korzinka

Client clusterization

Korzinka Analytics
Team

Speaker

- Marfukhujaev Saidamir, 28
- MSc in Business Intelligence & Analytics
- ex-Alif, Largest Islamic fintech
- Currently CDO at Korzinka Go & Head of Omnichannel Analytics
- I believe in data-driven approach and data democratization

Company

- Korzinka, dominant multi-format grocery retail chain
- Founded in 1996
- Korzinka Go is leading e-grocery service, founded in 2017
- More than 150 markets in every region of Uzbekistan
- 2 mln active identified cutomers

korzinka

Business Context & Problem

Context:

- Korzinka serves massive amount of identified customers monthly
- Customers differ in needs, behaviors, and preferences
- Active CRM, loyalty program, mobile app, and e-grocery delivery

Problem:

- Generic, mass communication is no longer effective
- Lack of behavioral understanding of customers
- Difficulty in prioritizing segments for growth, retention, and activation

Opportunity:

- Segment customers using transactional and behavioral data
- Drive personalization in campaigns, loyalty, and promotions
- Embed insights into core strategies: assortment, store clustering, CRM

Our approach

🔍 **Why Clustering?**

- Traditional segmentation (e.g., RFM, demographics) is predefined and static
- We needed a data-driven, flexible approach to capture complex behaviors
- Clustering helps discover natural patterns in customer behavior without prior assumptions

📊 **What Data Did We Use?**

- Transaction data: frequency, basket size, AOV, % discount usage, time of day
- Loyalty data: age, gender, app usage, card vs app behavior
- Category mix: % share of fresh, dairy, snacks, beverages, etc.

🧠 **How We Did It**

- Aggregated data at monthly level to capture patterns over time
- Used K-Means clustering (validated with silhouette scores & business logic)
- Reduced dimensionality for visualization using UMAP

Feature selection

🧾 **Core Features**

- **AOV** – Average Order Value
- **Avg. Monthly Spending** – Total monthly spend
- **Avg. Purchase Interval** – Days between orders
- **Monthly Frequency** – Avg. orders per month
- **Discount Share** – % of spending under promotion
- **Unique Markets Visited** – Store variety

🕒 **Temporal Behavior**

- **Weekday vs Weekend Share** – Revenue split
- **Time of Day Share** – Revenue split by 4 timeslots:
  - 06:00–11:00
  - 11:00–17:00
  - 17:00–22:00
  - 22:00–06:00

🧺 **Category Preferences**

Category revenue share across **19 groups**:

- Animal, Bakaleya, Bread, Car, Conservation, Dairy, Fruits, Healthy, Household, Hygeine, Kids, Meat, Non-Food, None, Ready-Food, Sweet, Tea, Vegetables, Water

Tools & Libraries

- Pandas (mainly working with dataframe),
- Numpy (for some computation optimizations),
- Sklearn (for label encodings, data scaling, PCA, KMeans algorithm implementation)
- Yellowbrick.cluster for visualizations of Elbow method
- Matplotlib for simple visualizations
- Airflow (automatization)
- Clickhouse (DWH)

What We Learned

🚫 **Demographics ≠ behavior**

- Dropped age and gender: didn’t explain enough variance
- Behavioral features were more predictive and actionable

🛒 **Channel share wasn’t meaningful**

- Online vs offline purchase share had low variance
- Didn’t contribute much to cluster separation

🤏 **Simpler is better**

- Complex engineered features added noise
- Best clusters came from well-aggregated, interpretable metrics

🔁 **Stability matters**

- Ran clustering multiple times with different seeds
- Chose features that gave consistent segments across runs

🧭 **Human validation is critical**

- Business sense-checks and marketing team input were key
- Final clusters only made it if we could act on them

Next steps and implications

🧩 **Embed segments into business workflows**

- Integrate segments into CRM, dashboards, and marketing tools
- Use them for targeted campaigns, retention, and reactivation

🛒 **Segment-based category & pricing strategies**

- Tailor assortments and planograms to match dominant store segments
- Design promo mechanics based on segment behavior (e.g. discount sensitivity)

📲 **Personalized loyalty program**

- Build segment-specific rewards (e.g. for frequency, category expansion)
- Move from one-size-fits-all to personalized experience

🔄 **Automate and iterate**

- Add clustering logic to DWH pipelines
- Regularly retrain and monitor segments as behavior evolves