

TheAnalyticsTeam

# Sprocket Central Pty Ltd

Data analytics approach

Selmir  
Kalender

# Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

# Introduction

## Identification & Recommendation Of Top 1000 Customers To Target From The Dataset

### Business Problem

- Sprocket Central Pty Ltd is a long-standing KPMG client whom specialises in high-quality bikes and accessible cycling accessories to riders.
- Their marketing team is looking to boost business by analysing their existing customer dataset to determine customer trends and behaviour.
- Using the existing 3 datasets (***Customer demographic, customer address and transactions***) as a labelled dataset, we will recommend which of the 1000 new customers should be targeted to drive the most value for the organisation.

### Data Overview

**Some of the more important insights this project will look into are:**

- RFM Analysis of regular customers
- Bike related purchases over the last 3 years
- Age groups of regular and new customers
- Wealth segmentation of regular and new customers
- Mobility of new customers
- Most profitable postcodes

# Data Exploration

## Data Quality Assessment

Data Quality Dimensions:

- **Accuracy:** Correct Values
- **Completeness:** Data Fields with Values
- **Consistency:** Values Free from Contradiction
- **Currency:** Values up to Date
- **Orderliness:** Data Containing Allowable Values
- **Relevancy:** Data Items with Value Meta-Data
- **Uniqueness:** Records that are Duplicated

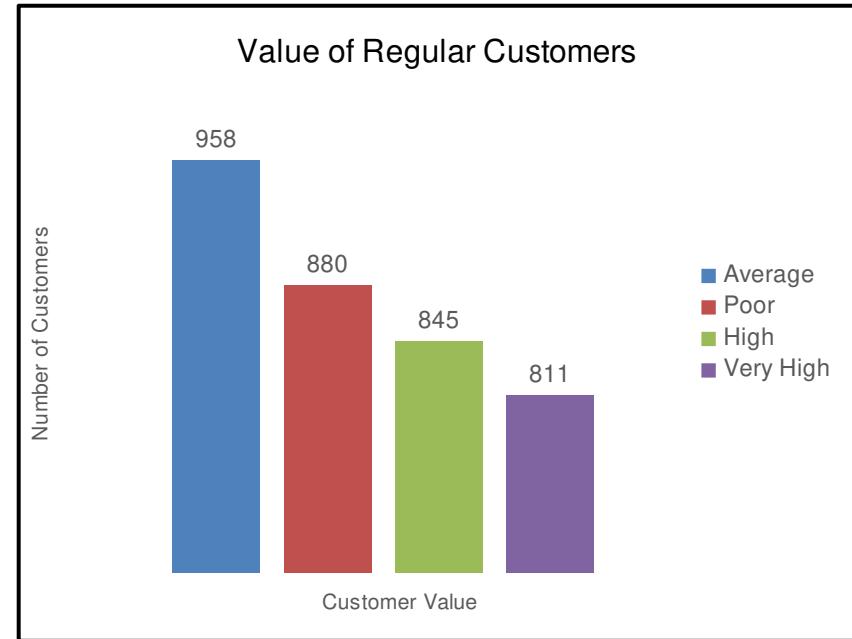
	<i>Accuracy</i>	<i>Completeness</i>	<i>Consistency</i>	<i>Currency</i>	<i>Orderliness</i>	<i>Relevancy</i>
Transactions	<ul style="list-style-type: none"><li>• Missing column 'profit'</li></ul>	<ul style="list-style-type: none"><li>• 'customer_id' incomplete</li><li>• 'brand' &amp; 'online_orders' have blank values</li></ul>			<ul style="list-style-type: none"><li>• Incorrect formats for 'list_price', 'product_sold_date'</li></ul>	
New Customer List		<ul style="list-style-type: none"><li>• 'Q', 'R', 'S', 'T' 'U' are missing column names</li></ul>		<ul style="list-style-type: none"><li>• Deceased customers filtered out</li></ul>	<ul style="list-style-type: none"><li>• 'postcode' has incorrect format</li></ul>	<ul style="list-style-type: none"><li>• 'Q', 'R', 'S', 'T' 'U' filtered out</li></ul>
Customer Demographic	<ul style="list-style-type: none"><li>• 'DOB' is inaccurate</li><li>• Missing column 'age'</li></ul>	<ul style="list-style-type: none"><li>• 'job_title' and 'customer_id' have blank values</li></ul>	<ul style="list-style-type: none"><li>• Gender specification inconsistent</li></ul>	<ul style="list-style-type: none"><li>• Deceased customers filtered out</li></ul>		<ul style="list-style-type: none"><li>• Column 'default' deleted</li></ul>
Customer Address		<ul style="list-style-type: none"><li>• 'customer_id' is incomplete</li></ul>	<ul style="list-style-type: none"><li>• 'state' is inconsistent</li></ul>			

A detailed description has been sent via Email

# Data Exploration

## RFM (Recency, Frequency, Monetary) Value of Regular Customers

- The majority and a total of **958** regular customers, had an RFM value between **144** and **311**, what makes them the customers with average value for Sprocket
- On the second place are the customers with poor ratings having an RFM value of **880** points, and high rated customers on third place having a value of **845** points
- The most valuable customers have RFM values of **411** and over, and a total number of **811** customers belong to that „Very High“ group



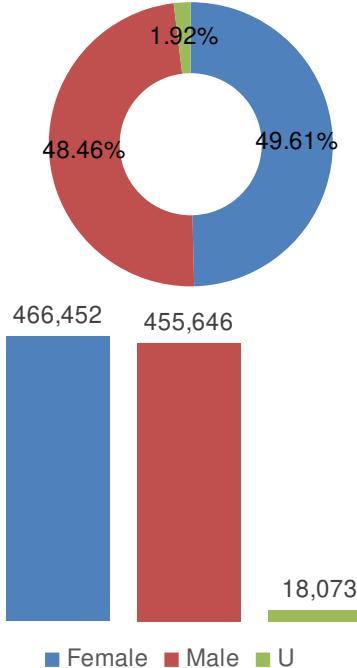
RFM Measures: MIN: 111 Q1: 144 MED: 311 Q3: 411 MAX: 444

# Data Exploration

## Bike Related Purchases in the last 3 Years

- Women make the majority of bike related purchases
- The bike related purchases that were made in the last 3 years, were mostly made by women **50%**, closely followed by men **48%**.  
The customers with unknown gender make approximately **2%** of bike related purchases.
- From the numerical perspective, women have purchased around **466,452** bike related purchases, **11,000** more than the second placed men (**455,646**)

Bike Related Purchases last 3 Yrs.  
(Regular Customers)

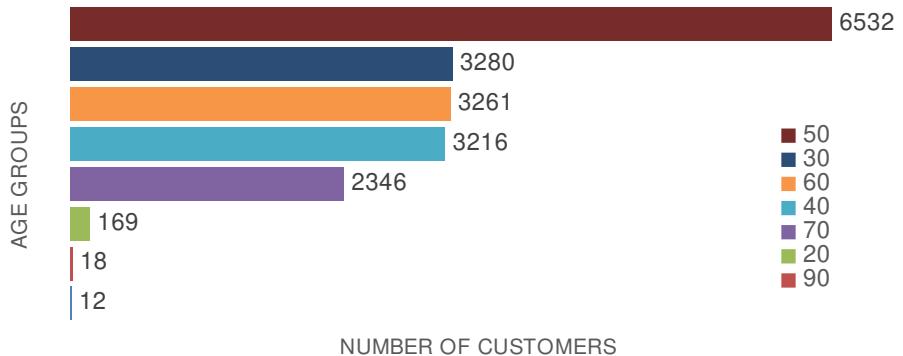


# Data Exploration

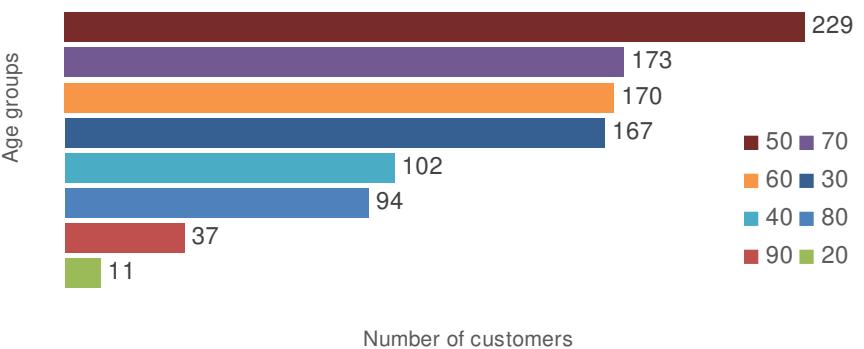
## Customer Age Groups

- The age group **50** makes the largest part of the regular as well as the new customers
- The age groups **30, 60** and **40** make about 50% less customers in the regular customers group than the age group **50**, whereas in the new customer group, the age groups **70, 60** and **30** make the second, third and fourth place
- Surprisingly the age group **20** isn't well presented neither in the regular customers group as well as the new customers group
- Unsurprisingly the age groups **80** and **90** make the bottom downs of the lists in both customer groups

Regular Customers Age Groups



New Customers Age Groups

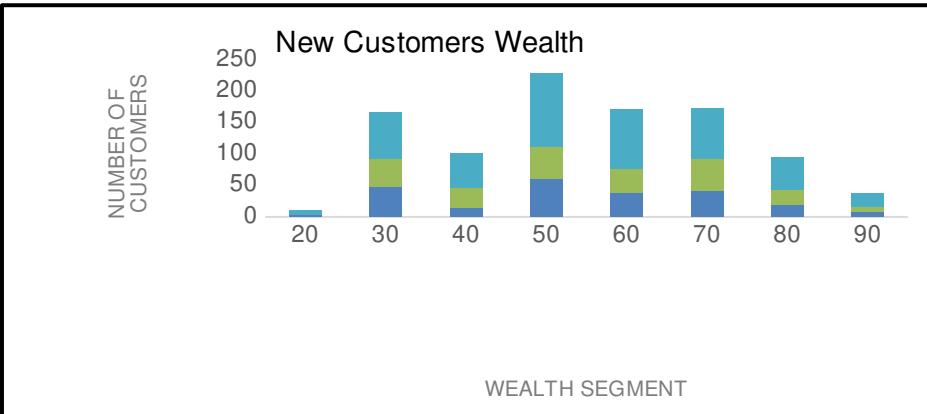


# Data Exploration

## Customer Wealth Segmentation by Age Group

The Mass Customers are most represented in all of the age groups

- On the second place is the **High Net Worth** group, mostly represented in the age group 50
- Affluent Customers** are on the third place, also mostly represented in the age group 50



# Data Exploration

## Customer Mobility

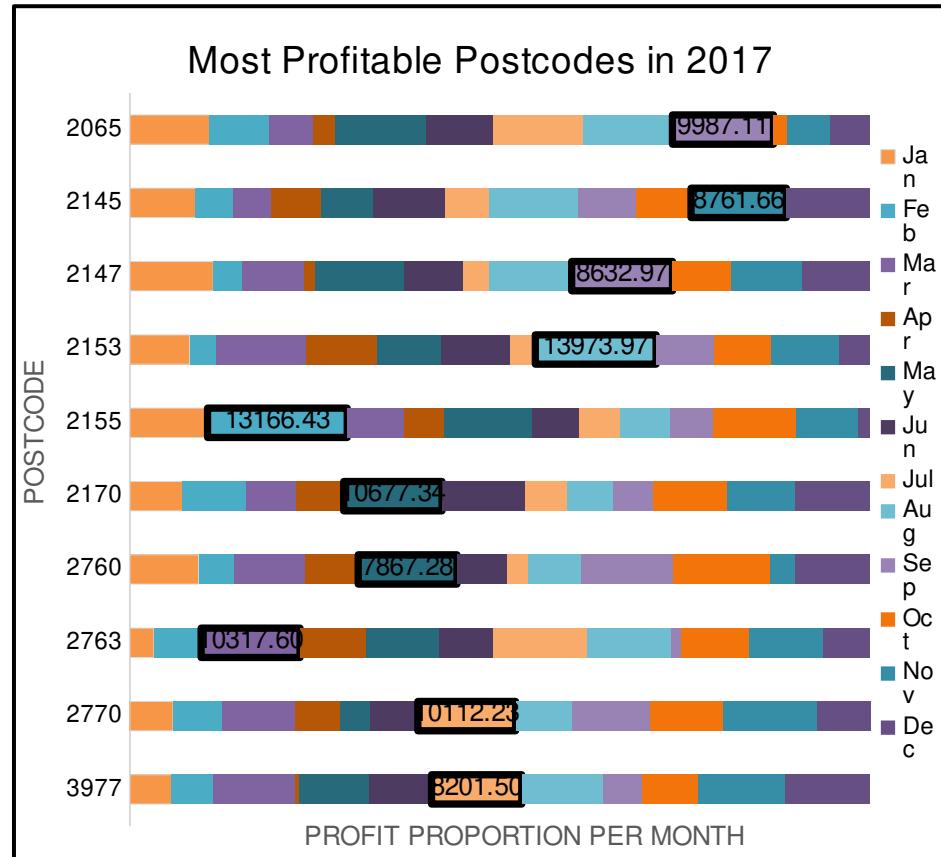
- Most new customers, **507** of them *do not* own a car, and **493** *own* one
- The majority of customers in New South Wales (NSW) *do not* own a car
- In the states Queensland (QLD) and Victoria (VIC), the majority of customers *own* a car



# Data Exploration

## Top 10 Most Profitable Months per Postcode

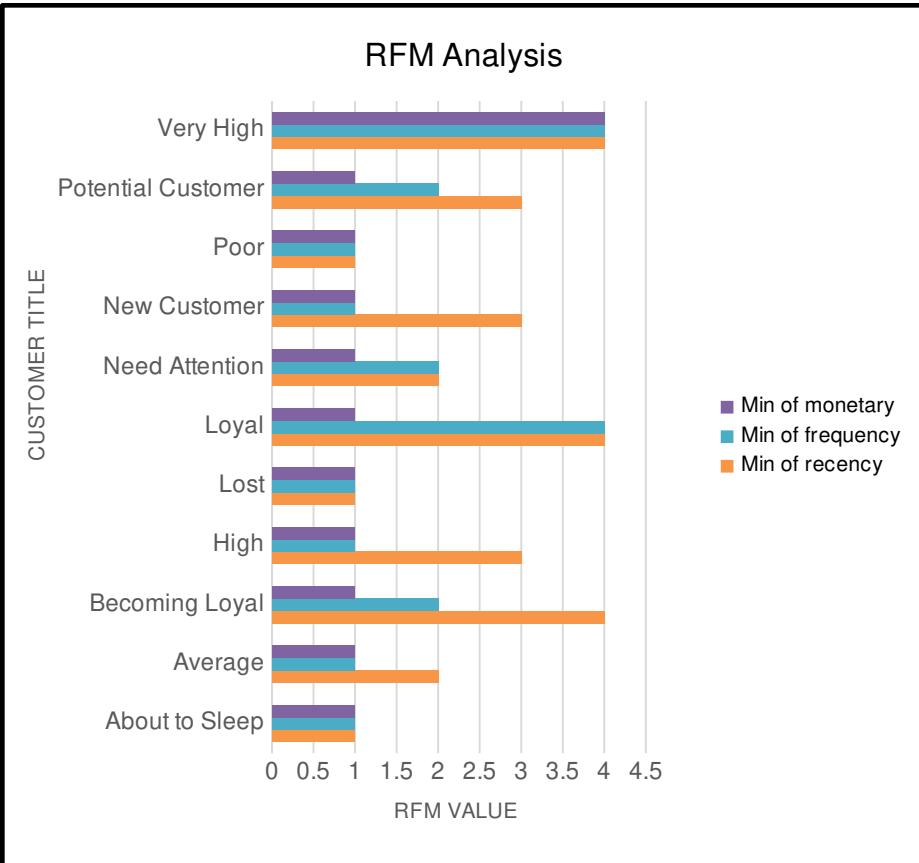
- The Postcode where the most profit is coming from, is marked with the number **2153**, in which the most profitable month (Aug.) has made **13.973,97 Dollars**, and a total of **85,628.39 Dollars** were made
- On the second place is the Postcode **2170**, which has made **80,420.06 Dollars** in total, with the most profitable month being May
- The least profitable Postcode is **2760** which has made a total profit of **58,934.17 Dollars**, and it also contained the least profitable month (May) with **7,867.28** in profit



# Model Development

## RFM Customer Classification

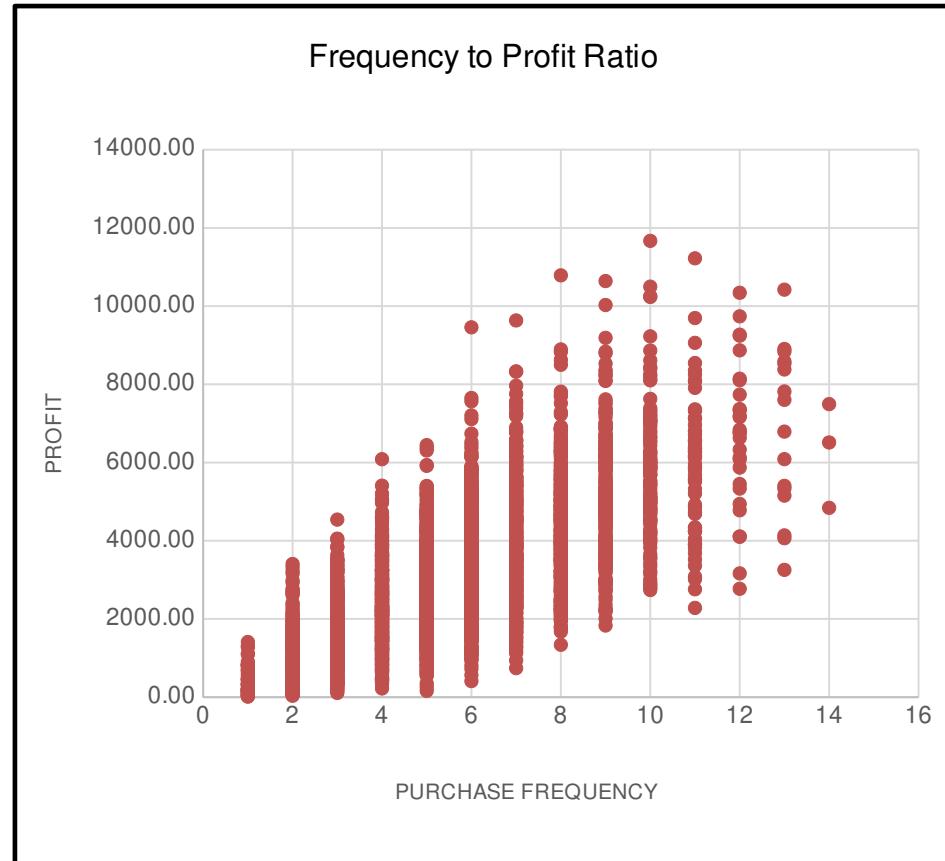
- The **RFM value** determines which customers a business should target to increase its revenue and value
- With the help of the RFM value, we can see which customers have a high level of engagement with the business in the three mentioned categories



# Model Development

## Scatter-Plot Formed on RFM Analysis

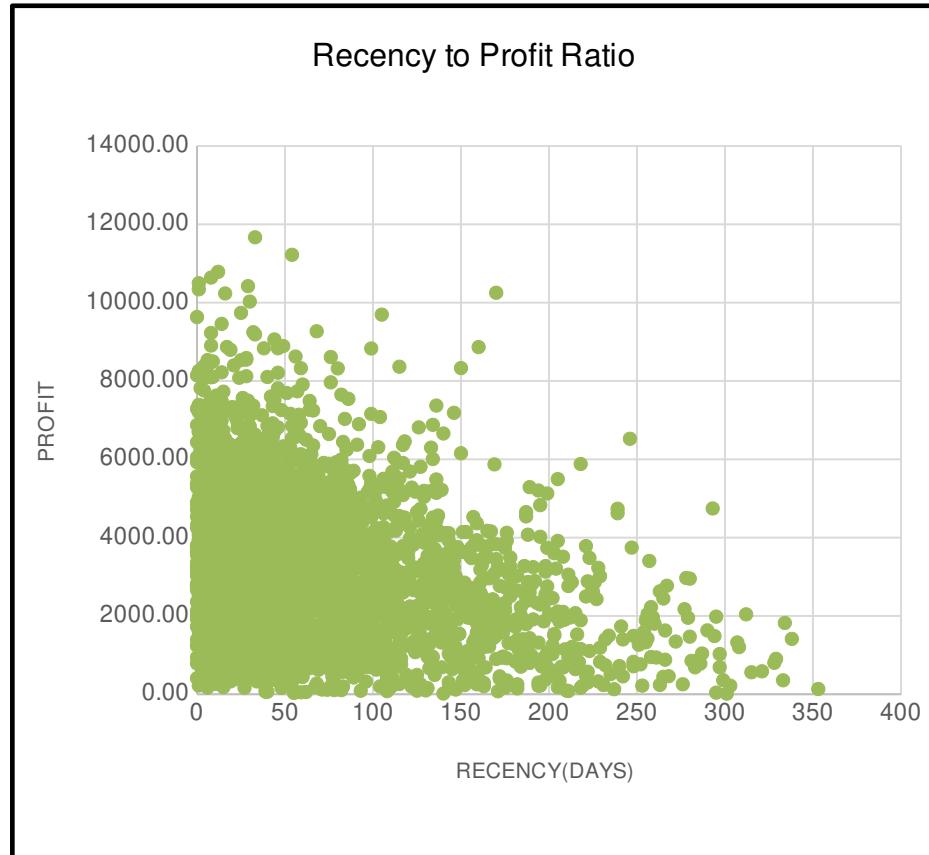
- The customers who are classified as **Very High, Loyal** and **Becoming Loyal** visit more frequently which correlates with increased revenue for the business
- As expected, there is a positive relationship between frequency and profit increase



# Model Development

## Scatter-Plot Formed on RFM Analysis

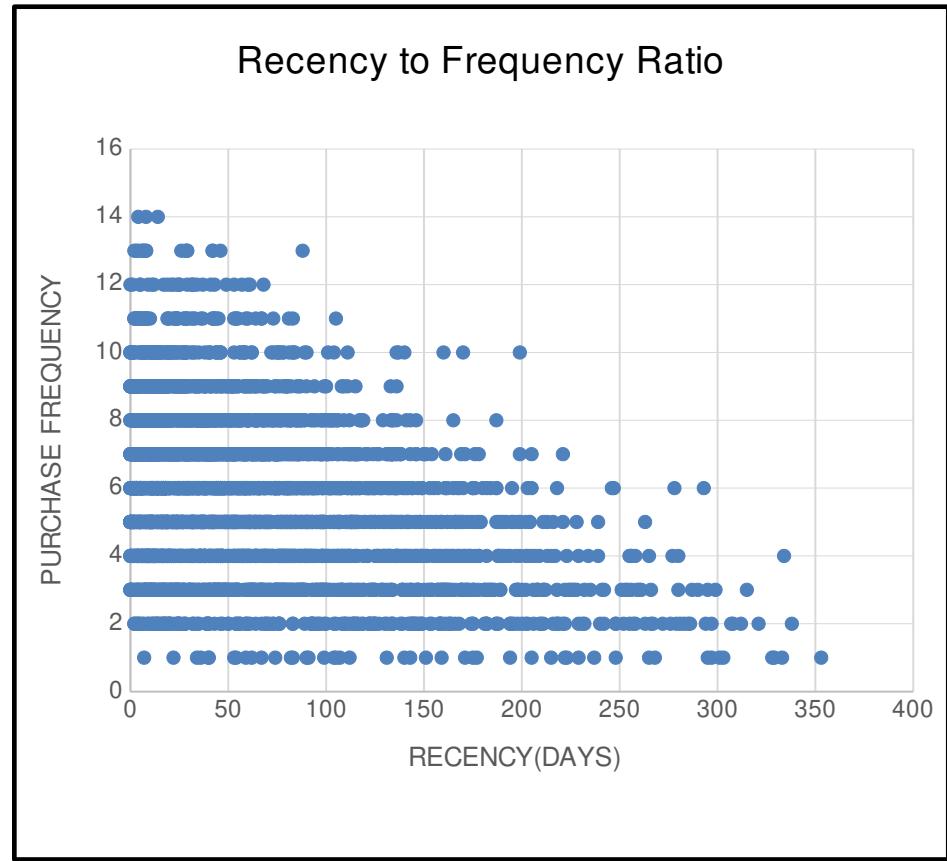
- Customers who have purchased more recently, have generated more revenue in comparison with customers whose purchase is a while back
- Almost all the way up to 100 days, the generated revenue was somewhat moderate
- After 180 – 200 days the revenue was very low



# Model Development

## Scatter-Plot Formed on RFM Analysis

- Customers who have visited more recently (0 – 50 days), will most likely visit more frequently in the future
- Very low frequency of 0 – 2 is correlated to high recency ratio
- The frequency is negatively correlated with recency, which means that more recent customers, bring value more frequently for the business



# Appendix

# Appendix

## Customer Group Definitions

Rank	Customer Group	Description	RFM Value
1	Very High	Most recent purchase. Buys often. Biggest spender.	444
2	Loyal	Most recent purchase. Buys often. Spends a lot of money.	433
3	Becoming Loyal	Relatively recent purchase. Bought a couple of times. Spends a lot of money.	421
4	New Customer	Recent purchase. Bought rarely. Spends average amount of money.	344
5	Potential Customer	Recent purchase. First buy. Spends small amount of money.	323
6	High	No purchases recently. RFM value larger than average.	311
7	Need Attention	Purchase a while ago. RFM value is below average.	224
8	Average	Purchase was long time ago. Frequency was high. Amount spent was high.	212
9	About to Sleep	Low recency. Low frequency. High amount spent.	124
10	Poor	Low recency. Low frequency. Small amount spent.	112
11	Lost	Very low RFM value.	111