MARKET BASKET ANALYSIS

- POOJITHA VIJJAPU

Overview

- **INTRODUCTION**
- **PROBLEM STATEMENT**
- **DATA : DESCRIPTION, CLEANING, EDA**
- **DATA EXTRACTION & MANIPULATION**
- **MARKET BASKET INTRODUCTION**
- **ANALYSIS**
- **HYPOTHESIS TESTING**
- **INTREPRETATION & RESULTS**
- **SCOPE OF IMPLEMENTATION AND IMPROVEMENT**

INTRODUCTION

- Market Bastet Analysis on Sales transaction of a Pharmacy
- Gain insights into customer behavior and identify opportunities for improving sales and marketing strategies

PROBLEM STATEMENT

The objective of this project is to understand associations and identify items frequently bought together which will help to :

- enhance customer experience
- manage inventory
- optimize marketing strategy.

Hypothesis to be tested in this project is that customers who purchase diabetic tablets are also likely to purchase multivitamins or other supplements.

DATA DESCRIPTION

Pharmacy Store Transactional Data

We are using raw data directly obtained from a Pharmacy Store in India.
The dataset consists of transactional data with a row for every item sold.

Timeframe

July 2022 to
April 2023

Raw Datasize

32 columns
331k rows

Immediate Cleaning

5 Columns
331k rows

DATA DESCRIPTION

	invno	invdate	itemcode	itemname	n_net_sales
0	199/22/S/37229	01/07/22	I09961	CONCOR AM 5MG-10	89.352
1	199/22/S/37230	01/07/22	I35602	PILORUTE CREAM-30GM	108.4815
2	199/22/S/37230	01/07/22	I01127	ALDACTONE-25MG TAB	51.327
3	199/22/S/37231	01/07/22	I13350	DULOREN-20MG TAB-10	67.5
4	199/22/S/37231	01/07/22	I13350	DULOREN-20MG TAB-10	202.5
5	199/22/S/37231	01/07/22	I24562	LIBEROCAL-PLUS TAB-10S	925.704
6	199/22/S/37232	01/07/22	I30156	NELATON CATHETER-1S	125

DATA CLEANING

MISSING VALUES

There were no missing values

WHOLESALE BILLS

Removed Transactions containing wholesale B2B sales.

RETURNS

Removed transactions with item returns

Exploratory Data Analysis

328310
Rows

124164
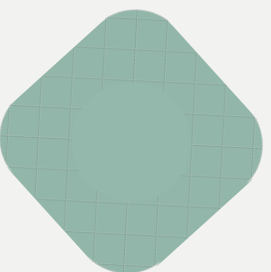Unique
Transactions

12963
Unique Items

Exploratory Data Analysis

Sales by week of the Day

Sales by Day of the Week

17.55%

17.81%

16.37%

15.01%   14.69%   14.80%

3.77%

Total Sales

Sunday   Monday   Tuesday   Wednesday   Thursday   Friday   Saturday
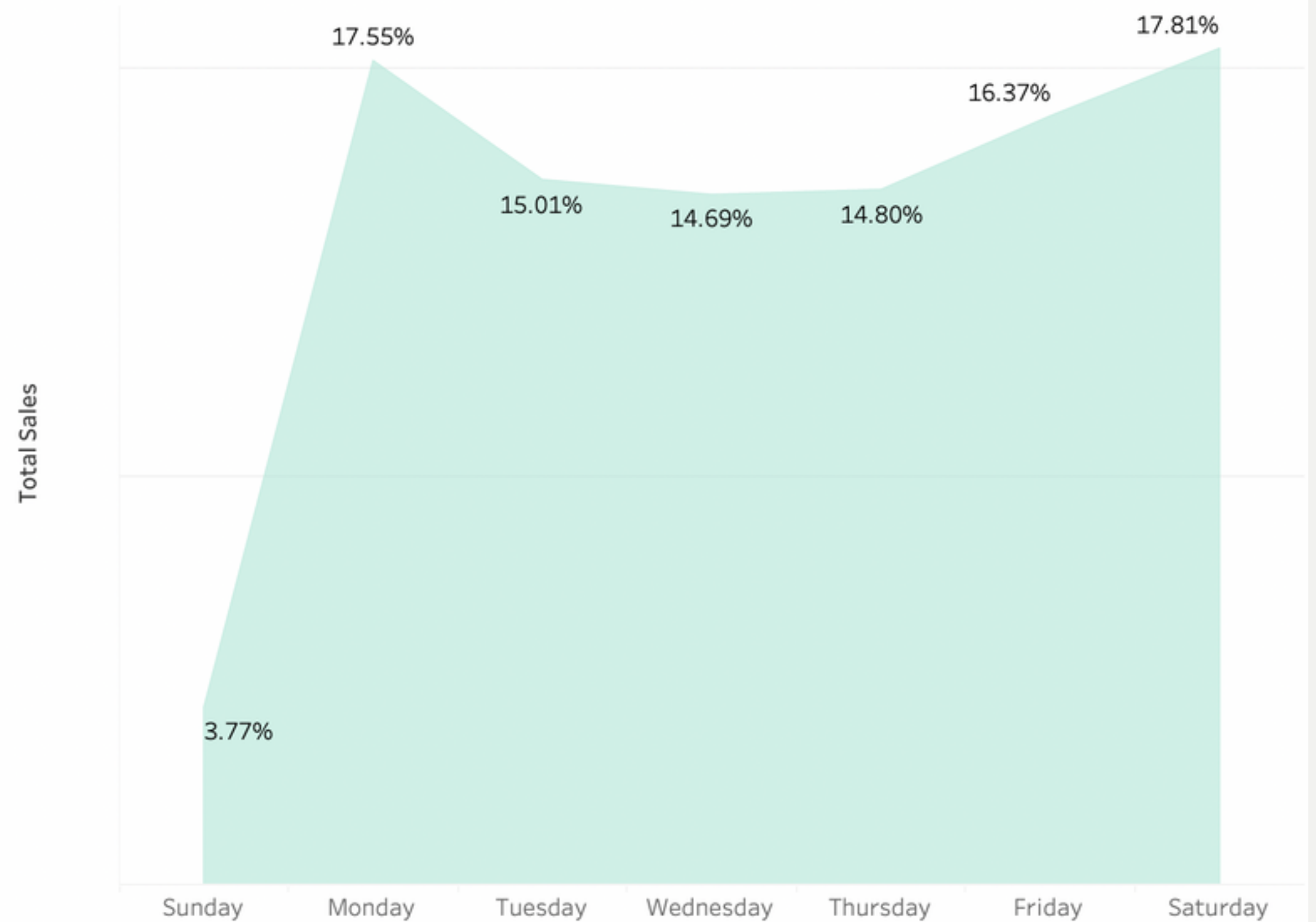
Exploratory Data Analysis

Number of items and number of transactions

Transaction Count for different number of Items
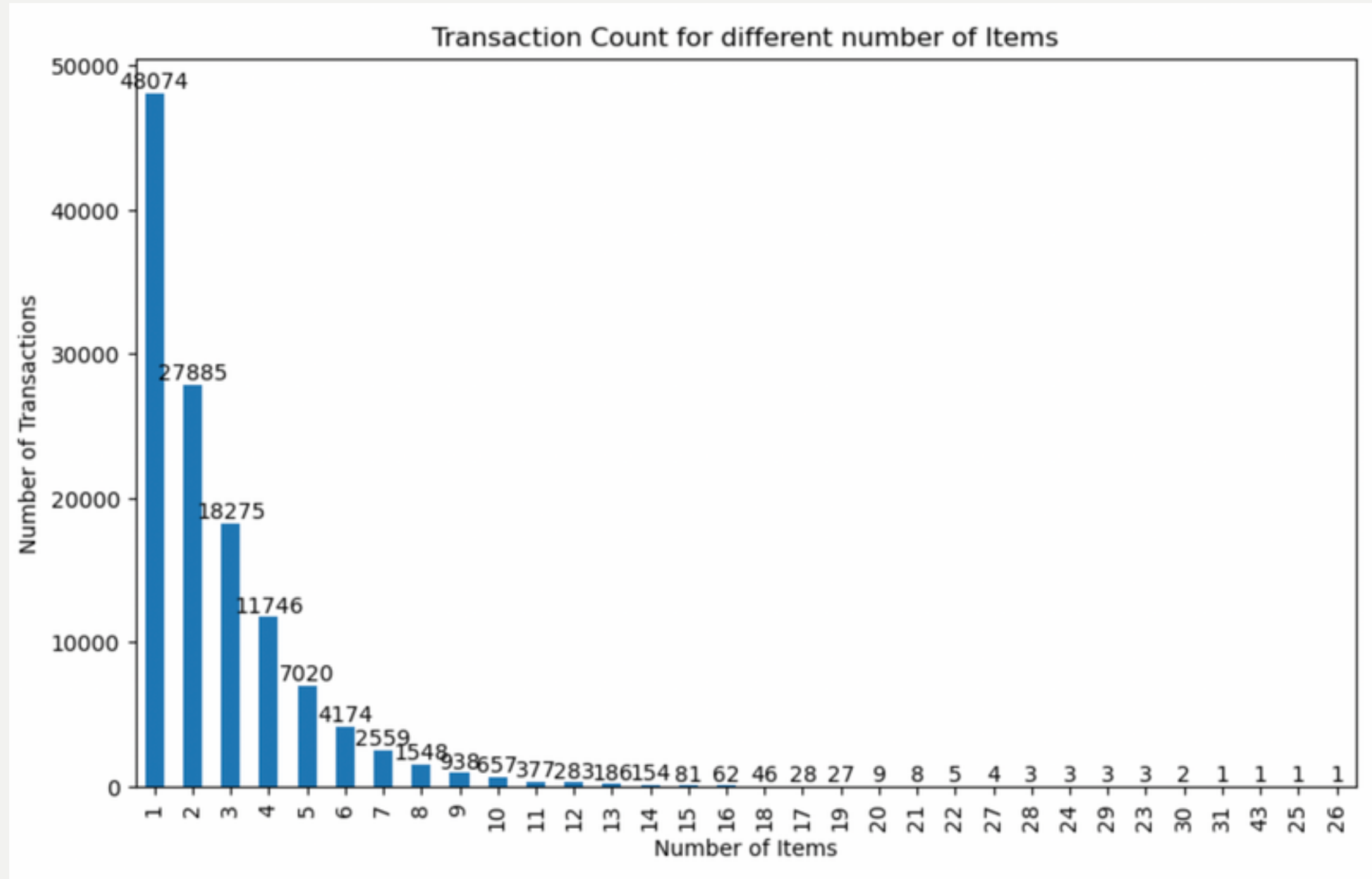
Exploratory Data Analysis

Most frequent items

Top Items

Itemcode

Count of Itemcode

364   2,219
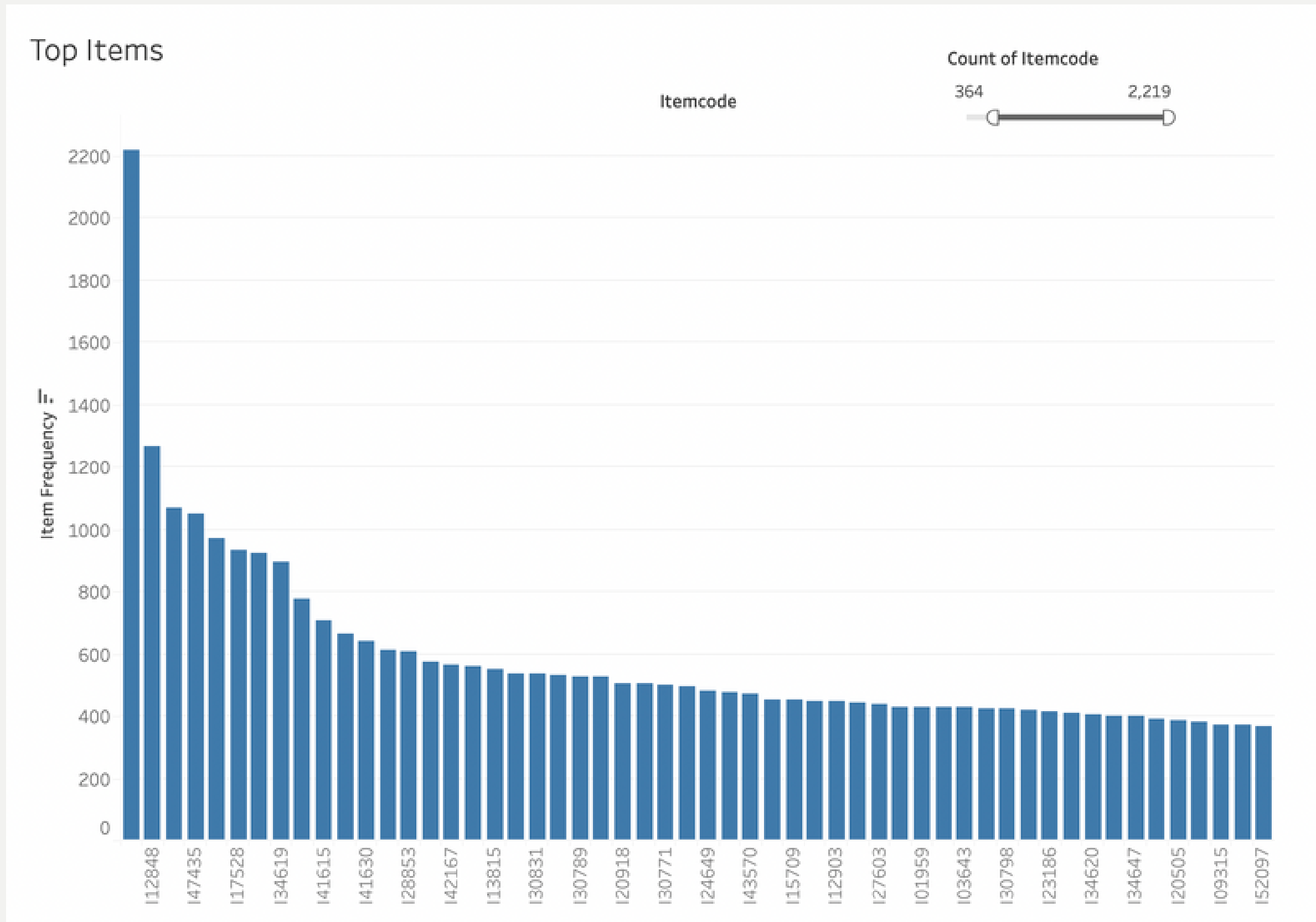
Data Extraction

ITEM COUNT

Removed <1 and >14

ITEM FREQUENCY

Top 50 items

$$\frac{28047}{5}$$

Data Manipulation

Converted into most suitable form

itemcode invno	I00179	I01815	I01959	I02972	I03643	I08274	I09315	I10106	I12592	I12600	...
199/22/S/100008	0	0	0	0	0	0	0	0	0	0	...
199/22/S/100011	0	0	0	0	1	0	0	0	0	0	...
199/22/S/100012	0	0	1	0	0	0	0	0	0	0	...
199/22/S/100013	0	0	0	0	0	0	0	0	0	0	...
199/22/S/100015	0	0	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...

Market Basket Analysis

What?

It is a data mining technique that identifies associations and patterns in customer purchase behavior by examining the contents of their shopping baskets.
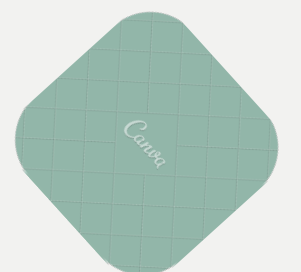
Why?

Helps sellers identify trends which will enable stock management, or marketing.

How?

SUPPORT

CONFIDENCE

LIFT

Market Basket Analysis

Support

It refers to the frequency or occurrence of an item or item(or item set) in the dataset.

Support(A) = (Number of transactions containing A) / (Total number of transactions)

Confidence:

Likelihood of one item or itemset occurring given the occurrence of another item or item set.

Confidence(A -> B) = (Number of transactions containing both A and B) / (Number of transactions containing A)

Lift:

Degree of association or dependency between two items or item sets.

Formula: Lift(A -> B) = (Support(A and B)) / (Support(A) * Support(B))

Market Basket Analysis

Rule : Bread -> Milk.

Support (Bread) = 0.8
Support(Milk)= 0.6
Support(Bread∩ Milk) = 0.4

Confidence(Bread -> Milk) =
(Number of transactions containing both Bread and Milk) /
(Number of transactions containing Bread)
= 2 / 4 = 0.5

Lift(Bread -> Milk) = (Support(Bread∩ Milk)) / (Support(Bread)
* Support(Milk)) = (0.4) / (0.8 * 0.6) = 0.83

Transaction 1	Bread	Milk	
Transaction 2	Bread	Eggs	
Transaction 3	Milk	Eggs	
Transaction 4	Bread	Milk	Eggs
Transaction 5	Bread	Juice	

Analysis

Using the apriori function from the mlxtend (Machine Learning Extensions) library performed the following Steps :

1. Use the metrics support and confidence, to find out the frequent itemset count

USED SUPPORT OF 0.001

Why?

Lower support value gave no results.

```
        support            itemsets
0      0.025093           (I00179)
1      0.019998           (I01815)
2      0.019236           (I01959)
3      0.022807           (I02972)
4      0.018189           (I03643)
..        ...                ...
123    0.001286    (I41620, I51032)
124    0.002714    (I41630, I51032)
125    0.001143    (I51032, I43570)
126    0.001381    (I47084, I47435)
127    0.001381    (I47435, I51032)

[128 rows x 2 columns]
```

Analysis

2. Use metric Lift or Confidence to find out the Association Rules.

USED LIFT AS
THE METRIC

Why?

Gives the degree of association or dependency between two items if they were independent of each other.

	antecedents_id	antecedents_name	consequents_id	consequents_name	lift	confidence
0	I12616	D10 ML DISPOVAN SYRINGE	I42167	NS100ML SODIUM CHLORIDE	17.599657	0.442464
1	I42167	NS100ML SODIUM CHLORIDE	I12616	D10 ML DISPOVAN SYRINGE	17.599657	0.517045
2	I10106	COR-3 TAB -15 S	I24977	LIVOGEN Z TAB-15S	13.446835	0.263789
3	I24977	LIVOGEN Z TAB-15S	I10106	COR-3 TAB -15 S	13.446835	0.266990
4	I30789	NEXPRO 40MG TAB-15S	I24977	LIVOGEN Z TAB-15S	9.984936	0.195876
5	I24977	LIVOGEN Z TAB-15S	I30789	NEXPRO 40MG TAB-15S	9.984936	0.230583
6	I10106	COR-3 TAB -15 S	I30789	NEXPRO 40MG TAB-15S	9.969057	0.230216
7	I30789	NEXPRO 40MG TAB-15S	I10106	COR-3 TAB -15 S	9.969057	0.197938
8	I54469	RECITA FORTE TAB<15S>	I01959	AMIXIDE-H TABS-10*S	9.822775	0.188953
9	I01959	AMIXIDE-H TABS-10*S	I54469	RECITA FORTE TAB<15S>	9.822775	0.321782
10	I12616	D10 ML DISPOVAN SYRINGE	I12600	D5 DISPOVAN SYRINGE-5ML	9.185099	0.165316
11	I12600	D5 DISPOVAN SYRINGE-5ML	I12616	D10 ML DISPOVAN SYRINGE	9.185099	0.269841
12	I09315	CLOPILET TAB $-15'S	I13809	ECOSPRIN 150MG TAB	8.794045	0.243697
13	I13809	ECOSPRIN 150MG TAB	I09315	CLOPILET TAB $-15'S	8.794045	0.149485
14	I42167	NS100ML SODIUM CHLORIDE	I12600	D5 DISPOVAN SYRINGE-5ML	7.576479	0.136364
15	I12600	D5 DISPOVAN SYRINGE-5ML	I42167	NS100ML SODIUM CHLORIDE	7.576479	0.190476
16	I30831	NICARDIA RETARD 20MG 15S	I02972	ARKAMIN 100 MCG	7.174720	0.163636
17	I02972	ARKAMIN 100 MCG	I30831	NICARDIA RETARD 20MG 15S	7.174720	0.169102
18	I41620	SHELCAL HD TAB 15 S	I24977	LIVOGEN Z TAB-15S	6.955955	0.136456
19	I24977	LIVOGEN Z TAB-15S	I41620	SHELCAL HD TAB 15 S	6.955955	0.162621
20	I15709	EVION 400MG-10	I24649	LIMCEE 500 MG TAB	6.356538	0.135593

Results

The Association Rules give us a list of related items.

These can be used in different scenarios by the stakeholder - for marketing, operations, stocking, etc

	antecedents_id	antecedents_name	consequents_id	consequents_name	lift	confidence
0	I12616	D10 ML DISPOVAN SYRINGE	I42167	NS100ML SODIUM CHLORIDE	17.599657	0.442464
1	I42167	NS100ML SODIUM CHLORIDE	I12616	D10 ML DISPOVAN SYRINGE	17.599657	0.517045
2	I10106	COR-3 TAB -15 S	I24977	LIVOGEN Z TAB-15S	13.446835	0.263789
3	I24977	LIVOGEN Z TAB-15S	I10106	COR-3 TAB -15 S	13.446835	0.266990
4	I30789	NEXPRO 40MG TAB-15S	I24977	LIVOGEN Z TAB-15S	9.984936	0.195876
5	I24977	LIVOGEN Z TAB-15S	I30789	NEXPRO 40MG TAB-15S	9.984936	0.230583
6	I10106	COR-3 TAB -15 S	I30789	NEXPRO 40MG TAB-15S	9.969057	0.230216
7	I30789	NEXPRO 40MG TAB-15S	I10106	COR-3 TAB -15 S	9.969057	0.197938
8	I54469	RECITA FORTE TAB<15S>	I01959	AMIXIDE-H TABS-10*S	9.822775	0.188953
9	I01959	AMIXIDE-H TABS-10*S	I54469	RECITA FORTE TAB<15S>	9.822775	0.321782
10	I12616	D10 ML DISPOVAN SYRINGE	I12600	D5 DISPOVAN SYRINGE-5ML	9.185099	0.165316
11	I12600	D5 DISPOVAN SYRINGE-5ML	I12616	D10 ML DISPOVAN SYRINGE	9.185099	0.269841
12	I09315	CLOPILET TAB $-15'S	I13809	ECOSPRIN 150MG TAB	8.794045	0.243697
13	I13809	ECOSPRIN 150MG TAB	I09315	CLOPILET TAB $-15'S	8.794045	0.149485
14	I42167	NS100ML SODIUM CHLORIDE	I12600	D5 DISPOVAN SYRINGE-5ML	7.576479	0.136364
15	I12600	D5 DISPOVAN SYRINGE-5ML	I42167	NS100ML SODIUM CHLORIDE	7.576479	0.190476
16	I30831	NICARDIA RETARD 20MG 15S	I02972	ARKAMIN 100 MCG	7.174720	0.163636
17	I02972	ARKAMIN 100 MCG	I30831	NICARDIA RETARD 20MG 15S	7.174720	0.169102
18	I41620	SHELCAL HD TAB 15 S	I24977	LIVOGEN Z TAB-15S	6.955955	0.136456
19	I24977	LIVOGEN Z TAB-15S	I41620	SHELCAL HD TAB 15 S	6.955955	0.162621
20	I15709	EVION 400MG-10	I24649	LIMCEE 500 MG TAB	6.356538	0.135593

PROJECT IMPLEMENTATION & IMPROVEMENT IDEAS

Bundle frequent itemsets
Sell the obtained frequent item sets together

Market
Market the prescription not needed items

Categorize
Categorize items for cleaner results

Use Customer ID
Utilizing customer ID might help us correctly identify the basket

Increase Frequent Itemset
Increase the size of frequent item set

THANK YOU!