User segmentation of a delivery service

Victor von Eisenhart-Rothe

Content of the presentation

Executive summary

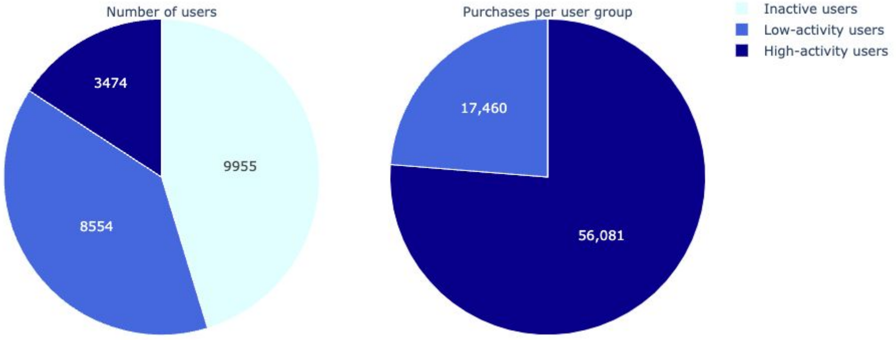
Goal: Support the marketing team (re)activate users through providing a meaningful user segmentation based on data on the consumption behaviour of customers.

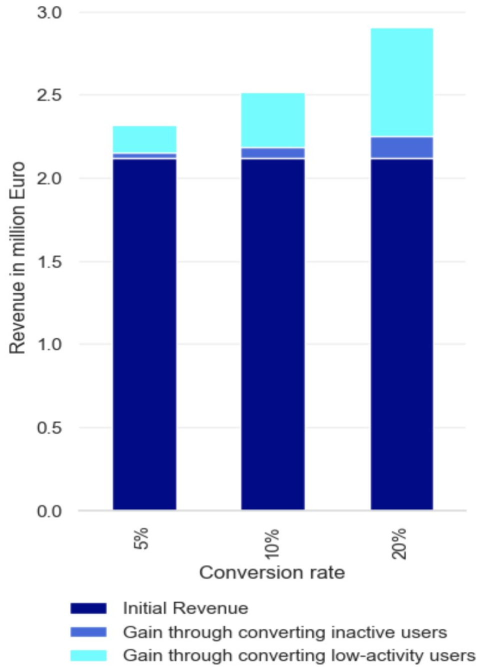
Outcome:

- **KPI:** With the amount of orders done by the users being the **core metric** that we want to improve, this metric was chosen to classify:
  - **Inactive** users (0 orders)
  - **low-activity** users (< 6 orders per year)
  - **high-activity** users (>= 6 orders per year)
- While only representing **16% of the users**, high-activity users **make up 74% of the revenue**.
- Characteristics of the different user types were carved out so that the marketing team can target users to **convert them to high-activity users**.
- The main distinguishing factor between high- and low-activity users is, that in addition to restaurants, **high-activity users order from grocery and retailer stores as well.**

What’s in it for the company?

- Most of the revenue generated comes from high-activity users
  → Aim to convert users to high-activity users
- **Focus:** What distinguishes high-activity users from other users?
- **KPI:** PURCHASE_COUNT





User Personas - Highko

Characteristics:

- **High-activity user** - Orders on a regular basis (**>=6 times per year**)
- Main revenue source of the company (74%)
- Orders mostly at **restaurants** (48%) but also uses **grocery** (21%) and **retail store service** (27%).
- Pet supply (0%) and merchandise (4%) are not in his interests
- Has his habits: Does not try out so many **distinct venues** (53 %)

Main insight for marketing: Generating users like Highko is our aim. He already uses the product and different services regularly so the marketing team should focus on nudging him with advertisements of his preferred services to further strengthen the habit of using the product.

User Personas - Lowla

Characteristics:

- **Low-activity user** - orders food every now and then, but not often (<6 times per year)
- Second revenue source of the company (26%)
- Spends a decent amount of money with each order (2.50 Euro more than high-activity users)
- Orders exclusively at **restaurants** (100%)
- Likes to try out **different venues** (90%)
- Does not know that it's also possible or never thought of ordering from other stores, such as grocery or retail store

Main message for marketing: In comparison to Highko, Lowla has never used any other service than restaurants yet! In fact, it is the main distinguishing aspect between the two. <u>The marketing team should inform Lowla about this service so that they can convert her into a high-activity user.</u>

User Personas - Average Joe:

Characteristics:

- Inactive user - registered for the **mobile app** but has **never ordered anything**
- Favorite food type is ‘**American**’, followed by ‘**Japanese**’ and ‘**Italian**’.
- Is most likely to ...
    - order around **dinner time.**
    - order at a **restaurant.**
    - use the **delivery** service rather than takeaway.

Main insight for marketing: Truth is, we don’t know a lot about Joe, as we don’t have meaningful customer data on him. <u>For marketing purposes, we try to activate him and target him with what works best for most customers - the average.</u>

How did I get to these insights?

Aim: Provide different user-personas to marketing, that resemble actual user-groups that can be targeted through advertisement specifically tailored to their characteristics to increase the amount of orders they make.

Main steps:

1. Data cleaning and transforming
2. Exploratory analyses
3. Further analyses with segmented user subsets

1. Data cleaning and transforming

- **Inspect the original dataset:**
  - pandas_profiling ProfileReport gives you an overview of the dataframe and allows you to inspect each variable
  - What needs to be cleaned or transformed?
- **Duplicates:** No duplicate rows → No further step required
- **Missing values:** Many NaNs that could be represented by "0"
- **Noise:** Unnecessary characters can be removed of strings of variables
- **PURCHASE_COUNT_BY_STORE_TYPE:** Dict. has to be exploded to columns
- **Drop columns** that don't contain valuable enough information for the business case
- **Inactive users:** No valuable information from inactive users → Split DF in subsets based on if they have purchased something already

2. Exploratory analyses

- Create a Profile Report with Pandas Profiling and inspect each variable
  - Inspect Means, Medians, Distributions and potential Outliers
  - Inspect PURCHASE_COUNT as the core metric to optimize for through marketing
  - First insights from the variables
    - PREFERRED_RESTAURANT_TYPES: Most represented value is 'American', followed by 'Japanese' and 'Italian'
    - STORE_TYPES: restaurants are most frequently used service, followed by retail_store and grocery
    - Device: 83% of orders are taken via mobile device

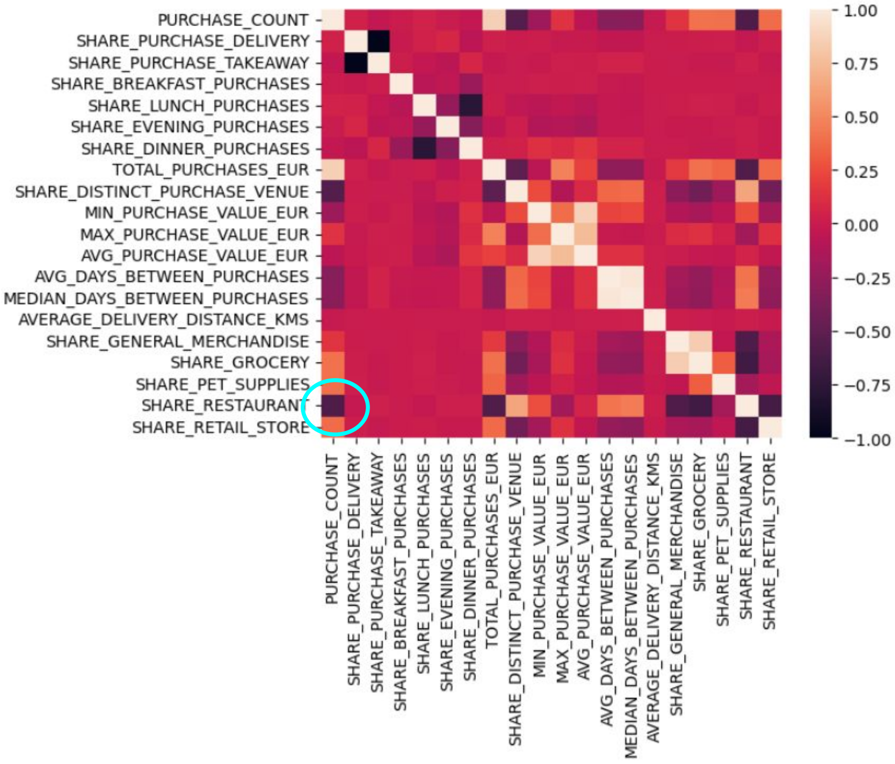
2. Exploratory analyses

- Create a Profile Report with Pandas Profiling and inspect each variable
  - Drop further columns that are not needed anymore
  - Inspect correlations
    - PURCHASE_COUNT: High correlations but not meaningful as it is obvious that counts of purchases grow together with PURCHASE_COUNT
      → Should be transformed to relative values
- Transform variables that count the types of purchases made so they are reported in percentages and can be used to investigate the relationship with PURCHASE_COUNT
  - e.g.: $SHAREDELIVERY = \frac{PURCHASECOUNTDELIVERY}{PURCHASECOUNT}$

2. Exploratory analyses

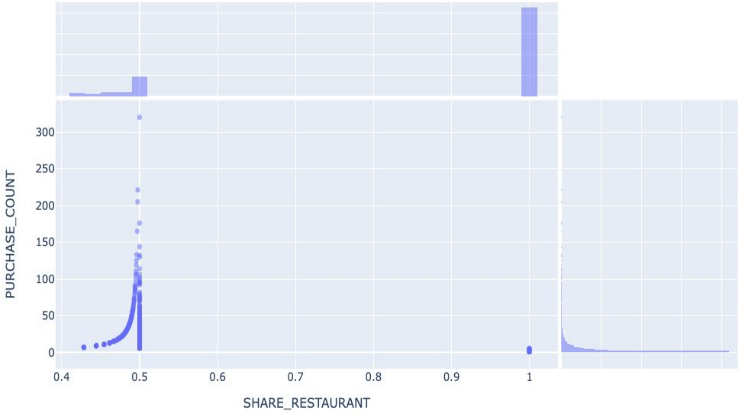
- Inspect correlations via heatmap
  - Not many meaningful correlations
  - However: SHARE_RESTAURANT and PURCHASE_COUNT are strongly correlated

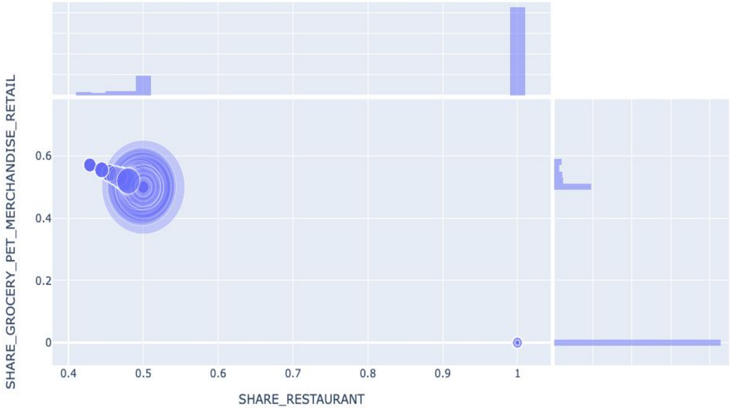
    → Should be further investigated



2. Exploratory analyses

- Scatterplot between PURCHASE_COUNT and SHARE_RESTAURANT
  - Behavioural difference up from 6 purchases
  - <u>Every user with more than 6 purchases also uses other services than restaurants</u>
- Scatterplot between SHARE_OTHER_SERVICES and SHARE_RESTAURANT, with dot size corresponding to PURCHASE_COUNT, to visualize the relationship.
- Chosen cut-off for user segmentation: PURCHASE_COUNT = 6





2. Exploratory analyses

- Segmentation approach: Every user with 6 or more purchases is considered a high-activity user.
    - Pro:
        - Clear behavioural pattern linked to segmented user groups that can be targeted by marketing team and is easy to interpret
        - Segmentation based on the core-metric PURCHASE_COUNT
            - Allows to identify key factors that make users purchase more
            - Allows to link potential revenue increases to marketing KPIs or OKRs
    - Contra:
        - Manually decided threshold → Maybe there are other patterns that could have been discovered algorithmically through supervised (with PURCHASE_COUNT as target) or unsupervised learning (KMeans).
            - But: Unlikely, given the scarcity of meaningful correlations across the heatmap

3. Further analyses with segmented user subsets

- Inspect differences between the groups:
  - Quantitative variables: T-tests & Plotting to interpret differences
  - Qualitative variables: Plotting
- Calculate and plot distribution of the groups and order volume of the groups as well as potential increases in revenue and order volume (Slide 4)