CUSTOMER SEGMENTATION

Clustering Customers & Generating a Customer Value Prediction Model

by Alina Aufenanger

CUSTOMER SEGMENTATION

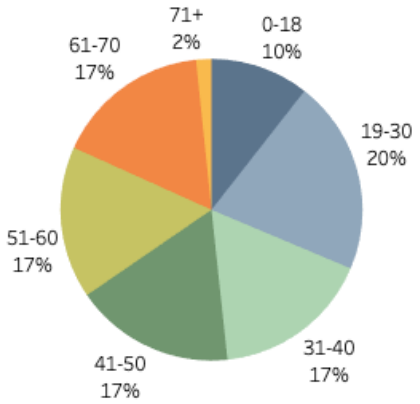
Introduction

Company	Data	First steps
TheLook - an E-Commerce clothing shop operating since 2019.	Stored in Google Data Base containing tables on Users, Orders, Order Items, Products, etc.	Getting better insights into the business by conducting EDA & customer segmentation.

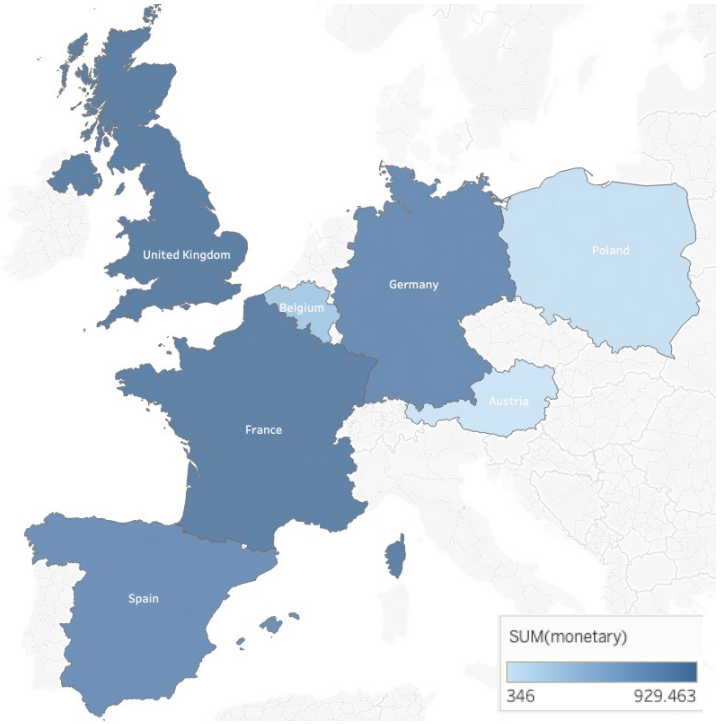
CUSTOMER SEGMENTATION

Types of Segmentation

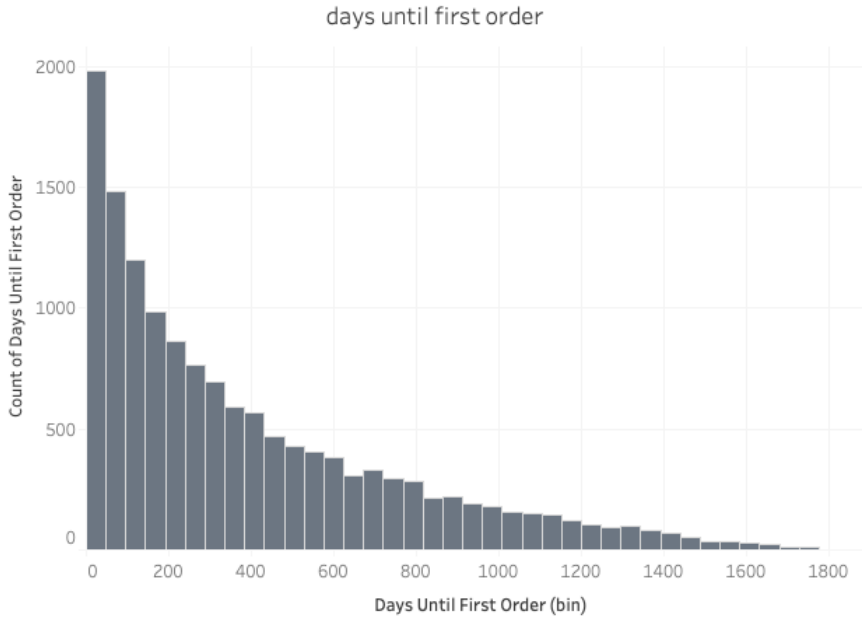
Demographic



Geographic



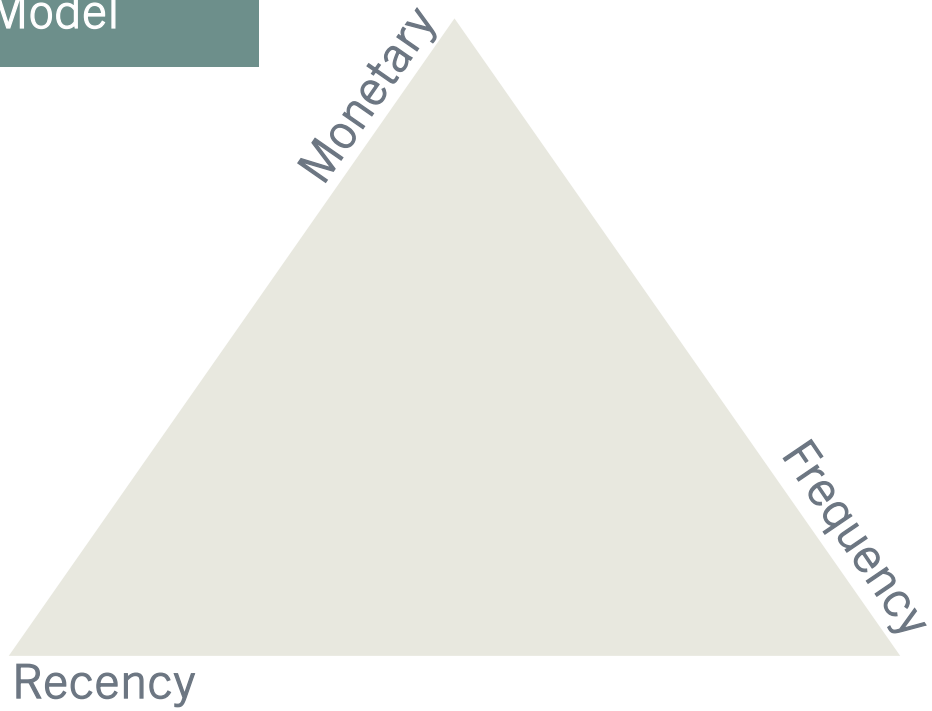
Behavioral



CUSTOMER SEGMENTATION

RFM vs K-Means Clustering

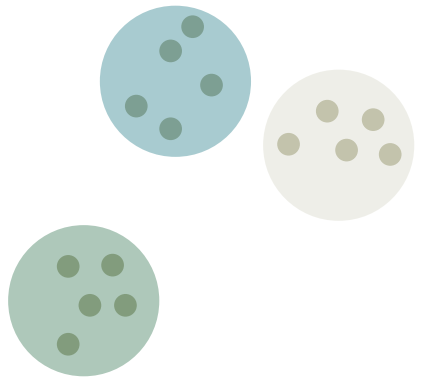
RFM-Model



- Rates customers on a scale of 1-5 for each RFM variable.
- Segments customers based on their total RFM score.
- Provides a straightforward and interpretable segmentation based on predefined score thresholds.

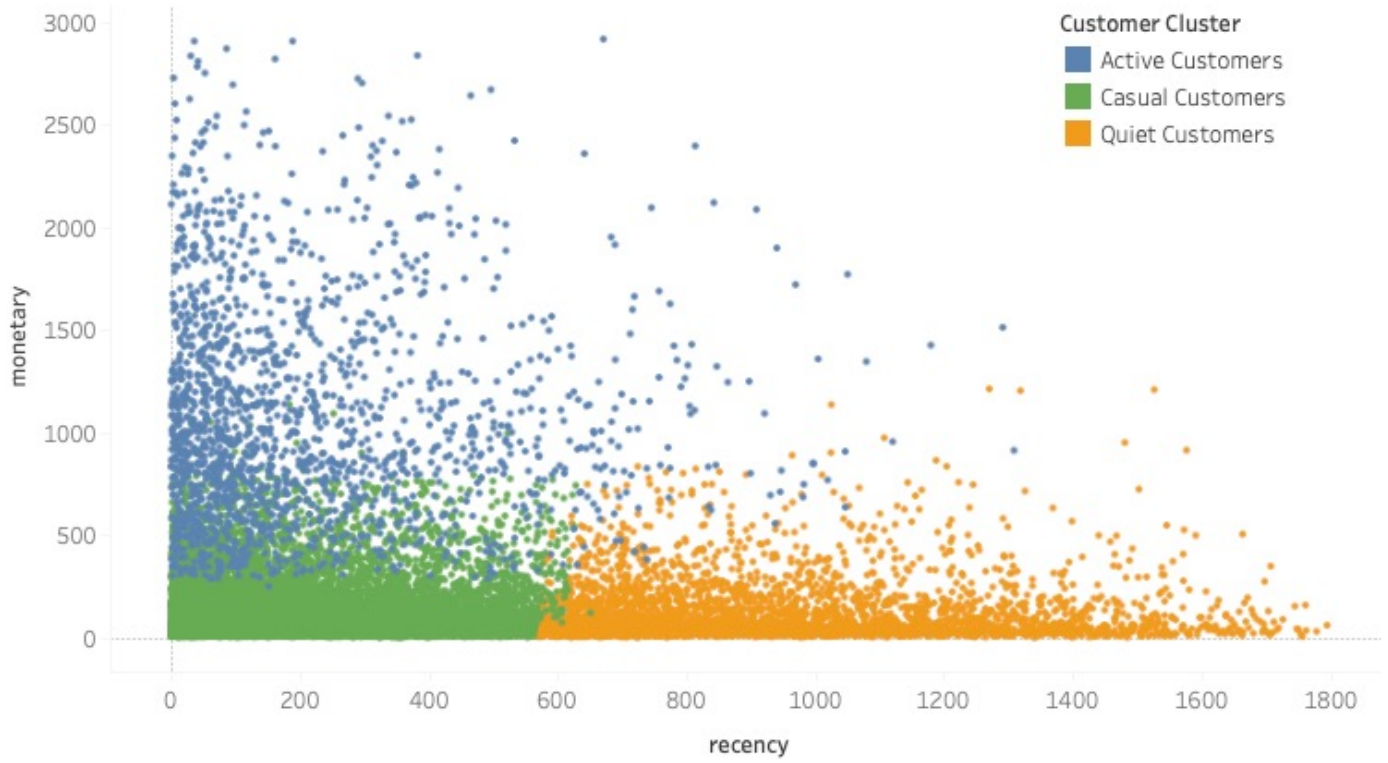
K-Means Model

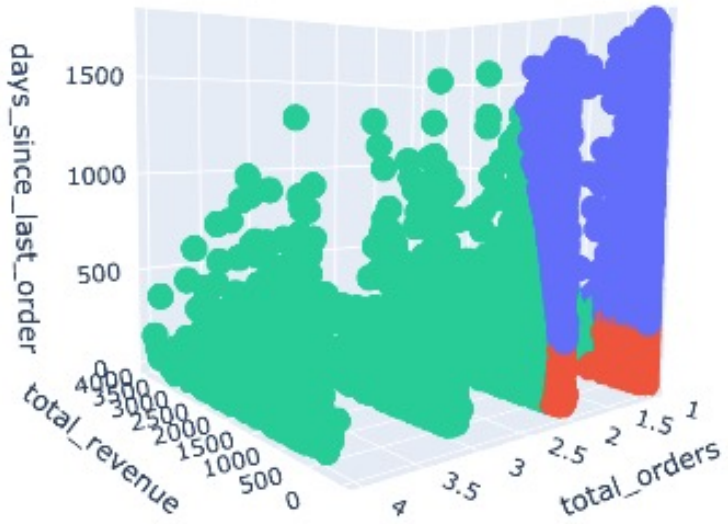
- Utilizes machine learning to group similar customers into 'k' clusters.
- Considers raw RFM variables as features for clustering.
- Allows for a more data-driven and adaptive segmentation based on patterns in the data.
- The decision to set 'k' to 3 is driven by business knowledge.



CUSTOMER SEGMENTATION

Customer Segmentation with K-Means





Active Customers: spend relatively higher amounts and have made multiple recent purchases
Casual Customers: recently made purchases but spent an average to low amount
Quiet Customers: haven't made a purchase in quite a while and have low to average spending

CUSTOMER SEGMENTATION

Customer Segments Overview

	Active Customer	Casual Customer	Quiet Customer
Total Customers	1,756	8,726	3,537
Average Revenue	$ 1,074	$ 157	$ 123
Average Revenue/Order	$ 334	$ 110	$ 97
Average Orders	3	1	1
Average Lifetime	508 days	102 days	43 days

CUSTOMER VALUE PREDICTION

Generating a customer centric revenue prediction model based on historical data

Feature Selection

Total Orders
Days since Registration (T)
Days between First & Last Order (LT)
K-Means Customer Cluster

Preprocessing

PowerTransform: T, LT
MinMaxScaler: all features

Model Selection

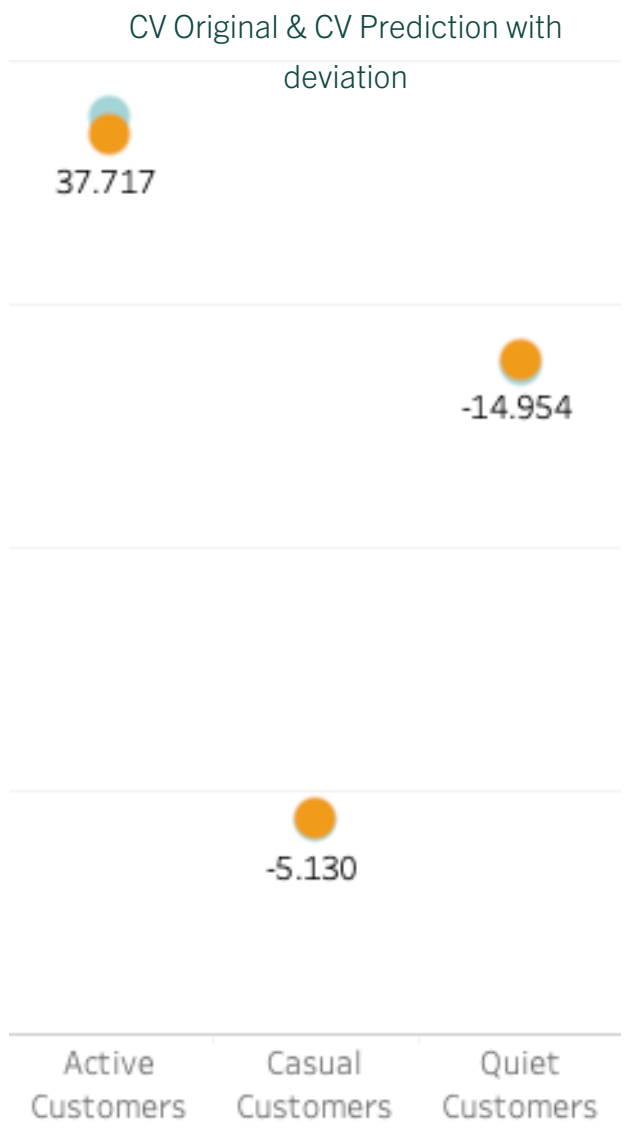
Linear Regression, Random Forest,
GradientBoosting, XGB Boost

Best Model

GradientBoostingRegressor
(min_samples_leaf=2, random_state=42,
subsample=0.8)

Best Feature: Total Orders (88%)

MAE Train 120, Test 121
R2 Train 0.74, Test 0.72



CONCLUSIONS

Usage

High-level business planning: setting revenue targets, developing marketing strategies, decision support

Limitations

Absence of concrete time component: model is based on historical data & its predictive power may diminish if there are significant changes in customer behavior.
Heavily depending on Total Orders Component

Next Steps

Scenario Analysis:

Conduct scenario analyses by inputting hypothetical changes to customer features into the model. This can help simulate the potential impact of business decisions on customer revenue and guide strategic planning.

Follow up on:

Product & Price Analysis, Implement Time Series Analysis

Thanks!

If you have any further questions let me know.

pictures on start & final slide taken from Zendesk Blog