
Data Science Capstone Project

Customer Segmentation

On E commerce Transaction Data

Segment E Commerce Customers

The retailer's invoice data provides insights into the groups of customers and products -

- *Segment a product into categories to enable improved product sales*
- *Segment a customer into categories to enable targeted digital marketing strategies*

Data acquisition

The European E-Commerce retailer's real life data is an intercontinental data set which contains all sales invoices occurring between 2010 and 2012. The data set contains a modest 8 attributes that contain the sales details for the retailer from various geographies.

The features in the dataset are -

- *InvoiceNo*
- *StockCode*
- *Description*
- *Quantity*
- *InvoiceDate*
- *UnitPrice*
- *CustomerID*
- *Country*

Exploratory Data Analysis

Choropleth of number of orders placed per geographic location. Europe appears to be the place from where most orders are placed.



Order nb.

15k

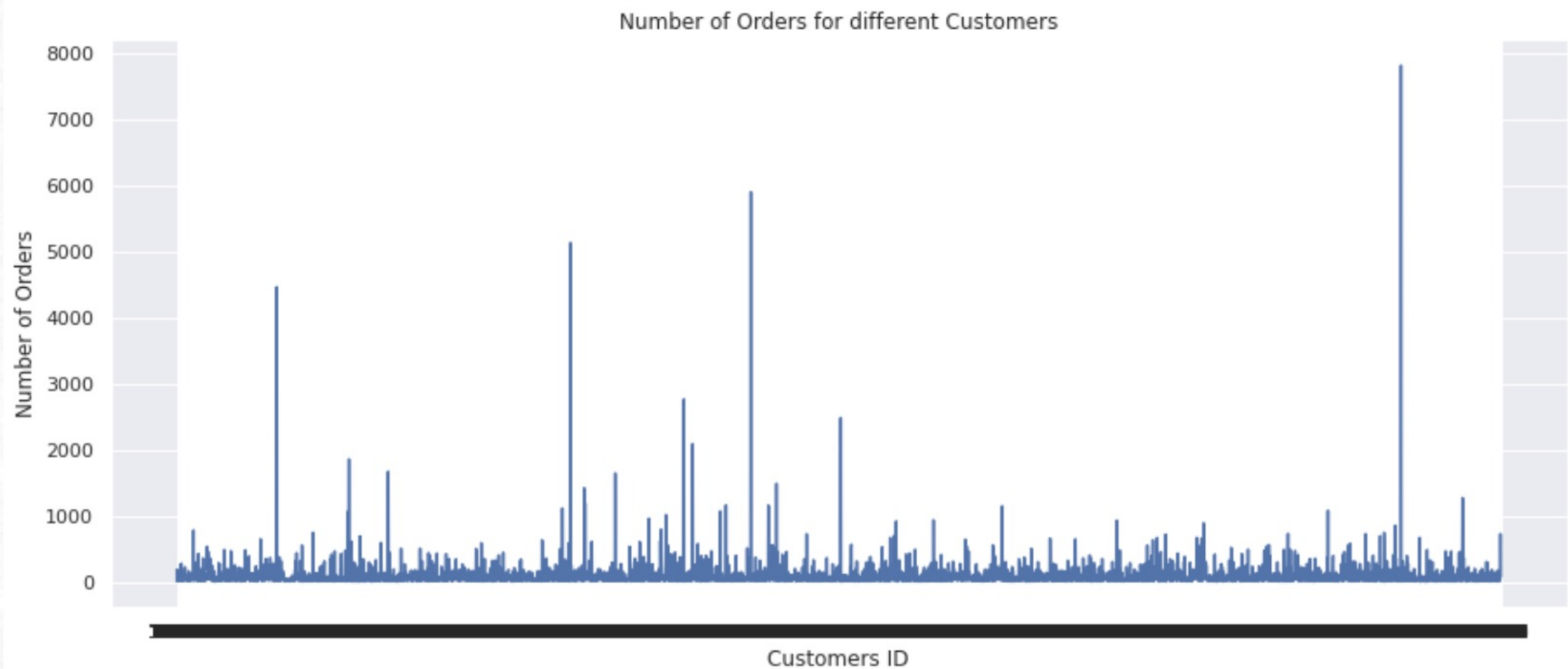
10k

5k

0

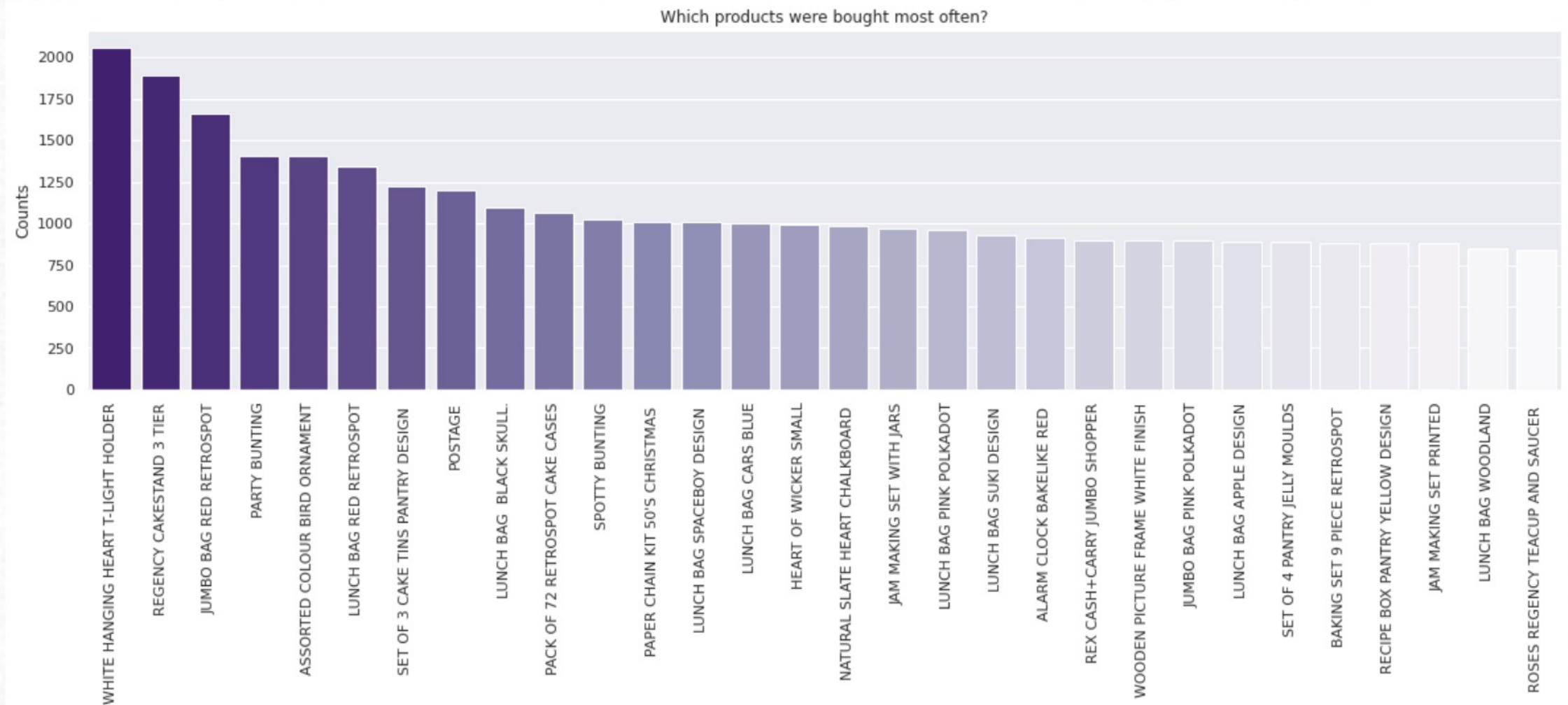
Exploratory Data Analysis

Money Spent by different Customers



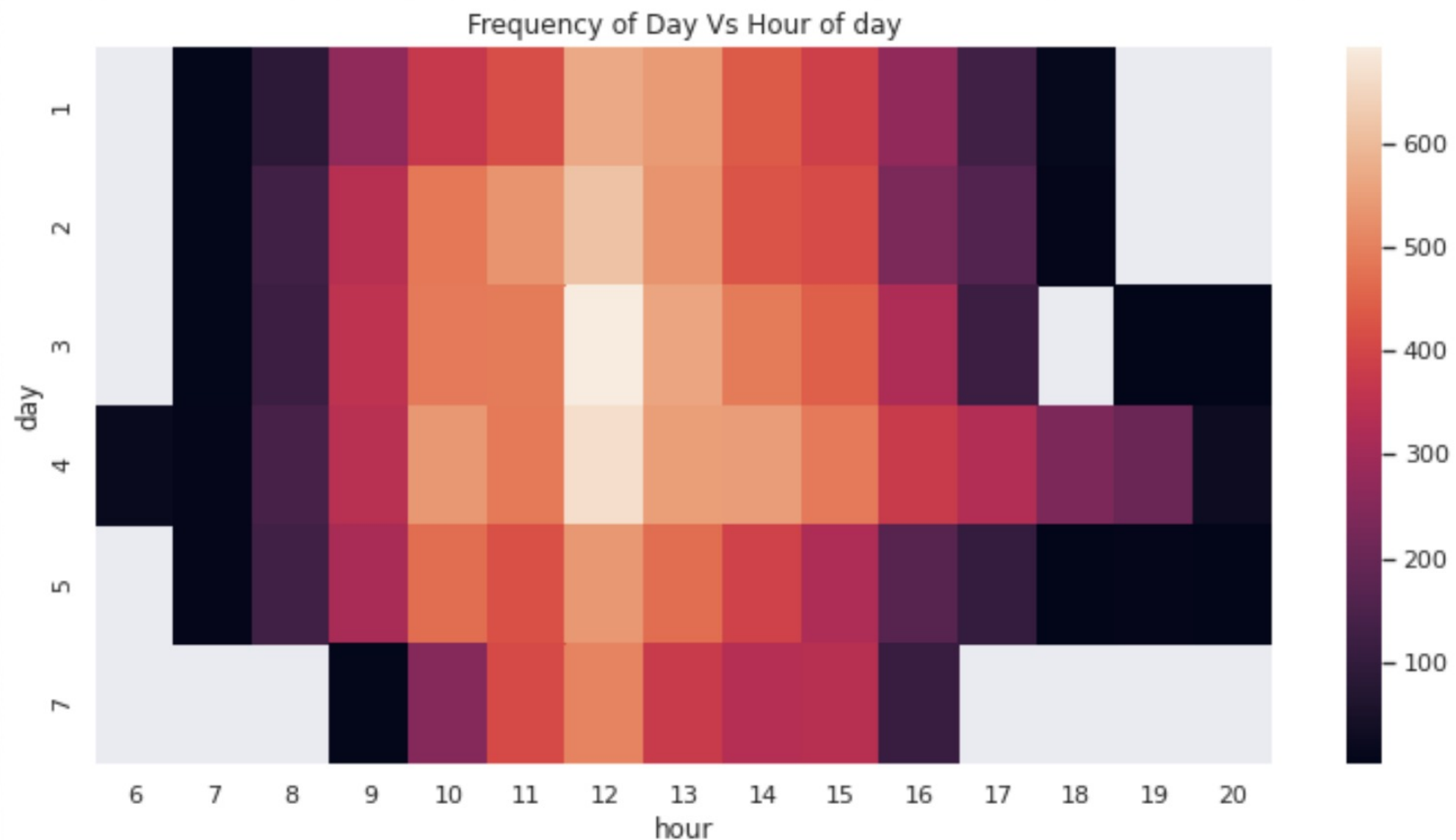
Exploratory Data Analysis

Visualize popular products among customers.



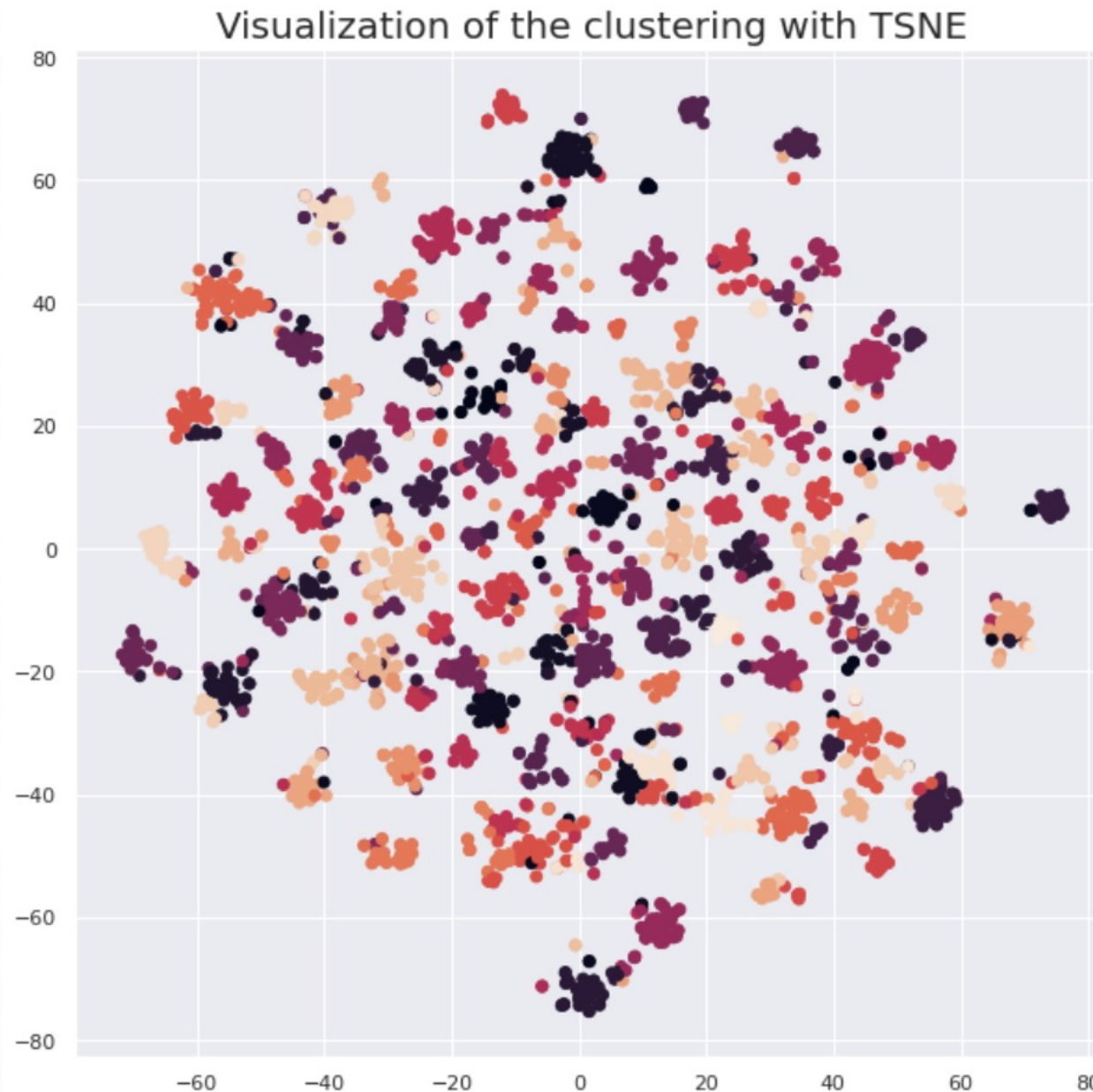
Exploratory Data Analysis

Heatmap of popular time periods when orders are placed.



Predictive ML Modeling: Clustering Products

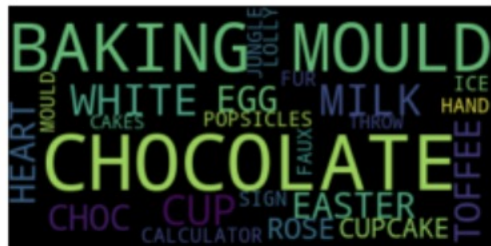
Clustering products based on their text in Description column



Predictive ML Modeling: Clustering Products

Word Cloud of most popular product description keywords among the product clusters

Cluster 98



Cluster 85



Cluster 45



Cluster 80



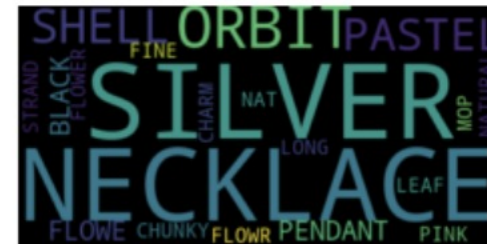
Cluster 71



Cluster 57



Cluster 20



Cluster 66



Cluster 29



Cluster 72



Cluster 49

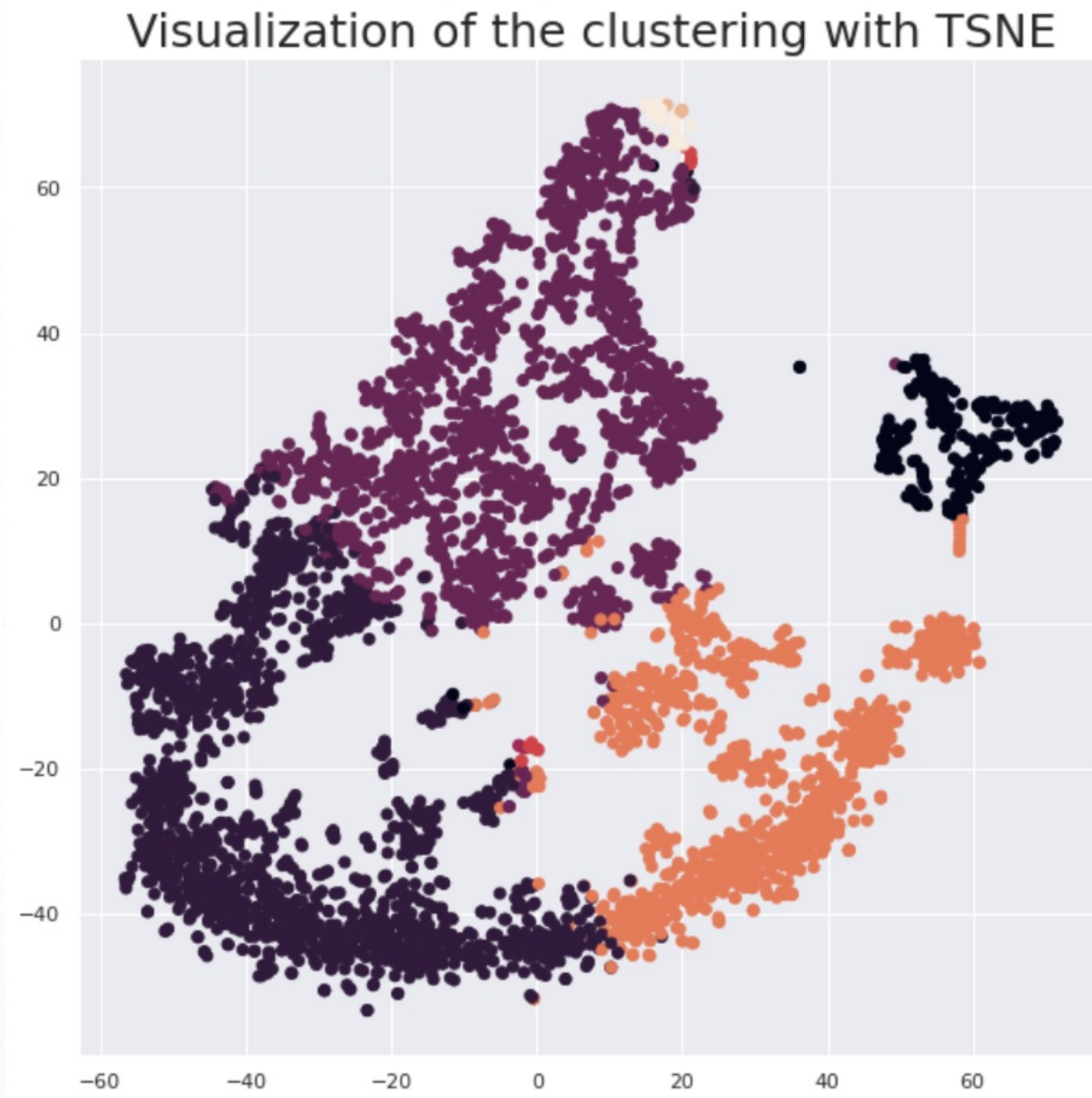


Cluster 55



Predictive ML Modeling: Clustering Customers

Customer cluster visualization with TSNE



Conclusion

- **Cluster 2:** VIP clients, High frequency with a lot of quantity (mean basket price of 513) bought on average and high monetary value
- **Cluster 7:** Very high purchase frequency with a mean basket price of 150 but good monetary value.
- **Cluster 4:** Very high basket price (huge quantity of products bought on average)
- **Cluster 0:** Good average customers
- **Cluster 6:** Good foreign customers
- **Cluster 1:** This cluster represents almost lost customers.
- **Cluster 5:** Highest monetary value but only one or two purchases over the year
- **Cluster 3:** Lost customers