**Market Basket Analysis**

INTRODUCTION

# The retailer wants to target customers with suggestions on itemset that a customer is most likely to purchase .I was given dataset contains data of a retailer; the transaction data provides data around all the transactions that have happened over a period of time. Retailer will use result to grove in his industry and provide for customer suggestions on itemset, we be able increase customer engagement and improve customer experience and identify customer behavior. I will solve this problem with use Association Rules type of unsupervised learning technique that checks for the dependency of one data item on another data item.

In this phase the design to innovation and data flow of market business analysis is going to be done.

DATASET

The data is obtained from <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

COLUMNS USED

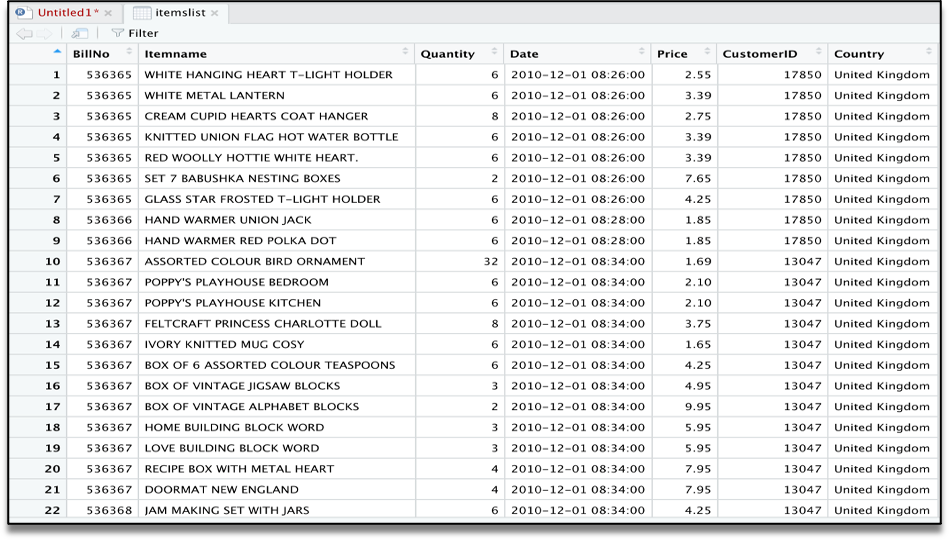
From Assignment-1\_Data data the following columns are used

# Bill no

* Item name
* Quantity
* Date
* Price
* Customer id
* Country

Data Pre-processing

Next, we need to upload Assignment-1\_Data. xlsx to R to read the dataset.Now we can see our data in R.

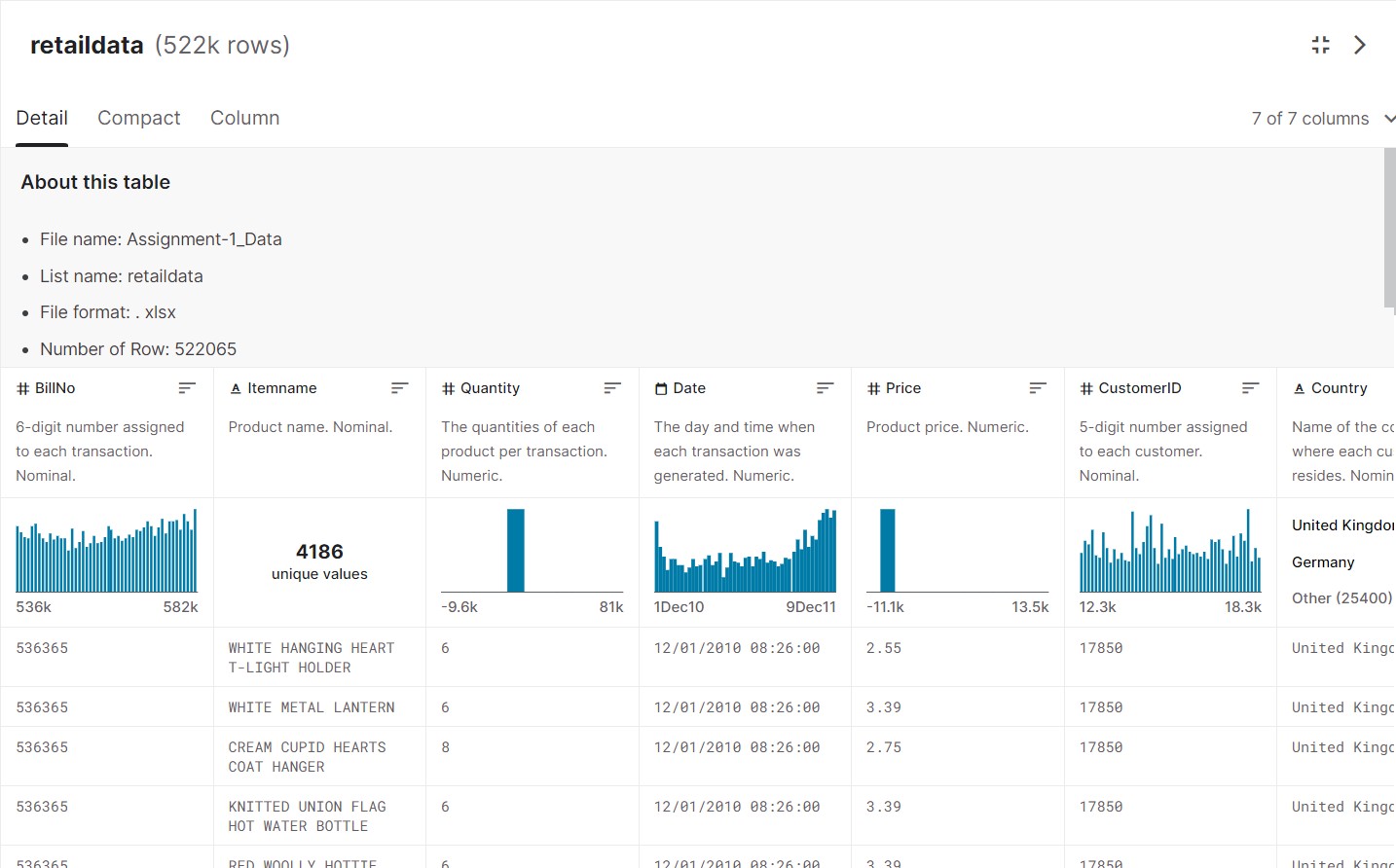


# The summary gives us some useful information:

* Density tells the percentage of non-zero cells in a sparse matrix. In other words, total number of items that are purchased divided by a possible number of items in that matrix. You can calculate how many items were purchased by using density: 18193x7698x0.002291294=337445
* Summary will show us most frequent items.
* Element (itemset/transaction) length distribution: It will gave us how many transactions are there for 1-itemset, 2-itemset and so on. The first row is telling you a number of items and the second row is telling you the number of transactions.

For example, there is only 1546 transaction for one item, 860 transactions for 2 items, and there are 419 items in one transaction which is the longest.

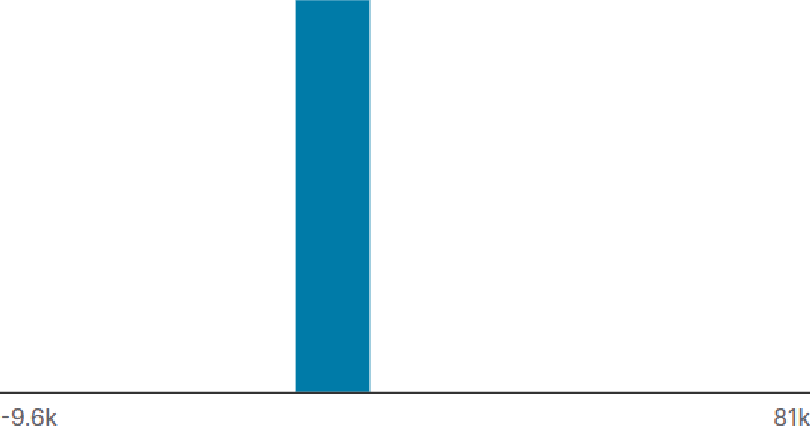
* **LOAD THE TRANSACTION DATASET AND PREPROCESS THE DATA FOR ASSOCIATION ANALYSIS.**



retaildata (522k rows)

Detail Compact Column It Quantity

Tne quantities of each product per transaction. Numeric.



0 Date

Tne day and time when eacn transaction was generated. Numeric.

retaildata (522k rows)

Deta il Compact Column 4t Price



It Custo merlD

5— digit numDer assigned to each customer. Nomi nal.

Valio B Mismatches B MissingB

Mean

Std. Deviation Quantiles



›

7 of 7 columns w

5ZZk 100%

0 0%

0 0%

10.1

161

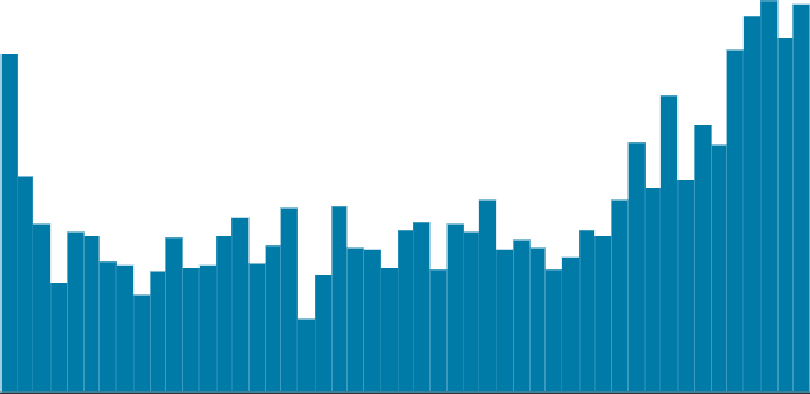


1 25%

3

10 75%

B1k Max



Valio B Mismatches B

MissingB

522k 100%

0 0%

0 OF

Minimum Mean

Maximum

1Dec10 4JuI11

9 Dec11



7 of7columns

retaildata (522k rows)

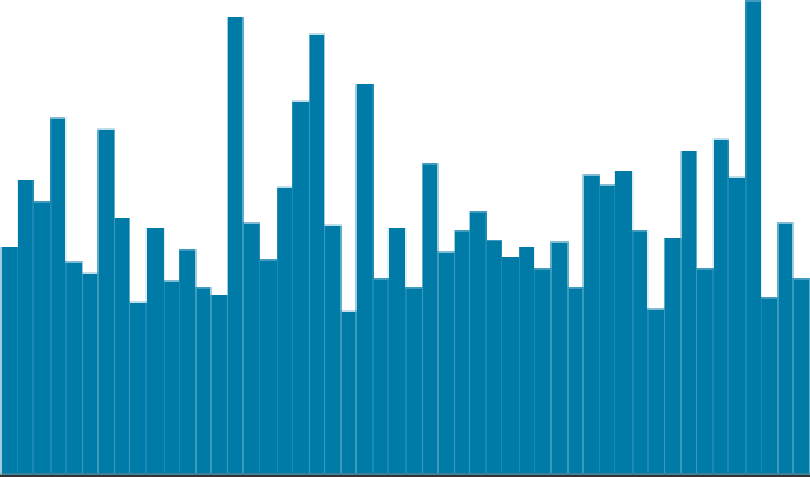
Detail Compact Column

:: ›

7 of 7 columns v

-11.1k

It CustomerlD

5-digit number assigned to each customer. Nominal.

12.3k

13.5k

18.3k

Valid E Mismatched a

N

Mean

Std. Deviation Quantiles

2.08 50%

4.13 75%

13.5k Max

388k 74%

0 0%

134k 26%

15.3k

1.72k

12.3k Min 25%

16.8k 75%

* 1. k Max

\* Country

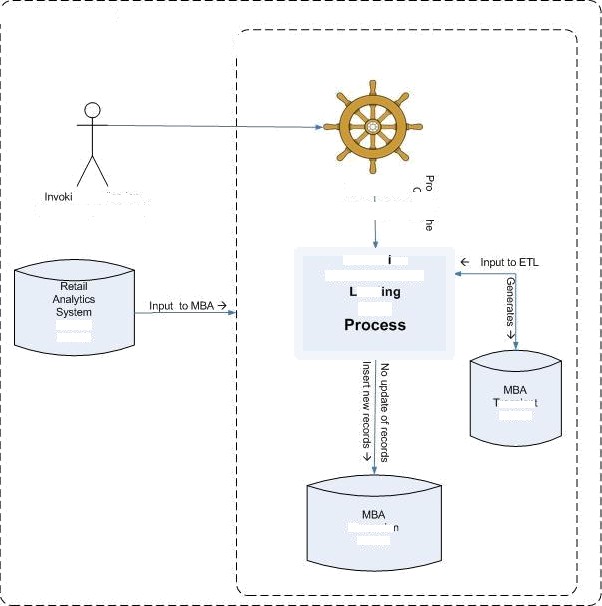
Name of the country where each customer resides. Nominal.

United Kingdom  Germany

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | 522k | 100% |
| MismatchedB | 0 | 0% |
| \*\* | Missing N | 0 | 0% |
|  | Unique  Most Common | 30  United King... | 93% |

Other (25400)

MBA ETL Process flow Diagram



Retail Analytics

Market Basket Analysis Schema

{Retail Analytla Bafct\)

Service Manager

0ou Tables

rarsien

tables

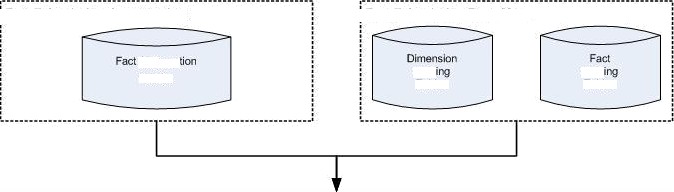
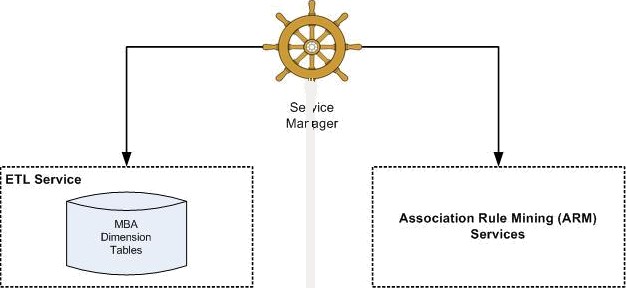
Ulmanslo Tables

Extract on

Transformation

(ETL)

Ratail Analytics DATABASE



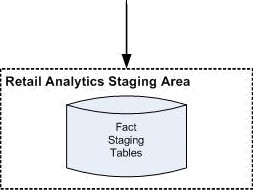
, Retail Analytics Aggregation

, Retail Analytics Data Mart

Tablea

Stag Tables

Stag TBbles



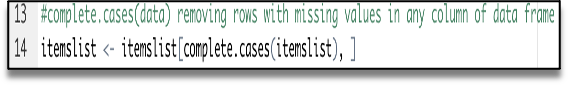
Market Basket Analysis Database Schema

I"""""""""""""”""""""""""""".

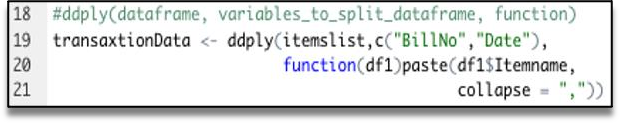
, t

Promotion Baseline (BL) Services

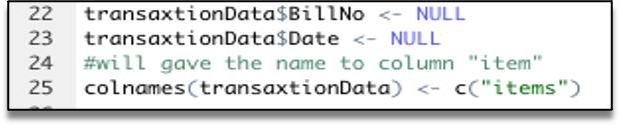
After we will clear our data frame, will remove missing values.



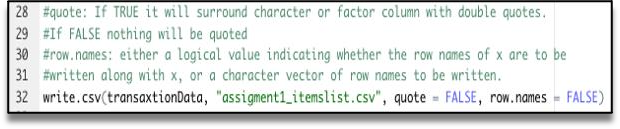
To apply Association Rule mining, we need to convert dataframe into transaction data to make all items that are bought together in one invoice will be in one row. Below lines of code will combine all products from one BillNo and Date and combine all products from that BillNo and Date as one row, with each item, separated by (,)



We don’t need BillNo and Date, we will make it as Null. Next, you have to store this transaction data into .csv



This how should look transaction data before we will go to next step.





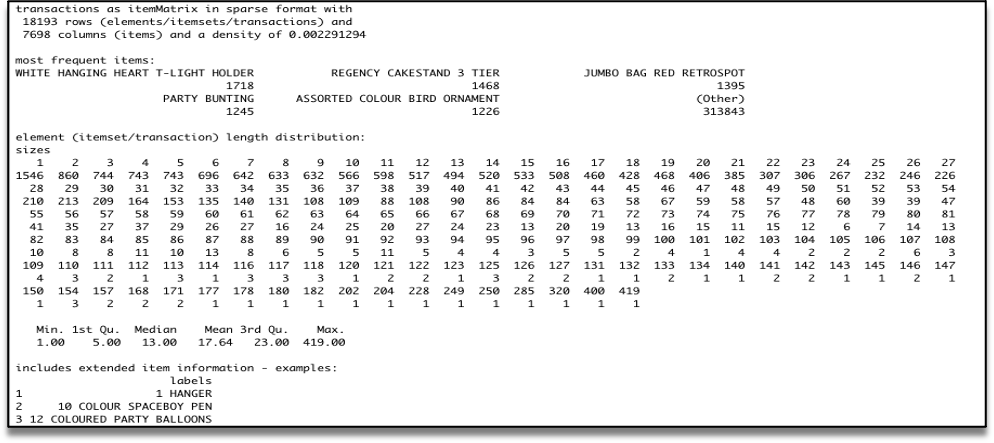
At this step we already have our transaction dataset, and it shows the matrix of items which bought together. We can’t see here any rules and how often it was purchase together. Now let’s check how many transactions we have and what they are. We will have to have to load this transaction data into an object of the transaction class. This is done by using the R function read.transactions of the arules package. Our format of Data frame is basket.

image

Let’s have a view our transaction object by summary(transaction)

image

We can see 18193 transactions (rows) and 7698 items (columns). 7698 is the product descriptions and 18193 transactions are collections of these items.



The summary gives us some useful information:

* + - Density tells the percentage of non-zero cells in a sparse matrix. In other words, total number of items that are purchased divided by a possible number of items in that matrix. You can calculate how many items were purchased by using density: 18193x7698x0.002291294=337445
    - Summary will show us most frequent items.
    - Element (itemset/transaction) length distribution: It will gave us how many transactions are there for 1- itemset, 2-itemset and so on. The first row is telling you a number of items and the second row is telling you the number of transactions.

For example, there is only 1546 transaction for one item, 860 transactions for 2 items, and there are 419 items in one transaction which is the longest.

Let’s check item frequency plot, we will generate an itemFrequencyPlot to create an item Frequency Bar Plot to view the distribution of objects based on itemMatrix (e.g., >transactions or items in >itemsets and >rules) which is our case.

