

CAPSTONE PROJECT

Target customer prediction for direct marketing campaigns at Kroger

By Chau Tran

PROJECT'S STRUCTURE FOR PIPELINE REPRODUCING

- Original data folder: Kroger_Retail_Dunnhumby
- Data after aggregation and profiling: DATA_CLEAN
- Notebooks:
 - + Data aggregation and profiling works:
 - 1_aggregation_campaigns_products.ipynb
 - 2_aggregation_households.ipynb
 - + Exploratory Data Analysis (EDA):
 - 3_eda_households_during_camp18.ipynb
 - 4_eda_households_beforeCamp18.ipynb
 - 5_eda_households_full.ipynb
 - 6_eda_retention_analysis.ipynb
 - + Modelling and evaluation:
 - 7_modelling.ipynb
- Detailed documentation folder: Documentation_Presentation

PROJECT SUMMARY

This Capstone project investigates different customer segmentations at Kroger, how their marketing campaigns impact on the overall customer engagement and forecasts the target groups for their marketing campaigns.

Direct marketing has been one of the efficient strategies for retail business to retain customer loyalty and to improve revenue against their competitors. By combining technology and data, delivering outstanding and personalised shopping experiences are the key for retailers to encourage consumers coming back to their stores.

Business question

Can we identify target customers who will redeem coupons offered by direct marketing campaigns?

Data question

By using demographic and history purchase data, can we predict target customers who will redeem coupons offered by direct marketing campaigns?

Stakeholder

Josue Hernandez - Marketing Manager at Kroger

Delivery

Identifying correctly characteristics of target customers will allow marketing team to focus on those who most likely purchase promoted products, limit their budgets to the highest profit potential people and personalise customer experiences.

Key findings achieved from the analysis include:

- Married people with family component of 2 adults and kids, who own a house with age between 25-34 or 45-54 and have income between 35-74k are those who likely redeem the coupons of the studied campaign.
- CUSTOMERS with COUPON REDEMPTION has 6% higher retention rate and 27% higher average purchase than the others.
- 'Champions' customer segment has the highest redemption rate of 25%.

The dataset was moderately imbalanced with minority group ratio of 26%. Oversampling technique (SMOTE) was used to balance the dataset.

7 classifier models were created and validated to predict the target customers: LogisticRegression, RandomForest Classifier, Gradient Boosting Classifier, XGBClassifier, KNeighborsClassifier, GaussianNB and AdaBoost. Five models LogisticRegression, RandomForest Classifier, Gradient Boosting Classifier, XGBClassifier, AdaBoost gave high performance and were tuned before applying on the test set. The final model with highest accuracy (84%) was andomForest Classifier.

Dataset description

- 102 weeks from 2009 fiscal year
- 2,595,732 transaction records
- 2500 unique customers
- 801 customers with demographic details
- 276,484 unique baskets in
- 92,353 unique products
- 30 mail campaigns sent to 1584 customers
- 1135 unique coupons