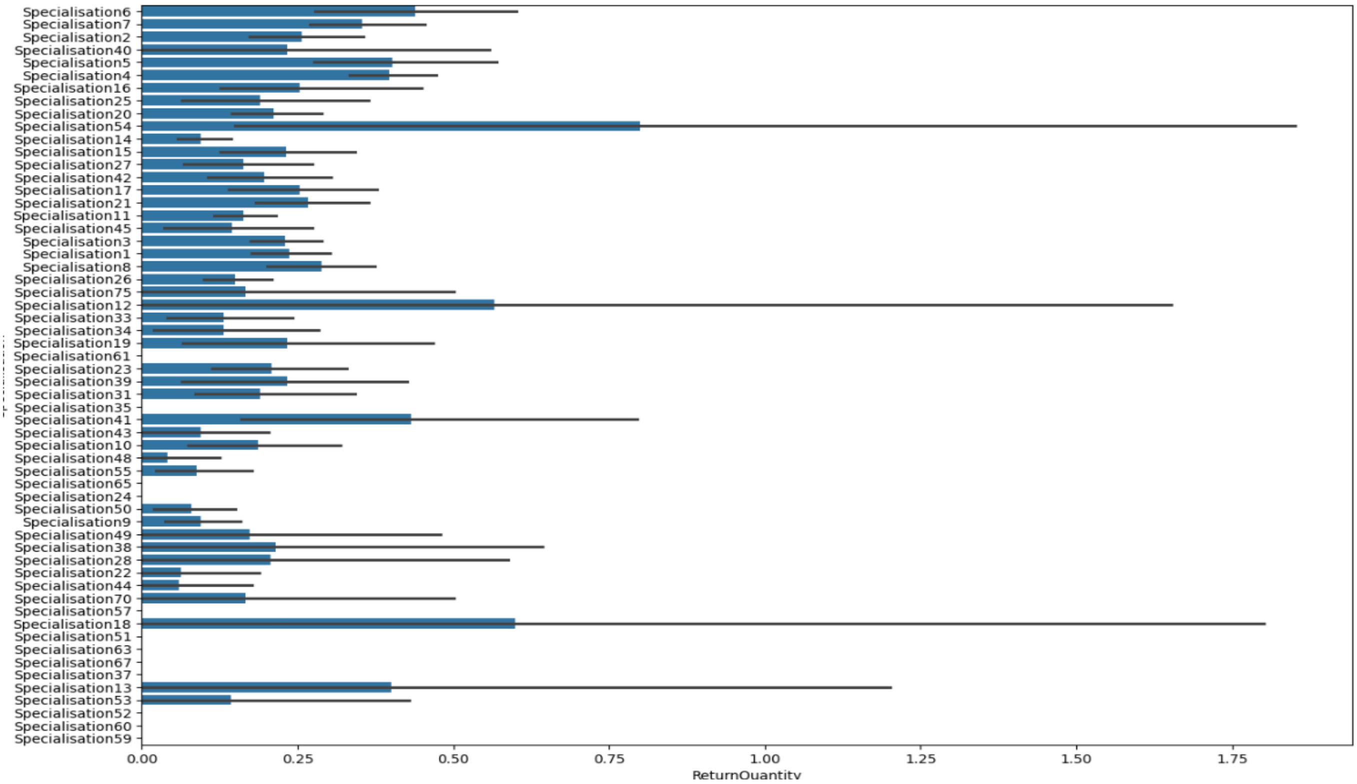

Fig – 13 : Specialization vs Return Qunatity.

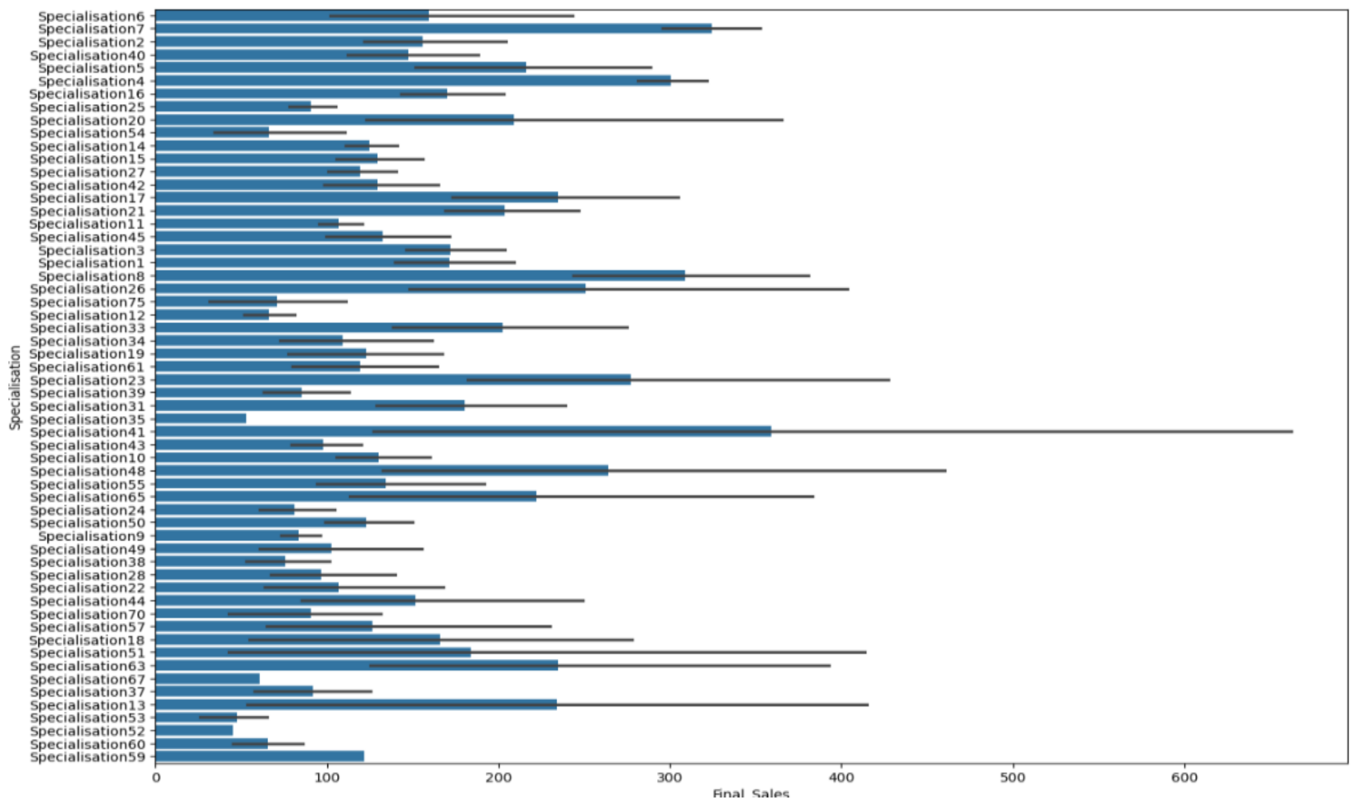

Fig - 14 : Specialization vs Final Sales.

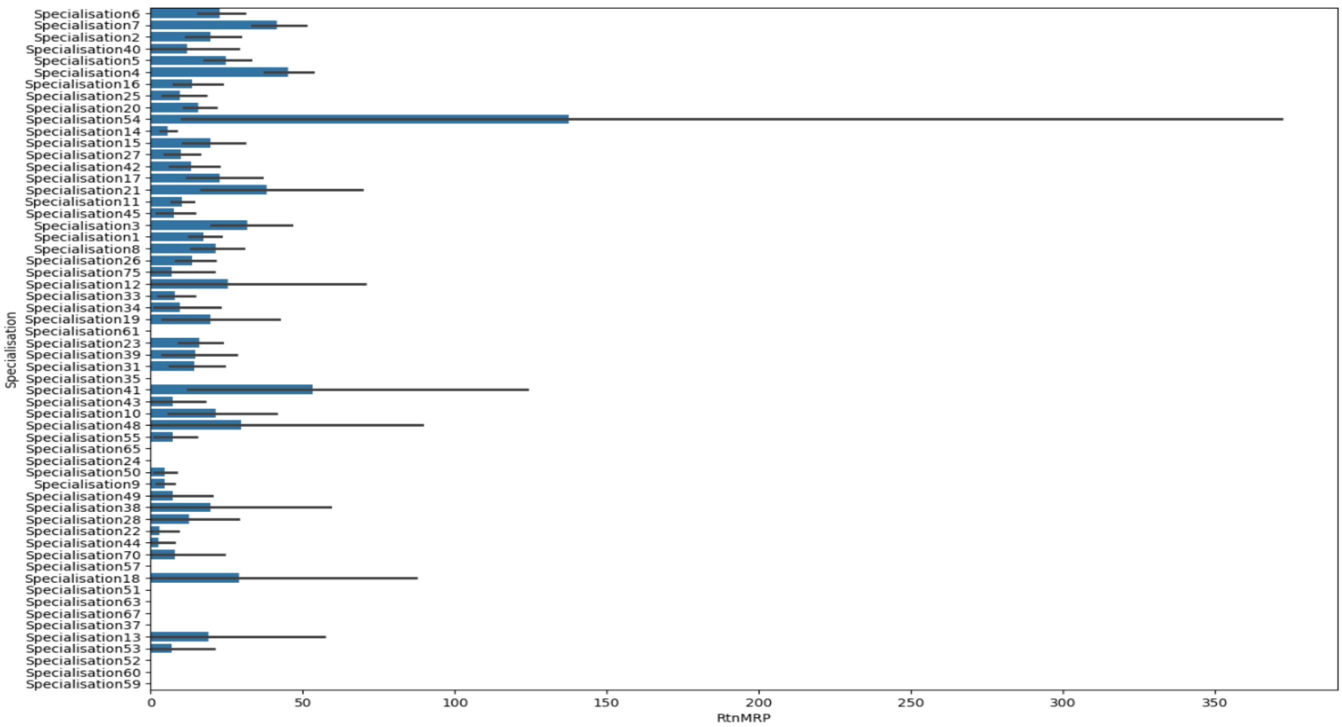

Fig – 15 : Specialization vs RtnMRP.

Similarly, using PostgreSQL using different functions like SUM (), AVG (), COUNT (), etc. Also, can process datasets.

4. Future Scope and Insights:

I. The future scope of medical inventory optimization and bounce rate is to leverage data-driven approaches and advanced machine learning techniques to forecast drug demand, optimize inventory levels, and reduce waste and costs. Some of the **benefits** of this approach are:

- Minimize drug shortages and stockouts, which can lead to improved patient care and satisfaction, as well as reduced bounce rate.
- Maximize the availability and utilization of drugs, which can increase sales and profits, as well as customer loyalty and retention.
- Reduce inventory costs and waste, which can improve cash flow and sustainability, as well as reduce the environmental impact of expired or unused drugs.

II. Statistical Insights:

- Unprocessed data shows the noise in all four-moment business decision like mean, standard deviation, variance, range, skewness, and kurtosis.
- Specialisation4 and Specialisation7 with Department1 have a higher number of returns of FORM1 {Formulation}.
- Specialisation54 contains highest RtnMRP : MRP of returned drug (Quantity included) as well as Highest ReturnQuantity
- Specialisation41 contains highest Final_Cost and Final_Sales
- Specialisation8 contains highest Quantity Similarly after go throughing all possible combination.
- Cleaning the data led to more stable distributions, essential for accurate decision-making.
- Specific subcategories like "INJECTIONS" and "TABLETS & CAPSULES" and the "Form1" formulation were identified as having high return counts.
- Seasonal trends in hospital operations, with varying revenue and demand across different months, were noted.

III. Business Insights

- Here we can understand that around 30 plus percent (approx. 30.548%) of customer in the data set based on a situation where they returned medicine with a final_sales value of zero this means that a significant portion of our customer did not get the medicine they needed which could lead to dissatisfaction of a customer so to improve business criteria we can increase our revenue and it's important to reduce the bounce rate by ensuring customer service at is best.

5. Recommendations:

After analysing whole research, in recommendations section, there are some improvements in dataset,

- Thorough examination of the "TABLETS & CAPSULES" and "INJECTIONS" subcategories to pinpoint and solve the root causes of high return rates.
- A thorough evaluation of the "Form1" formulation to pinpoint areas in need of modification or replacement.
- Assessment and potentially revaluation of suppliers, particularly for products related to Department1 and "Form1" formulation.
- To maximize stock levels and prevent returns, Department 1 enhanced inventory control.
- Development of efficient return management practices in Department1.
- Implementation of specialisation-specific strategies, particularly in Specialisation4 and Specialisation7, to address higher return rates.
- Seasonal resources should be deliberately allocated to maximize operations and resource usage during peak and off-peak seasons.

6. Conclusion:

Key points after the research:

- The initiative is designed with specific recommendations that, when implemented, target key areas of concern.
- By focusing on these identified problem areas, the primary goal is to address and mitigate issues effectively.
- A significant reduction in the bounce rate is one of the core objectives of this initiative.
- Lowering the bounce rate is crucial as it reflects directly on the hospital's ability to retain its patients.
- Increasing customer satisfaction is another vital goal of the initiative.
- Improved customer satisfaction will enhance the patient experience, contributing to better healthcare outcomes.
- By ensuring higher levels of satisfaction, the hospital expects to see a positive impact on patient loyalty.
- Enhanced overall performance of the hospital is a direct consequence of successfully implementing these changes.
- Improved performance will also contribute to the hospital's financial health, increasing its revenue.
- Thus, this initiative aims to holistically improve both the qualitative and quantitative metrics of the hospital's operations.

References:

[1]. Khan,Muztoba.(2014). Transportation Cost Optimization Using Linear Programming. International Conference on Mechanical, Industrial and Energy Engineering 2014 26-27 December 2014, Khulna, BANGLADESH. ICMIEE-PI-140224.

[2]. Vamsikrishna, A. & Raj, Vigneshwar& Sharma S G, Divya. (2021). Cost Optimization for Transportation Using Linear Programming. https://doi.org/10.1007/978-981-16-0976-3_2.

[3]. Uzorh, Engr &Nnanna, Innocent. (2014). Supply Chain Management Optimization Problem.

https://doi.org/10.13140/RG.2.2.19682.58566.

[4]. Oba, Kenneth &Abere, Abiye. (2023). Optimization of Shipping Cost of Sand for Selected Construction Projects of Julius Berger. 12. 16-20.

https://doi.org/10.5923/j.ijcem.20231201.02.

[5]. Çerkini, Bashkim&Bajrami, Roberta &Kosova, Robert & Shehu, Valentina. (2015). Transportation Cost Optimization. Academic Journal of Interdisciplinary Studies. https://doi.org/10.5901/ajis.2015.v4n2s1p42.

[6]. Qi, Chenyang & Ren, Jiaying & Su, Jin. (2023). GRU Neural Network Based on CEEMDAN–Wavelet for Stock Price Prediction. Applied Sciences. 13. 7104. https://doi.org/10.3390/app13127104.

[7]. Alkiayat, Mohammad. (2021). A Practical Guide to Creating a Pareto Chart as a Quality Improvement Tool. Global Journal on Quality and Safety in Healthcare. 4. https://doi.org/10.36401/JQSH-21-X1.

[8]. Basha, Mir Mohammed Junaid & V.S, Navya & Wani, Sonali & Gogi, Vivekanand. (2020). Study of Inventory Management in Pharmaceuticals: A Review of COVID-19 Situation. International Journal of Innovative Science and Research Technology. 5. 366-371. https://doi.org/10.38124/IJISRT20AUG257.

[9]. Bhojwani, Divya & Bhojwani, Khushboo. (2019). A study to review drug inventory and pharmacy management with reference to I.V. & injectables at a tertiary municipal care hospital with 1800 bedded hospital. 342-350.

[10]. Binsar, Faisal. (2020). Drug Stock Optimization Based on Consumption Patterns for Hospital Formulary Using Deep Learning Approach. International Journal of Advanced Trends in Computer Science and Engineering. 9. 2689-2697. https://doi.org/10.30534/ijatcse/2020/31932020.

[11]. Tuga, Mauritsius. (2019). Bank Marketing Data Mining using CRISP-DM Approach. International Journal of Advanced Trends in Computer Science and Engineering. 8. 2322-2329. https://doi.org/10.30534/ijatcse/2019/71852019. .

[12]. Mariscal, Gonzalo & Marbán, Oscar & Fernández, Covadonga. (2010). A survey of data mining and knowledge discovery process models and methodologies. Knowledge Eng. Review. 25. 137-166. https://doi.org/10.1017/S0269888910000032.

[13]. Zolnoori, Maryam & Fung, Kin & Patrick, Timothy & Fontelo, Paul & Kharrazi, Hadi & Faiola, Anthony & Wu, Shirley & Eldredge, Christina & Luo, Jake & Conway, Mike & Zhu, Jiaxi & Park, Soo & Xu, Kelly & Moayyed, Hamideh. (2019). The PsyTAR Dataset: From Patients Generated Narratives to a Corpus of Adverse Drug Events and Effectiveness of Psychiatric Medications. Data in Brief. 24. 103838. https://doi.org/10.1016/j.dib.2019.103838.

[14]. Panigrahi, Chhabi & Panda, Bijayalaxmi & Pati, Bibudhendu. (2022). Exploratory Data Analysis and Sentiment Analysis of Drug Reviews. Computación y Sistemas. 26. https://doi.org/10.13053/cys-26-3-4093.

[15]. Vibin Ravi Kumar, Pallavi Waghmare, Sampath Bukya, Bharani Kumar Depuru, Dr. Ilankumaran Kaliamoorthy, Forecasting Drug Demand for Optimal Medical Inventory Memangement: A Data Driven Approach with Advanced Machine Learning Techniques. ISSN No.: - 2456-2165. https://ijisrt.com/assets/upload/files/IJISRT23SEP159.pdf.

[16]. Shreya.S, Ankita Yadav, Aivarajan S.N, Bharani Kumar Depuru, Efficient Transportation Cost Minimization Strategies in Supply Chain Management: A Comprehensive Analysis and Optimization Framework by using Solver Methodology & Python Programming. ISSN No.:- 2456-2165. https://www.ijisrt.com/assets/upload/files/IJISRT23AUG1773.pdf.