USE CASE STUDY REPORT

Group No.: Group 15
Topic : Pharmacy Supply Chain Database Management System
Student Names: Rutuja Dhanawade and Snehal Rajwar

Executive Summary

The primary objective of this study was to design and implement a relational database that is industry-ready for the Pharmaceutical Industry to show the complete cycle of getting raw material from suppliers to the end product bought by the customer. This project will make it convenient for Walgreens to create a supply chain ecosystem linking various Customers to Pharmacies, Pharmacies to the Drug Distributors and Drug Manufacturers to Raw Material suppliers with ease.

This model can help Pharmacies maintain a steady flow of Drugs and overcome the problem of out-of-stock scenario by the statistics generated which will make sure that customer always receives the product when needed and in process based on these patterns sales for all others in the chain increases as well.

This also helps in linking the distributors and suppliers and Pharmacies in the same area together to make it convenient and faster to supply respective products to each other cutting down on the time needed for transportation overall helping cut cost for them which can help in reducing the prices of the drugs sold to the customer. Overall saving everyone money.

Another issue which the project aids in solving is to help the pharmacies with store layout & product placement based on three parameters i.e. Age, Gender and on the combination of drugs sold together. This not only can help increase the sale for the Pharmacies but also make the customer shopping experience smoother which again can result in increasing the amount of returning customers.

The EER and UML diagrams were modeled, followed by the mapping of the conceptual model to a relational model with the required primary and foreign keys. This database was then implemented fully MySQL and a prototype with four tables and three relationships were implemented on MongoDB NoSQL database to study the feasibility of this database in a NoSQL environment.

The created database is an immense success, and by connecting it to R, the analytics capabilities are immense, some of which have been shown in the study. These queries are immensely helpful in optimizing the flow of products right from the Raw Material supplier to the hand of the Customer along with maintaining a profitable scenario for each entity involved.

I. Introduction

The global pharmaceutical manufacturing market size was valued at USD 405.52 billion in 2020 and is expected to grow at a compound annual growth rate (CAGR) of 11.34% from 2021 to 2028 . Optimizing the process of supply for this industry can help save millions of dollars for the industry and all the stakeholders involved.

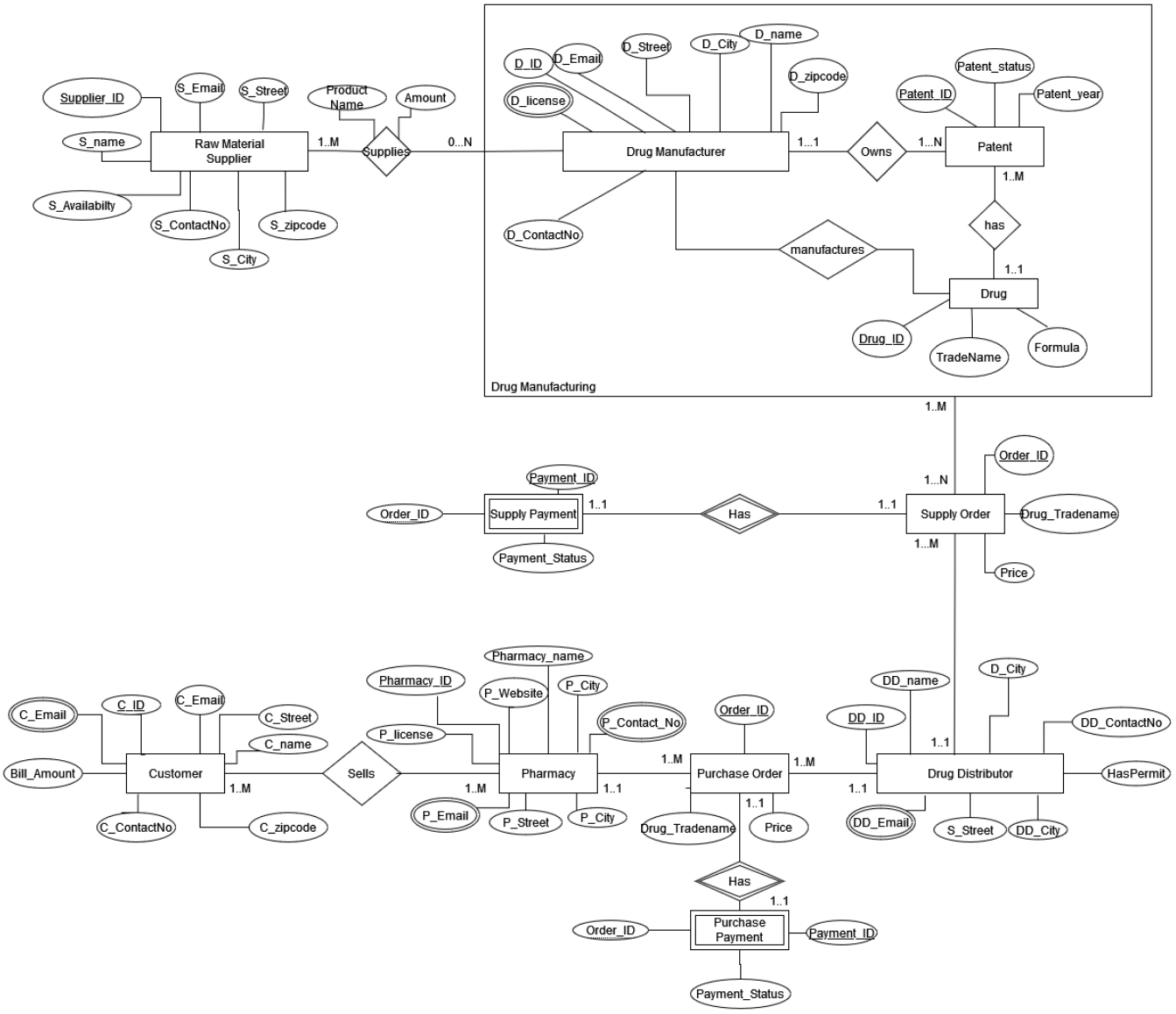
The top priority in any health system is delivering medicine as a strategic product. In the present context of a health-conscious society, management of pharmaceutical supply chains has become more complex because it involves the life-saving interest of human beings and requires the participation of different stakeholders such as pharmaceutical manufacturers, wholesalers, distributors, customers, information service providers, and regulatory agencies. Limited research is available in the area of pharmaceutical supply chains. The Council of

Supply Chain Management Professionals defines supply chain management as planning and management of all activities involved in sourcing, procurement, conversion, and all logistics activities. There are a variety of aspects of evaluating in the supply chain; eliminating bottlenecks, balancing between tiniest material cost and transportation, optimizing manufacturing flow, maintaining the right mix and location of factories and warehouses, vehicle routing analysis, dynamic programming and efficient use of capacities, inventories, and labors are of main aspects of supply chain optimization.

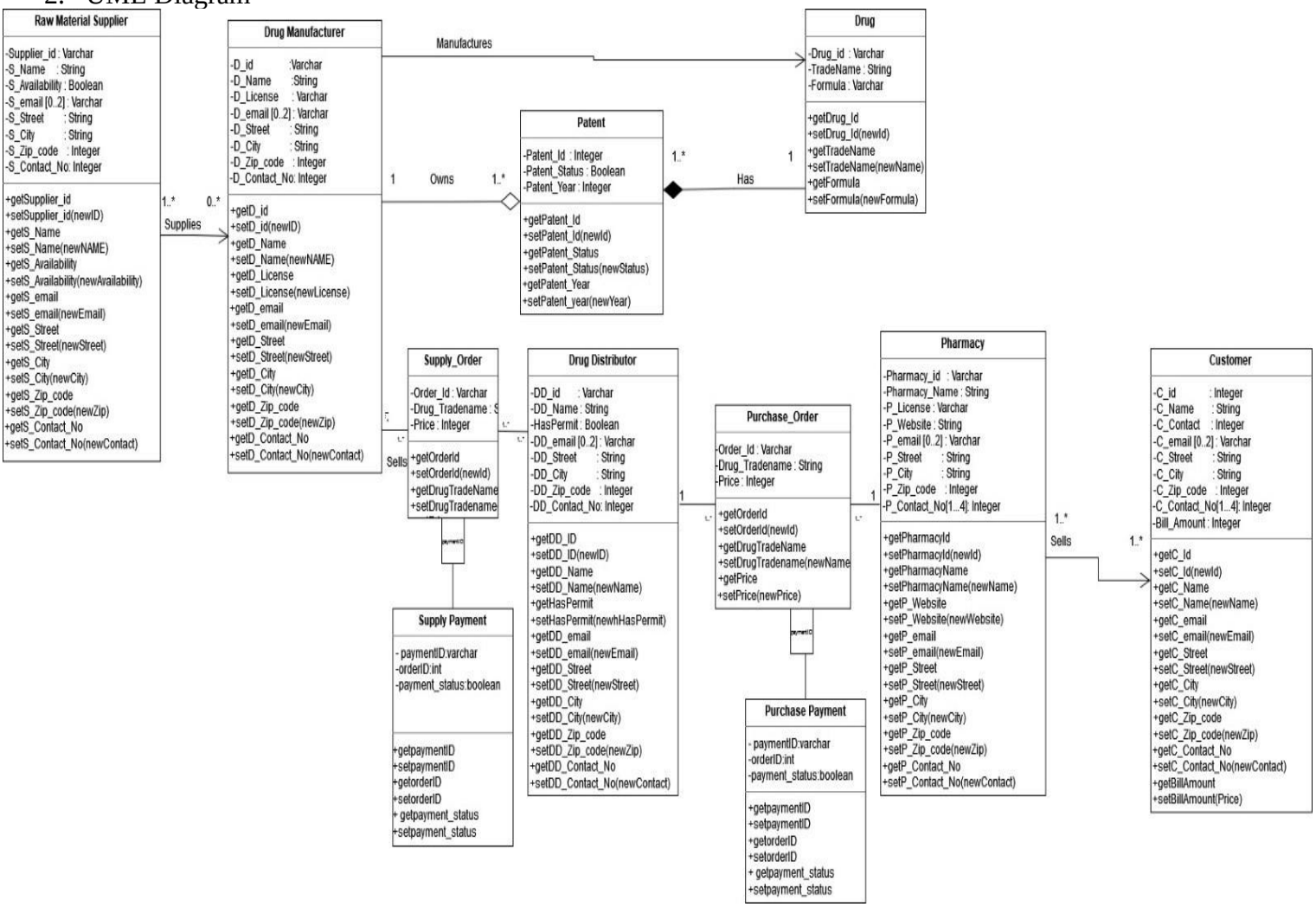
This database model aids all stockholders need to institute the right configuration and adaptability to create best practices and to overcome the obstacles in a continually changing environment. The pharmaceutical supply chain model shown aims to provide medicines in the right quantity, with the acceptable quality, to the right place and customers, at the right time and with the optimum cost to be consistent with the health system's objectives, ensuring benefits for its stockholders.

II. Conceptual Data Modeling

1. EER Diagram



2. UML Diagram



III. Mapping Conceptual Model to Relational Model

1. Raw Material Supplier(Supplier_ID,S_Name,S_Availability,S_ContactNo,S_City,S_Street,S_Email, S_ Zipcode)
2. Supplies(*Supplier_ID*,*D_ID*,Product Name,Amount)
3. Drug Manufacturer(D_ID,D_Name, D_ContactNo,D_City,D_Street,D_Email,D_Zipcode)
4. License(D_License, *D_ID*)
5. Patent(Patent_ID,Patent_Status,Patent_year,*D_ID*,*Drug_ID)*
6. Drug(Drug_ID,TradeName,Formula,*D_ID)*
7. Drug_Manufacturing(*D_ID*,*Patent_ID*,*Drug_ID*,*Order_ID*)
8. Supply_Order(Order_ID,Drug_Tradename,Price, Date)
9. Supply_Payment(Order_ID,Payment_ID, Payment_Status, Date)
10. DrugDistributor(DD_ID,DD_Name,DD_City, DD_Street, DD_Zipcode)
11. DrugDistributorEmail(DD_Email,*DD_ID*)
12. Purchase_order(Order_ID,Drug_Tradename,Date,Price,*DD_ID, Pharmacy_ID)*
13. Purchase Payment*(Order_ID* ,Payment_ID,Payment_status,Date)
14. Order_ID cannot be a null value
15. Phramacy(Pharmacy_ID,Pharmacy_Name,P_Contact_No,P_City,P_Street,P_Email,P_Zipcode, P_License,P_website)
16. Sells(*C_ID* ,*Pharmacy_ID*)
17. Customer(C_ID,C_Name,C_ContactNo,C_City,C_Street,C_Email,C_Zipcode,Bill_Amount,Age, C_Gender,Drugs_purchased)

IV. Implementation of Relation Model via MySQL and NoSQL

1.MySQL Implementation:

Q1. Display the Pharmacy name, license, zipcode, website which sells a particular medicine at the cheapest rate. *(Such queries can be implemented to find pharmacies which sell medicine below or above average price)*

select s.Pharmacy_ID, pl.Pharmacy_Name, cp.Drugs_Purchased, cp.Unit_Price, cp.Quantity, cp.Date, pl.P_License, p.P_Zipcode, p.P_Website from pharmacy p, pharmacy_license pl, customer_purchase cp, sells s
where p.Pharmacy_ID=s.Pharmacy_ID **and** p.P_License=pl.P_License **and** cp.C_ID=s.C_ID
andcp.Unit_Price=
(**select min**(cp1.Unit_Price) **from** customer_purchase cp1 **where** cp1.Drugs_Purchased =cp.Drugs_Purchased);

Pharmacy_ID	Pharmacy_Name	Drugs_Purchased	Unit_Price	Quantity	Date	P_License	P_Zipcode	P_Website
11641481	BROADWAY APOTHECARY	Lactaid Tablets	11	1	2021-07-18	1154414597	97401	HTTP://WWW.BROADWAYAPOTHECARY.
11641481	BROADWAY APOTHECARY	Hand Sanitizer	10	5	2021-10-06	1154414597	97401	HTTP://WWW.BROADWAYAPOTHECARY.
11643507	RALEY'S - 317	Boston Conditioning Solution	1	3	2021-10-21	1487887303	95376	HTTP://WWW.RALEYS.COM
11643507	RALEY'S - 317	Pataday Eye Drops (2.1ml, 1ml)	10	1	2021-07-17	1487887303	95376	HTTP://WWW.RALEYS.COM
11644749	PUCCI'S LEADER PHARMACY	O-Tussin DM Syrup	75	1	2020-01-06	1780687277	95816	HTTP://WWW.PUCCIRX.COM

Q2. Find the Drug Distributor's Name, City and Drug Manufacturer's name, City which supply drugs to Pharmacies located in 'Creswell' City

select distinct dm.D_Name as Drug_Manufacturer_Name, da.D_City as Drug_Manufacturer_City, dm.D_Zipcode,dd.DD_Name as Drug_Distributor_Name ,dda.DD_City as Drug_Distributor_City, dd.DD_Zipcode, pl.Pharmacy_Name, pa.P_City as Pharmacy_City, p.P_Zipcode **from** drug_manufacturer dm, dmanufacturer_address da, drug_distributor dd, drug_distributor_address dda , pharmacy p, pharmacy_address pa, pharmacy_license pl, supplyorder so, purchase_order po
where so.D_ID=dm.D_ID **and** so.DD_ID=dd.DD_ID **and** po.DD_ID=dd.DD_ID **and** po.Pharmacy_ID=p.Pharmacy_ID **and** dm.D_Zipcode=da.D_Zipcode **and** dd.DD_Zipcode=dda.DD_Zipcode **and** p.P_Zipcode=pa.P_Zipcode **and** p.P_License=pl.P_License **and** pa.P_City='Creswell';

Drug_Manufacturer_Name	Drug_Manufacturer_City	D_Zipcode	Drug_Distributor_Name	Drug_Distributor_City	DD_Zipcode	Pharmacy_Name	Pharmacy_City	P_Zipcode
Access Near West	Chicago	60653	SOLAS PHARMACEUTICALS LTD	MAPLE VALLEY	94538	PHARMACY EXPRESS - 048	CRESWELL	97426
Danny Boy 2 Unlimited Pat	Chicago	60610	GMAC HEALTH LTD	GLENDALE	94015	PHARMACY EXPRESS - 048	CRESWELL	97426
Access Peterson	Chicago	60615	GMAC HEALTH LTD	GLENDALE	94015	PHARMACY EXPRESS - 048	CRESWELL	97426

Q3 Find the pharmacies whose payment to the drug distributor is pending and display the amount to be paid by that pharmacy.

select po.POrder_ID, pp.Payment_ID, pl.Pharmacy_Name, pp.Payment_Status, sum(po.Total_Price) as Pending_Payment_Amount **from** pharmacy p, pharmacy_license pl,purchase_payment pp, purchase_order po
where p.Pharmacy_ID=po.Pharmacy_ID **and** po.POrder_ID=pp.POrder_ID **and** pl.P_License=p.P_License **and** pp.Payment_Status="FALSE"
group by po.POrder_ID;

POrder_ID	Payment_ID	Pharmacy_Name	Payment_Status	Pending_Payment_Amount
12482002	2	CVS - 3937	FALSE	135896.6
12482005	5	WALGREENS - 13596	FALSE	22400
12482007	7	WALGREENS - 13596	FALSE	16240
12482008	8	PHARMACY EXPRESS - 048	FALSE	26275
12482011	11	CVS - 3937	FALSE	6545

Q4. Count the number of pharmacies depending on a certain pin code.

```
select p.P_Zipcode, count(pl.Pharmacy_Name) as
No_Of_Pharmacies
from pharmacy p, pharmacy_address pa,
pharmacy_license pl
where p.P_License=pl.P_License and
p.P_Zipcode=pa.P_Zipcode
group by p.P_Zipcode
order by No_Of_Pharmacies desc;
```

P_Zipcode	No_Of_Pharmacies
94115	9
95758	9
95823	9
94538	8
94040	8

Q5. Display Medicines consumed by male & female separately

```
select C_ID,Drugs_Purchased,Quantity from
customer_purchase
where C_ID IN
        (select c_id from customer
        where C_Gender="Male");
```

C_ID	Drugs_Purchased	Quantity
1	Guaifenesin Syrup	1
1	Mucinex	1
1	Mucinex DM	1
100	Loperamide (2mg tabs)	1
103	Chloraseptic Spray	1
103	Cold-Eeze Zinc Lozenges	3

```
select  C_ID,Drugs_Purchased,Quantity from
customer_purchase
where C_ID IN
        (select c_id from customer
        where C_Gender="Female");
```

C_ID	Drugs_Purchased	Quantity
10	Calamine Lotion	1
10	DeepSea Saline DropsNeti Pot	1
10	Hand Sanitizer	2
10	Vitamin C (100mg)	10
101	Azithromycin	3
101	Calcium Citrate + D (Citracal)	5

Q6. Depending on the age and gender of the customer, find the count of Drugs consumed

```
select  c.C_Age, c.C_Gender,
count(cp.Drugs_Purchased)
 as Count_of_Medicines
from customer c, customer_purchase cp
where c.C_ID=cp.C_ID
group by C_Age, C_Gender;
```

C_Age	C_Gender	Count_of_Medicines
19	Male	19
30	Female	9
20	Male	5
23	Female	14
49	Female	5

Q7. Calculate total sales of all the drugs Pharmacies in the year 2021 and sort it in descending order

```
select p.Pharmacy_ID, pl.Pharmacy_Name, pa.P_City, sum(Total_Price) as Total_Sales_in_$
from pharmacy p, pharmacy_license pl ,pharmacy_address pa, sells s, customer_purchase cp
where p.P_License=pl.P_License and p.P_Zipcode=pa.P_Zipcode and s.Pharmacy_ID=p.Pharmacy_ID
and s.C_ID=cp.C_ID and cp.Date like "%2021%"
group by p.Pharmacy_ID
ORDER BY Total_Sales_in_$ desc;
```

Pharmacy_ID	Pharmacy_Name	P_City	Total_Sales_in_$
11700840	LUCKY SUPERMARKETS - 758	SAN JOSE	4276.3
11697960	KAISER PERMANENTE OAKLAND MEDICAL CEN...	OAKLAND	1995
11709980	BI-MART - 644	FLORENCE	1650
11708622	VALLEY HEALTH CENTER - SUNNYVALE	SUNNYVALE	1592.88
11695438	TRINITY PROFESSIONAL HEALTH MART PHARM...	WEAVERVILLE	1469

Q8. Find the details of pharmacies whose license is expiring in 2025.

select pl.P_License,pl.Pharmacy_Name,pl.Expiry_Date, pa.P_City from pharmacy_license pl, pharmacy_address pa, pharmacy p **where** Expiry_Date like "%2025%" **and** pl.P_License=p.P_License **and** p.P_Zipcode=pa.P_Zipcode;

P_License	Pharmacy_Name	Expiry_Date	P_City
1356370399	SAFEWAY	2025-04-07	STOCKTON
1952413874	PHARMACA INTEGRATIVE PHARMACY	2025-04-07	NOVATO
1215954631	WAL-MART - 1881	2025-04-27	ANTELOPE
1285655258	MAI-TRAM PHARMACY	2025-04-28	SAN FRANCISCO
1154414597	BROADWAY APOTHECARY	2025-07-28	EUGENE
1659513026	BI-MART - 680	2025-08-24	SISTERS

2.NoSQL Implementation:

The collections were created with the help of Mongo shell and MongoDB Compass

Q1. Find the Drug Manufacturers located in Ohio and Seattle

db.Drug_Manufacturer.find({ City:{$in:["Ohio","Seattle"]}});

```
},
{
  _id: ObjectId("61b0391727d86324c6793ac1"),
  D_ID: 'D62D34',
  D_Name: 'Austin CBC',
  D_Contact: 544707130,
  D_Email: 'AustinBC@manu.com',
  D_Zipcode: 60608,
  City: 'Seattle',
  Street: '345 A 25th street'
},
{
  _id: ObjectId("61b0391727d86324c6793ac4"),
  D_ID: 'D62D37',
  D_Name: "Charlie's",
  D_Contact: 2347071331,
  D_Email: 'charlies@manu.com',
  D_Zipcode: 60614,
  City: 'Ohio',
  Street: '345 A 48th street'
},
```

Q2. Display the patent details where status is granted

db.Patent.aggregate({$match:{Patent_Status:{$eq:" True"}}});

```
Pharmacy_SCM> db.Patent.aggregate({$match:{Patent_S
[
  {
    _id: ObjectId("61b031df27d86324c6793ab1"),
    Patent_ID: 'PPP85',
    Patent_Status: 'True',
    Patent_year: 2019,
    Drug_ID: 16271
  },
  {
    _id: ObjectId("61b031df27d86324c6793ab3"),
    Patent_ID: 'PPP88',
    Patent_Status: 'True',
    Patent_year: 2019,
    Drug_ID: 16273
  },
  {
    _id: ObjectId("61b031df27d86324c6793ab5"),
    Patent_ID: 'PPP90',
    Patent_Status: 'True',
    Patent_year: 2019,
    Drug_ID: 16275
  },
  {
```

Q3. Display the patents which we applied before 2018

db.Patent.aggregate({$match:{Patent_year:{$lt:2018}}});

```
Pharmacy_SCM> db.Patent.aggregate({$match:{Patent_
[
  {
    _id: ObjectId("61b031df27d86324c6793ab4"),
    Patent_ID: 'PPP89',
    Patent_Status: 'False',
    Patent_year: 2016,
    Drug_ID: 16274
  },
  {
    _id: ObjectId("61b031df27d86324c6793ab6"),
    Patent_ID: 'PPP91',
    Patent_Status: 'True',
    Patent_year: 2016,
    Drug_ID: 16276
  },
  {
    _id: ObjectId("61b031df27d86324c6793ab8"),
    Patent_ID: 'PPP93',
    Patent_Status: 'False',
    Patent_year: 2016,
    Drug_ID: 16278
```

Q4. Find the average price of each Drug

This query is performed in 2 ways : Aggregation Pipeline & MapReduce Pipeline

1. Aggregation Pipeline

```
db.Drug.aggregate([
        {$group:{_id:"$Drug_ID",value:{$avg:"$Pr
ice"}}},
        {$out:"avg_Drug_Price"}
]);
```

```
Pharmacy_SCM> db.avg_Drug_Price.find()
[
  { _id: '206473', value: null },
  { _id: '4907', value: null },
  { _id: '159672', value: null },
  { _id: '97013', value: 54.5 },
  { _id: '33019', value: 532.5 },
  { _id: '33020', value: 20 },
  { _id: '33021', value: 60 },
  { _id: '208087', value: null },
  { _id: '31947', value: null },
  { _id: '33022', value: 100 },
  { _id: '66736', value: 46 },
  { _id: '163740', value: null }
]
Pharmacy_SCM>
```

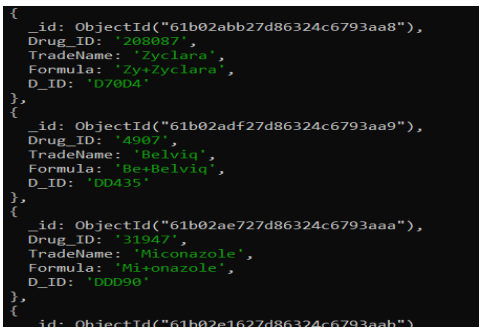
2. MapReduce Pipeline

```
var mapFunction1=function(){
emit(this.Drug_ID,this.Price);
};

var reduceFunction1 = function(drugId,price){
return Array.avg(price);
};

db.Drug.mapReduce(
mapFunction1,
reduceFunction1,
{out:"mapreduce_avg"}
)
```

```
DeprecationWarning: Collection.mapReduce() is
See https://docs.mongodb.com/manual/core/map-r
{ result: 'mapreduce_avg', ok: 1 }
Pharmacy_SCM> db.mapreduce_avg.find();
[
  { _id: '4907', value: NaN },
  { _id: '163740', value: NaN },
  { _id: '206473', value: NaN },
  { _id: '66736', value: 46 },
  { _id: '97013', value: 54.5 },
  { _id: '33019', value: 532.5 },
  { _id: '33020', value: 20 },
  { _id: '208087', value: NaN },
  { _id: '31947', value: NaN },
  { _id: '33022', value: 100 },
  { _id: '33021', value: 60 },
  { _id: '159672', value: NaN }
]
Pharmacy_SCM>
```

Q5. Find all the records in Drug collection

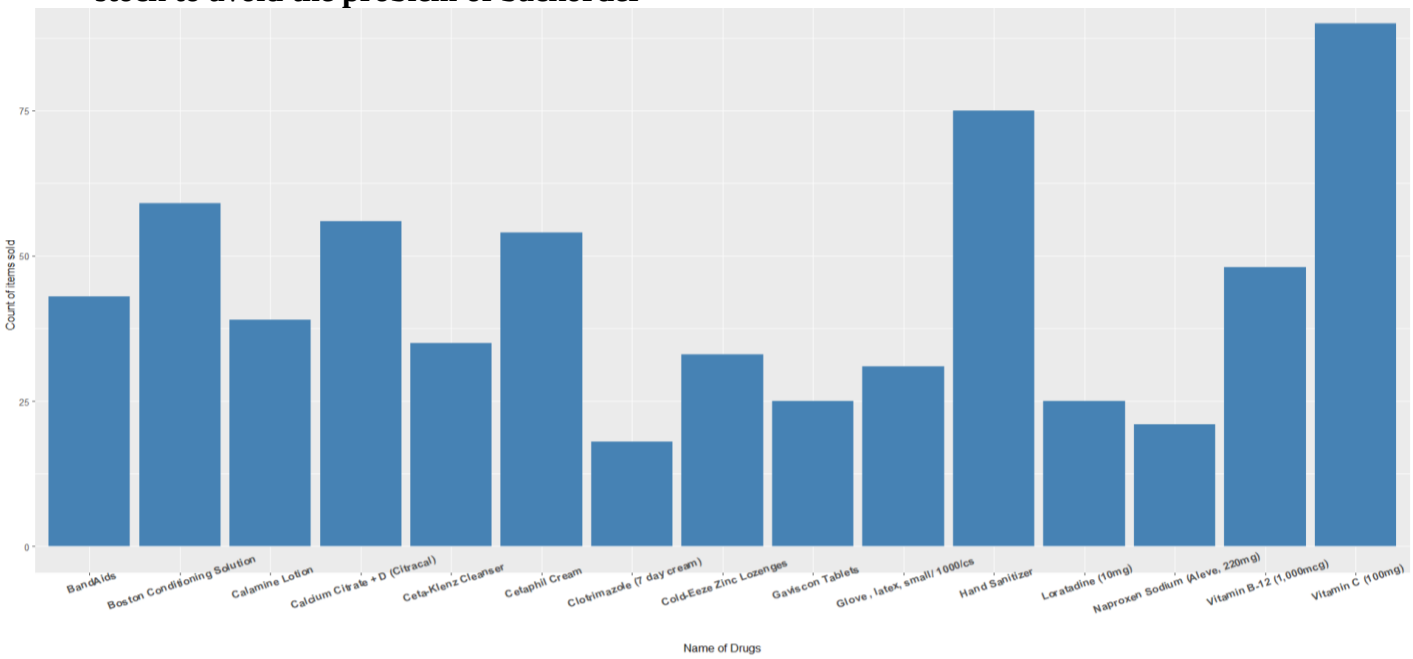
```
db.Drug.find();
```



V. Database Access via R

The database is accessed using R language. The 'RMySQL' library in R which provides all functions for connection and SQL Queries is used. The dbConnect() is used for connection and dbSendQuery() is used to query the data. The result set is fetched using dbFetch() function. Wherever required, the result set list is converted into dataframe using as.data.frame() function for further analysis.

1. **Top 20 Most sold Drugs/Medicines are fetched and plotted, so that these medicines are always in stock to avoid the problem of backorder**



2. **Market Basket Analysis to find the Most Frequently together purchased drugs**

Market basket analysis is a data mining technique which helps retailers to understand customer purchasing patterns by analyzing items which are frequently bought together and therefore predict items which can be bought with a given set of items. For example, if a customer is buying Mucinex and Vicks (cold medicines), there are high chances that he/she might buy an Inhaler as well. So a retailer can showcase these items in together and always keep them in stock in order to increase the profit.

12 Most Frequently together bought itemsets

```
> inspect(rules[1:12])
     lhs                                                  rhs                          support confidence coverage lift count
[1]  {Omega 3 Fish Oil (1,000mg)}                      => {Vitamin A (10,000 IU)}      0.033   1          0.033    25.2 5
[2]  {Omega 3 Fish Oil (1,000mg), One-a-day Womenâ€™s Vitamin} => {Vitamin A (10,000 IU)}      0.026   1          0.026    25.2 4
[3]  {One-a-day Womenâ€™s Vitamin, Vitamin A (10,000 IU)} => {Omega 3 Fish Oil (1,000mg)} 0.026   1          0.026    30.2 4
[4]  {Calcium Citrate + D (Citracal), Ceta-Klenz Cleanser} => {Cetaphil Cream}          0.026   1          0.026    11.6 4
[5]  {Afrin Nasal Spray}                               => {Hand Sanitizer}             0.020   1          0.020     6.9 3
[6]  {Vagisil Wash}                                    => {Rephresh}                   0.020   1          0.020    50.3 3
[7]  {Rephresh}                                        => {Vagisil Wash}               0.020   1          0.020    50.3 3
[8]  {E-Contra One Step}                               => {Clotrimazole (3 day cream)} 0.020   1          0.020    50.3 3
[9]  {Clotrimazole (3 day cream)}                      => {E-Contra One Step}          0.020   1          0.020    50.3 3
[10] {Milk of Magnesia}                                => {Mineral Oil}                0.013   1          0.013    75.5 2
[11] {Mineral Oil}                                     => {Milk of Magnesia}           0.013   1          0.013    75.5 2
[12] {Chlorpheniramine (4mg)}                          => {Delsym}                     0.013   1          0.013    30.2 2
> plot(rules[1:12], method="graph",measure = "count",engine = 'igraph' , shading = "confidence" )
```

Figure 1

On performing market basket analysis on the customer purchase data fetched from database, above pattern is observed for most frequently together purchased medicines.

In **Figure 1**, consider rule 4, which indicates that if 'Calcium Citrate' and 'Ceta-Klenz Cleanser' is bought then customers tend to buy 'Cetaphil Cream' also along with it as it is observed with a confidence of 1 (100%) and count of 4 which means such a purchase (including this combination of medicines) occurred 4 times in our database. **Figure 2** shows graphical representation of the same.

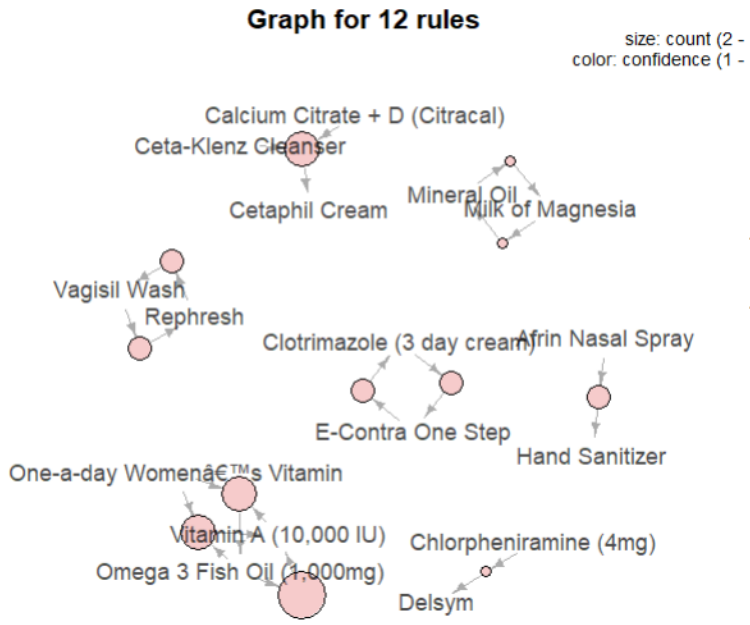


Figure 2

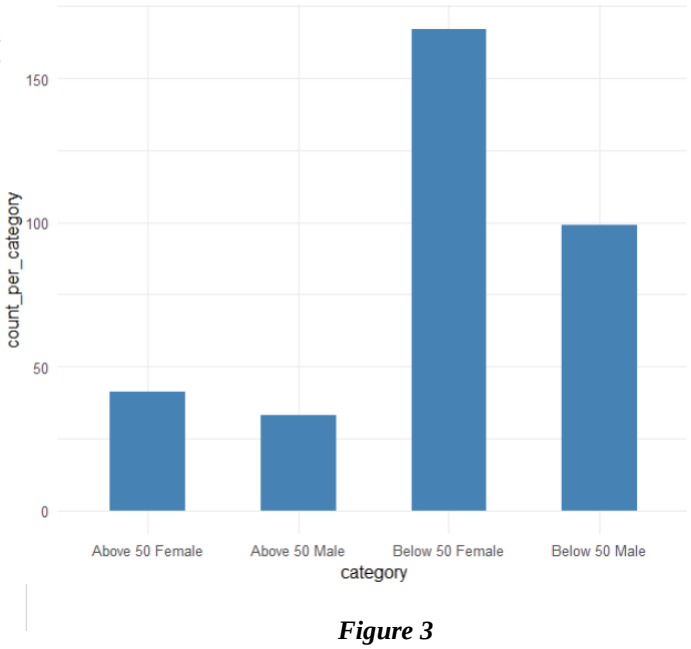


Figure 3

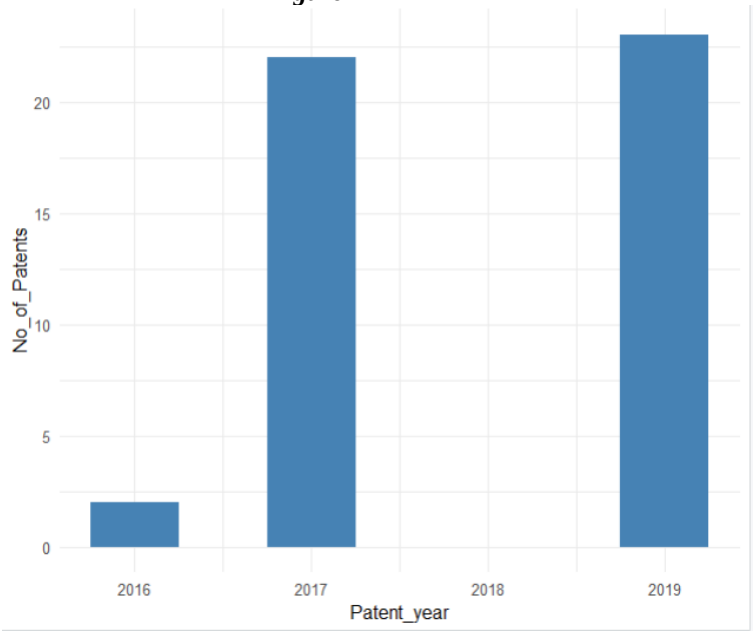


Figure 4

Figure 3 : Displays the number of medicines consumed by male and female customers above and below 50 years of age.

Figure 4 : Displays the number of successfully generated patents of drugs every year.

Figure 5 : Depicts a scenario where the R code prompts user to enter pin code to indicate the drug distributors and pharmacies located in that location.

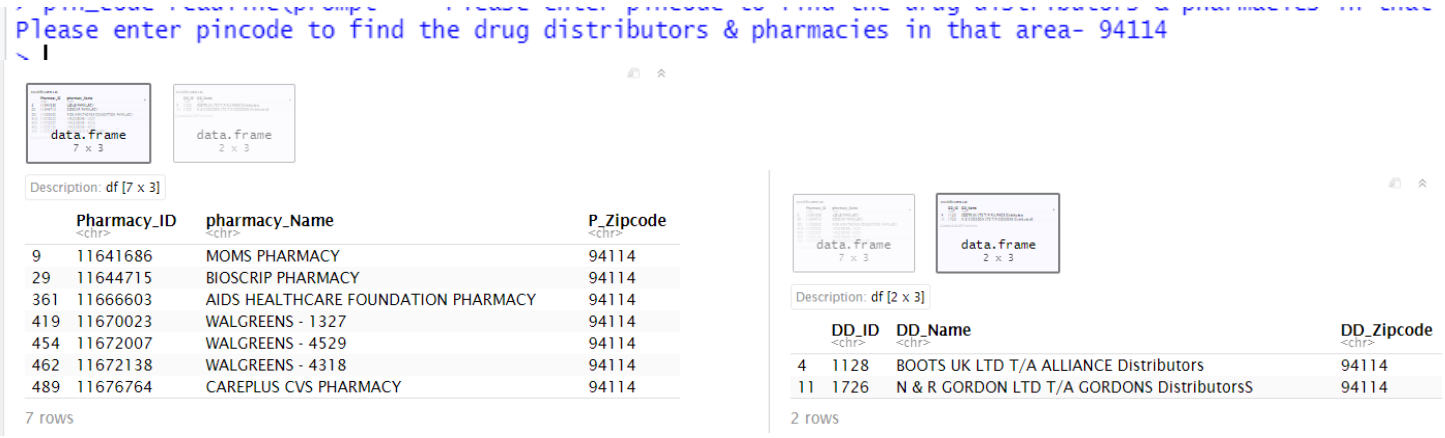


Figure 5

VII. Summary and recommendation

Our Pharmaceutical Supply chain model is an attempt to provide a complete solution for connecting from grass root level of drug generation that is raw material to the in hand product making the process smooth and cost efficient for all the stakeholders involved .We are doing this by providing analysis and various statistics such as drugs sold together , most drugs sold, drug sale based on demographic etc.

Some drawbacks of the model:-

- In order to make sure that their is no redundancy in the database and the information is organized in a concise manner we normalized our relational model which ended up making our data very spread out which lead a lot of joins which made the task a tedious process.
- The increase in tables can also lead to a maintenance overhead.
- We have majorly used SQL for considering the size of the project the use of NoSQL for the model will allow for high-performance, agile processing of information at massive scale. It can store unstructured data across multiple processing nodes, as well as across multiple servers. As such, the NoSQL distributed database infrastructure has been the solution of choice for some of the largest data warehouses.