Online Retail

Ghadi Khalid

AIM OF PROJECT

- **Extract insights and provide it to decision-makers to improve company marketing and increase sales.**
- **Showcase benefit of using a association rules to the company.**

RESEARCH QUESTIONS

COUNTRY	QUANTITY	PROFIT	PRODUCTS
How many customers from different country , dose profit change?	How dose quantity trend change based on date?	How was the profit for this year based on months and days?	What is most sold products in the store?

I addressed more question in the report it's written with **bold** font

DATA SOURCE

This analysis is on Online_Retail_II dataset provided by **UCI**

- The dataset contain transactions occurring for a UK-based and registered online shop
- The store mainly sells unique all-occasion gift-ware

DATA COLUMNS

Variable	Description
InvoiceNo	6-digit integral number uniquely assigned to each transaction.
	If this code starts with the letter ‘c’, it indicates a cancellation
StockCode	Product (item) code
Description	Product (item) name
Quantity	quantities of each product (item) per transaction
InvoiceDate	Invice date and time
UnitPrice	Product price per unit in sterling (Â£)
CustomerID	Customer number/Id
Country	country where a customer resides

01

CLEAN DATA

Remove NA ,Negative numbers and fix data structure

02

EDA

Explor data and analysis profit of store.

03

INTERACTIVE VISUALIZATION

Provide interactive plot instead of dashboard

04

MARKET BASKET ANALYSIS

Provide most bought products

01

CLEAN DATA

CHALLENGE

- Negative values
- NA values
- Wrong columns type

RESULT

- Change negative values to 0
- Keep rows with NA values
- Change column type

```
[1] 525461      8
[1] TRUE
    Invoice   StockCode Description    Quantity InvoiceDate        Price Customer ID      Country
          0           0        2928           0           0           0      107927            0
```

I tried to use impute to replace Description NA but failed

```
tibble [525,461 x 9] (S3: tbl_df/tbl/data.frame)
 $ Invoice        : num [1:525461] 489434 489434 489434 489434 489434 ...
 $ StockCode      : chr [1:525461] "85048" "79323P" "79323W" "22041" ...
 $ Description    : chr [1:525461] "15CM CHRISTMAS GLASS BALL 20 LIGHTS" "PINK CHERRY LIGHTS" "WHITE C
 $ Quantity       : num [1:525461] 12 12 12 48 24 24 24 10 12 12 ...
 $ InvoiceDate    : POSIXct[1:525461], format: "2009-12-01 07:45:00" "2009-12-01 07:45:00" "2009-12-01
 $ Price          : num [1:525461] 6.95 6.75 6.75 2.1 1.25 1.65 1.25 5.95 2.55 3.75 ...
 $ Customer ID    : num [1:525461] 13085 13085 13085 13085 13085 ...
 $ Country        : Factor w/ 40 levels "Australia","Austria",..: 37 37 37 37 37 37 37 37 37 37 ...
 $ All_Units_Price: num [1:525461] 83.4 81 81 100.8 30 ...
```

02

EDA

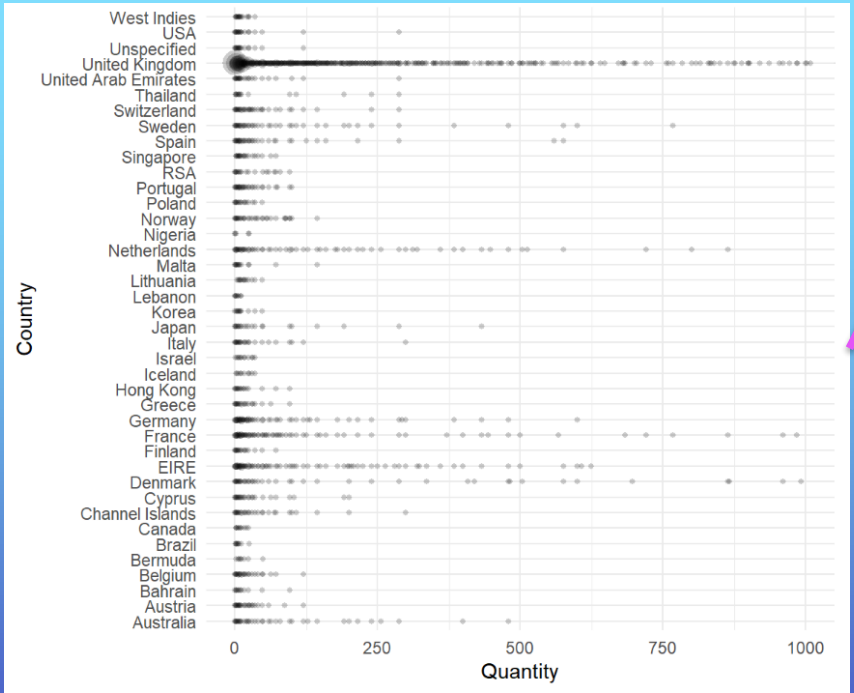
CHALLENGE

- Dataset Size [525,461 rows]
- Combine Invoices

RESULT

- Create sample to test code first
- Create different variables grouped by Invoices

What is total orders for each country with total quantity?



Show 10 entries Search:

	Country	No.Invoice	Total_Quantity
1	Australia	654	20189
2	Austria	537	6697
3	Bahrain	107	1025
4	Belgium	1054	12075
5	Bermuda	34	2798
6	Brazil	62	189
7	Canada	77	894
8	Channel Islands	906	11991
9	Cyprus	554	4625
10	Denmark	428	229690

Showing 1 to 10 of 40 entries Previous 1 2 3 4 Next

What is total number of customers and total sales for each country?

Show 10 entries Search:

	Country	Number_of_Customers	Total_Sales
1	Australia	654	31446.8
2	Austria	537	13414.33
3	Bahrain	107	2355.65
4	Belgium	1054	24557.08
5	Bermuda	34	1253.14
6	Brazil	62	268.27
7	Canada	77	1216.66
8	Channel Islands	906	24546.32
9	Cyprus	554	11389.75
10	Denmark	428	50906.85

Showing 1 to 10 of 40 entries Previous 1 2 3 4 Next

How many Invoices in data?

Here is each Invoices with it total quantity and price

Show 10 ▼ entries

Search:

	Invoice	Total_Quantity	Total_Price
1	489434	166	505.3
2	489435	60	145.8
3	489436	193	630.33
4	489437	145	310.75
5	489438	826	2286.24
6	489439	219	426.3
7	489440	16	50.4
8	489441	102	344.34
9	489442	275	382.37
10	489443	120	285.06

Showing 1 to 10 of 24,222 entries

Previous 1 2 3 4 5 … 2423 Next

03

INTERACTIVE VISUALIZATION

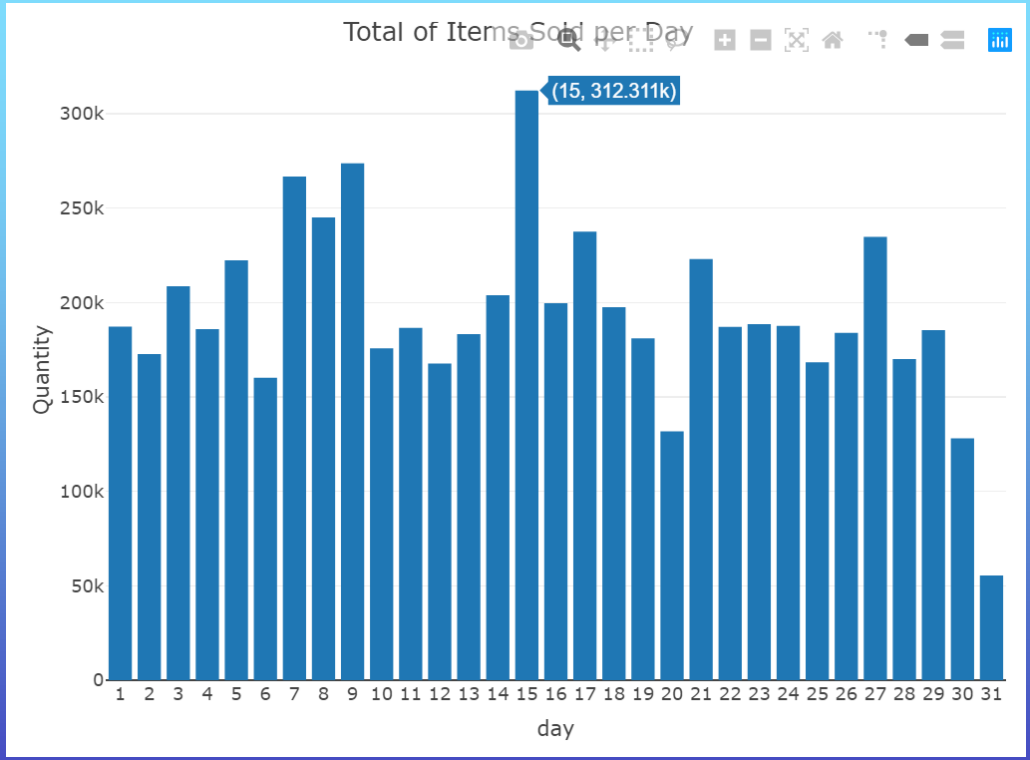
CHALLENGE

- Dashboard
- Prepare data to plot
- Try different packages[highcharter]

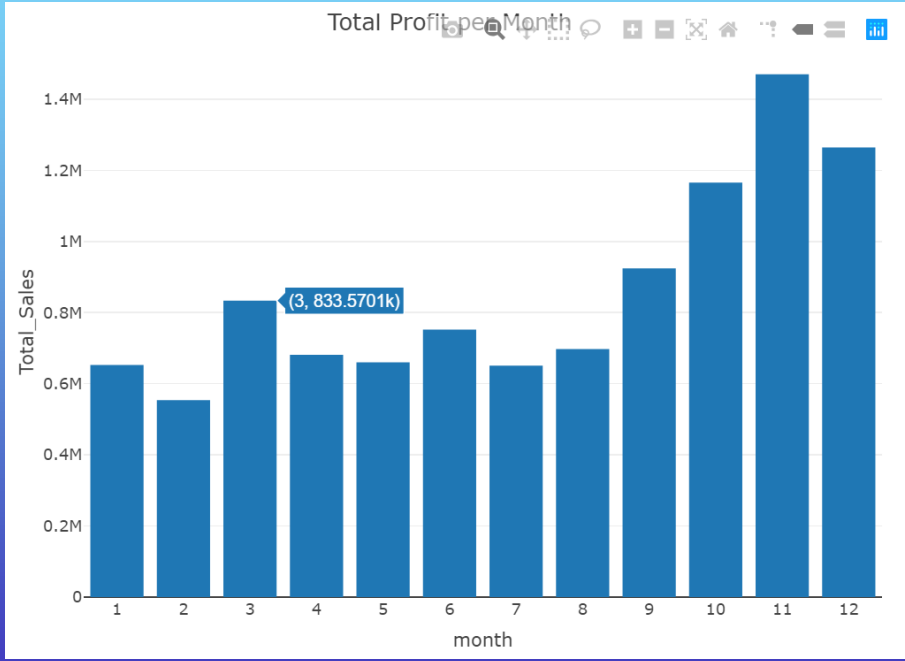
RESULT

- Instead of Dashboard I used interactive plot packages & included it in report
- Make different variables that I wanted to plot
- Used plotly package

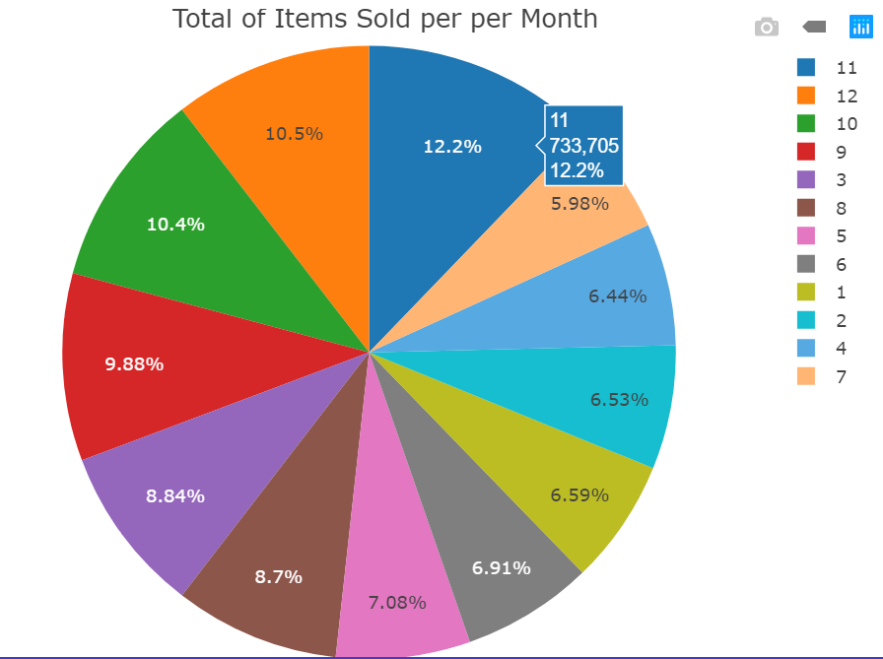
Total items sold per day?



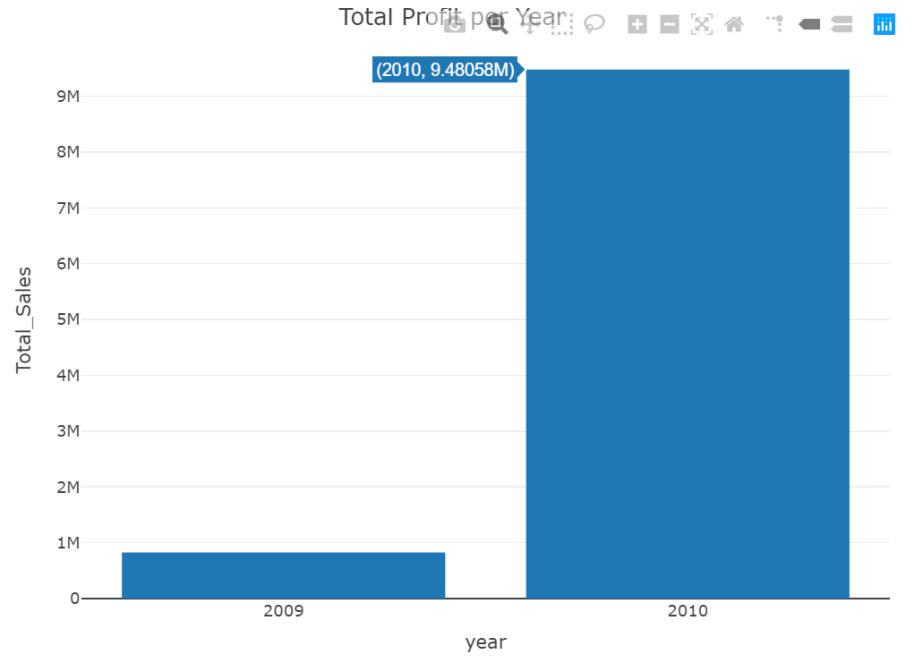
Monthly Performance of Store[Total Profit]



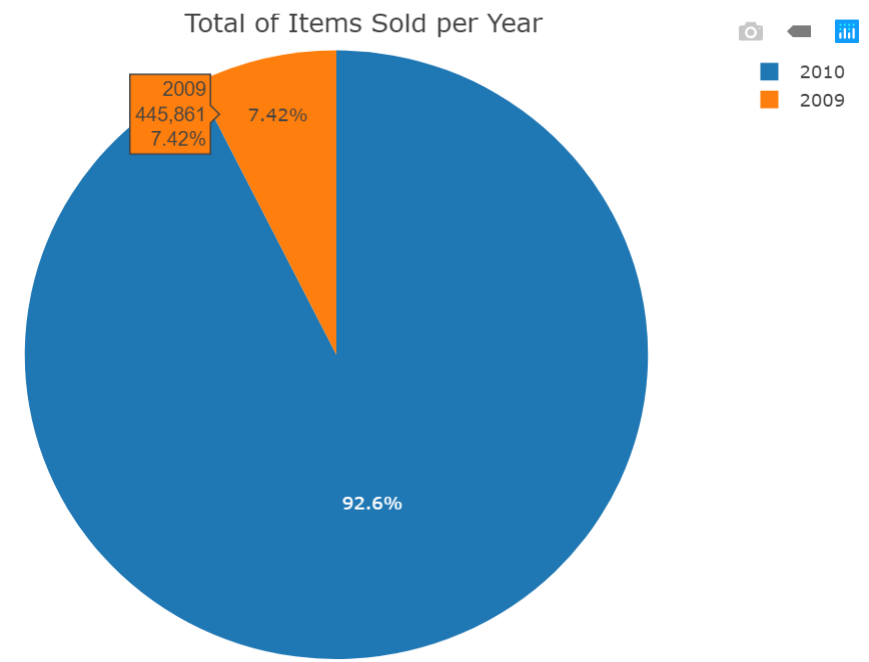
What is the amount of items soled every month?



What is total profit per year?



What is total items sold per year?



04

MARKET BASKET ANALYSIS

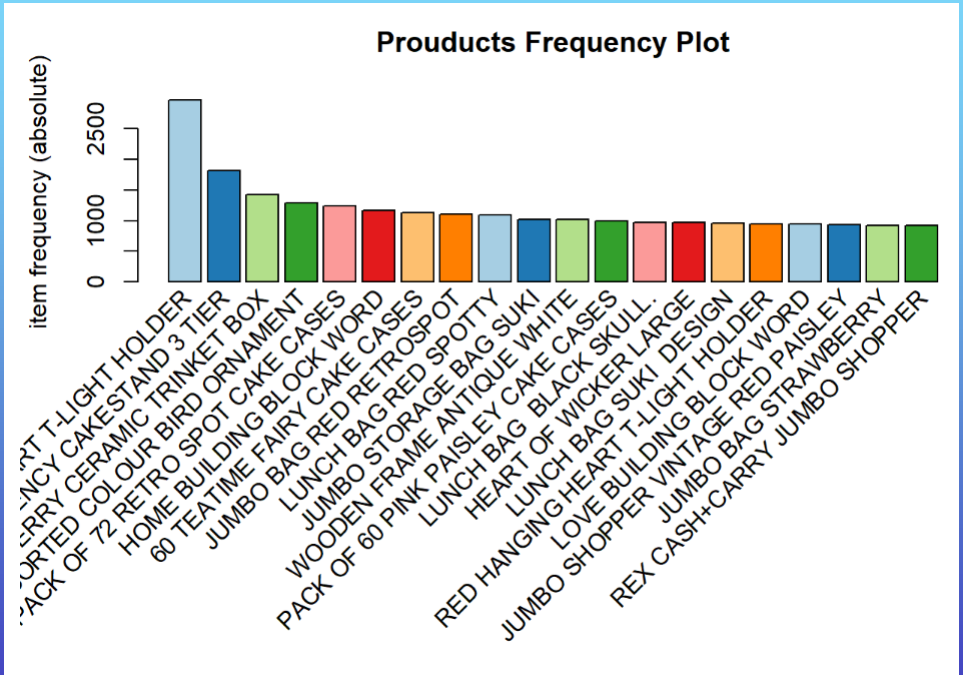
CHALLENGE

- Understand MBA
- Visualize results so that it’s easy to understand

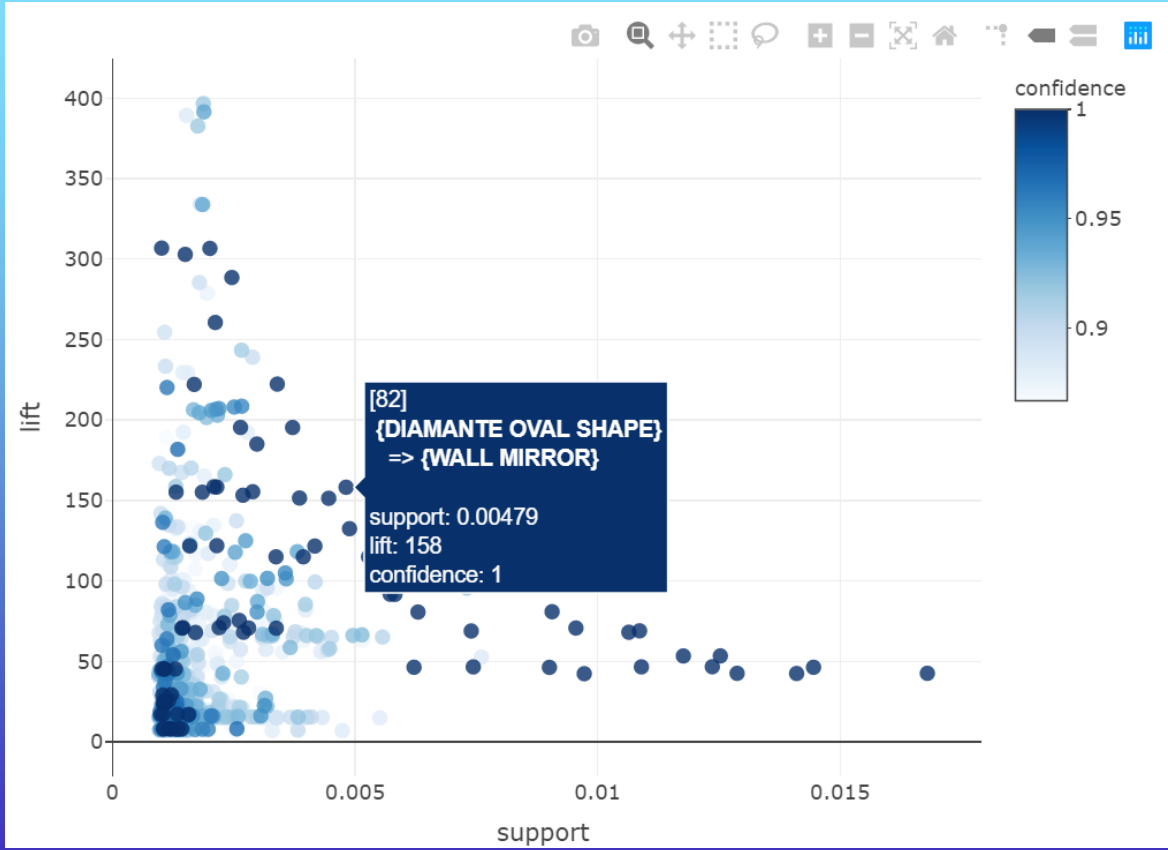
RESULT

- Extract popular products
- Extract Association Rules
- Change column type

What is top 20 sold items in store?

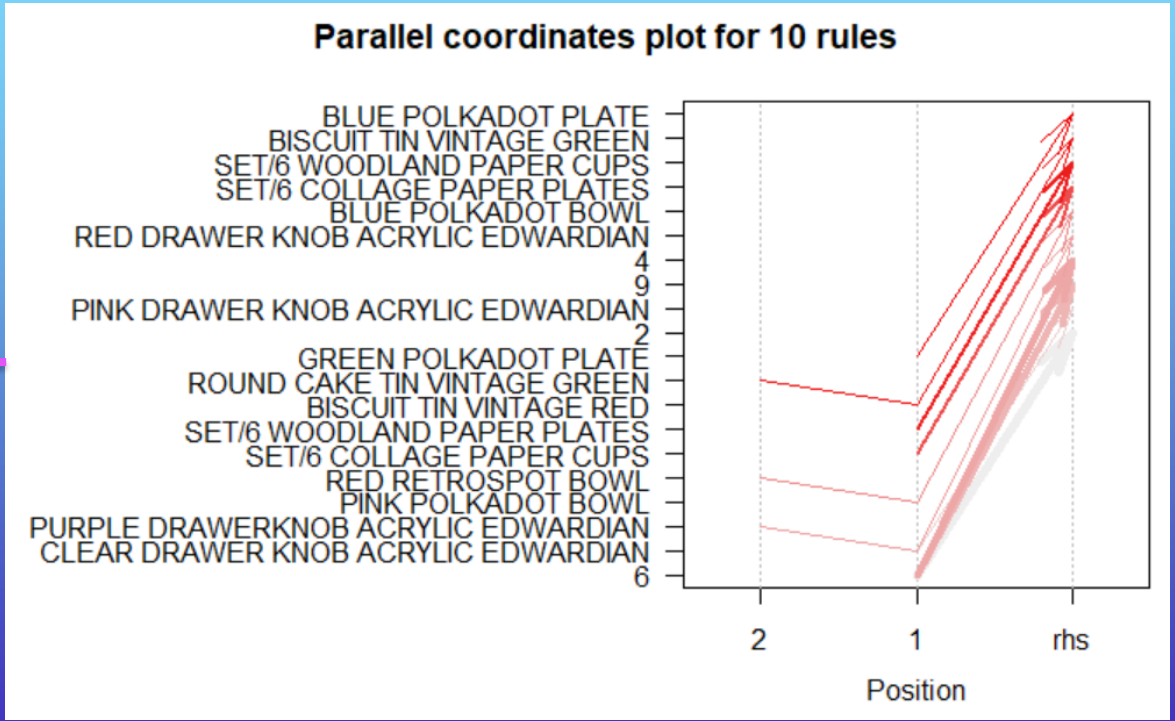


What is This best 1000 rules?

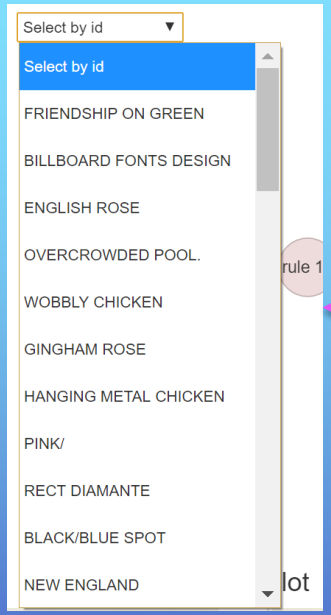


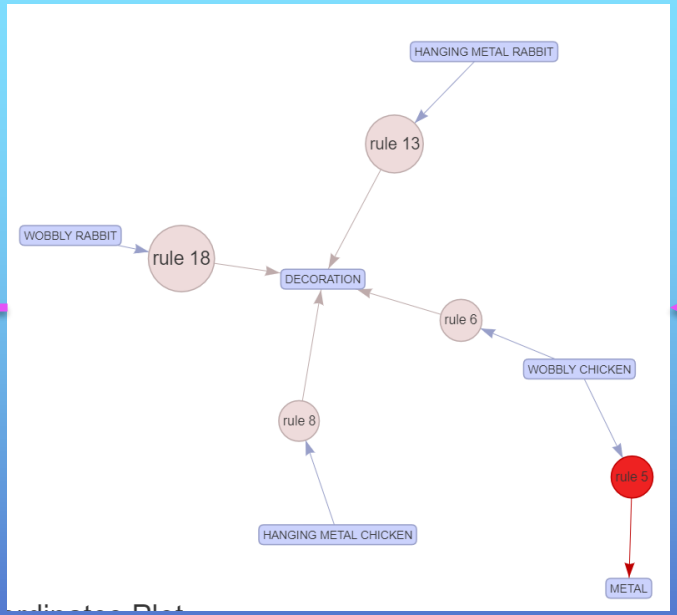
The darker the color the stronger the rule become.
ROUND CAKE TIN VINTAGE GREEN has rule with BISCUIT TIN VINTAGE GREEN.
o, who buy ROUND CAKE TIN VINTAGE GREEN will definitely buy ROUND CAKE TIN VINTAGE GREEN as well.

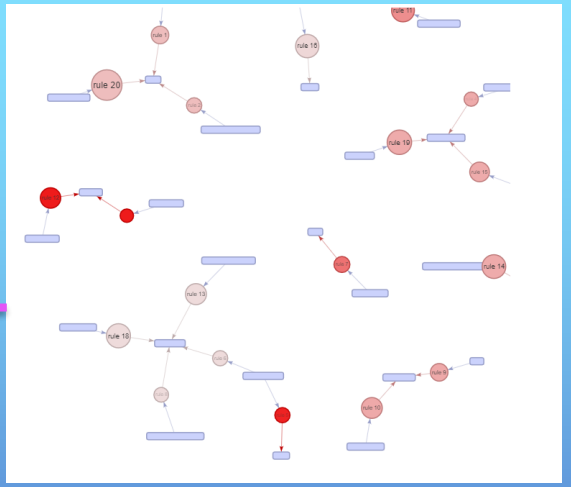
I Tried this but the result kept on change & have the number shown
So it's fail



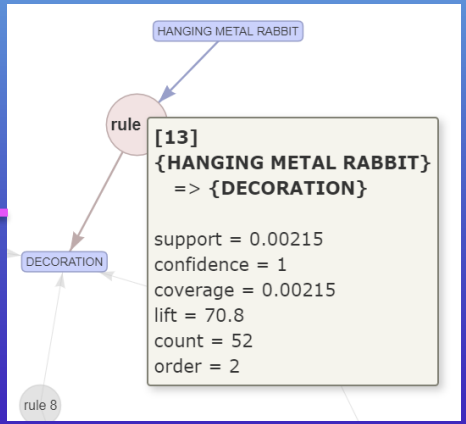
Showcase top 20 rules extracted from data?







Show all rule details



CONCLUSION

This report is to provide an explanation for small/medium companies for why & how can data analysis help provide insights to benefit the business.

FUTURE DEVELOPMENT

- Customer Behavior Analysis
- Save Interactive Plots
- Recommendation Engine
- Create website to showcase the association rules for products

APPLICABILITY

This analysis can be done on any store transaction

THANK YOU