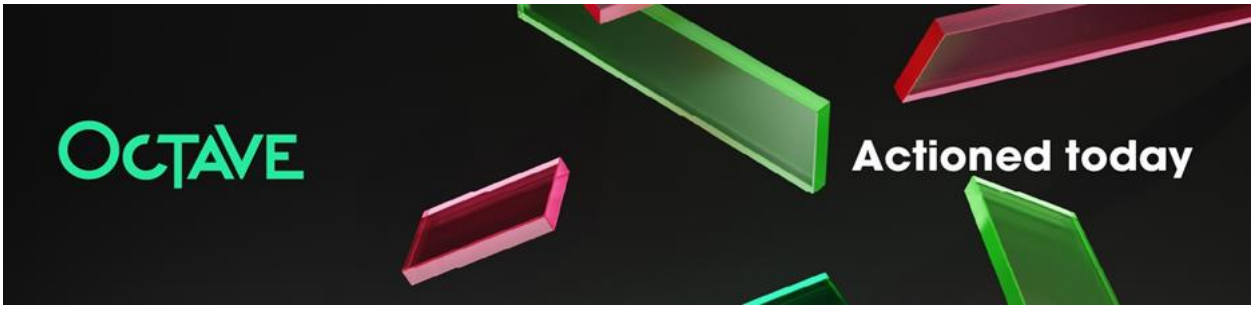
Data Storm 4.0

Round 01 - Kaggle Competition
Case Study

OCTAVE &

Rotaract Club of University of Moratuwa,

April 24, 2023



Store Profiling

1.1 Business Problem

Store profiling is done to analyze the performance of stores based on their sales and customer behavior. The purpose of store profiling is to identify stores that are performing well and those that are not, so that appropriate action can be taken to improve their performance. to allocate resources, such as equipment, marketing, and staffing, to improve the performance of low-rated stores and to optimize the process of item range decision.

1.2 Use Case Definition

A Beverages Company XYZ produces a diverse range of over 40 beverage items and distributes them to over 100 stores. Currently, Beverages Company XYZ adopts a uniform approach towards all its stores and lacks a data-driven decision-making process when it comes to making critical decisions such as resource allocation, distribution and introducing new beverage items. The company makes such decisions qualitatively based on experience rather than leveraging data to optimize these decisions.

However, in recent times, these approaches were found to be inefficient, and they wanted to improve their decision-making process through advanced analytics and machine learning techniques by performing a store profiling on the existing and new stores based on their sales and customer behaviors.

The initial solution they developed was to classify the existing outlets as High, Moderate, and Low based on salesperson's point of view and now they seek an advanced analytics solution to automate this process for the new stores. They have been collecting various data for the past 6 months at store level and store-customer-transaction level. The task is to design and develop an advanced analytics solution that can be used by Beverages Company XYZ to perform store profiling and enhance their decision-making process.

1.3 Data Sources

You are provided with the following data source files to develop a data analytics solution.

1.3.1 Historical Transaction Data Set

The historical transaction data set consists of 488,788 records which are collected from each customer purchase. The granularity of the data set is at the *date - customer - invoice – item* level. The data set can be found in *Historical-transaction-data.csv*. Please refer to the *Data-dictionary.xlsx* for more details about the attributions.

1.3.2 Store Info Data Set

The store info data set consists of 124 stores with shops space and store profile. Data set can be found Store-info.csv. The granularity of the data set is at store level. Please refer to the *Data-dictionary.xlsx* for more details about the attributions.

1.3.3 Data Dictionary

Data dictionary file contains the information about data attributes (*Data-dictionary.xlsx)*

1.4 Deliverables and Evolution metrices

In this competition, you are required to submit the following.

1.4.1 Analytical Solution

Analytical solution to identify the profile of given outlets as High, Moderate or Low. You are required to submit a csv file, which contains the ***Shop_id*** and the ***Shop_profile (High, Moderate, Low)*** using the given ***Testing-data*** set.

You can use F1 metric for evaluation – 60 Points

F1 Score is given by;

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

where;

$$Precision = \frac{True\ Positive.}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

1.4.2 Technical Report

Report of your solution with clearly defined steps, features, feature engineering steps, modelling approaches, evaluation metrices, all the necessary plots/figures and interesting business findings that you can derive from this analysis **– 40 Points.**

(a) Discuss any additional attributes that you can come up with to perform store profiling?

(b) List down at least 5 interventions that the Beverages Company XYZ can take to enhance their decision-making process using the above store profiling mechanism.