**CRM ANALYSTICS - Kaggle.com**

**Data Profiling: Cleaning; analysis and reporting**

**Type of data: Customer sales dataset**

**Datatype: Numeric, Text, Date**

**Other Content**

**Variable Description**

InvoiceNo: Invoice number that consists 6 digits. If this code starts with letter 'c', it indicates a cancellation.

StockCode: Product code that consists 5 digits.

Description: Product name.

Quantity: The quantities of each product per transaction.

InvoiceDate: Represents the day and time when each transaction was generated.

UnitPrice: Product price per unit.

CustomerID: Customer number that consists 5 digits. Each customer has a unique customer ID.

Country: Name of the country where each customer resides.

**Process involved in the data analysis**

Software: Python

Analysis: Exploratory Data Analysis; Recency Frequency and Monetary Analysis; Market Penetration Analysis

**DATA CLEANING - OUTLIERS 541909(Total data columns before cleaning)**

1. Remove InvoiceNo with “c” as the start code- cancelled: 9288
2. Remove negative quantity value: 1336
3. Remove negative unit price: 2
4. Remove blank customerID: 133359
5. Total data columns removed: 143985
6. Total dataset read on python script: 3183392
7. Remove duplicate data using python code: 5192

**Total dataset used for analysis: 3141856**

**DATA ANALYSIS**

1. CUSTOMER ANALYSIS (Customer Segmentation analysis using RFM Model Analysis)
2. MARKET ANALYSIS(Geographical penetration and purchasing Frequency)