Data Science - Technical Assignment

Uriel Feigelsohn
March 24
ZoomInfo

Agenda

- Problem presentation
- Data Exploration
  - Main Features, Correlations, Pairplots
- Analysis
  - PCA, Feature Importance, and main factors to churn
- Insights and Recommendations
  - Churn Model and Metrics, Tiers and Triggers- New KPis ideas (Proactively CX), User comparison

Churn

Goal: Identify key factors and Reduce customer churn rate by implementing targeted retention strategies based on the insights.

Data: 5630 rows, 16% churn rate, 19 features, ~5%nulls in some columns, no Duplicates, looks like a chunk of a bigger file.

Index: Customer ID

Main Features - Churn

Usage

- Tenure
- Hours spend on App
- Number of Devices Reg

User

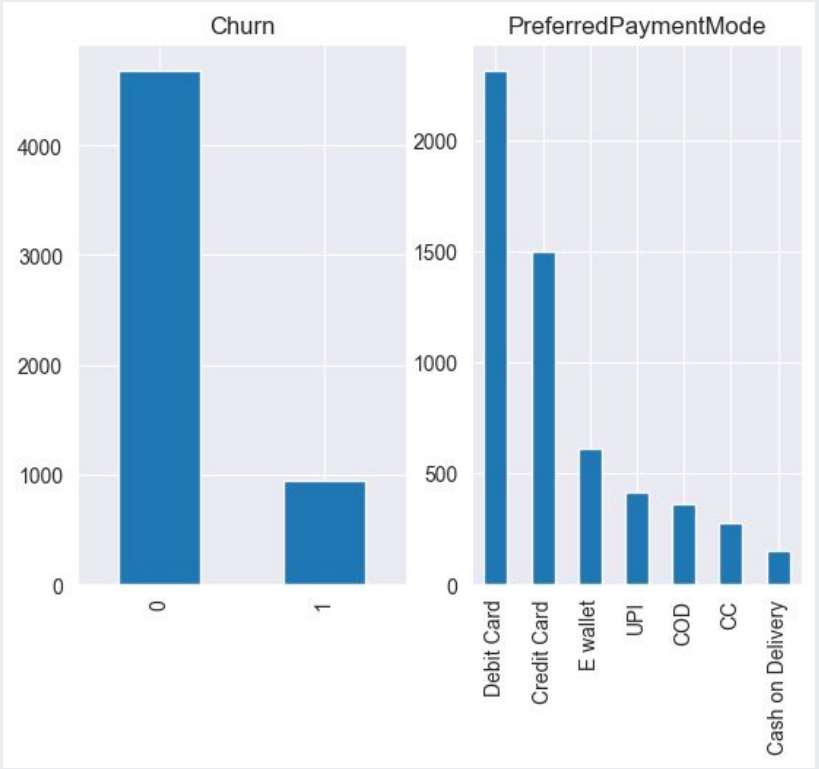
- City
- Gender
- Marital Status
- Number Of Address
- Satisfaction Score
- Complain

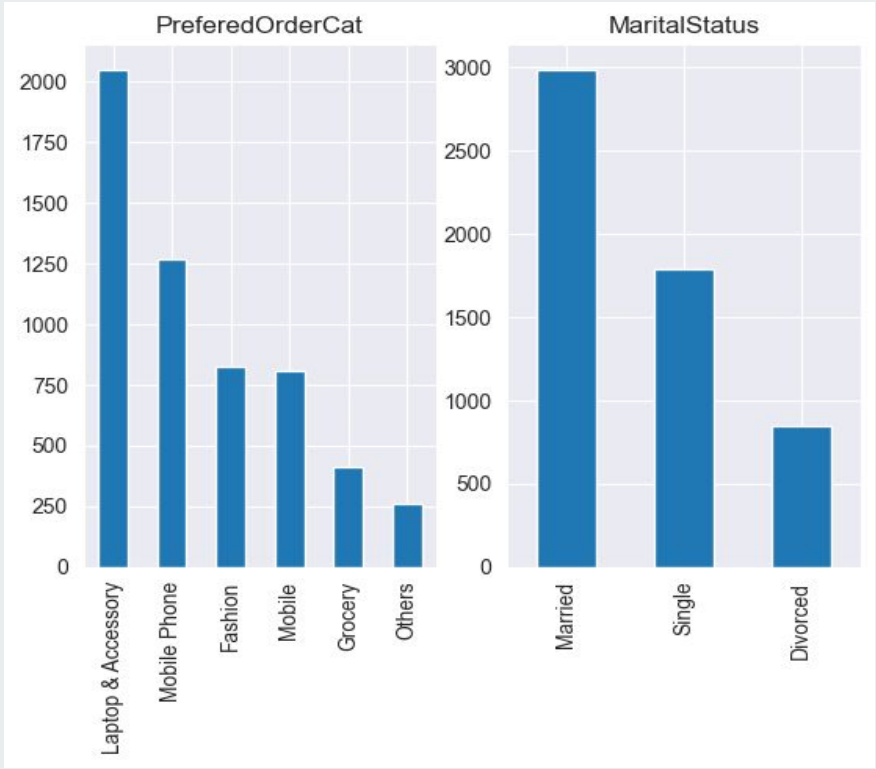
Orders

- Days Since Last Order
- Cashback Amount
- Order Count
- Coupon Used
- Order Hike y/y
- Prefered Device
- Payment mode
- Prefered Order category

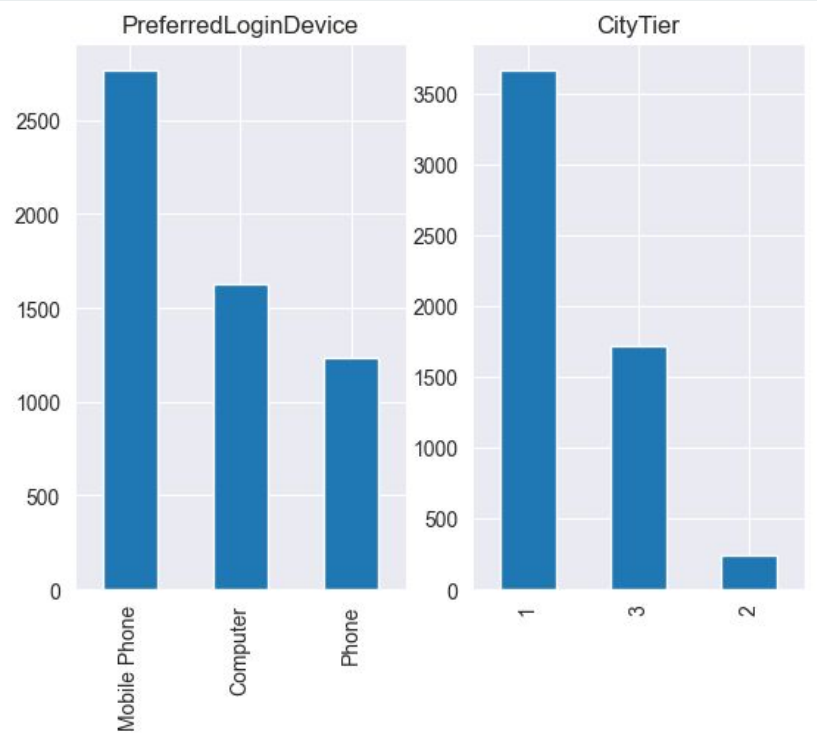
How they perform vs. Churn and between them

Categorical Features- presentation





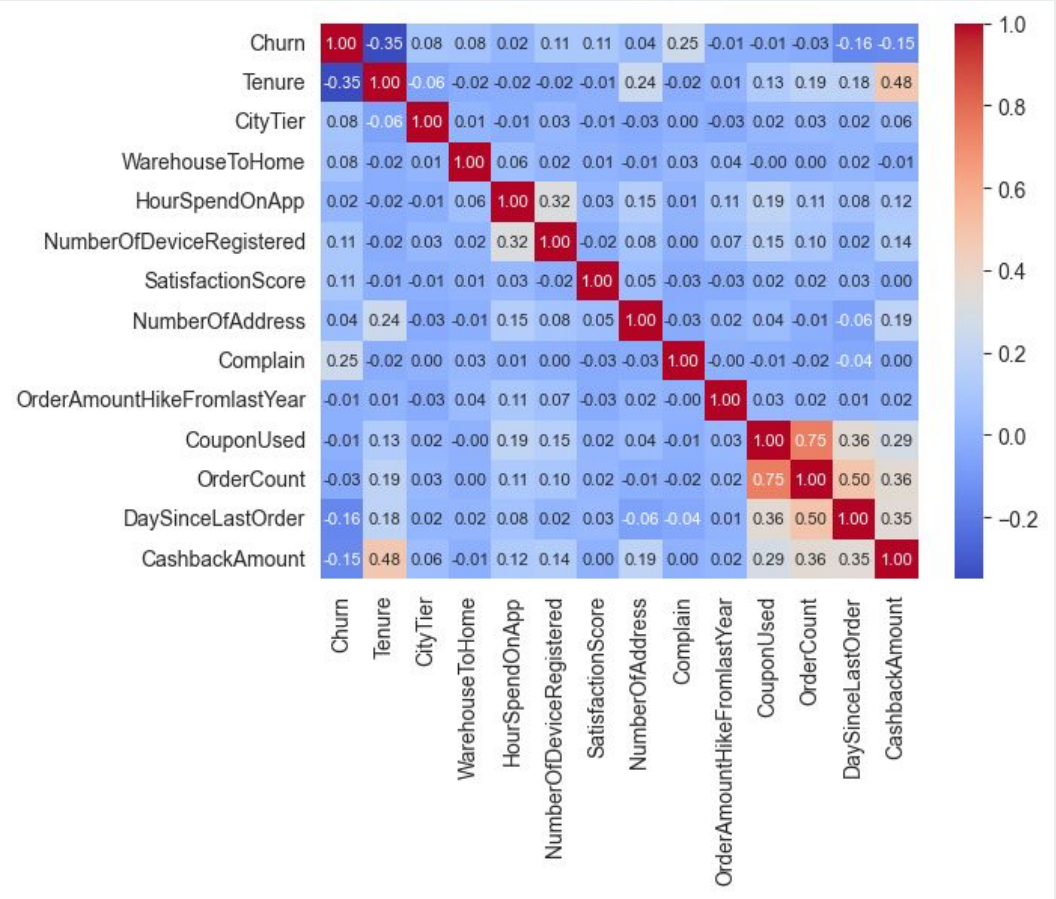
Categorical Features- presentation



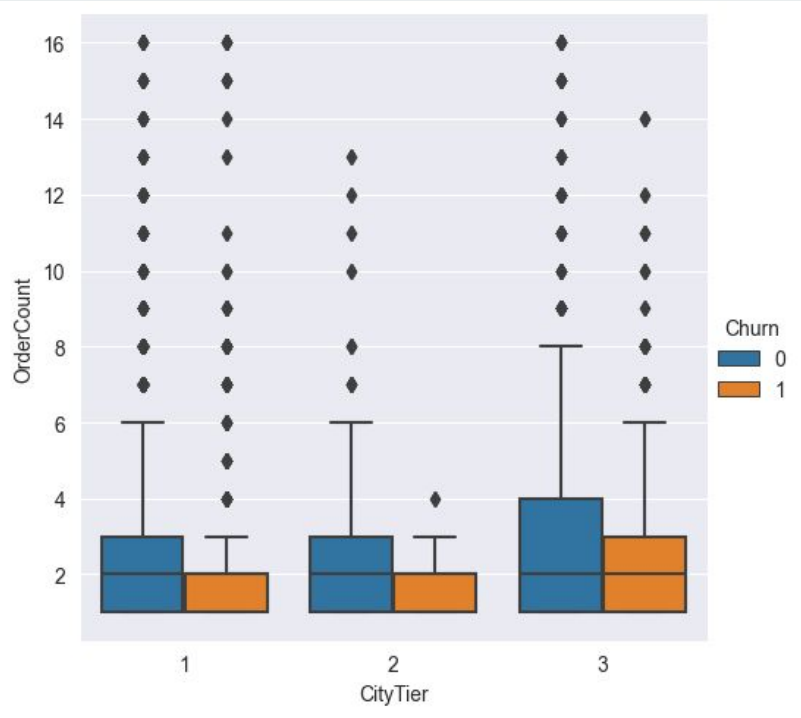
Correlations

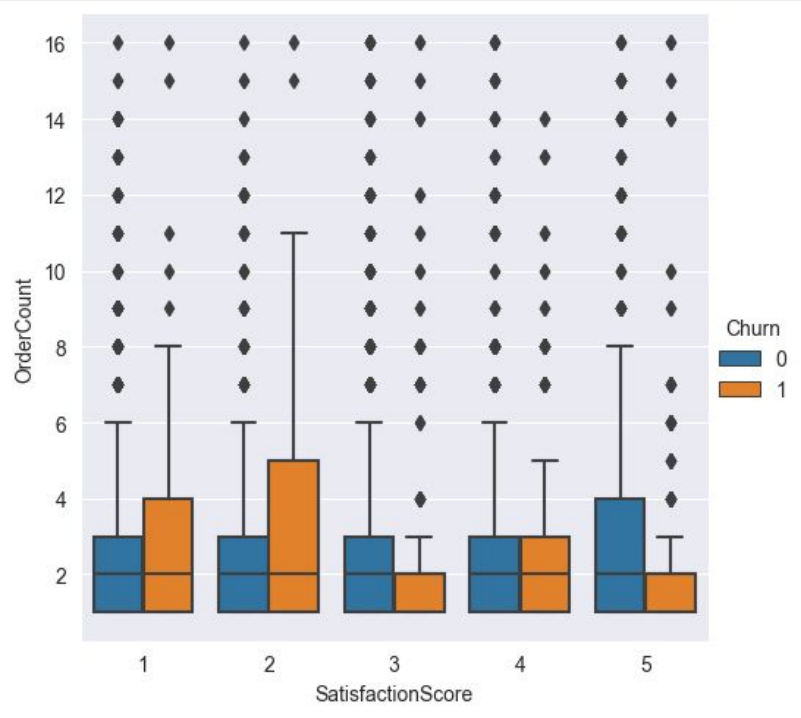
- Churn vs.
  - Tenure
  - Complain
  - Cashback
  - Days since last order

- High correlation between all the Order values

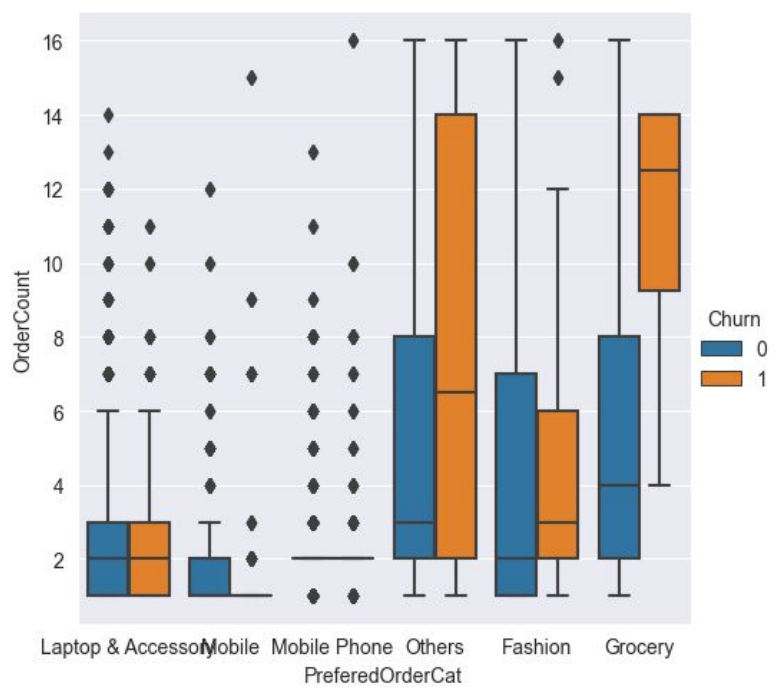


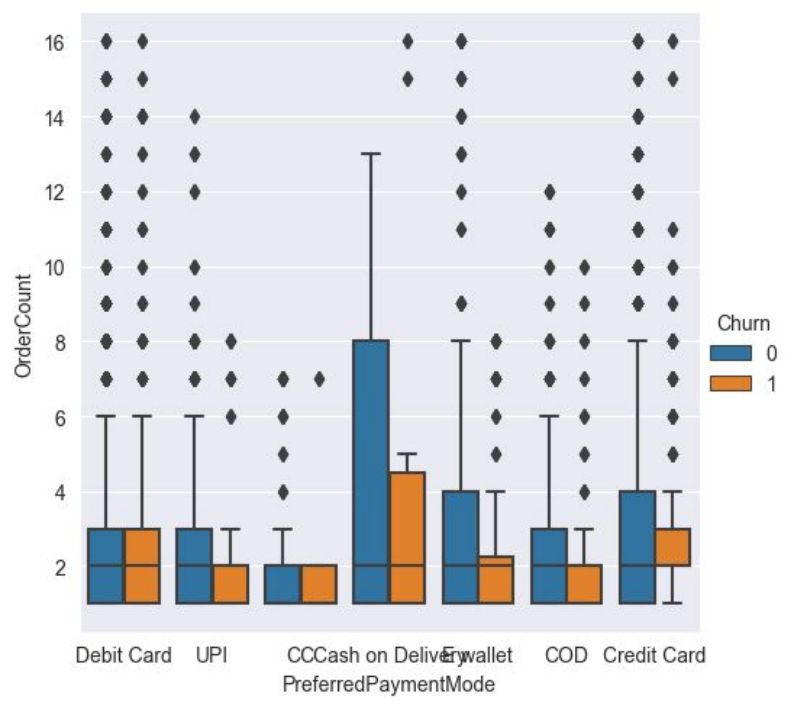
Categorical Features by Churn%





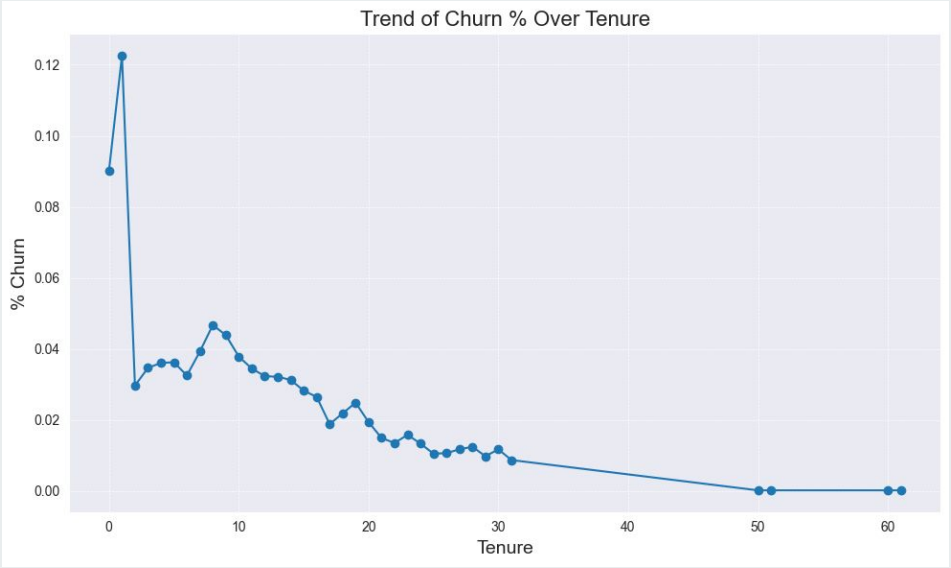
Categorical Features by Churn%





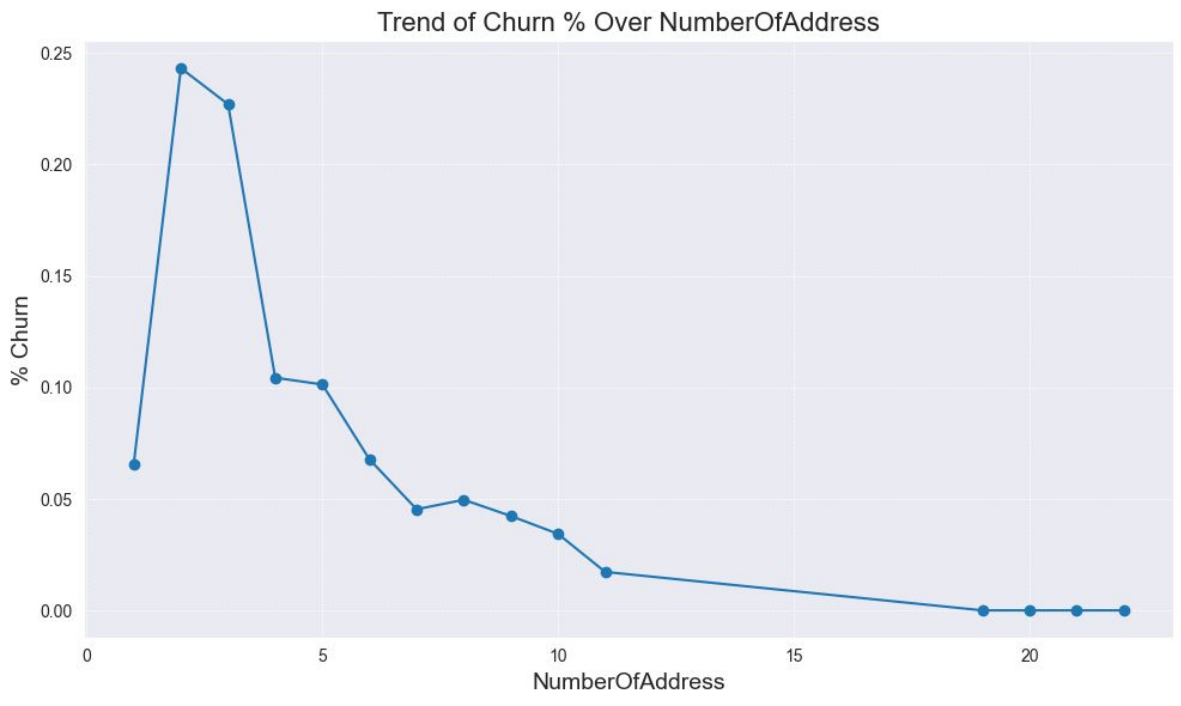
Interesting relationships with Churn

- Tenure churn cycle



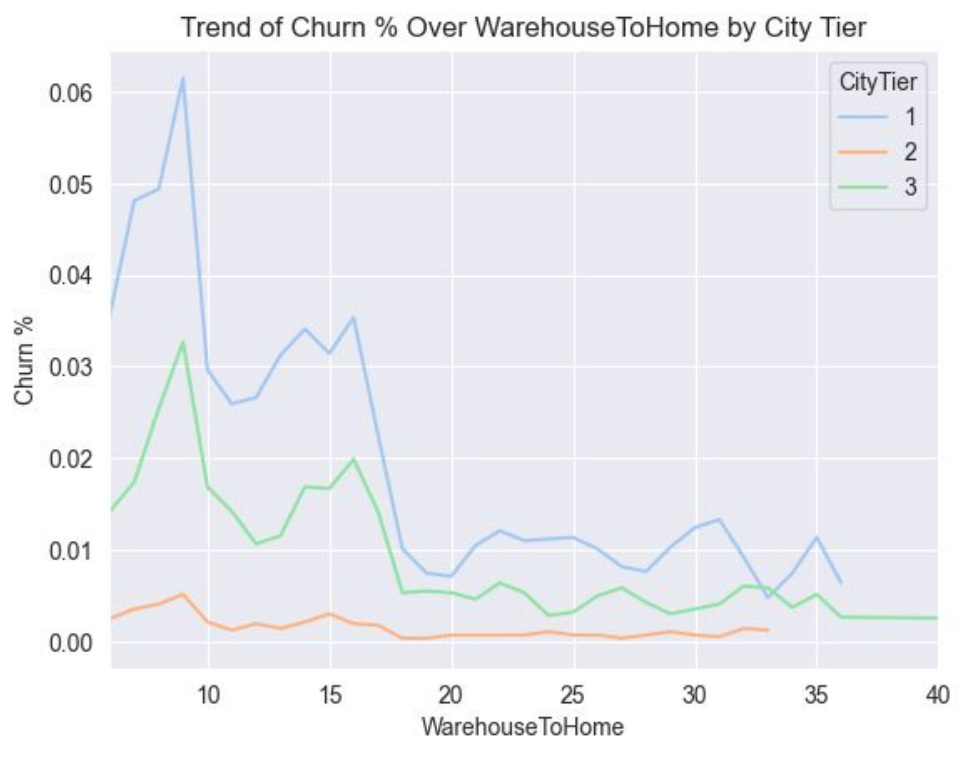
Interesting relationships with Churn

- Retention by engagement



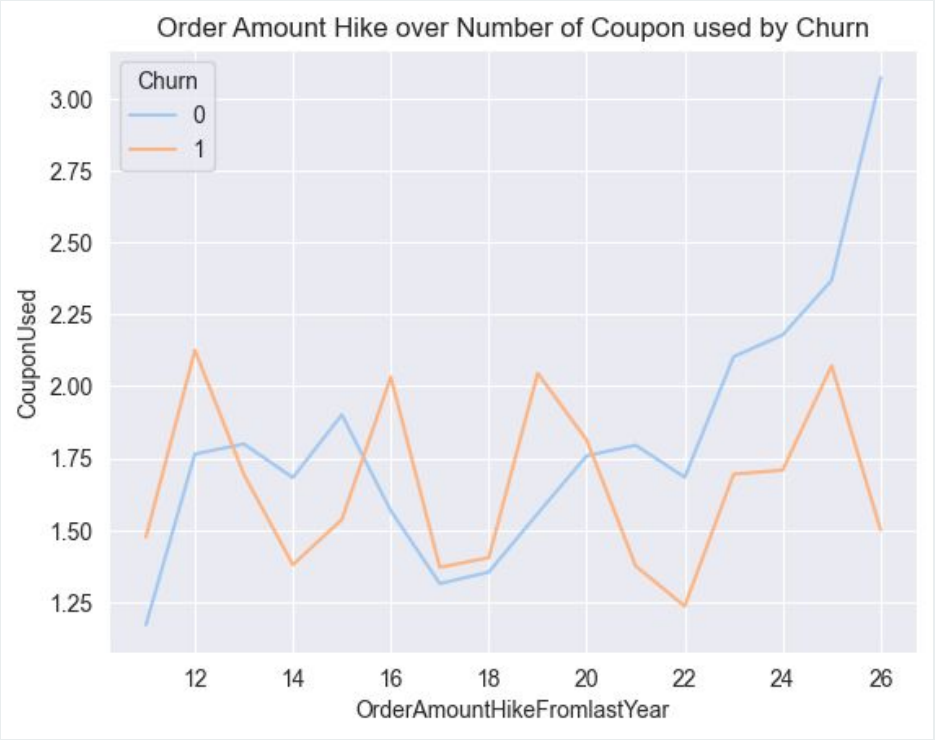
Interesting relationships with Churn

- City differentiation

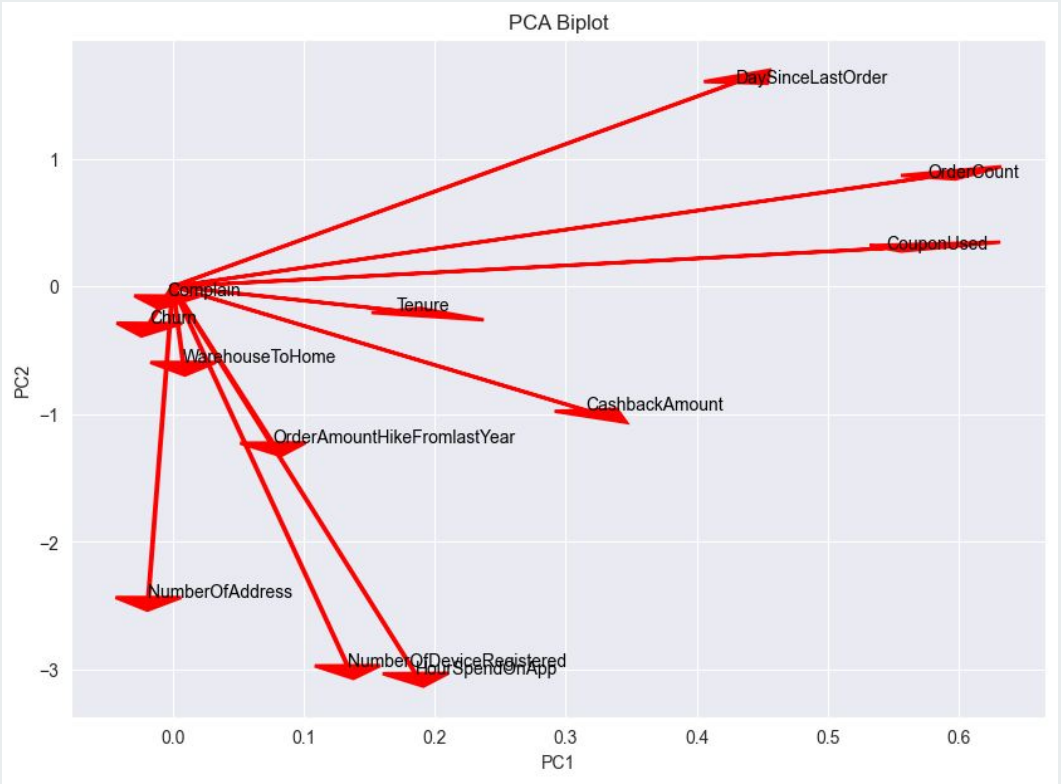


Interesting relationships with Churn

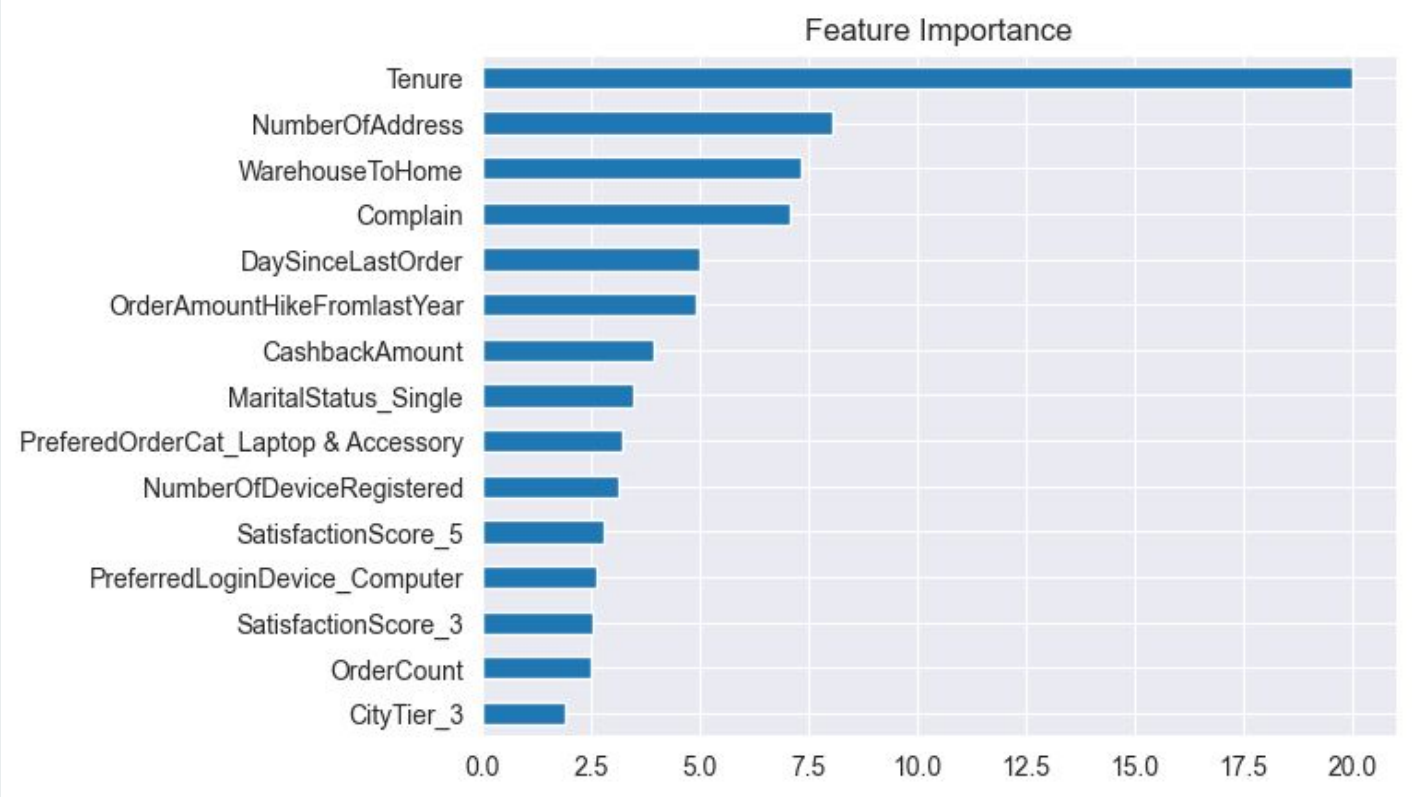
- Coupons effect



Interesting relationships between features- PCA

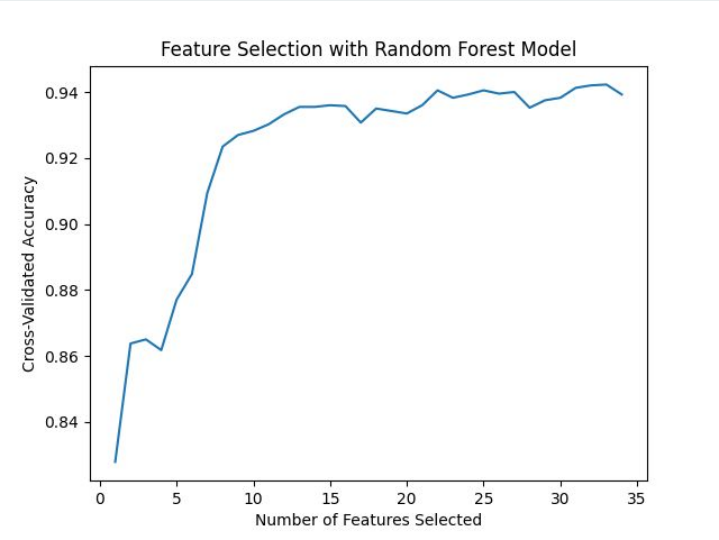


Key Features- XGB Classifier



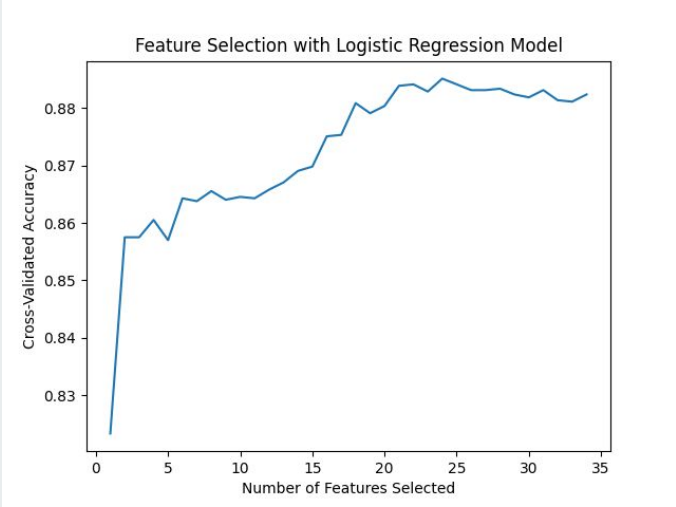
Churn Model

- Imbalanced data - ~16% Churn rate
- Pipeline:
  - Feature Engineering
  - Normalization
  - One hot encoder
  - Select Best features
  - Over Sampling
  - Classifier



Recommendation- Churn Model

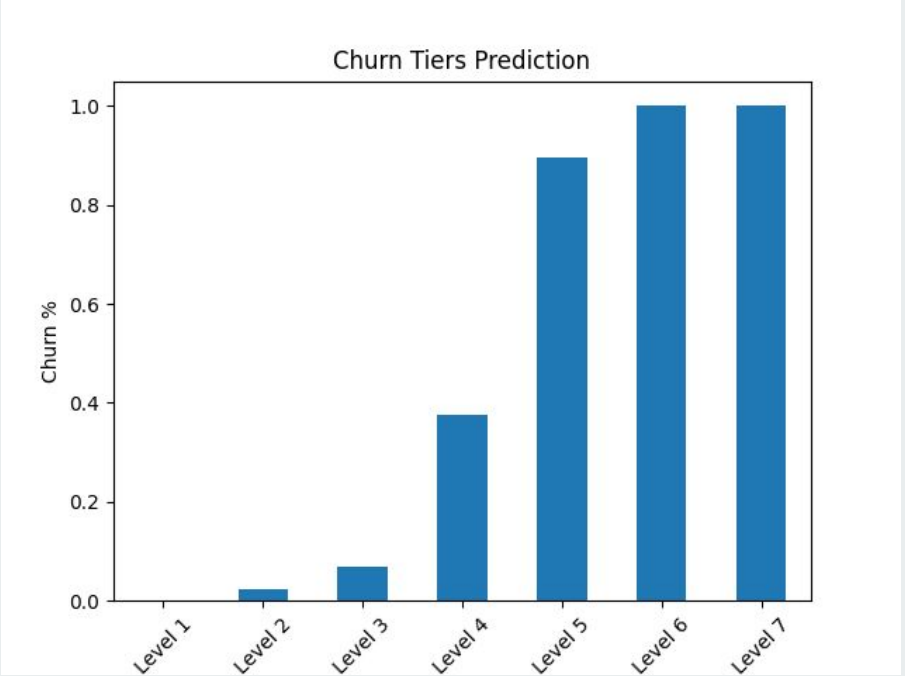
- Tree based approach perform well better than logistic regression (started from simple models, tree approach worked better, hinting about a non-strict linearity relationship).
- Metrics (Imbalanced data):
    - F1 (Minority class)
    - Macro f1 (Same weight for all classes)



Insights- Proactive cx

- Tiers churn: Proactively reach users that Move between churn tiers.

  Each level represent a predicted range of Probabilities to churn (X) vs. the actual Churn Rate (Y)



Potential Retention Strategies:

New KPIs, Signals and triggers

- Personalized offers and incentives- Coupons, target/untarget city and distance/ target high hike users, offer coupon before dead end.
- Improved customer support and engagement based on triggers- Movement between Churn tiers
- Enhancing product/service features based on customer feedback- Satisfaction Score, categories.
- Predictive analytics to anticipate and prevent churn

KPIs

- Retention (Tenure cycles signals, hike,)
- Engagement- More locations added (5+ as a sweet spot), Devices
- Satisfaction Score Rate (3+)

To Do

- Model optimization
  - Fine-tune hyperparameters to improve model performance.
  - Explore different algorithms or ensemble methods for better predictive power.

- Feature Engineering
  - Add new features for new KPIs
  - Identify additional relevant features to enhance model accuracy.
  - Add new Recency features, such as last X days activity level.
  - Smarter Imputation (KNN, etc.)

- Robust validation in validation sets. Check performance over time (Different dates).

Thanks

Questions?