Customer Behaviour Analysis

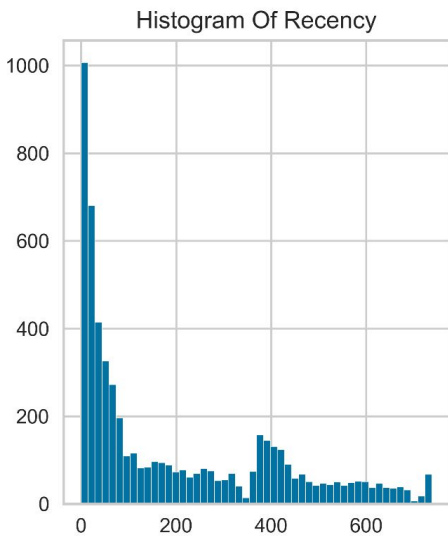## Ninos Kako
Capstone Project
Sprint 2

Customer Behaviour Analysis

- Data Cleaning

- RFM Metrics

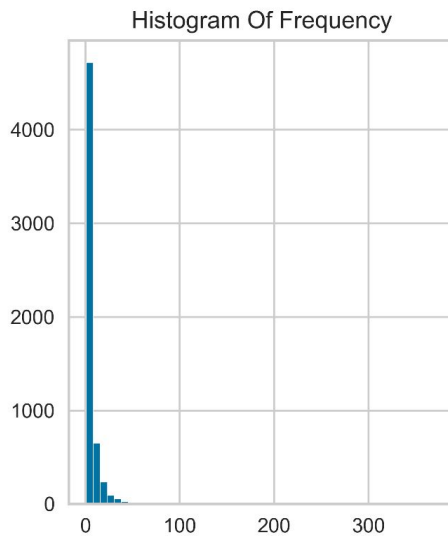- Clustering (K-means, Agglomerative, Gaussian Mixture)

Data Cleaning

- Inconsistent Item Descriptions, **67k** records.

- Customer ID, dropped 240k records.

- Dropped 'Other Charges'.

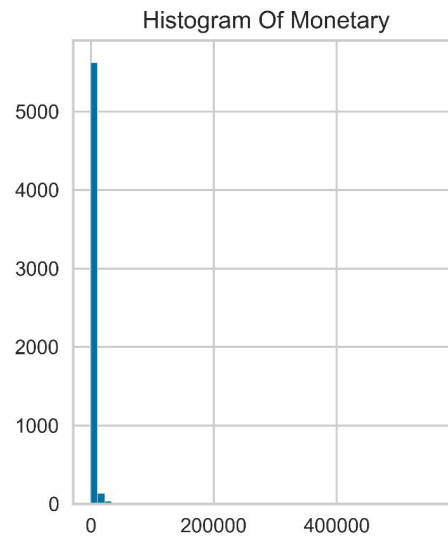- Dropped Duplicates.

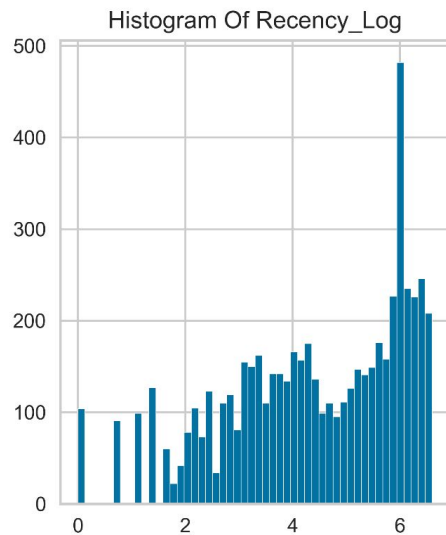- 'System Error' Duplicates.

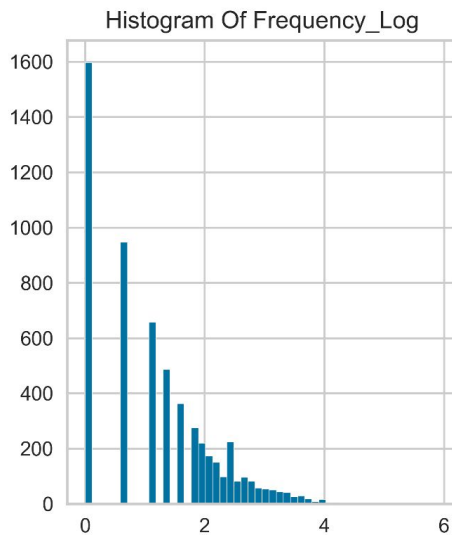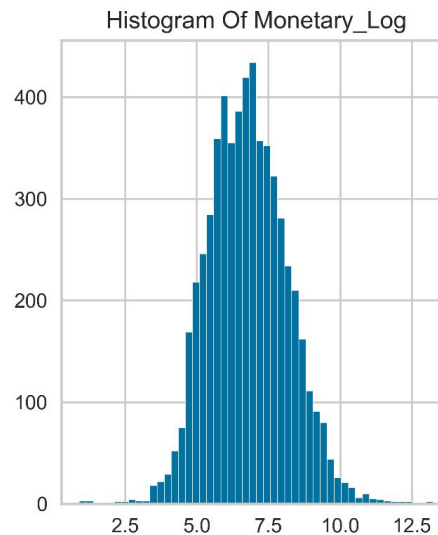# RFM Metrics

Recency   Frequency: Purchases   Monetary

Returns

RFM Metrics

Histogram Of Recency_Log   Histogram Of Frequency_Log   Histogram Of Monetary_Log

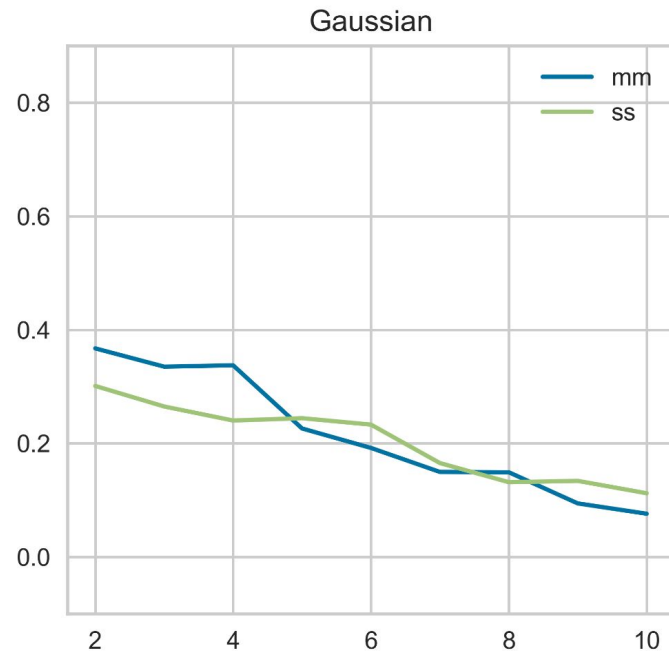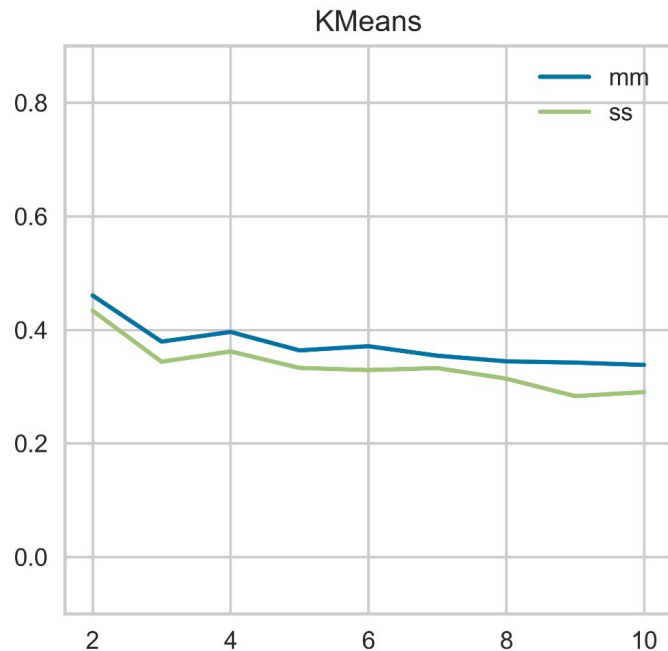Recency   Frequency: Purchases   Monetary
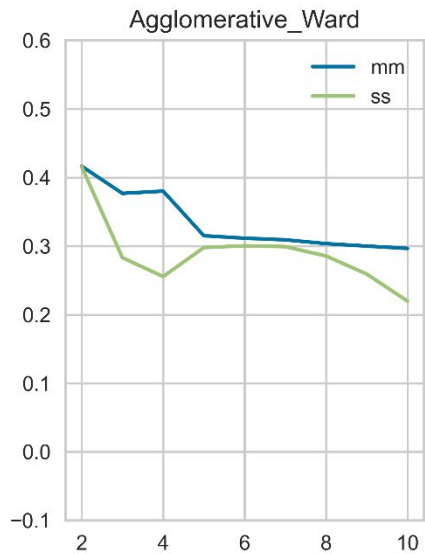
Returns

Clustering

- Scaling
  - Standard Scaler
  - MinMax Scaler

- K-means
- Gaussian Mixture Models
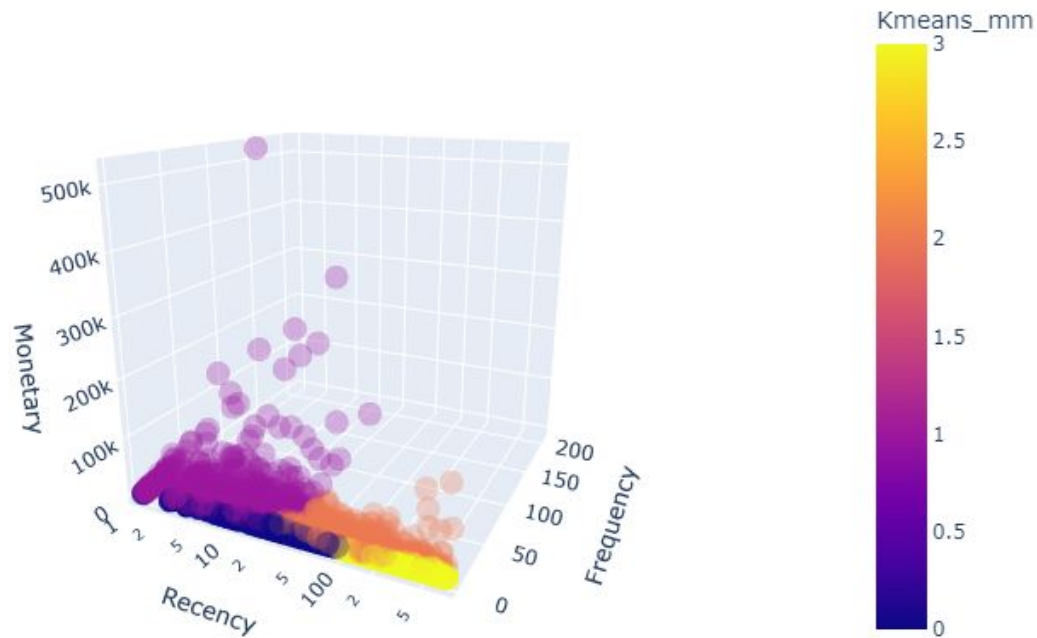- Agglomerative

K-means and Gaussian

Silhouette Scores

Agglomerative Clustering

K-means

- Normalized data
- 4 clusters

Customer Segments

Group	R	F	M	Total_Sales	Customers
New Customers	29.0	2.0	583.91	917740.00	1244
Old Infrequent	407.0	1.0	329.58	1063996.82	2183
Old Top Customers	98.0	6.0	1980.40	4031339.76	1367
Top Customers	8.0	12.0	4185.30	10341692.46	1039

Final Steps

Market Basket Analysis

- Understand customers' purchasing patterns
- Identify which items are frequently purchased together

Thank You!