April 6th , 2022

PY-Sales-Optimizer

(Proposal for report structure)

Benjamin Ebewein

Henry Ye

Michael Kroschel

Xiuli Dong

# Table of contents

[Table of contents 1](#_Toc99618576)

[Introduction 2](#_Toc99618577)

[1 Data description 2](#_Toc99618578)

[2 Data exploration & pre-processing 2](#_Toc99618579)

[2.1 Exploration of selected features 2](#_Toc99618580)

[A Transactions, orders & status 2](#_Toc99618581)

[B Price, discount amount & grand total 3](#_Toc99618582)

[C Consumption categories 3](#_Toc99618583)

[D Customer lifetime 4](#_Toc99618584)

[2.2 Further data pre-processing 4](#_Toc99618585)

[3 RFM segmentation 4](#_Toc99618586)

[3.1 Short introduction to RFM concept 4](#_Toc99618587)

[3.2 RFM scores & baseline model 4](#_Toc99618588)

[3.2.1 Customer scoring 4](#_Toc99618589)

[3.2.2 Baseline segmentation 5](#_Toc99618590)

[3.3 Clustering models & exploration 5](#_Toc99618591)

[3.4 Final model & segmentation 7](#_Toc99618592)

[4 Marketing Strategy 7](#_Toc99618593)

[4.1 WHO? – Target segments 7](#_Toc99618594)

[4.2 HOW? – Offer & discount strategy 8](#_Toc99618595)

[4.3 WHEN? – Windows of opportunity 8](#_Toc99618596)

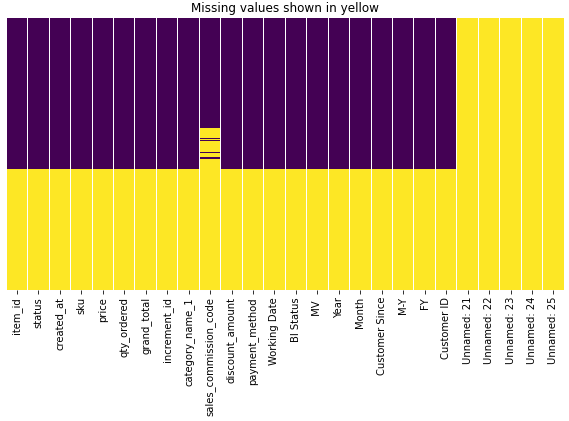
[Conclusion 9](#_Toc99618597)

# Introduction

* *Context of RFM methodology and its application/usage*
* *Project scope and goal*
* *Rough overview of the structure and content of the document*

# Data description

* *Origin of data and contextualization*
* *Structure of data-set*
* *Overview (Table format) with Name, Type, Description per variable*
* *General data quality (duplicates, missing values,..)*

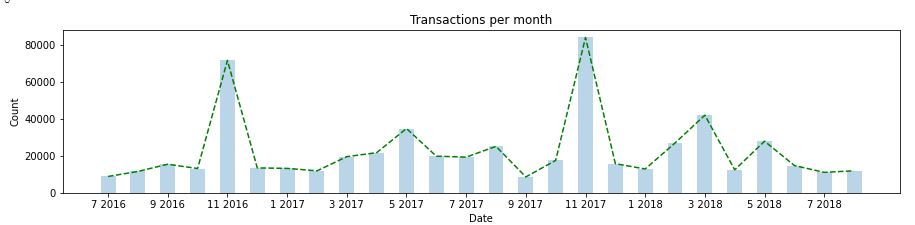


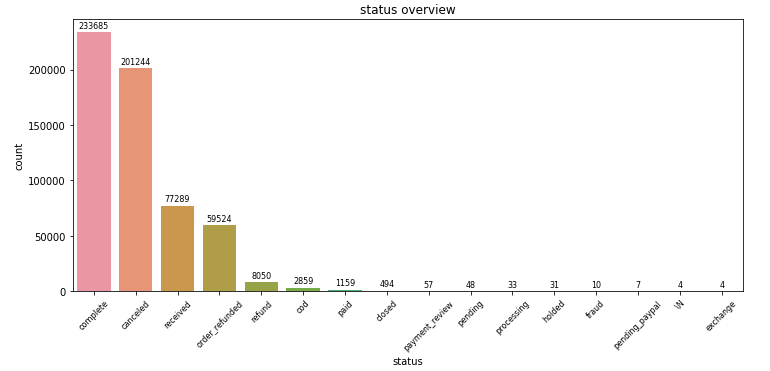
# Data exploration & pre-processing

## Exploration of selected features

### Transactions, orders & status

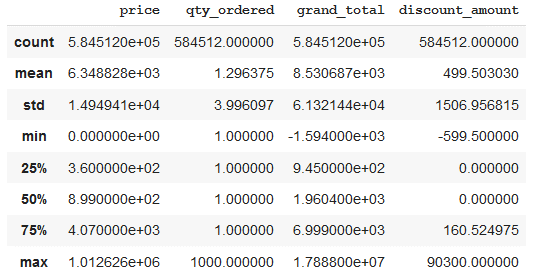
* *Summary of relevant insights and dependencies*
* *Implications for pre-processing*





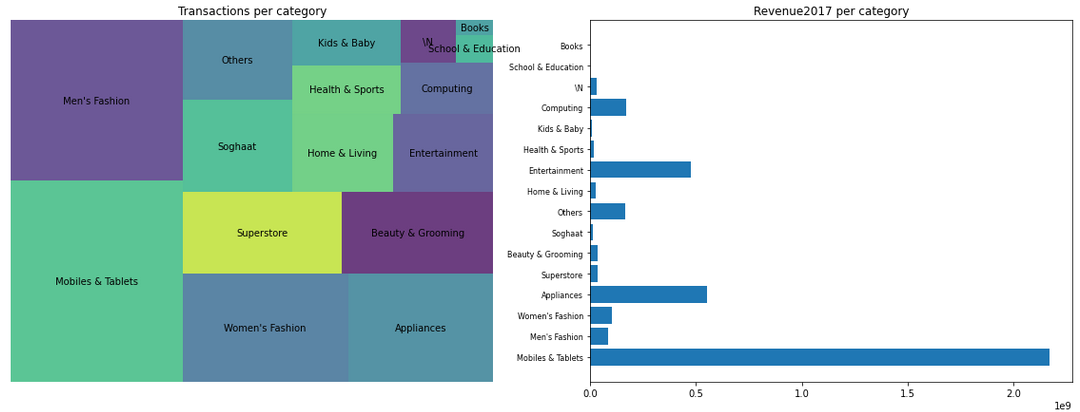
### Price, discount amount & grand total

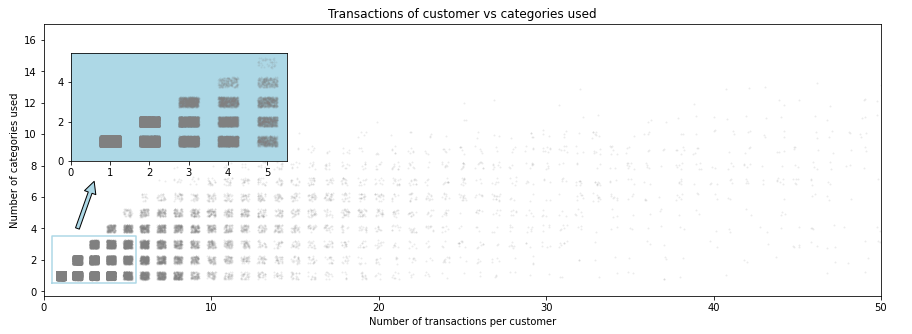
* *Summary of relevant insights and dependencies*
* *Implications for pre-processing (eg reasoning for creating new variable)*



### Consumption categories

* *Summary of relevant insights and dependencies*
* *Implications for pre-processing (eg reasoning for creating new variable)*





### Customer lifetime

* *Introduction of variable*
* *Reasoning for usage*

## Further data pre-processing

* *Summary of additional pre-processing steps*

# RFM segmentation

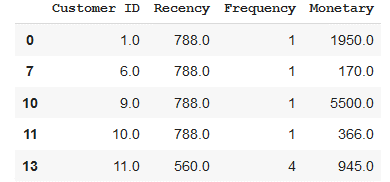
## Short introduction to RFM concept

* *Short summary of theory behind RFM: Principals, RFM scores and customer segments*

## RFM scores & baseline model

### Customer scoring

* *Description of scores, statistics and further pre-processing/cleaning*



Relative Frequency to be added in table left and resulting RFM scores

Additional visualizations of distribution and statistics of variables

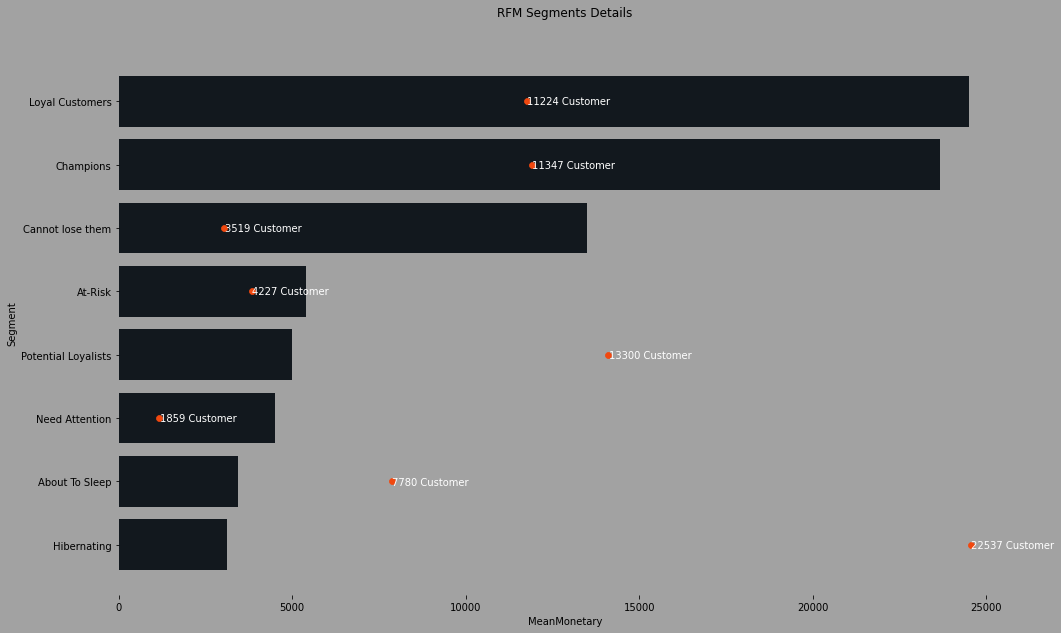
Distribution

* Recency
* Frequency
* Monetary

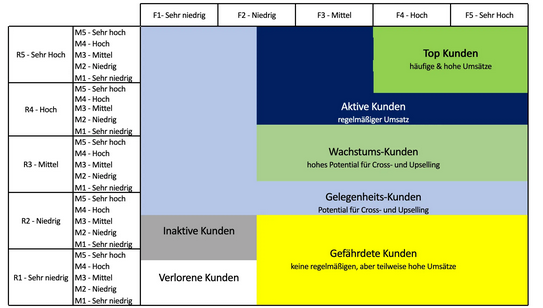
### Baseline segmentation

* *Segment customers according to “standard segmentation” from literature (see eg charts from Step4, but based on all 3 dimensions)*
* Champions: Bought recently, buy often and spend the most
* Loyal customers: Buy on a regular basis. Responsive to promotions.
* Potential loyalist: Recent customers with average frequency.
* Recent customers: Bought most recently, but not often.
* Promising: Recent shoppers, but haven’t spent much.
* Needs attention: Above average recency, frequency and monetary values. May not have bought very recently though.
* About to sleep: Below average recency and frequency. Will lose them if not reactivated.
* At risk: Some time since they’ve purchased. Need to bring them back!
* Can’t lose them: Used to purchase frequently but haven’t returned for a long time.
* Hibernating: Last purchase was long back and low number of orders. May be lost.



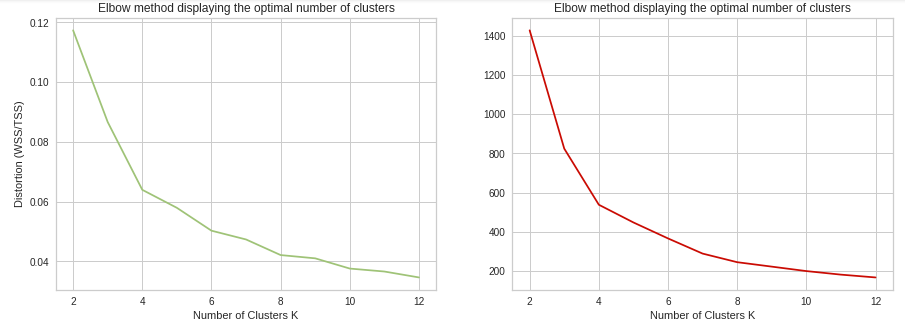


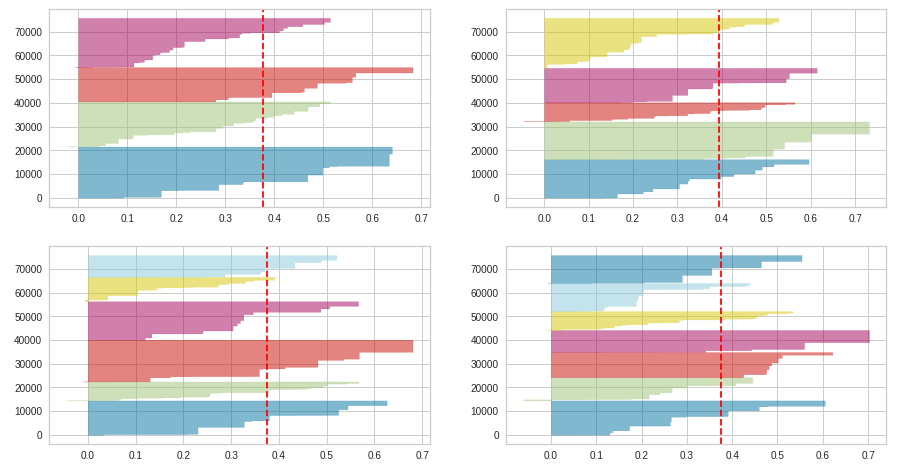
* *Evaluate segmentation according to different aspects (applicability, silhouette-scores, etc)*

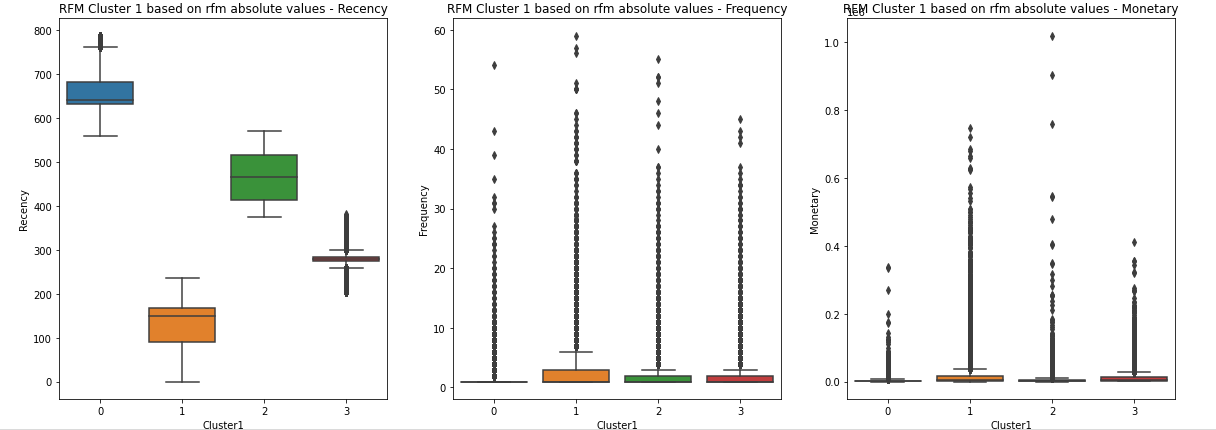


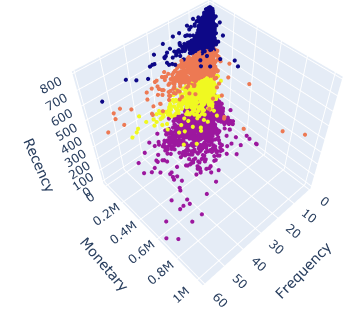
## Clustering models & exploration

* *Describe models we tested* 
  + *Absolute*
  + *Scores*
  + *Multiple models / joined models*
  + *Skewness absolute*
    - *New distribution (if applicable)*
    - *Best numbers of clusters: alternatively, a table to compare numbers of clusters for all models, the less is preferable for this measure only*
    - *Chart clusters vs customer segments*
    - *Average RFM per clusters*
    - *Relative important per RFM – heat map*
    - *Snake chart*
    - *3D visualisation etc.*
  + *Skewness absolute + lifetime value vs relative frequency*
  + *Comment on standardization*
* *Create overview of results and usability for our goal (use tables for comparison and easier reading)*
* *Suggestion to use visualization only for 1 model (and use back-up for additional ones). Maybe we can put several models in one graph (eg elbow-graph)*



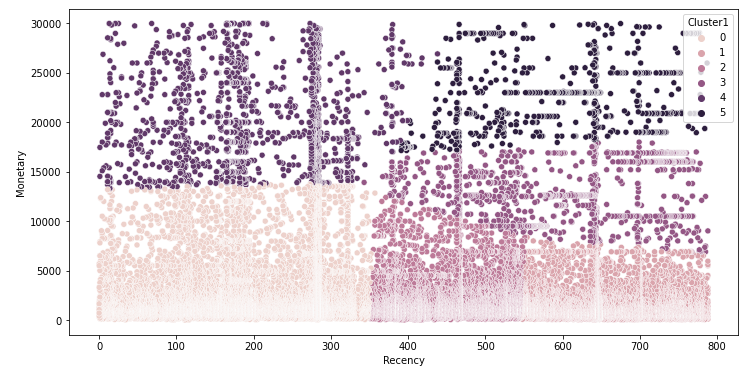


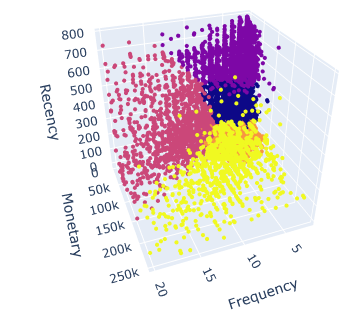




## Final model & segmentation

* *Explain why we have chosen this approach*
* *Explain in more details concept and results and why useful.*
* *Also use Evaluation measures like above*
* *Describe clusters*



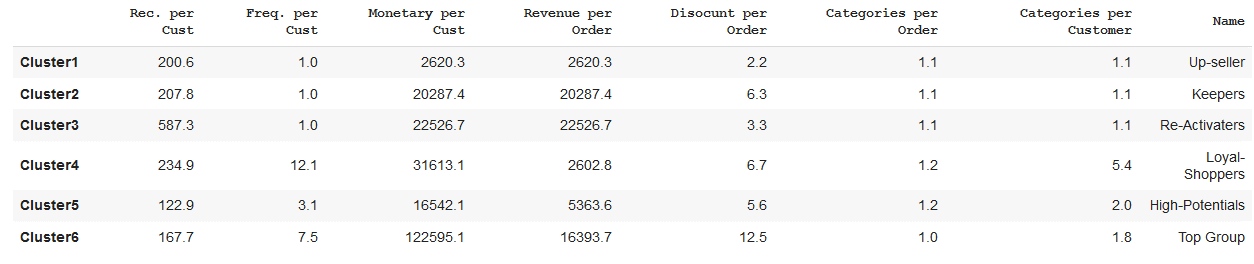


# Marketing Strategy

* *Describe objectives of our marketing strategy (right message to right person in right time)*

## WHO? – Target segments

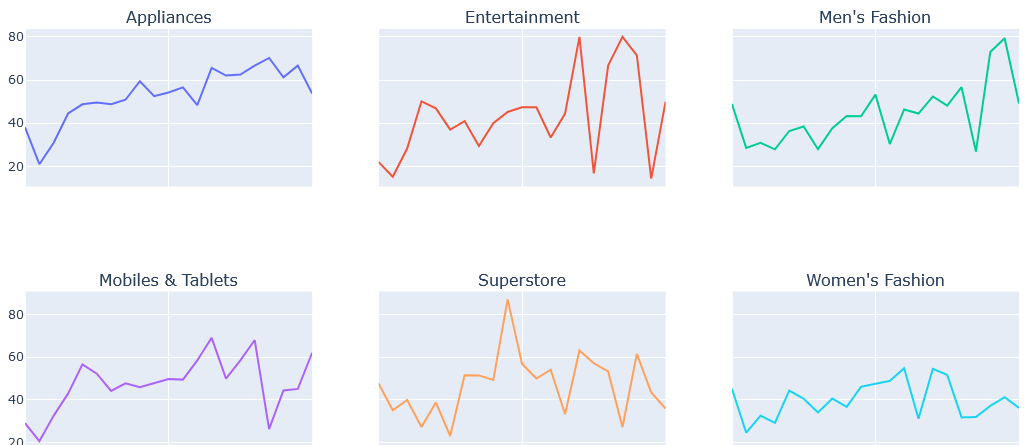
* *Give rational for selected target segments and description*



## HOW? – Offer & discount strategy

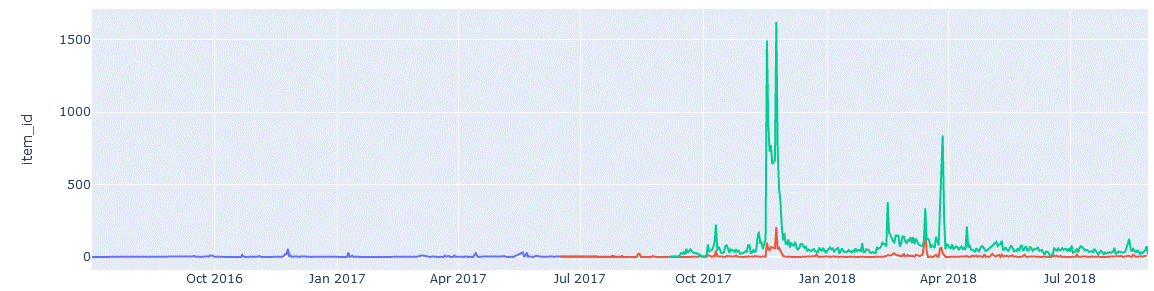
* *Explain offer (product /message) strategy and discount strategy*

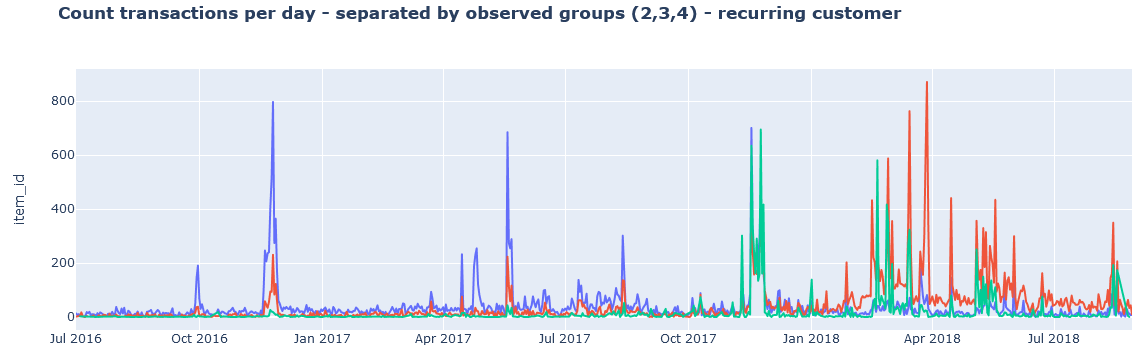


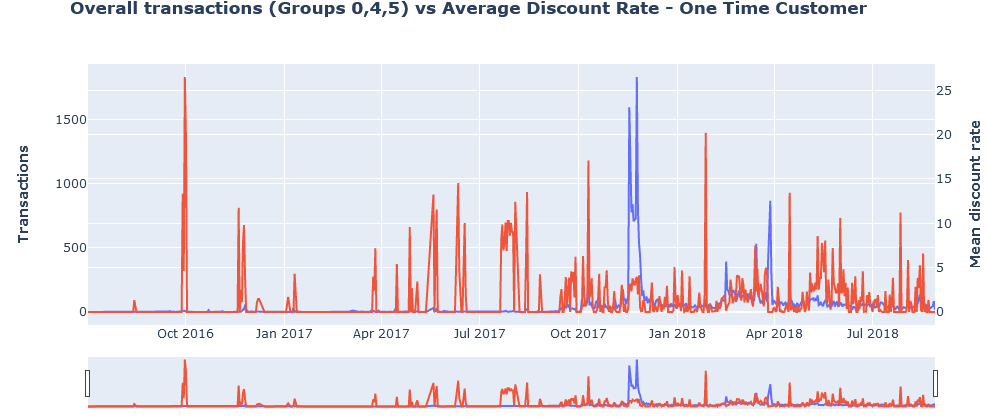


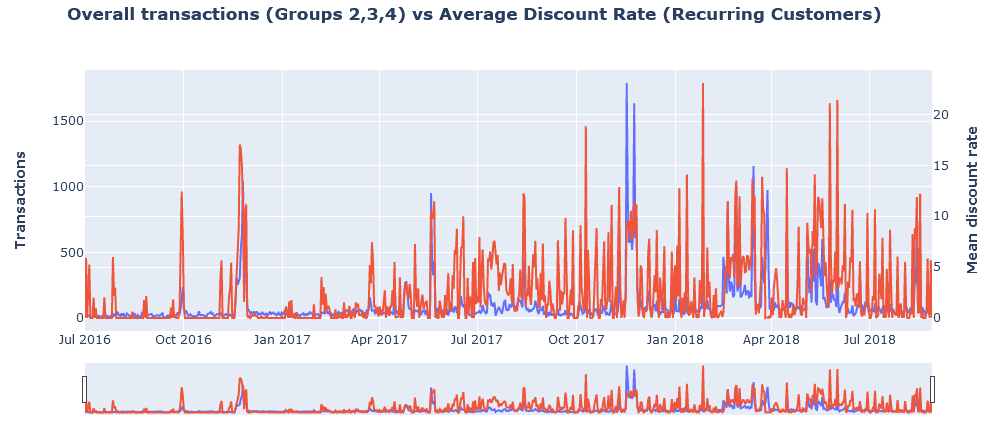
## WHEN? – Windows of opportunity

* *Explain findings on best timing for activation*









# Conclusion

* *Wrap-up*
* *Recap, did we achieve our goals?*
* *Qualitative evaluation of usability of approach*
* *Summary of learnings*
* *Outlook*