

The Wolt logo is a blue circle with the word "Wolt" in white, italicized script. It is centered within a larger, light gray circle. The background of the slide is white with decorative elements: a blue circle in the top left, an orange circle in the middle left, a green line in the top right, and a green line and an orange circle in the bottom left.

Wolt

Assignment for Data Analyst Intern

Dilhara Liyanaaratchi



Agenda

- Task
- Road Map
- Insights of the Analysis
- Steps of the process
- Approach used
- Pros and Cons

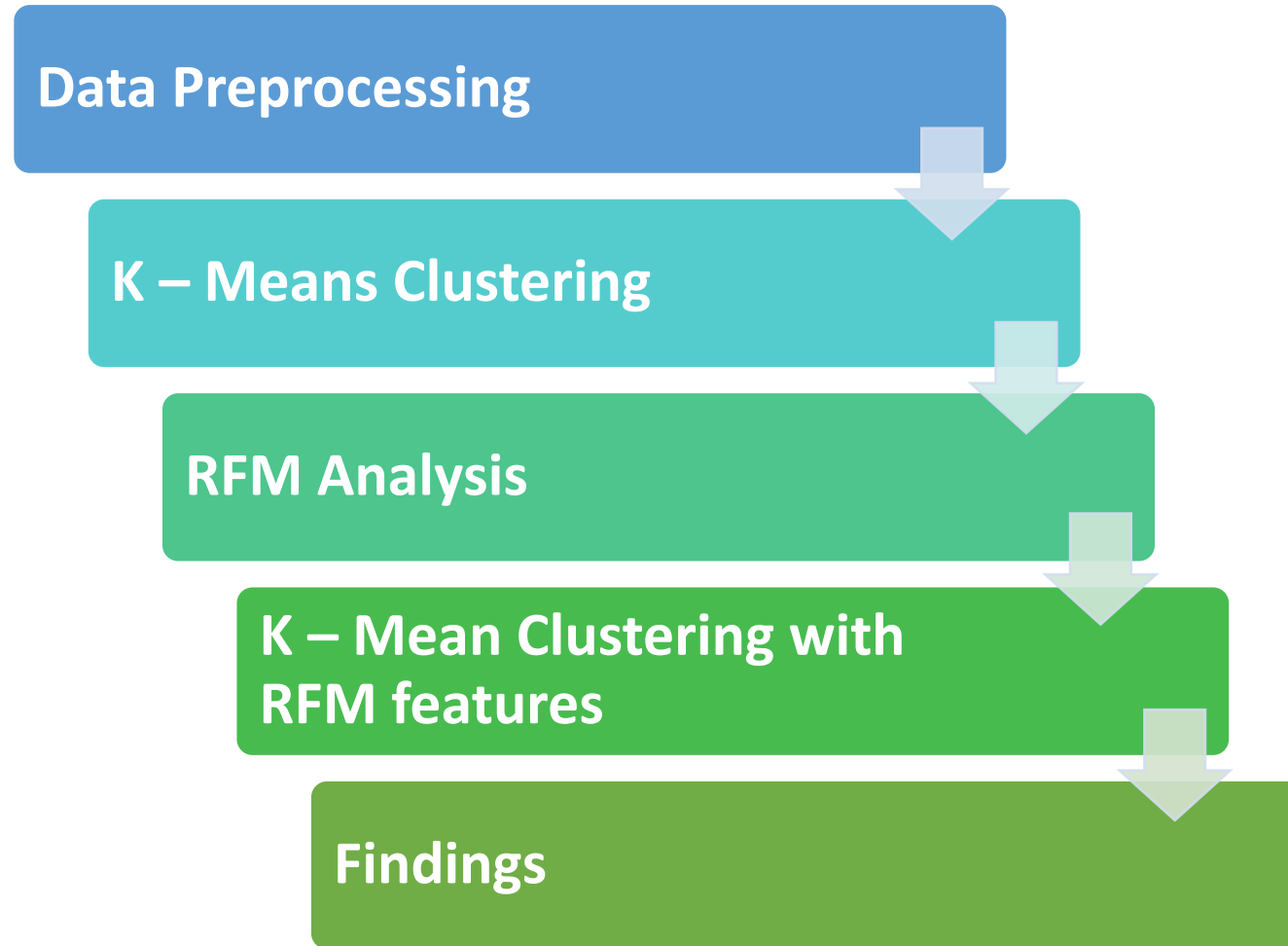


Task

The objective of the task is to identify the target user groups through customer segmentation for better understanding of the available users.

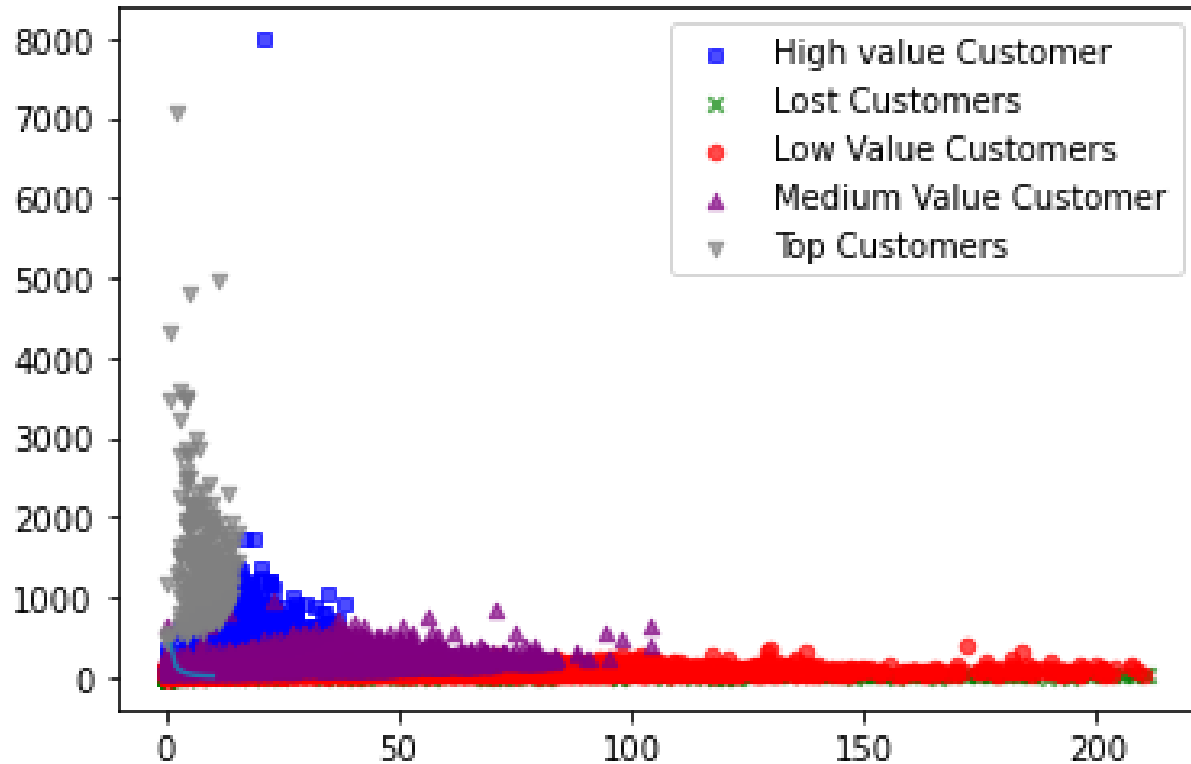


Road Map

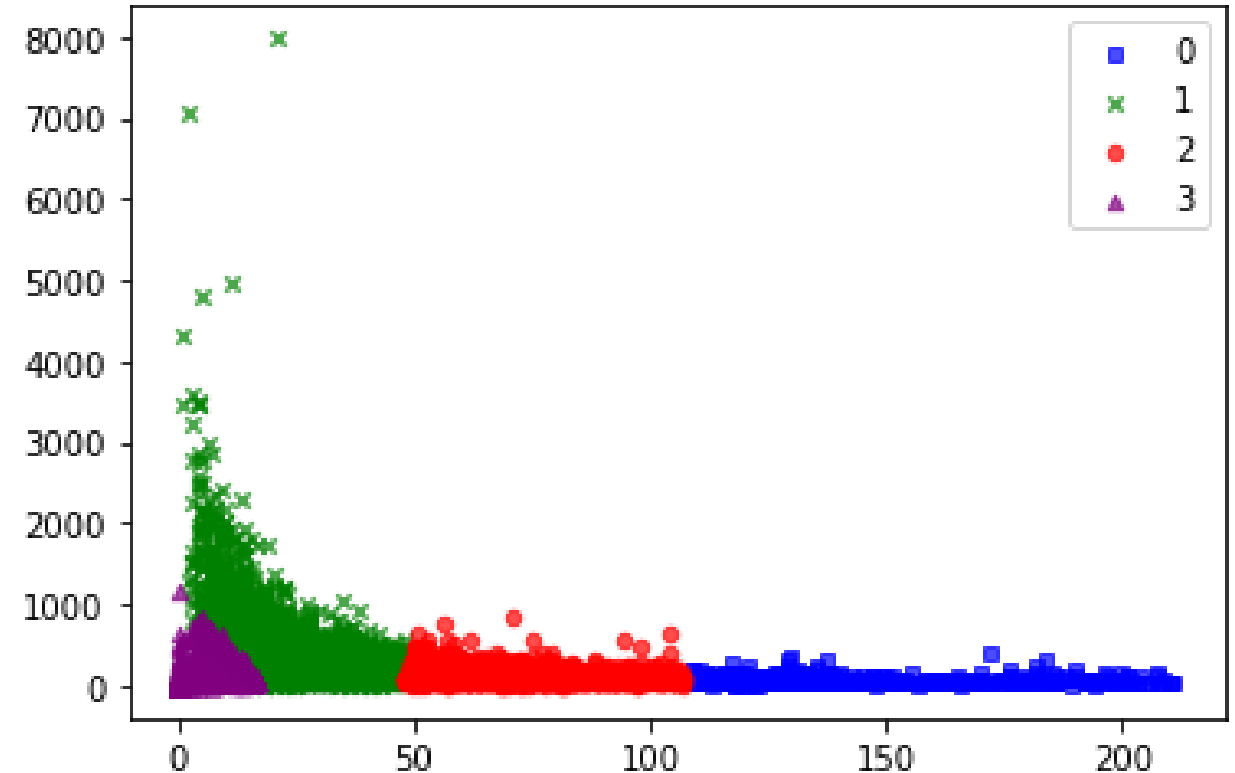


Insights of the Analysis

User Groups under RFM Analysis



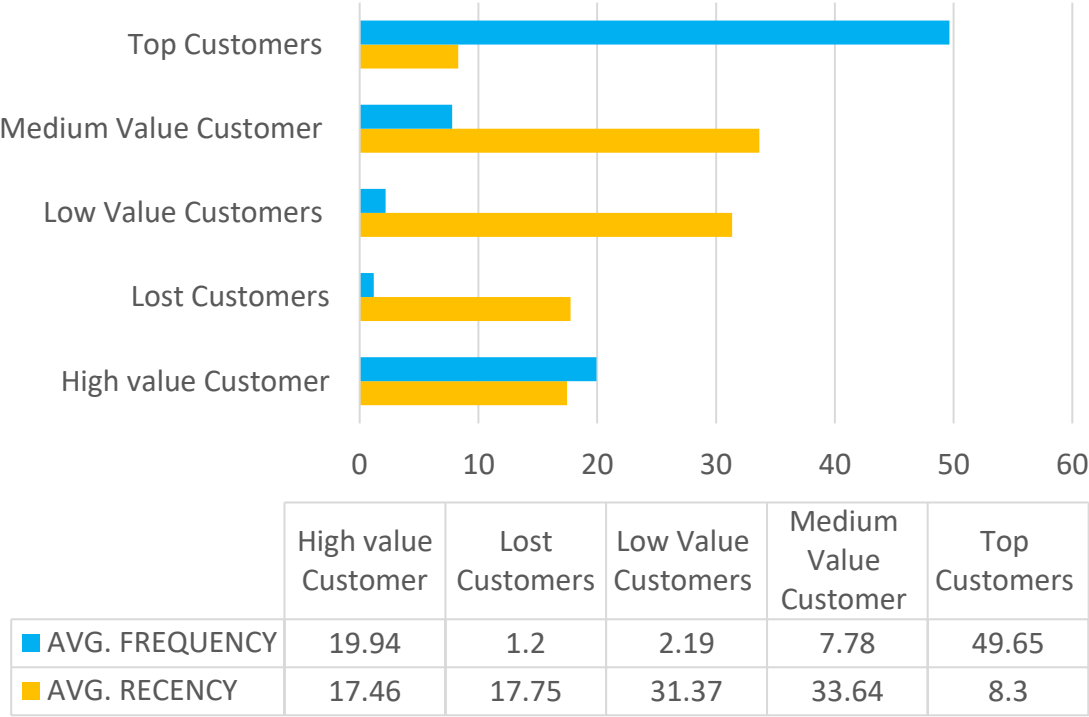
User Groups under K-Means Clustering with RFM Features



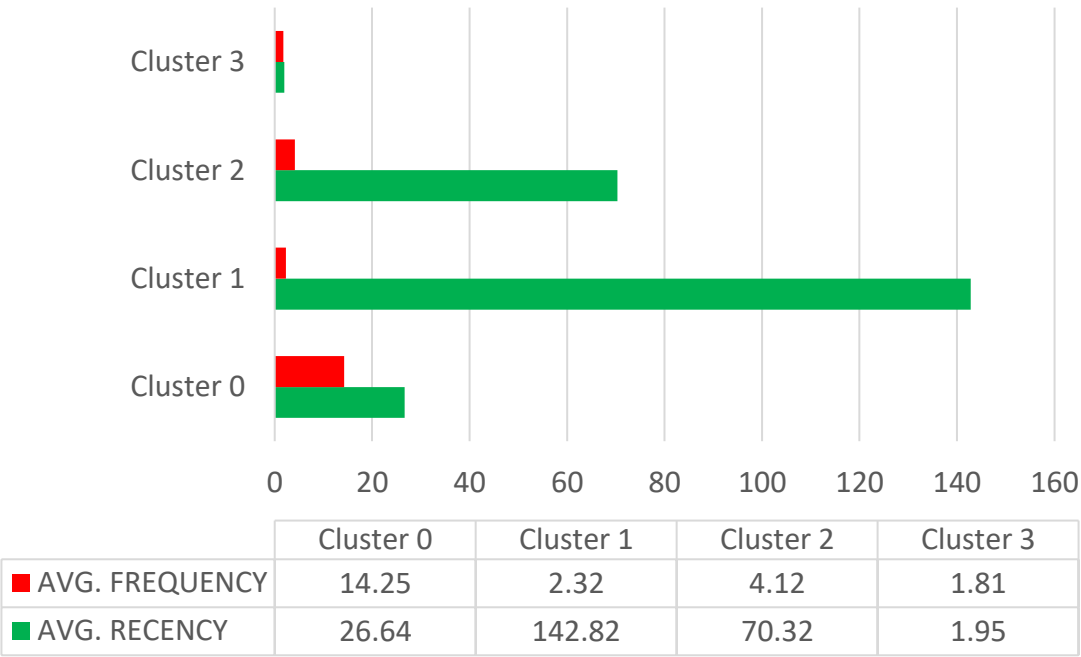
Findings: As per the derived user groups, based on the above customer segmentation approach, in order to understand the user groups, following the results of K – Means clustering with RFM features would be much advisable

Findings: The best combination of recency and frequency is, having a low recency value and high frequency value, hence the cluster 1 would be the best users

Average recency and frequency of the user groups based on RFM Analysis

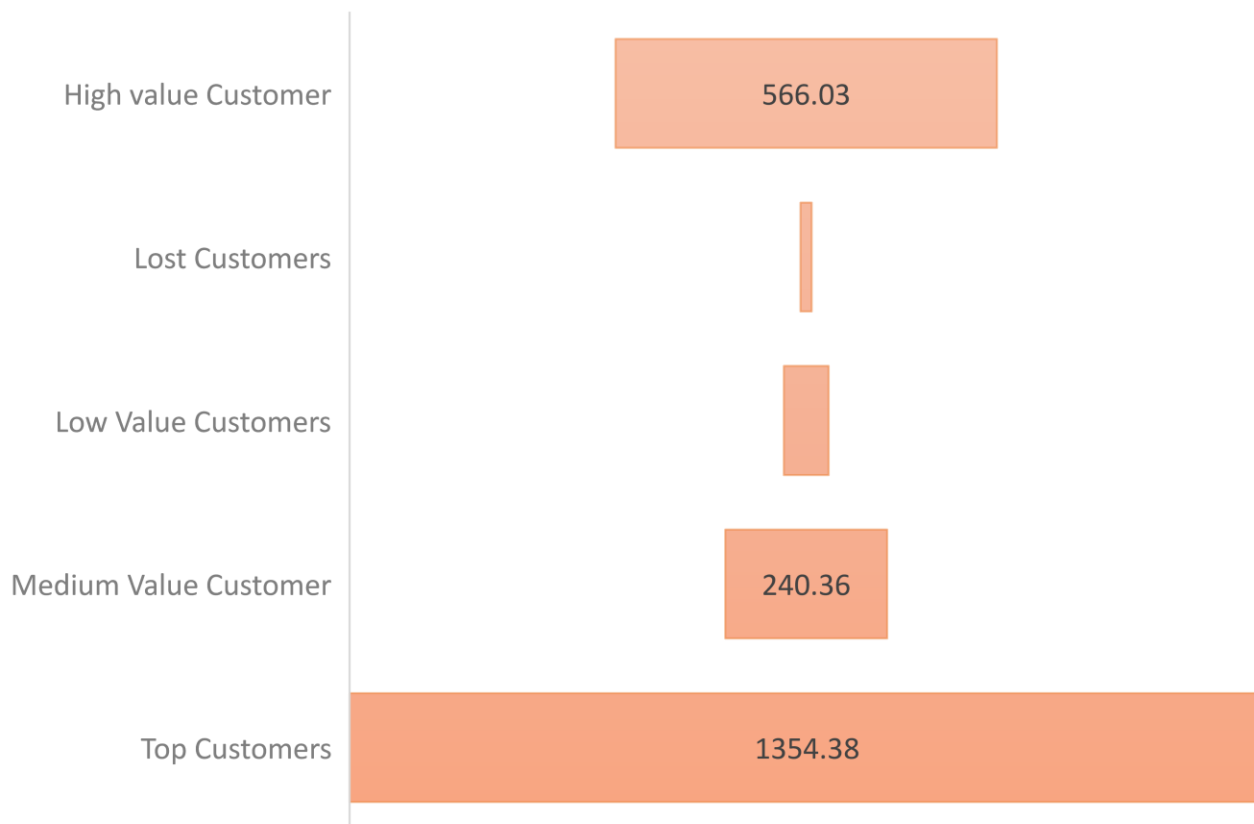


Average recency and frequency of the user groups based on K-Mean with RFM Features

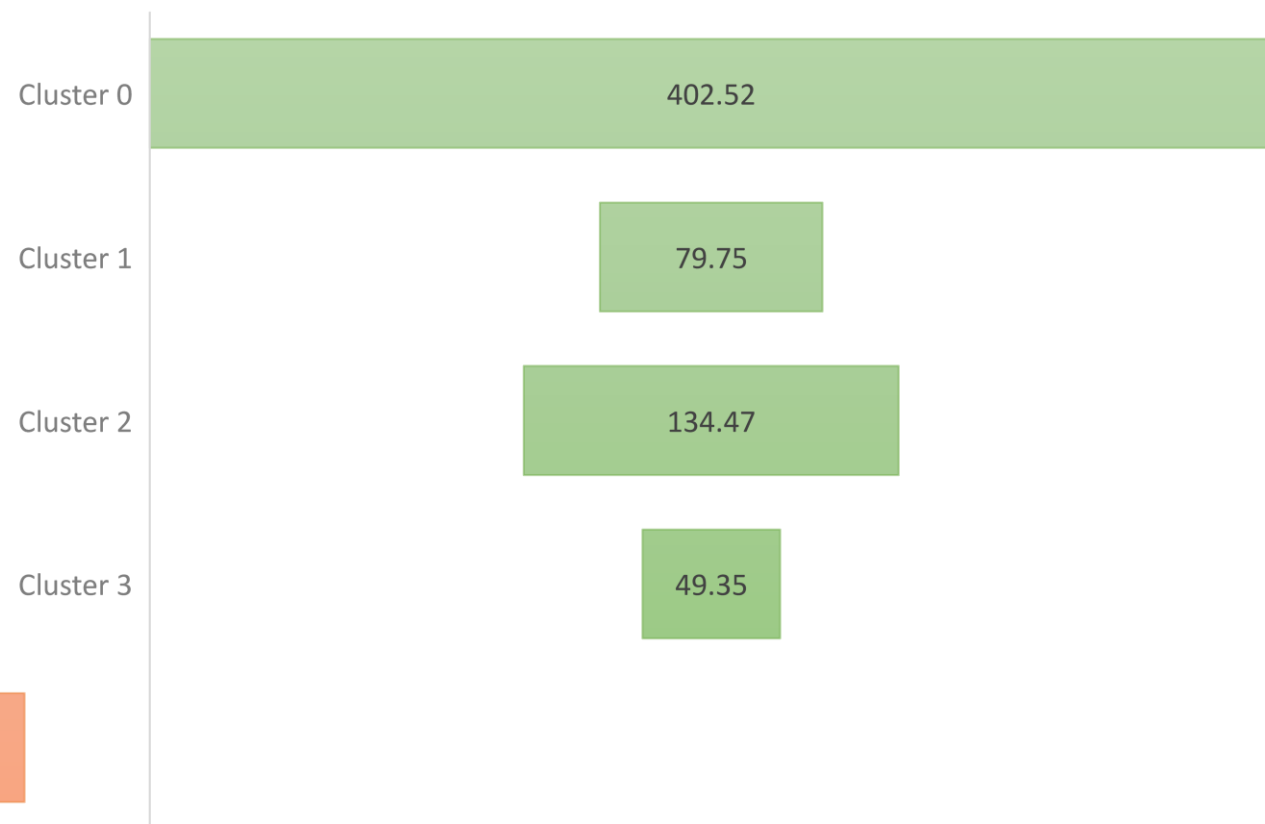


Findings: It was identified that, most influential attributes for Total Purchase in EUR are Lunch orders, Dinner orders, Restaurants, Grocery and Retail orders, and orders that are made from ios devices. And this behavior quite similar towards the total purchase orders as well.

Average Monetary of the user groups based on RFM Analysis



Average Monetary of the user groups based on K-Means clustering with RFM features



Findings: As per the RFM analysis, the users has distributed as 337 as Top users, 1053 High value users, 2544 medium users and 6272 and 1822 as low value and lost users. Whereas, as per the K-Means clustering, the best users cluster is, cluster 0 which is 3736 users and number of users belongs clusters 1, 2 and 3 are 577, 2043 and 5670 users respectively.

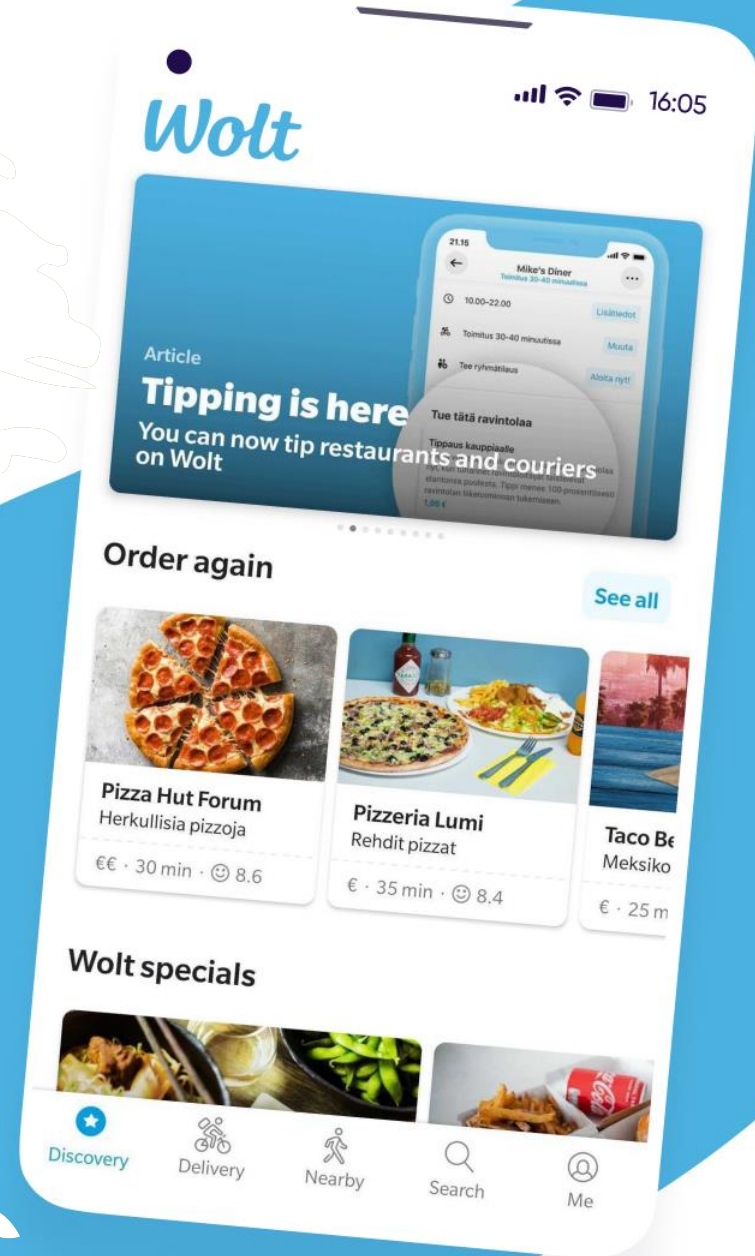
Steps of the process

- Data pre-processing – this includes data cleaning, removing outliers, removing highly correlated attributes, normalizing the data etc.
- Applied K – Means clustering with the attributes – clear clusters weren't identified
- Used RFM analysis – used, total purchase in EUR, Number of orders, average days between orders were used for M, F and R factors.
- Performed K-means clustering for the RFM attributes.



Approached Used

- *As the objective was to identify the different user groups based on the similarities such as, ordering habits and demographic and other variables, K- Means clustering approach was used. As this approach would make it easier the business to come up with tailored strategies each target groups.*
- *As the RFM Analysis, is trendy and effective customer segmentation approach that uses now a days to measure the loyalty of the customer, it was decided to include this approach as well.*
- ***As the final approach, both customer segmentation methods were combined and used for a better results***





Pros & Cons of the Approach

- *When it comes to the K – Means clustering, it is sensitive for the outliers which can be led to misleading results, but still, as it is an accurate way to interpret customer segmentation this was used, after dealing with outliers.*
- *The RFM analysis doesn't consider the behavior of the customers apart from the purchase content, but still provides insights into which customers are more valuable.*



THANK YOU !

