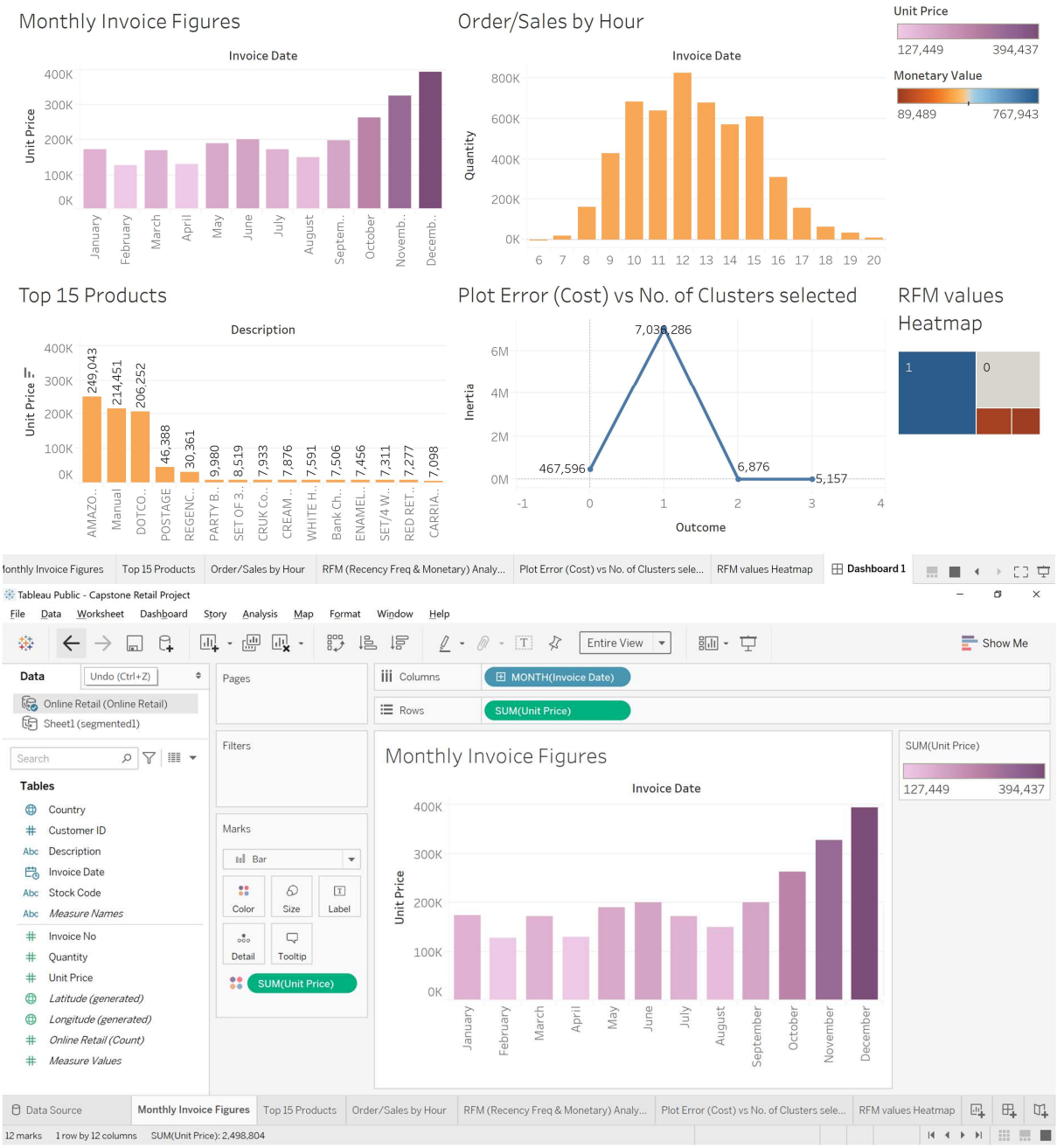
Retail Project

CountMeIn (github.com)

All activities have been completed and submitted as part of python code, results and code PDF and the Tableau outputs. I have attached all the outputs at the end of the file too.

All tasks that have been completed are marked as **Done**

1. **Tableau Outputs**
   Rajesh Nagarjunan - Profile | Tableau Public



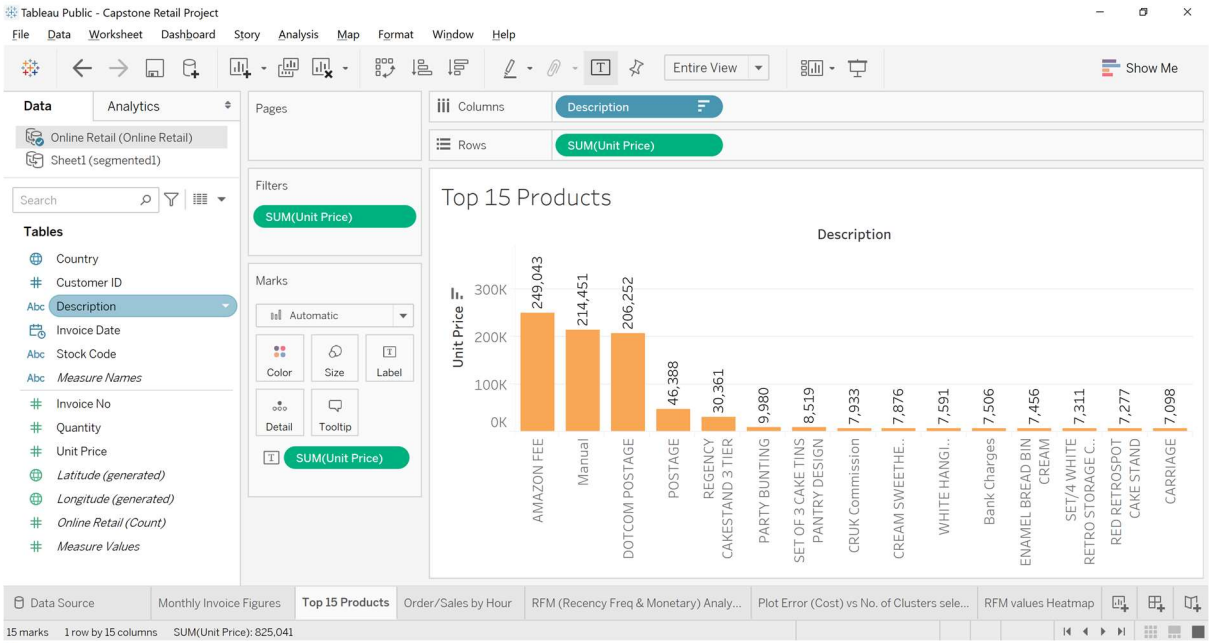


Tableau Public - Capstone Retail Project
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Entire View
Show Me
Data
Analytics
Online Retail (Online Retail)
Sheet1 (segmented1)
Search
Tables
Country
Customer ID
Description
Invoice Date
Stock Code
Measure Names
Invoice No
Quantity
Unit Price
Latitude (generated)
Longitude (generated)
Online Retail (Count)
Measure Values
Pages
Filters
SUM(Unit Price)
Marks
Automatic
Color
Size
Label
Detail
Tooltip
SUM(Unit Price)
Columns
Description
Rows
SUM(Unit Price)
Top 15 Products
Description
Unit Price
300K
200K
100K
0K
249,043
214,451
206,252
46,388
30,361
9,980
8,519
7,933
7,876
7,591
7,506
7,456
7,311
7,277
7,098
AMAZON FEE
Manual
DOTCOM POSTAGE
POSTAGE
REGENCY CAKESTAND 3 TIER
PARTY BUNTING
SET OF 3 CAKE TINS PANTRY DESIGN
CRUK Commission
CREAM SWEETHE..
WHITE HANGI..
Bank Charges
ENAMEL BREAD BIN CREAM
SET/4 WHITE RETRO STORAGE C..
RED RETROSPOT CAKE STAND
CARRIAGE
Data Source
Monthly Invoice Figures
Top 15 Products
Order/Sales by Hour
RFM (Recency Freq & Monetary) Analy...
Plot Error (Cost) vs No. of Clusters sele...
RFM values Heatmap
15 marks
1 row by 15 columns
SUM(Unit Price): 825,041

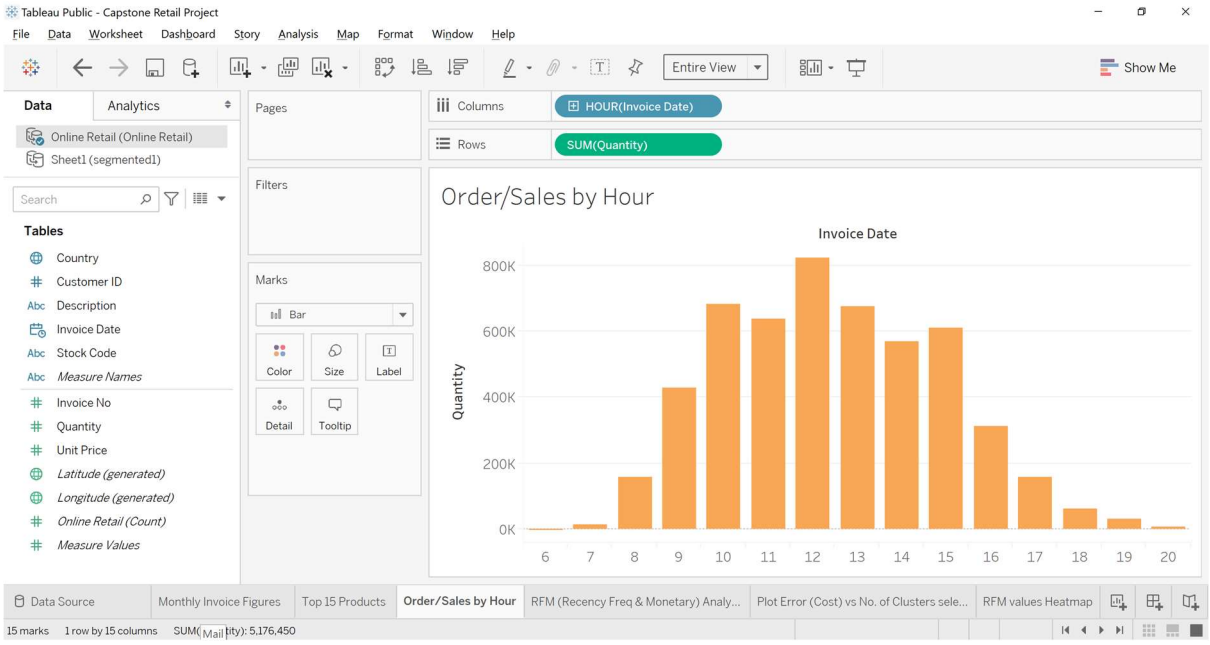


Tableau Public - Capstone Retail Project
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Entire View
Show Me
Data
Analytics
Online Retail (Online Retail)
Sheet1 (segmented1)
Search
Tables
Country
Customer ID
Description
Invoice Date
Stock Code
Measure Names
Invoice No
Quantity
Unit Price
Latitude (generated)
Longitude (generated)
Online Retail (Count)
Measure Values
Pages
Filters
Marks
Bar
Color
Size
Label
Detail
Tooltip
Columns
HOUR(Invoice Date)
Rows
SUM(Quantity)
Order/Sales by Hour
Invoice Date
Quantity
800K
600K
400K
200K
0K
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Data Source
Monthly Invoice Figures
Top 15 Products
Order/Sales by Hour
RFM (Recency Freq & Monetary) Analy...
Plot Error (Cost) vs No. of Clusters sele...
RFM values Heatmap
15 marks
1 row by 15 columns
SUM(Quantity): 5,176,450

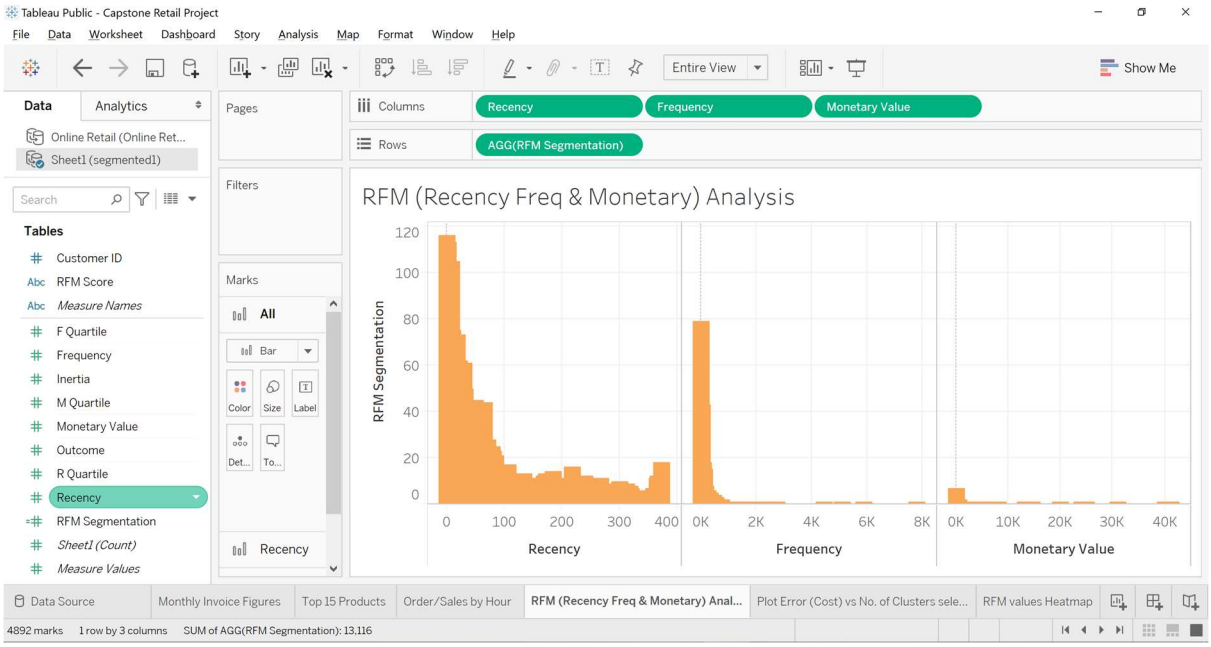


Tableau Public - Capstone Retail Project
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Entire View
Show Me
Data
Analytics
Online Retail (Online Ret...
Sheet1 (segmented1)
Search
Tables
Customer ID
RFM Score
Measure Names
F Quartile
Frequency
Inertia
M Quartile
Monetary Value
Outcome
R Quartile
Recency
RFM Segmentation
Sheet1 (Count)
Measure Values
Pages
Filters
Marks
All
Bar
Color
Size
Label
Det...
To...
Recency
Columns
Recency
Frequency
Monetary Value
Rows
AGG(RFM Segmentation)
RFM (Recency Freq & Monetary) Analysis
RFM Segmentation
120
100
80
60
40
20
0
0
100
200
300
400
Recency
0K
2K
4K
6K
8K
Frequency
0K
10K
20K
30K
40K
Monetary Value
Data Source
Monthly Invoice Figures
Top 15 Products
Order/Sales by Hour
RFM (Recency Freq & Monetary) Anal...
Plot Error (Cost) vs No. of Clusters sele...
RFM values Heatmap
4892 marks
1 row by 3 columns
SUM of AGG(RFM Segmentation): 13,116

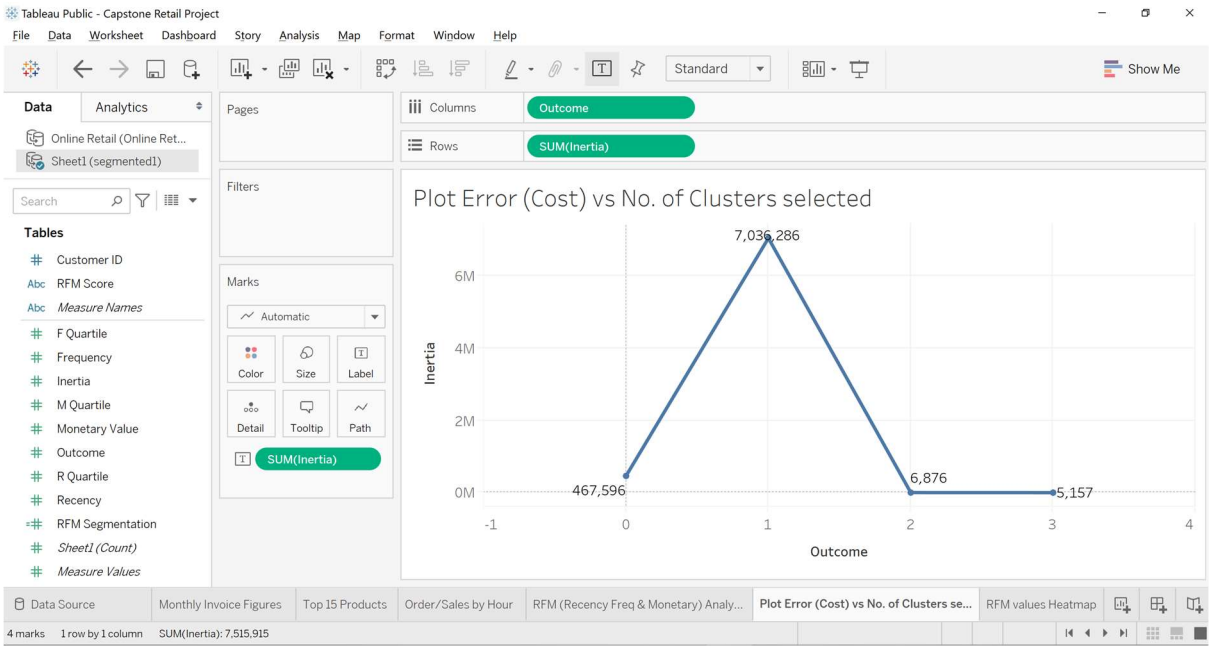
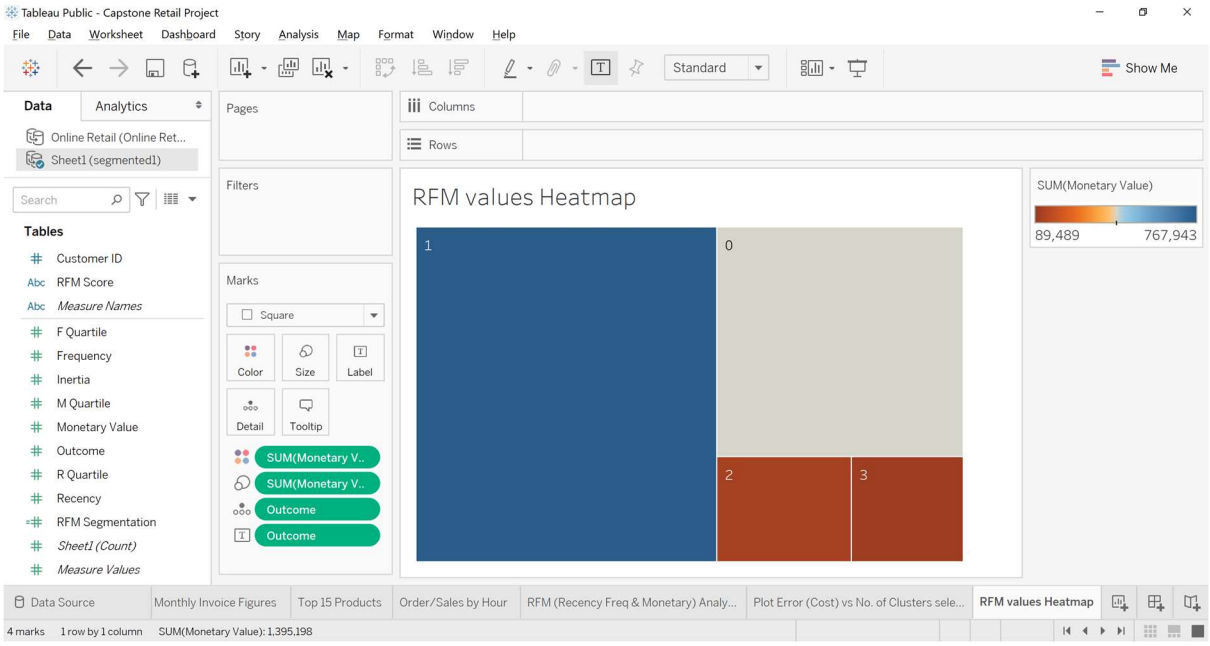


Tableau Public - Capstone Retail Project
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Standard
Show Me
Data
Analytics
Online Retail (Online Ret...
Sheet1 (segmented1)
Search
Tables
Customer ID
RFM Score
Measure Names
F Quartile
Frequency
Inertia
M Quartile
Monetary Value
Outcome
R Quartile
Recency
RFM Segmentation
Sheet1 (Count)
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Path
SUM(Inertia)
Columns
Outcome
Rows
SUM(Inertia)
Plot Error (Cost) vs No. of Clusters selected
7,036,286
6,876
5,157
467,596
6M
4M
2M
0M
Inertia
-1
0
1
2
3
4
Outcome
Data Source
Monthly Invoice Figures
Top 15 Products
Order/Sales by Hour
RFM (Recency Freq & Monetary) Analy...
Plot Error (Cost) vs No. of Clusters se...
RFM values Heatmap
4 marks
1 row by 1 column
SUM(Inertia): 7,515,915



2. Project Tasks and Completion Status

Project Task: Week 1

Data Cleaning:

1. Perform a preliminary data inspection and data cleaning. - **Done**

a. Check for missing data and formulate an apt strategy to treat them. **Done**

b. Remove duplicate data records. **Done**

c. Perform descriptive analytics on the given data. **Done**

Data Transformation:

2. Perform cohort analysis (a cohort is a group of subjects that share a defining characteristic). Observe how a cohort behaves across time and compare it to other cohorts.

a. Create month cohorts and analyze active customers for each cohort. **Done**

b. Analyze the retention rate of customers. **Done**

Project Task: Week 2

Data Modeling :

1. Build a RFM (Recency Frequency Monetary) model. Recency means the number of days since a customer made the last purchase. Frequency is the number of purchase in a given period. It could be 3 months, 6 months or 1 year. Monetary is the total amount of money a customer spent in that given period. Therefore, big spenders will be differentiated among other customers such as MVP (Minimum Viable Product) or VIP. **Done**

2. Calculate RFM metrics. **Done**

3. Build RFM Segments. Give recency, frequency, and monetary scores individually by dividing them into quartiles. **Done**

b1. Combine three ratings to get a RFM segment (as strings). **Done**

b2. Get the RFM score by adding up the three ratings. **Done**

b3. Analyze the RFM segments by summarizing them and comment on the findings. **Done**

Note: Rate “recency" for customer who has been active more recently higher than the less recent customer, because each company wants its customers to be recent. **Done**

Note: Rate “frequency" and “monetary" higher, because the company wants the customer to visit more often and spend more money

Project Task: Week 3

Data Modeling :

1. Create clusters using k-means clustering algorithm. **Done**

a. Prepare the data for the algorithm. If the data is asymmetrically distributed, manage the skewness with appropriate transformation. Standardize the data. **Done**

b. Decide the optimum number of clusters to be formed. **Done**

c. Analyze these clusters and comment on the results. **Done**

Project Task: Week 4

Data Reporting:

1. Create a dashboard in tableau by choosing appropriate chart types and metrics useful for the business. The dashboard must entail the following:

a. Country-wise analysis to demonstrate average spend. Use a bar chart to show the monthly figures **Done**

b. Bar graph of top 15 products which are mostly ordered by the users to show the number of products sold **Done**

c. Bar graph to show the count of orders vs. hours throughout the day **Done**

d. Plot the distribution of RFM values using histogram and frequency charts **Done**

e. Plot error (cost) vs. number of clusters selected **Done**

f. Visualize to compare the RFM values of the clusters using heatmap **Done**