Behavioral Clustering and Transition Forecasting of Consumer Data for Marketing Insights

SeongYeon Moon[1]

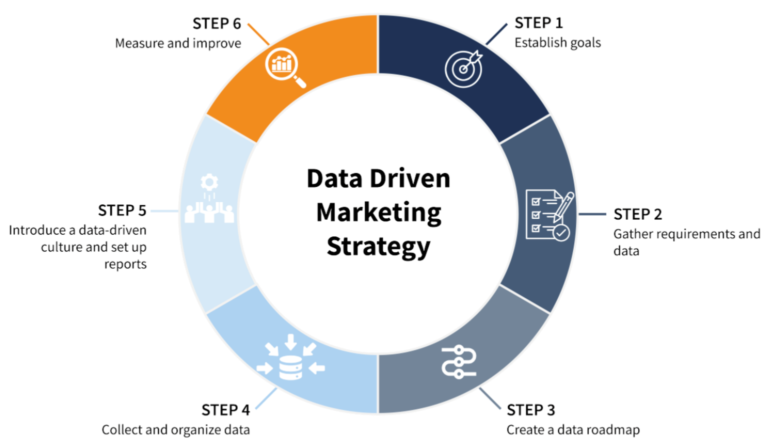
Hispanic Language and Literature[1], Seoul National University

danielmoon@snu.ac.kr

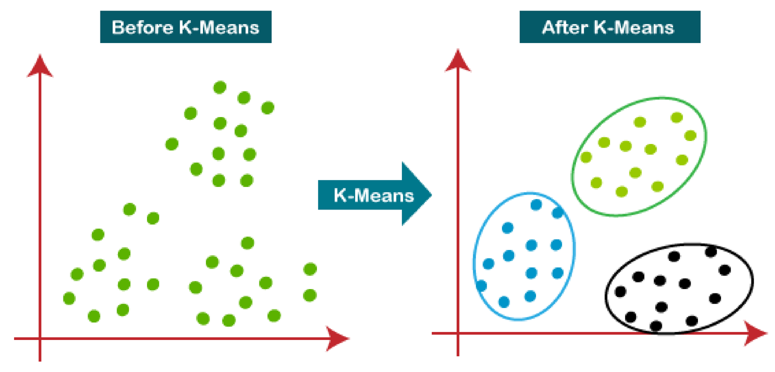
https://github.com/danielmoon2001/AI_final_project

Background and Methodology

- Traditional customer management techniques lack granularity and fail to address behavioural heterogeneity within the consumer base
- Data-driven consumer analytics enable e-commerce platforms to improve customer retention and purchase rates

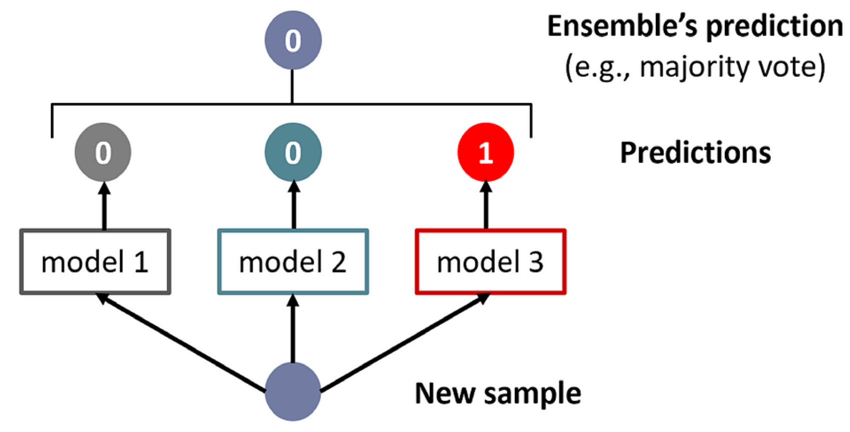


- Raw data was extracted from server-side transactional logs spanning over 4 years
- Features related to purchasing and behavior patterns were engineered and selected
- K-means Clustering Algorithm was used for consumer base segmentation



T. Kansal, S. Bahuguna, V. Singh and T. Choudhury, "Customer Segmentation using K-means Clustering," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, 135-139.

- Segment transition modeling was used to simulate consumer base evolution over time
- Forecasts for the upcoming season were conducted using ensemble predictions



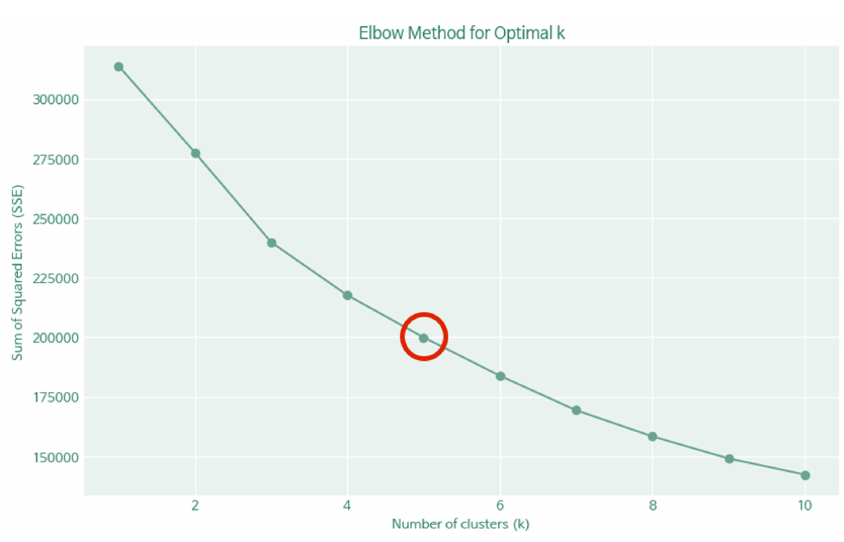
Abdolreza Mosaddegh, Amir Albadvi, Mohammad Mehdi Sepehri and Babak Teimourpour, "Dynamics of customer segments: A predictor of customer lifetime value", Expert Systems with Applications, Volume 172, 2021, ISSN 0957-4174.

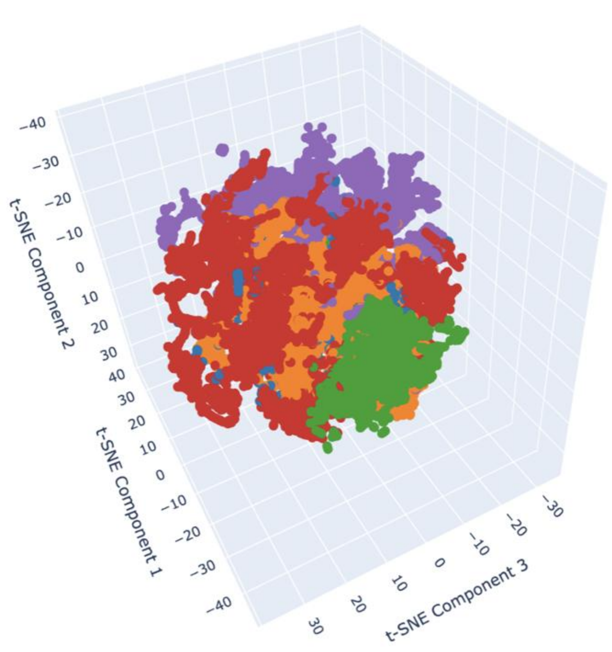
Research Goals

- **Segment consumer base using Clustering methods**
  - Utilize engineered features and K-means Clustering
- **Forecast cluster patterns**
  - Derive marketing insights for differentiated retention strategies

Customer Segmentation

- **Designed behavior-defining features per customer**
  - purchase_counts: total number of purchases made
  - days_since_joined: total number of days since joining platform
  - agree_to_marketing: ordinal encoding of marketing agreement levels
  - regular_prop: proportion of subscription orders
  - review_group: ordinal binning of review participation rates
- **Performed Clustering and selected the total number of clusters**
  - Utilized elbow-method techniques and t-SNE visualizations to determine optimal number of clusters

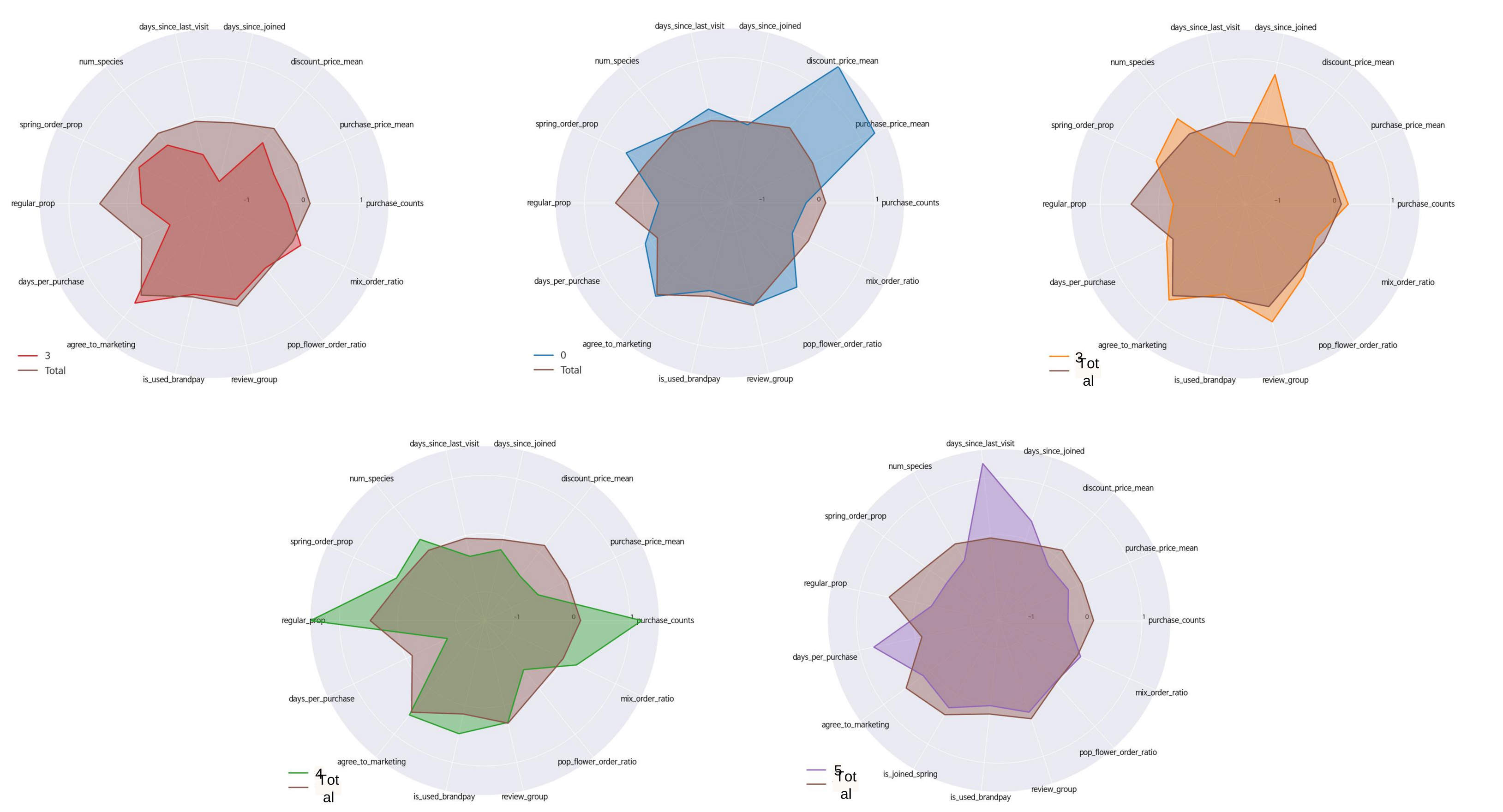




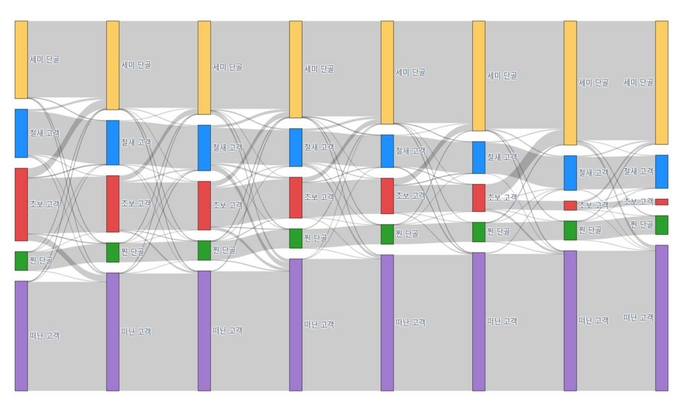
- **Finalized cluster definitions**
  - Reorganized consumer typology into five distinct segments:
  - Newcomers / Opportunists / Semi-Regulars / Loyalists / Dormant

Transition Modeling

- **Analyzed cluster characteristics and transition patterns**
  - Identified key differentiating factors between clusters



- Discovered probable transition paths among clusters to predict future transition inflow/outflow (average transition rate: 5.5%)
- Loyalists showed strong stability, whereas inflow from Newcomers and Opportunists led to Semi-Regulars being projected as the next dominant segment (at 28.6%)



Conclusions

- **Demonstrates the utility of unsupervised segmentation**
  - ML-based segmentation techniques can help to detect structurally distinct user types from behavior data
- **Temporal Modeling uncovers lifecycle-aware actionable insights**
  - Provides evidence-based expectations for lifecycle trajectories
  - Suggests opportunities for precision-targeted retention strategies