

Customer Behavior Prediction Model

Using RFM and Logistic Regression

Latifah Al-Rasheedi



Bussiness Problem



Since **marketing campaigns** consume a lot of companies' resources, companies nowadays understand the importance of analyzing their customers' behavior and its impact on the campaign's **success**.

Therefore I decided to build a model that predicts the next purchase day of the customer, which will help companies prepare special offers to their customers based on their level.



Project steps

01

Find suitable dataset

02

Data Cleaning

03

Feature Engineering

04

Data Visualization

05

RFM method

06

Data Modeling

The background features a large, irregular watercolor shape in shades of light blue and teal, centered on the page. Below this shape is a lighter green watercolor wash. Scattered around these shapes are numerous small, dark blue dots of varying sizes. A thin, dark blue line starts from the top right, curves around the central shape, and ends in a small, stylized loop at the bottom right.

01

Data Description

The data used is an E-Commerce dataset from **Kaggle**, which contains actual transactions from UK retailers

Transactions period:	From 01/12/2010 to 09/12/2011
Number of features:	8
Number of observations:	541,908
Features data types:	Categorical: Invoice_Number , Stock_Code, Description, Invoice_Date, Country Numerical: Customer_ID ,Unit_Price, Quantity

The background features a large, irregular watercolor splash in shades of light blue and teal, centered behind the text. Below this splash is a lighter green watercolor wash. Scattered around the central splash are several clusters of small, dark blue dots. A thin, dark teal line curves from the top right, looping around the right side of the central splash and ending in a small, stylized, wavy shape at the bottom right.

02

Data Cleaning

Cleaning Process

Columns

- **Change data type:**
invoice Date to date time type.
customer ID to integer type.
- **Make columns names**
lowercased.

Rows

- **Change 'description' to lower**
case.

Handling Null and duplicates

**Null
Values**

Description: 1454
Customer ID: 135080

Drop

**Duplicated
values**

Number of duplicates: 5225

Drop

**Cancelled
orders**

Zero quantity = Cancelled orders

Drop

The background features a large, irregular watercolor shape in shades of blue and green. To the left and bottom-left, there are clusters of small, dark blue dots. A thin, dark blue line curves from the top right towards the bottom right, ending in a small, stylized, wavy shape.

03

Feature Engineering

New Features added

Feature Splitting

Invoice_date

2010-12-01 08:26:00



Invoice_date	year	month	day	hour
2010-12-01 08:26:00	2010	12	1	8

Creating Features

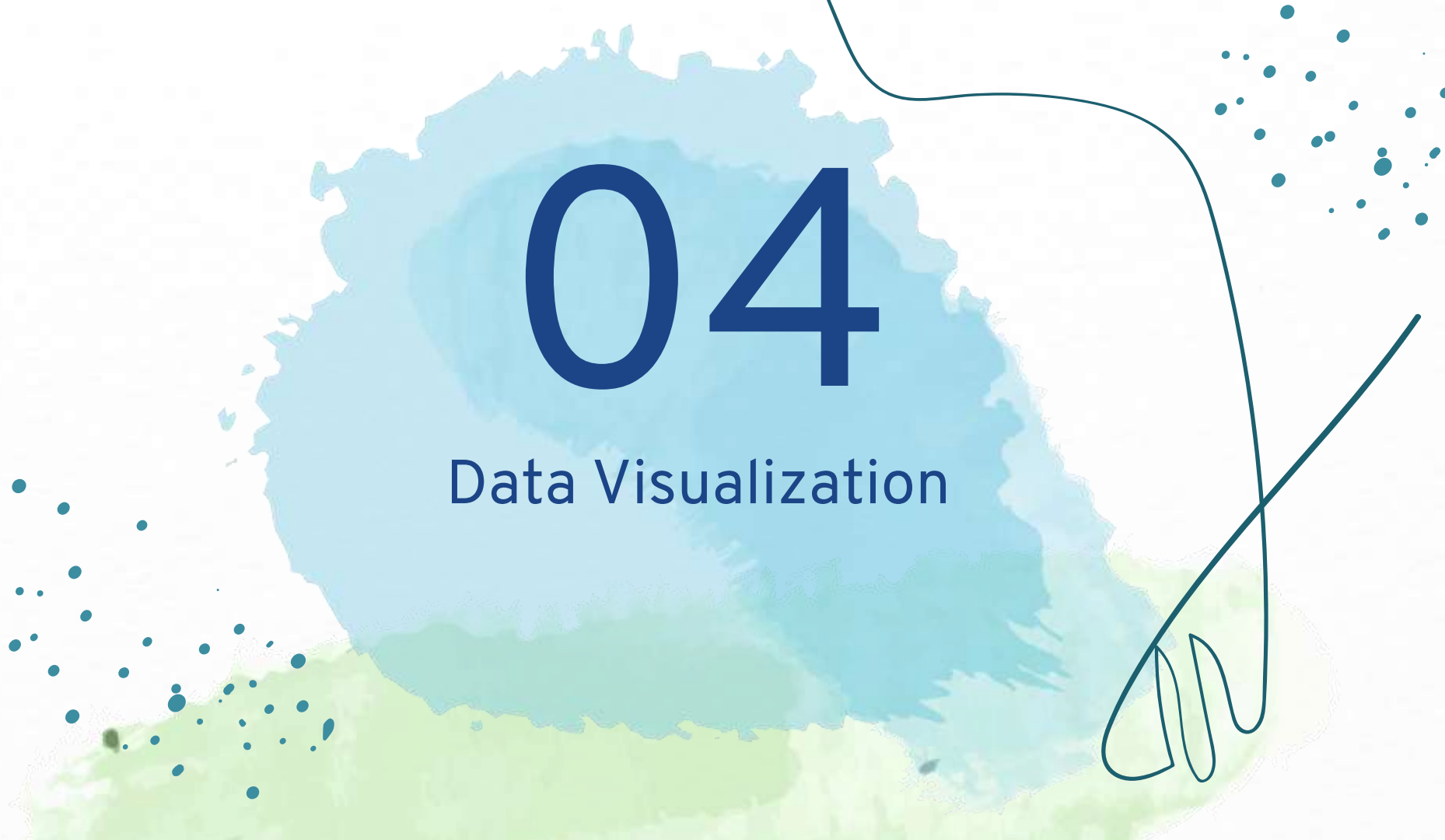
Quantity X Unit_Price


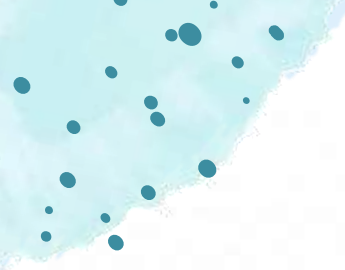


Total spent

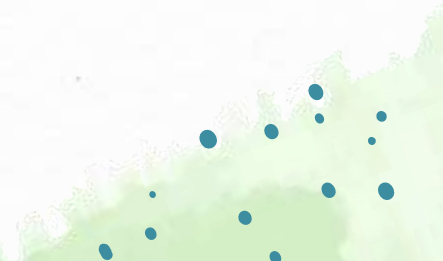
04

Data Visualization

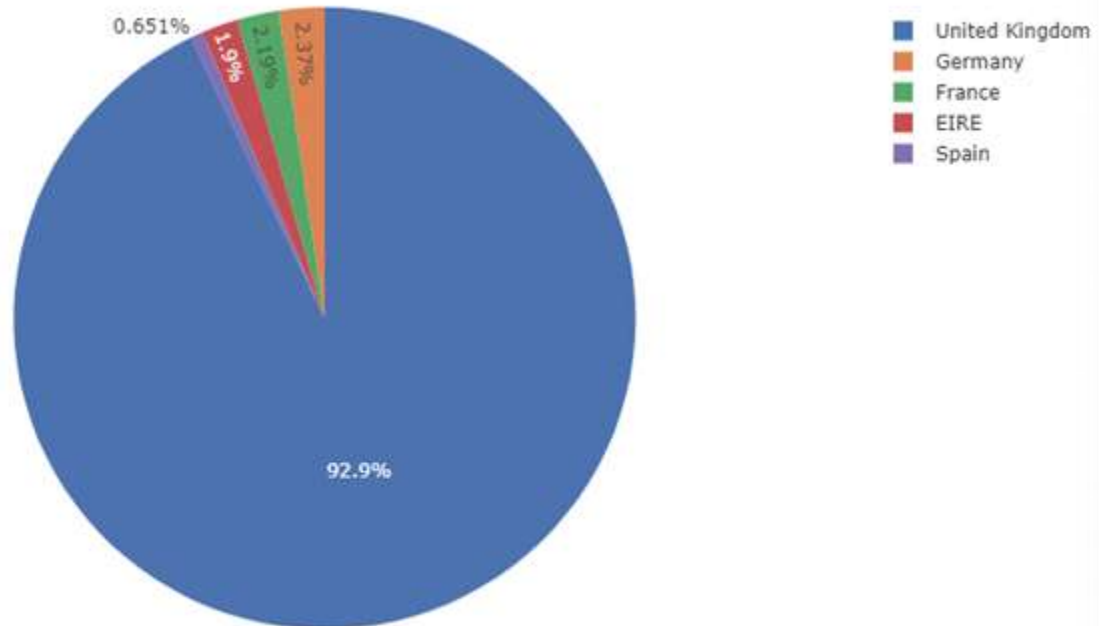




Let's answer some
business questions

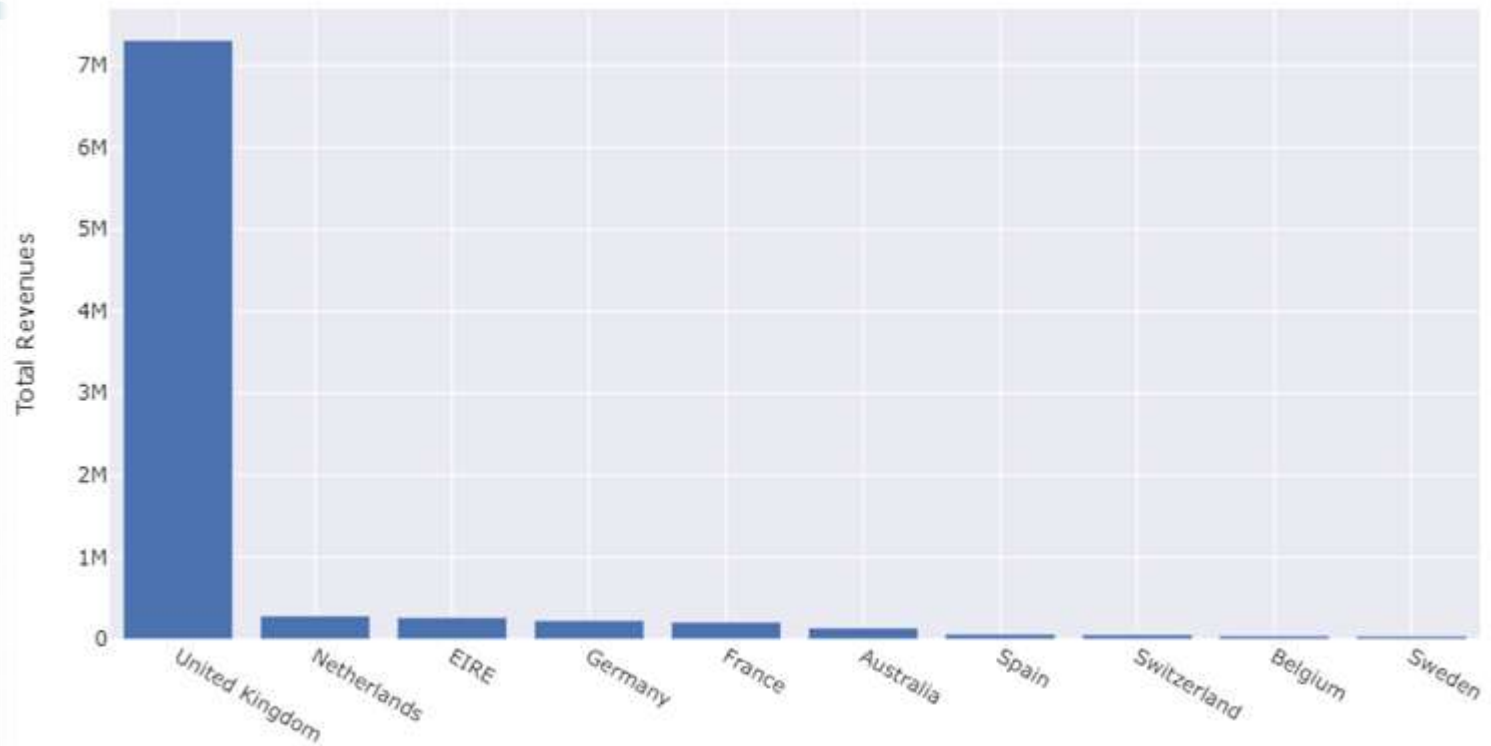


Top 5 Countries



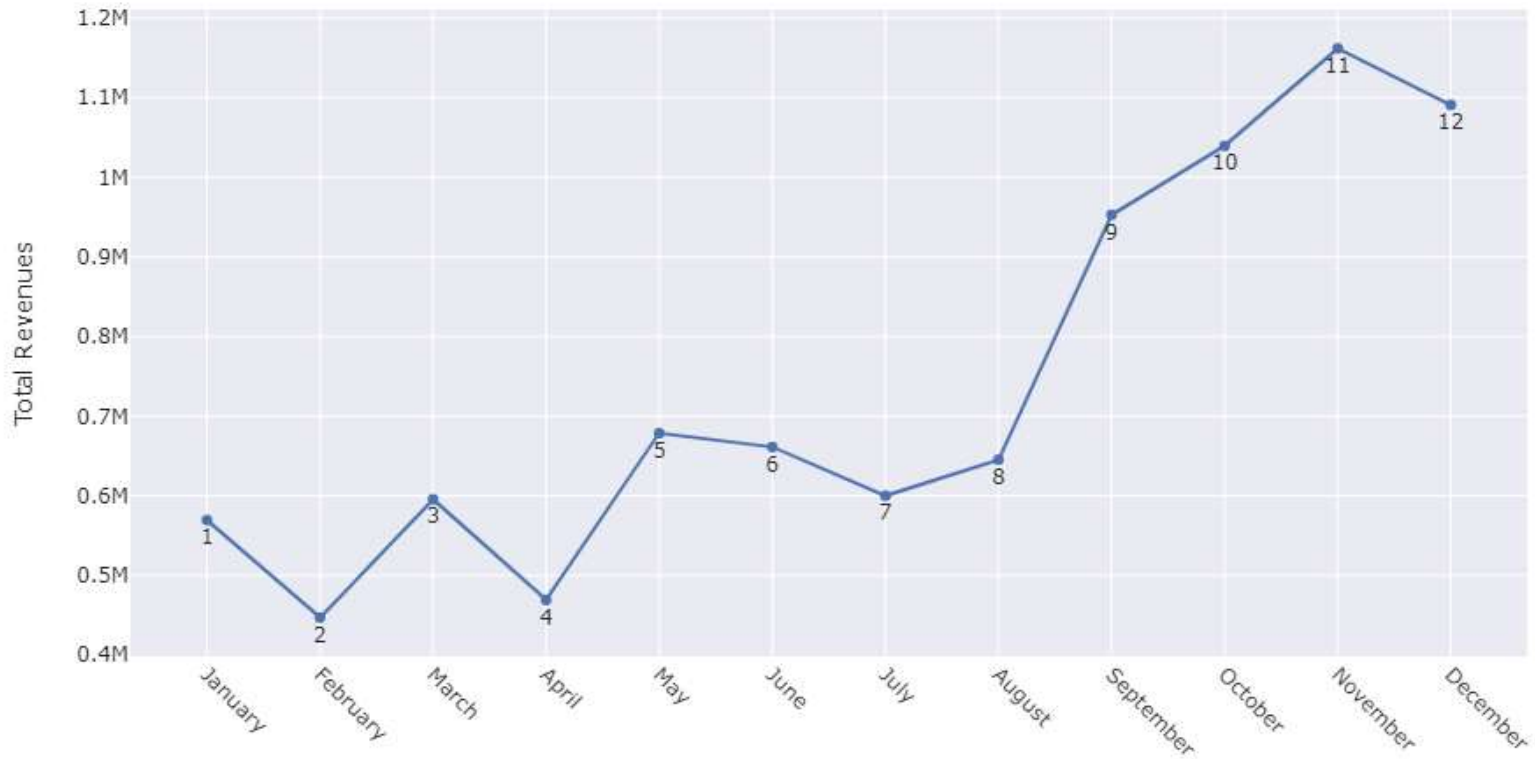
Which country with the highest number of customers?

Total Revenues By Country



Which countries with the highest revenues?

Total Revenues By Month



Which month ranked the highest revenues?



What time is the best for launching a new campaign/Advertisement?

The background features a large, irregular watercolor splash in shades of light blue and teal, centered behind the text. Below this splash is a lighter green watercolor wash. Scattered around the central splash are several clusters of small, dark blue dots. A thin, dark teal line curves from the top right, looping around the right side of the central splash and ending in a series of three loops at the bottom right.

05

RFM METHOD

RFM (Recency, Frequency, Monetary) analysis

It is a customer segmentation technique that uses past purchase behavior to divide customers into groups!

RECENCY

Days since last purchase

FREQUENCY

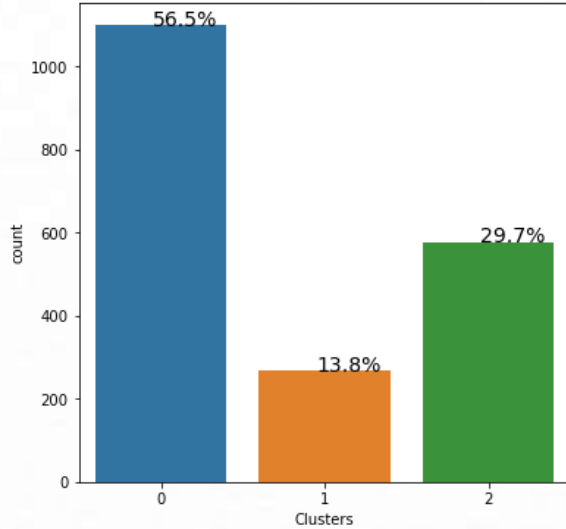
Total number of purchases

MONETARY

Total money this customer spent

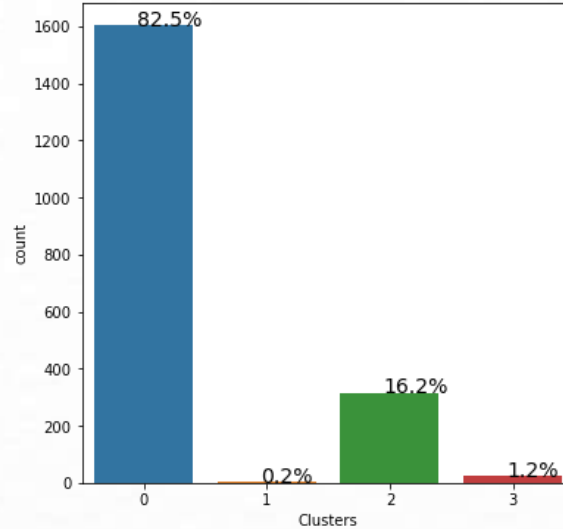
RFM Clusters

Recency Clusters



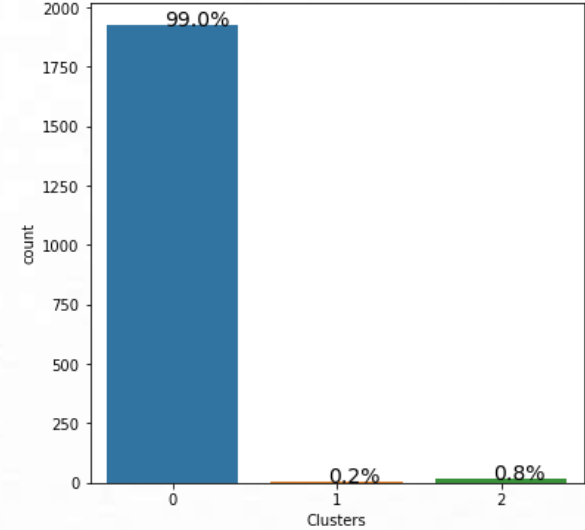
56% of the customers are active

Frequency Clusters



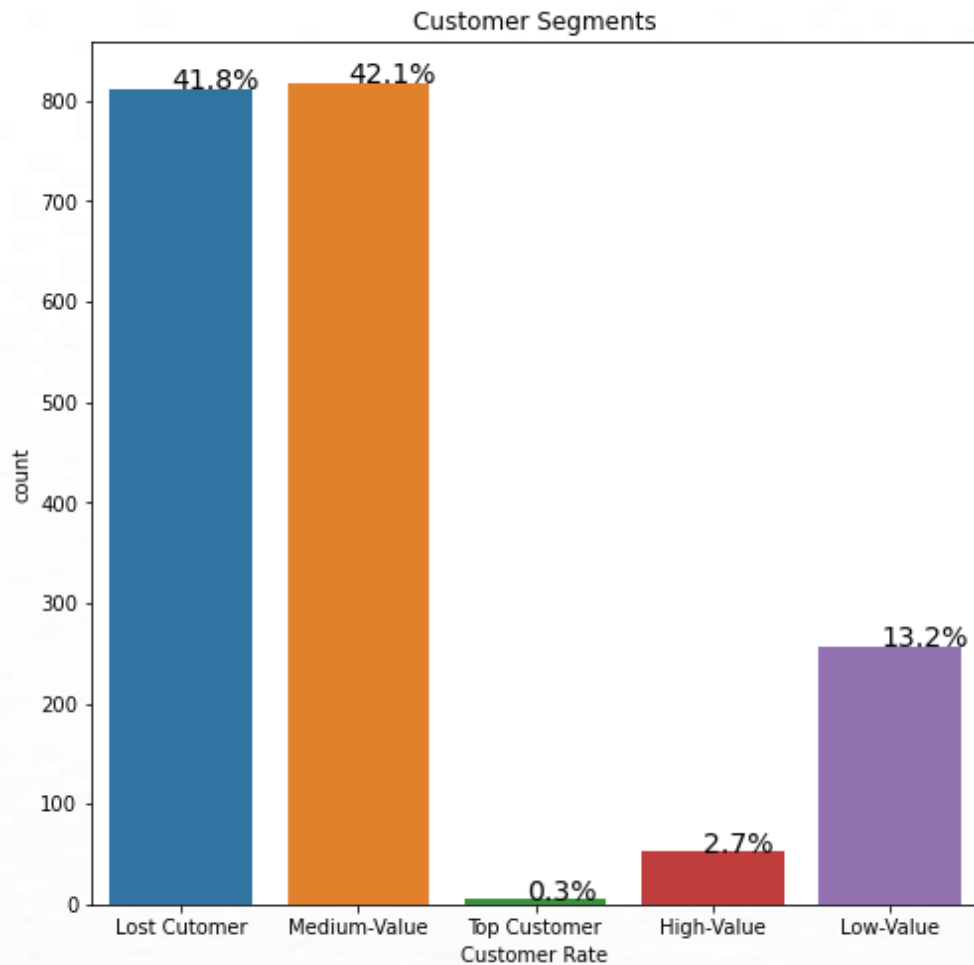
82% of the customers with few number of orders

Monetary Clusters



99% of the customers spent the least

RFM Segmentation



The background features a large, irregular watercolor shape in shades of blue and green. To the left and top right, there are clusters of small, dark blue dots. A thin, dark blue line curves across the right side of the image, ending in a small, stylized, wavy shape.

06

Data Modeling

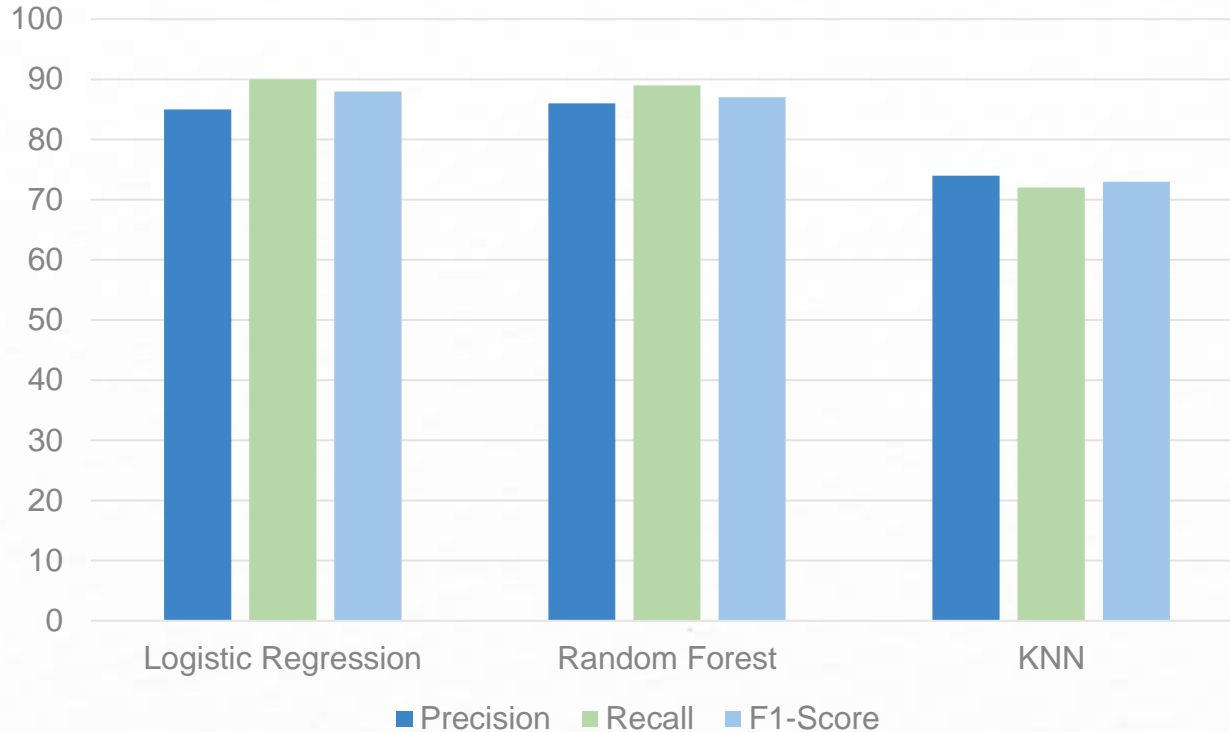
Models used

My target

Day Range: it 0 if a customer took more than 3 months to purchase. It 1 if customer purchased within 3 months.

Model	Accuracy
Logistic Regression	88%
Random Forest	88%
KNN	75%

Customers who purchased within 3 months



I chose logistic regression model, since it has the highest accuracy and F1-score!



Thank you!