EXPLORATORY ANALYSIS AND CLUSTERING OF WHOLESALE DATA

Unsupervised Machine Learning Project

Lara Onipede

Project goals

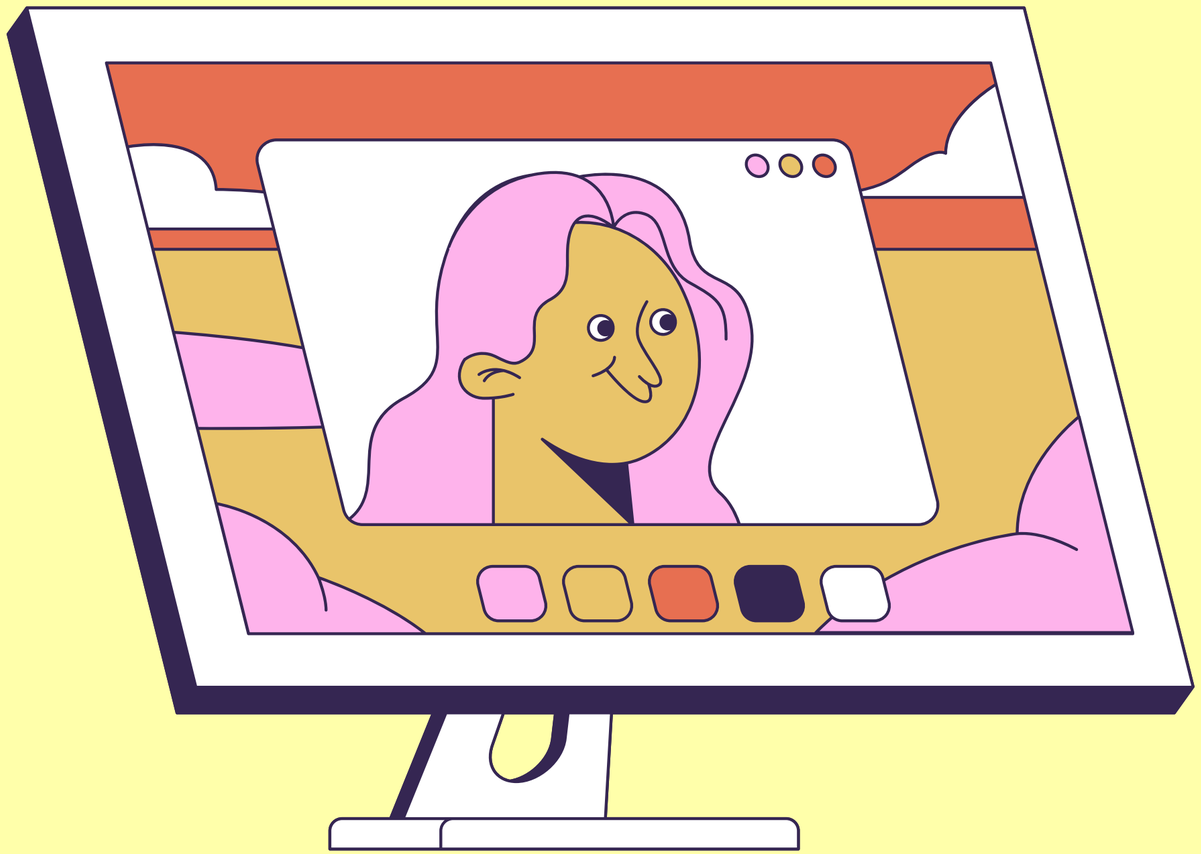
Perform exploratory data Analysis on the wholesales data dataset to draw insights about data

preprocess data for modelling

Use unsupervised learning models on the wholesale data dataset, specifically Kmeans and hierarchical clustering

Carry out a principal component analysis

Dataset



Wholesale Data, refering to clients of a wholesale distributor. it includes annual spending in monetary units on diverse product categories

440 rows 8 columns

Exploratory Data Analysis

Data Cleaning

The wholesale data, is pretty clean, with no missing values

Outlier Detection

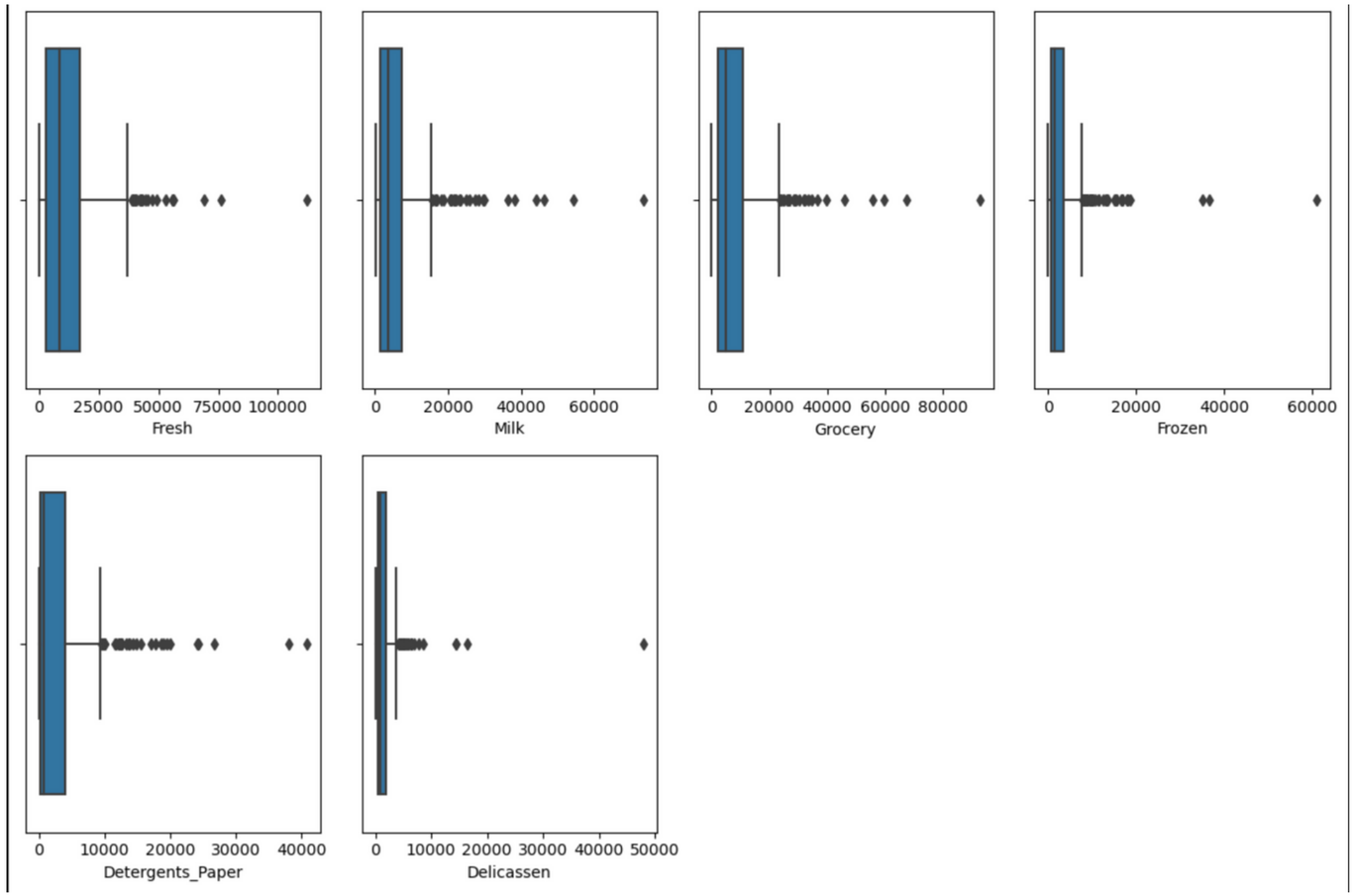
Data contained outliers that were revealed with box plots

Visualizations

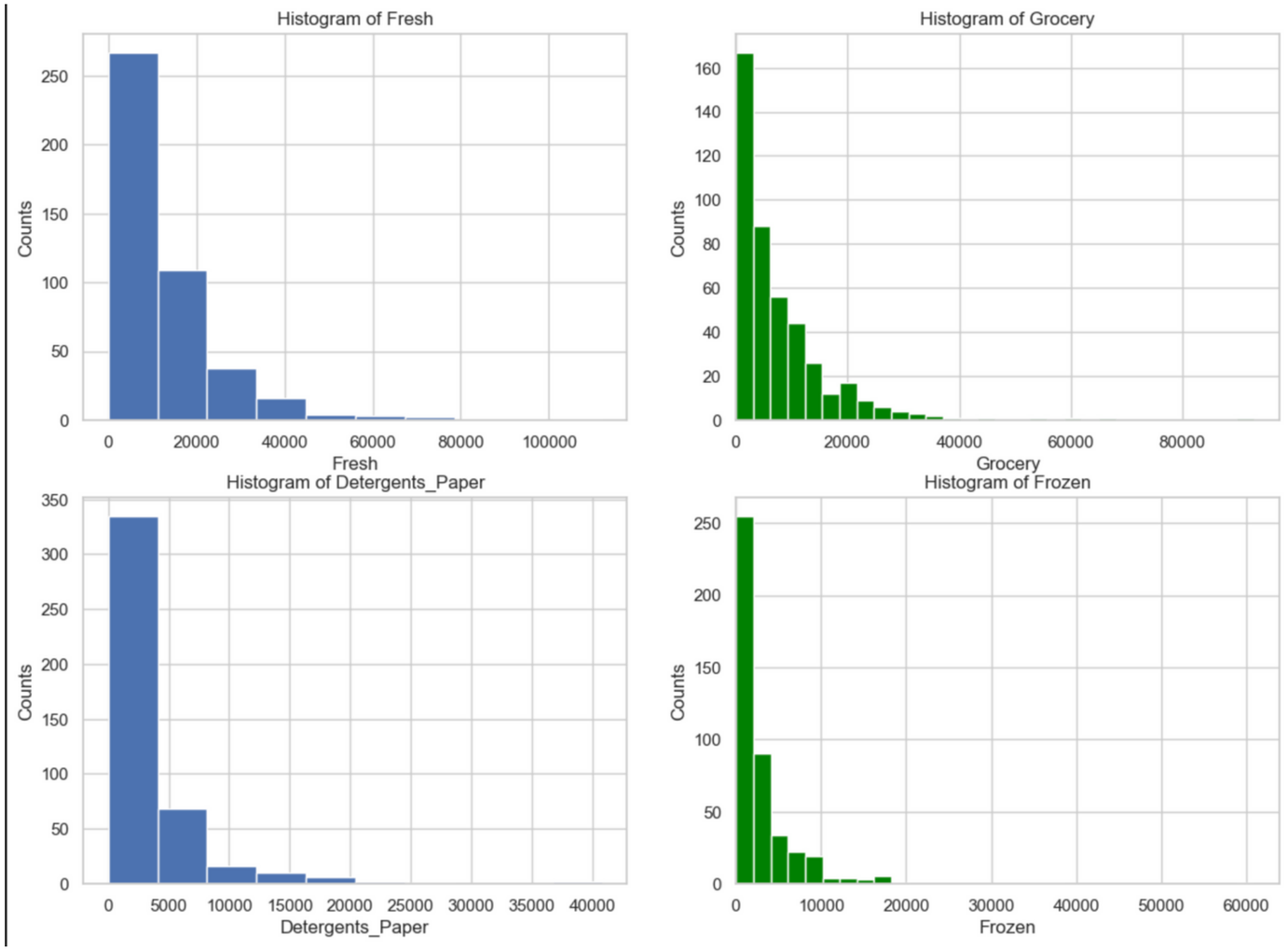
distributions, feature counts and other visualization techniques were used to gain insight from data

Visualize information with charts

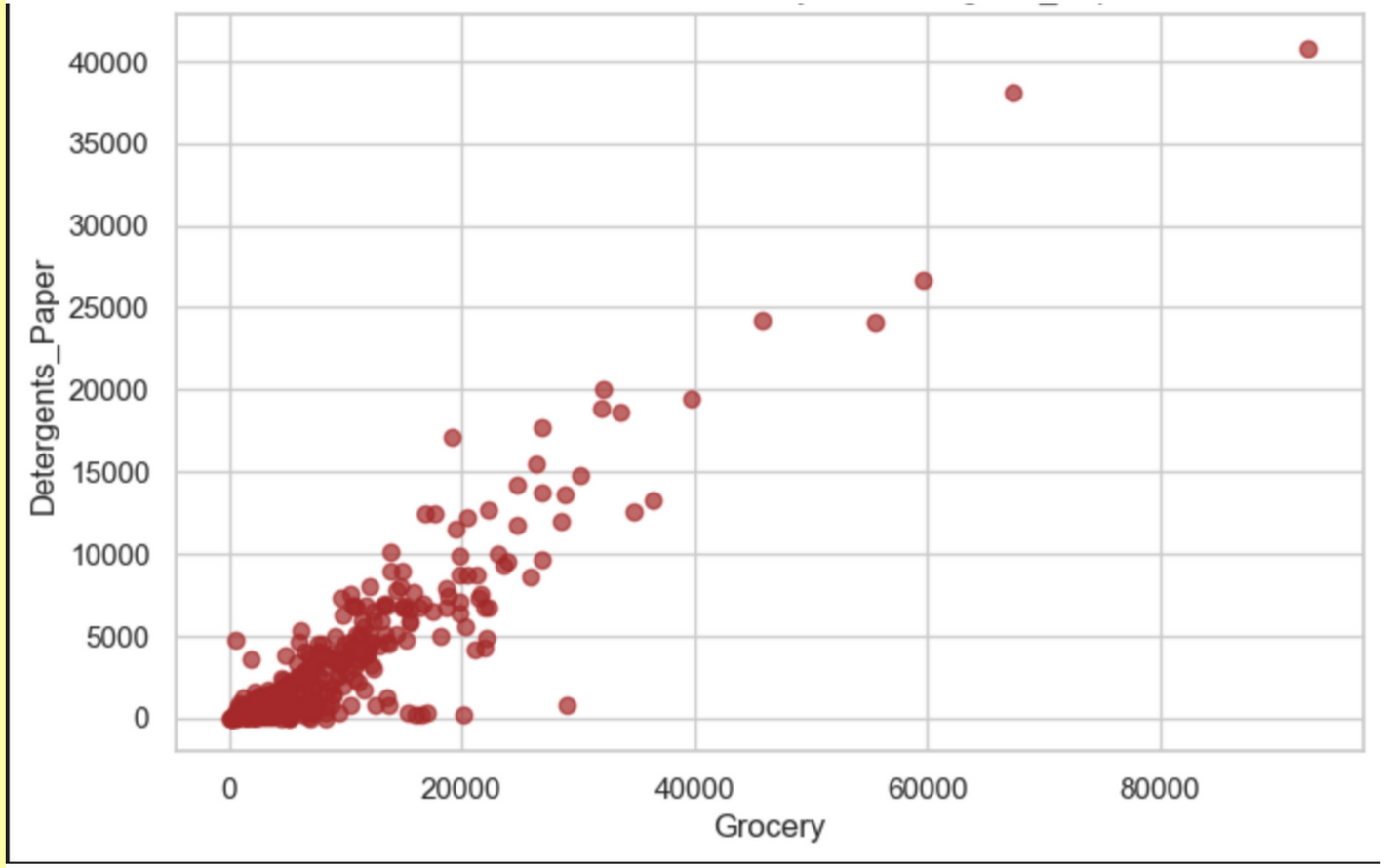
Boxplots showing the presence of outliers



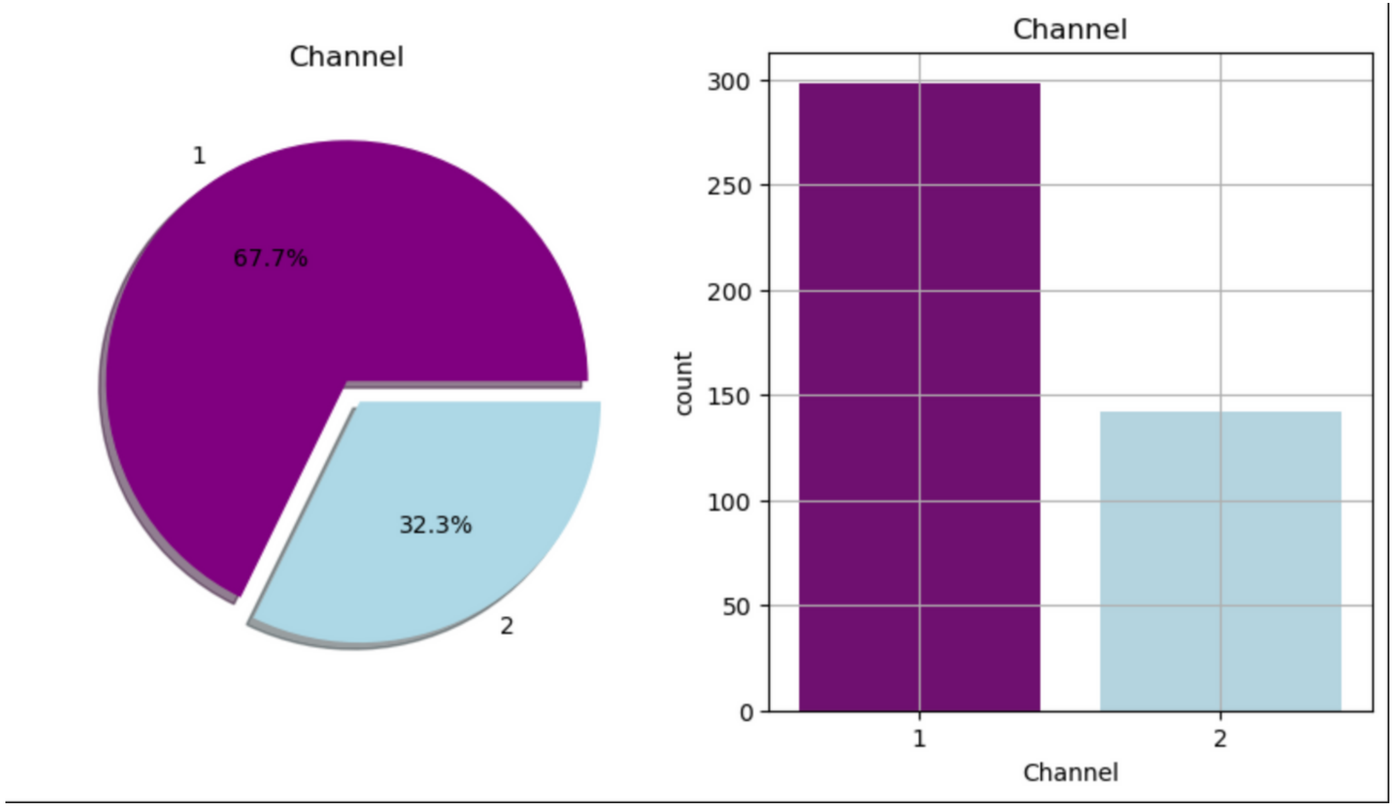
Histogram showing the distribution of features



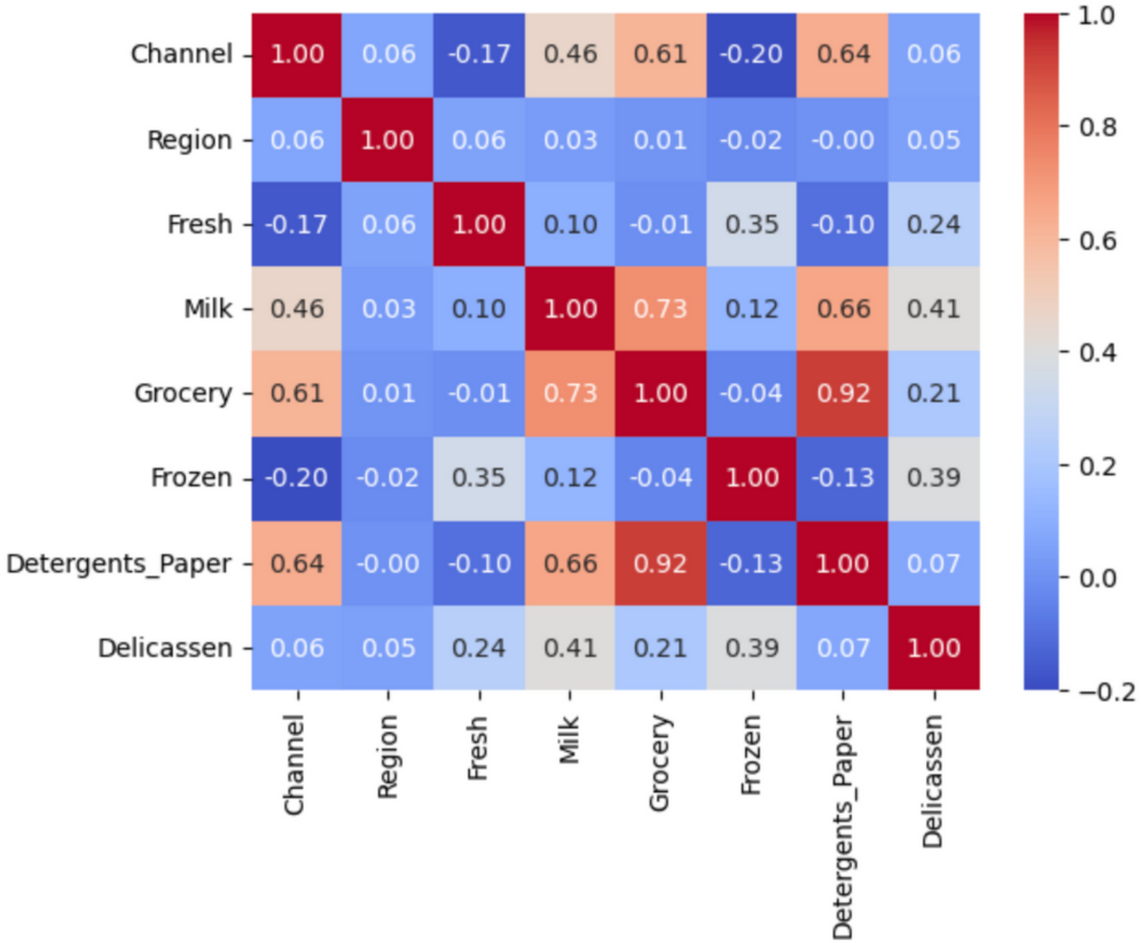
Scatter plot showing relationship between features (Detergents_paper and Grocery)

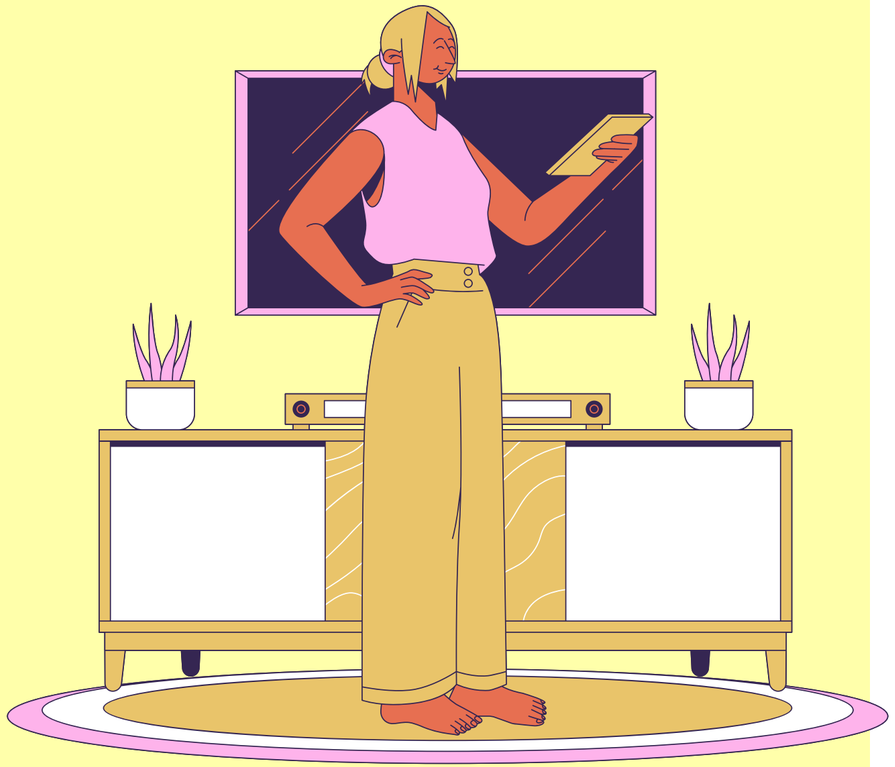


Count plot of channels

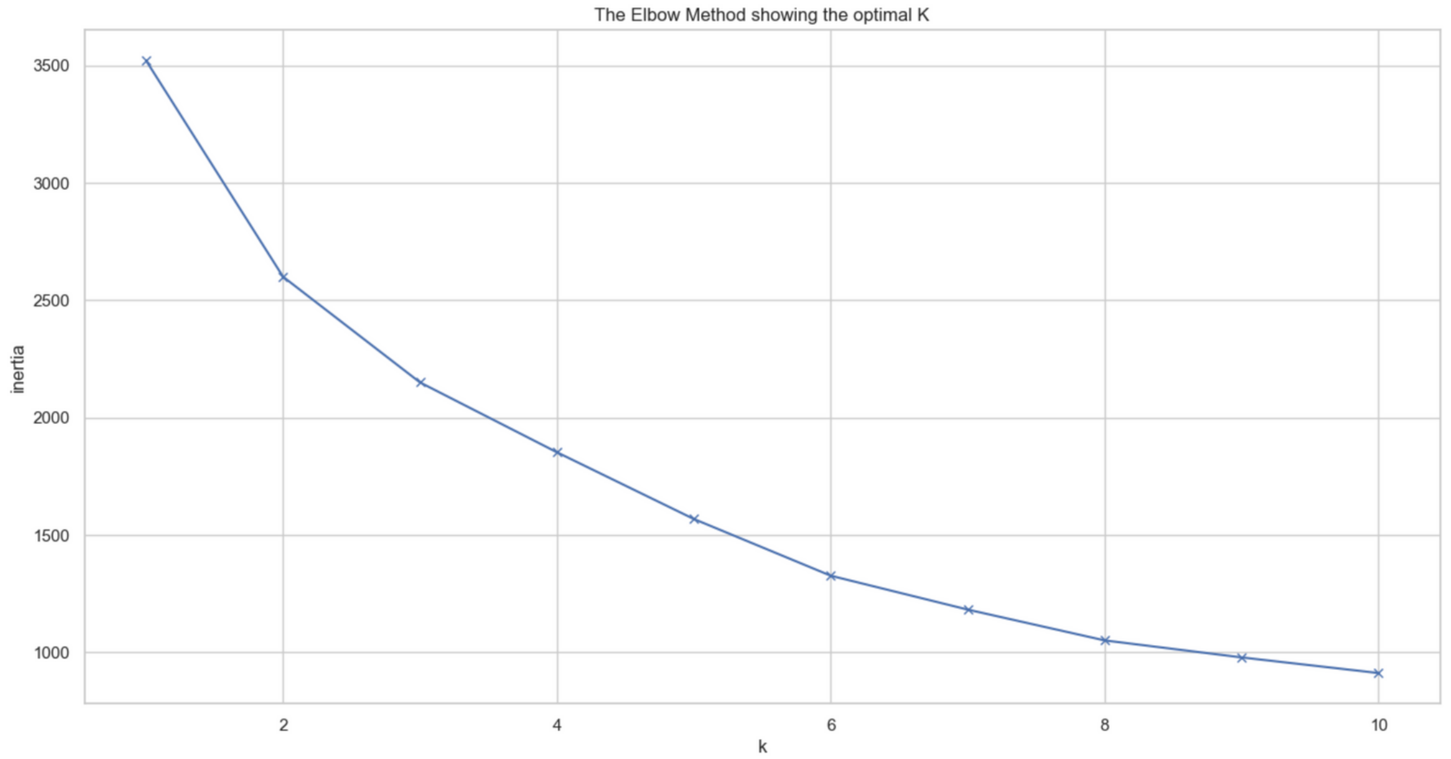


Correlation Matrix showing relationships between variables



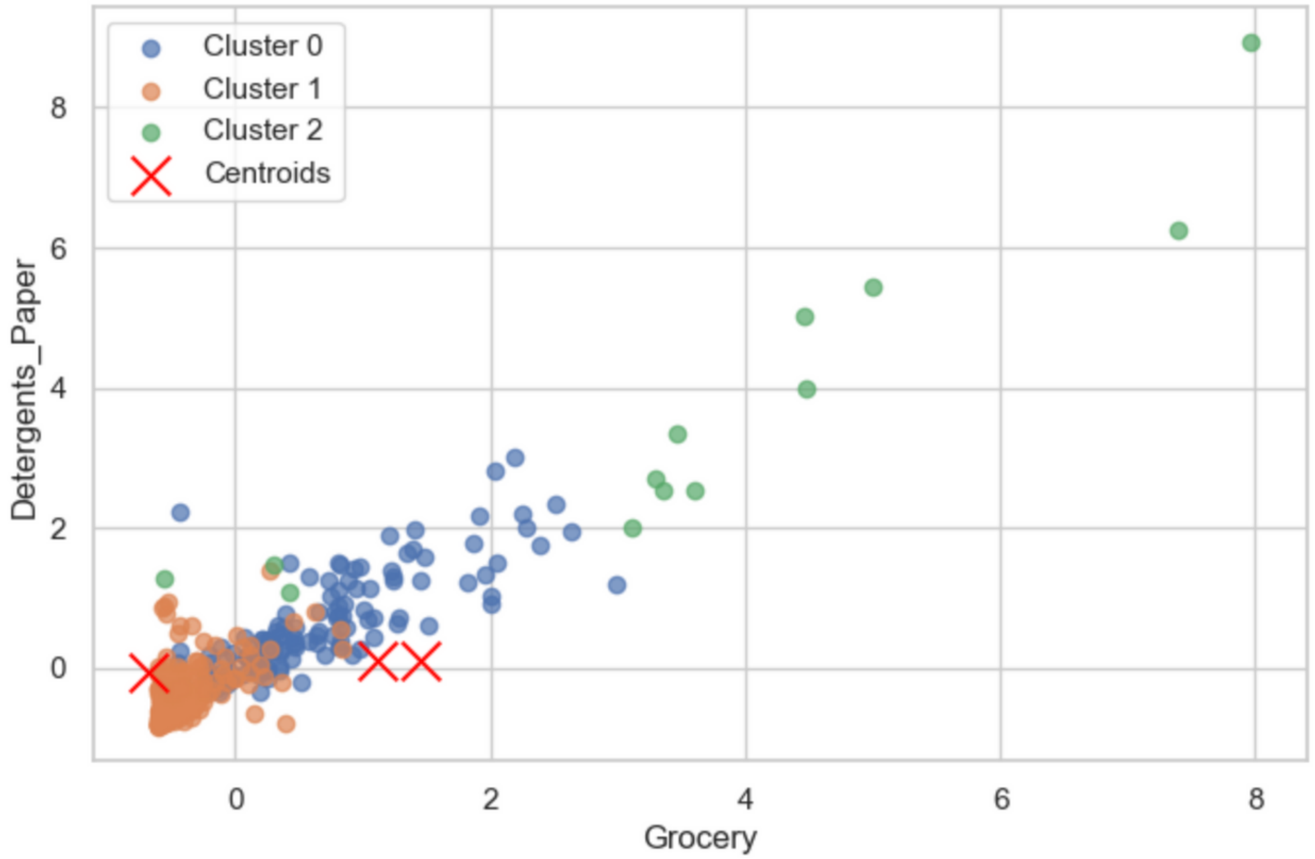


KMeans Elbow Method Giving Insight into optimal value of k

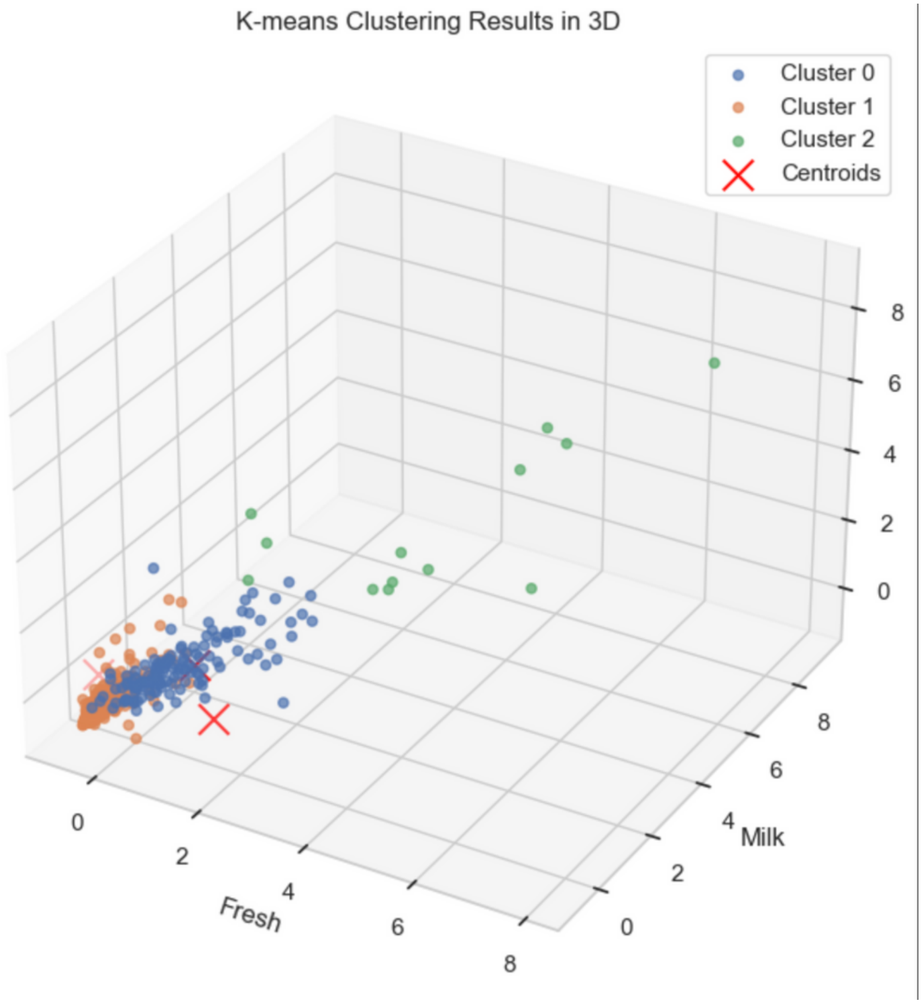


Visualize information with charts

KMeans Clustering



3D KMeans Clustering



Conclusion

The products can be divided into 3 optimal clusters, however, 2 of these are the most predominant.

With some advanced feature engineering, a better clustering model can be built from the dataset

Thank You