Algorithm Avengers

To Buy or Not to Buy: Investigating Customer Patterns in the Online Shopping Experience

By Harry Kim, Matthew Ohanian, Bennett Cohen, Pieter van Tol

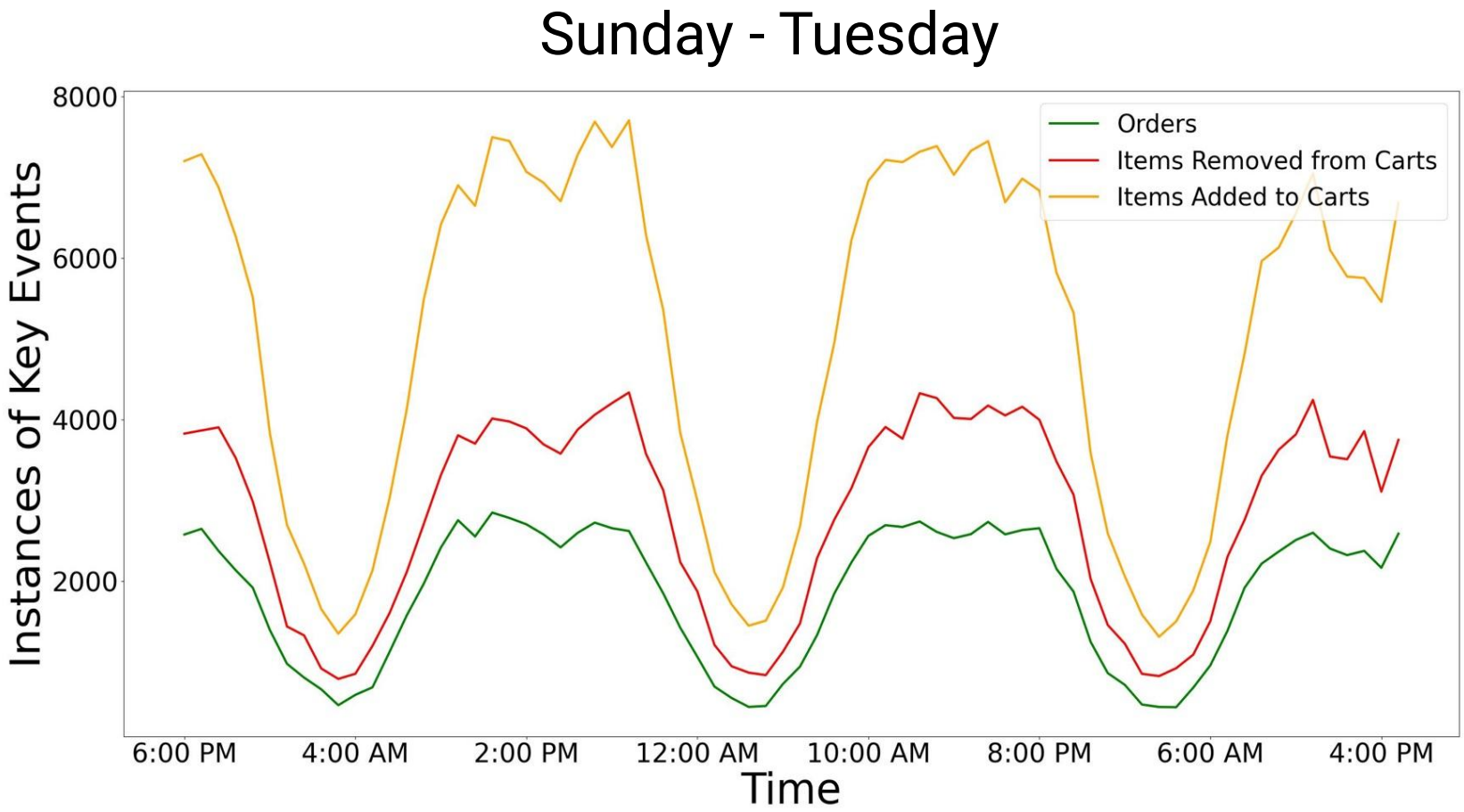
The Data

7-day web data from Fingerhut between 12/10/2022 ~ 12/17/2022

In that timespan ~

- 2,040,966 unique visits
- 577,004 registered users

Peek into the data



Research Question

What patterns exist among customers who add to their shopping cart and actually make a purchase versus customers who abandon cart?

How we tackled the data

Focused on extracting information from variables that fulfilled the following:

- Described the **customer** or the **features of their visits**
- **Scalable** to each customer visits

Example of Data Extraction

shopduration = max(hit_time_gmt) - min(hit_time_gmt)

OrderTRUE = as.logical(any(ordernumber > 0))

AddtoCart = sum(str_detect(clickaction, "Add to Cart") |

str_detect(eventlist, ",12,"))

Target Variables

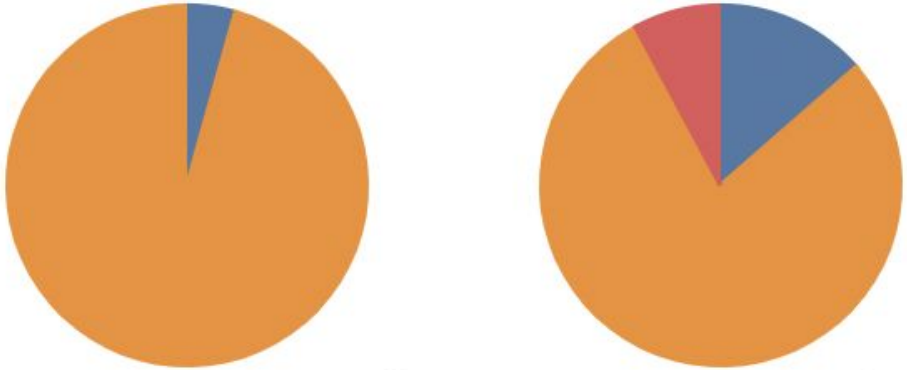
Add To Cart: Placing item in Cart and triggering Add To Cart Log

Abandoning Cart: Add to Cart True and no Order/Purchase

OrderTRUE: If the user had an order number

Cleaning Data

Limitations of Unregistered Users



User Actions

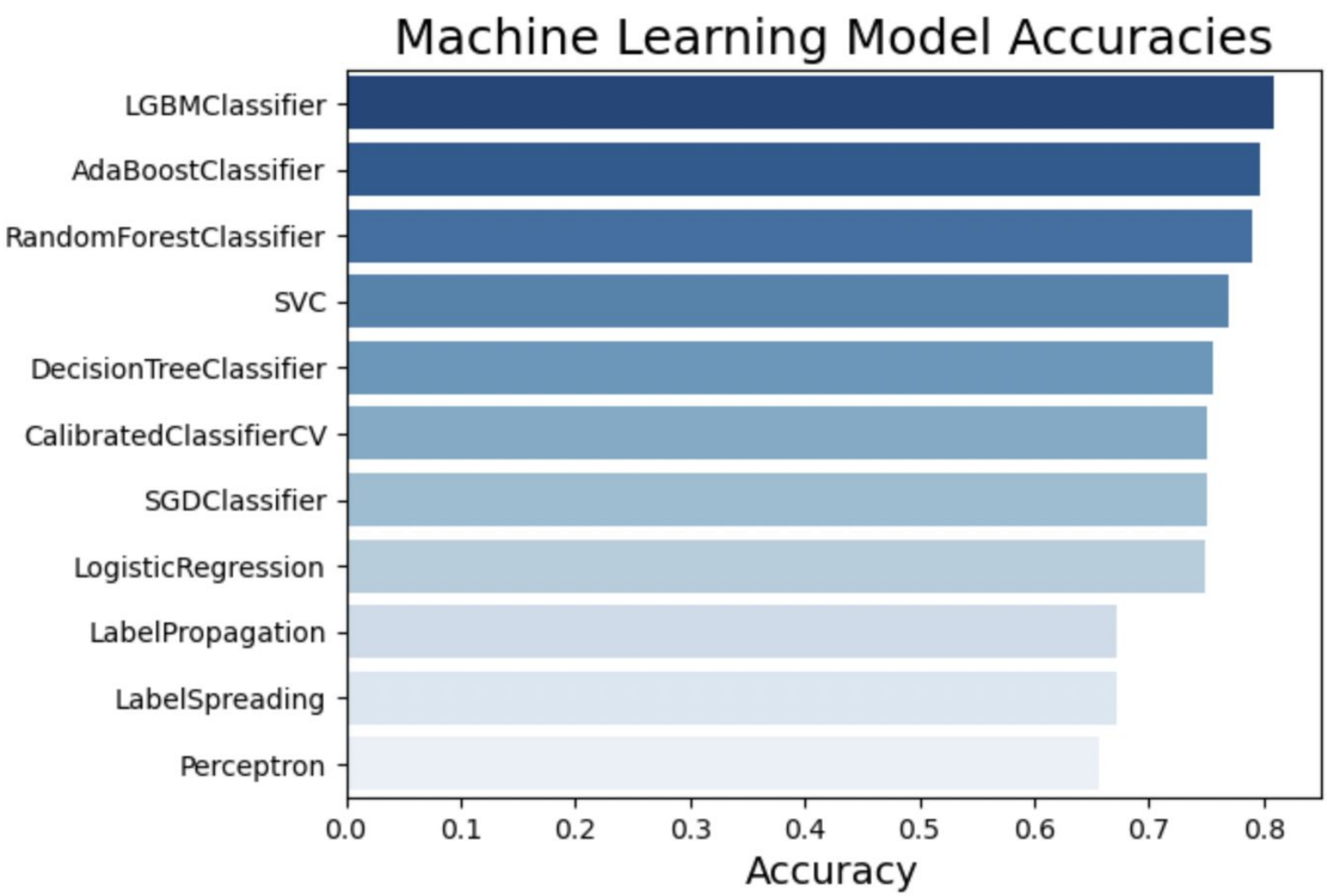
- Added to Cart
- Not Added to Cart
- Ordered from Cart

Machine Learning

- Supervised Classification Model
- Target variable:
  - Abandoned cart
  - Purchased from cart

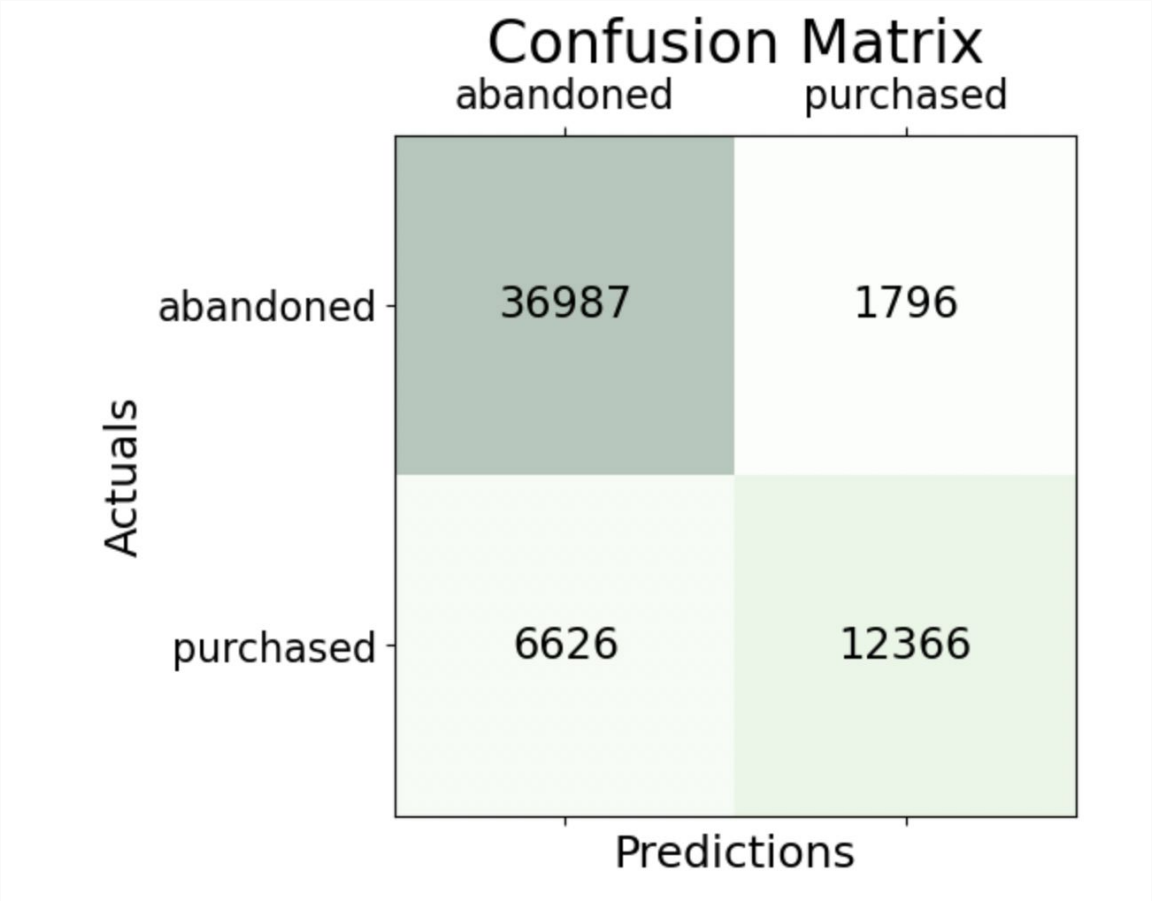
Top Performing Models

- Light Gradient Boosting
- Random Forests
- Adaptive Boosting
- Support Vector Machines

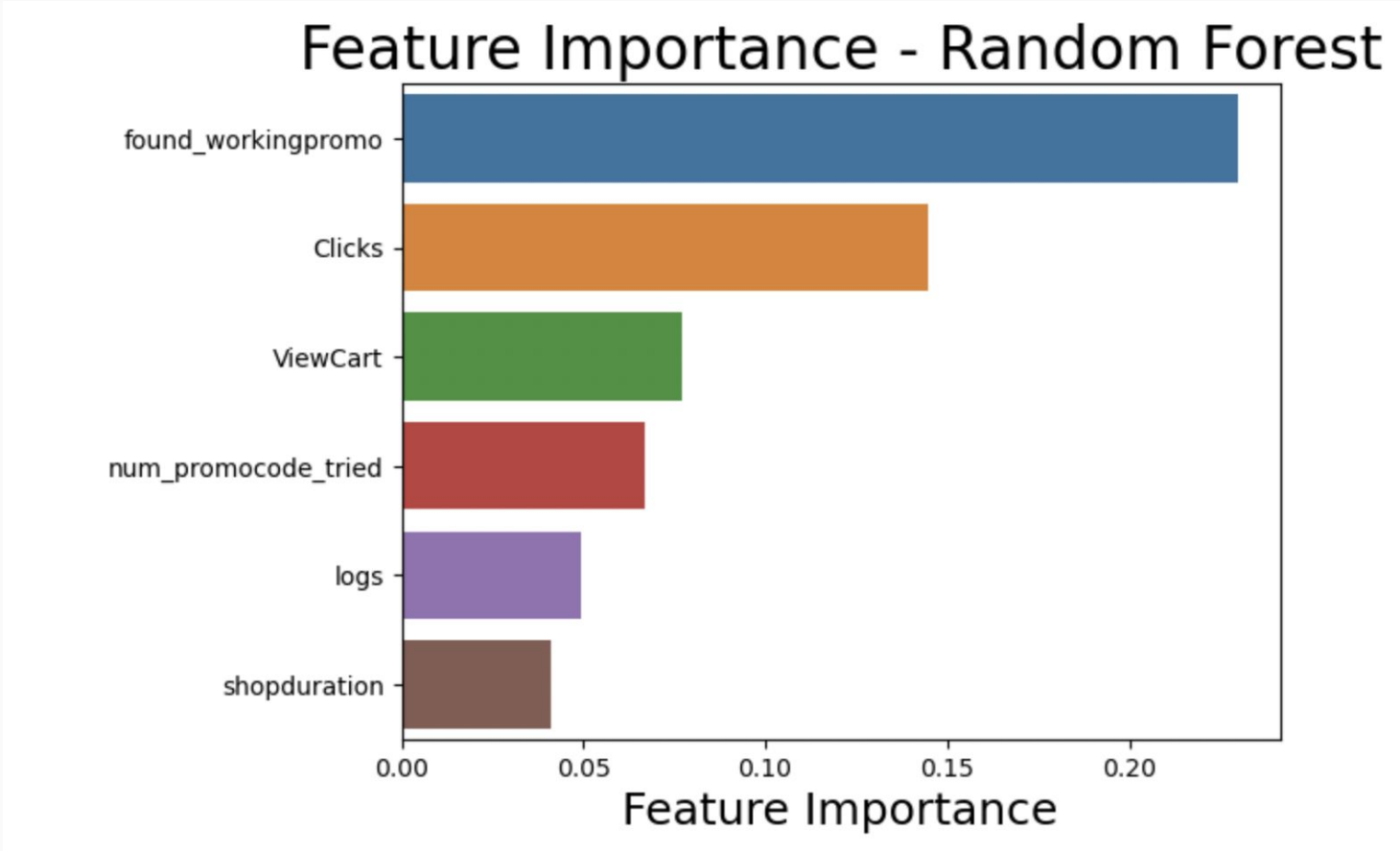

Machine Learning Model Accuracies
LGBMClassifier
AdaBoostClassifier
RandomForestClassifier
SVC
DecisionTreeClassifier
CalibratedClassifierCV
SGDClassifier
LogisticRegression
LabelPropagation
LabelSpreading
Perceptron
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Accuracy

Random Forests

Test Accuracy: 85.42

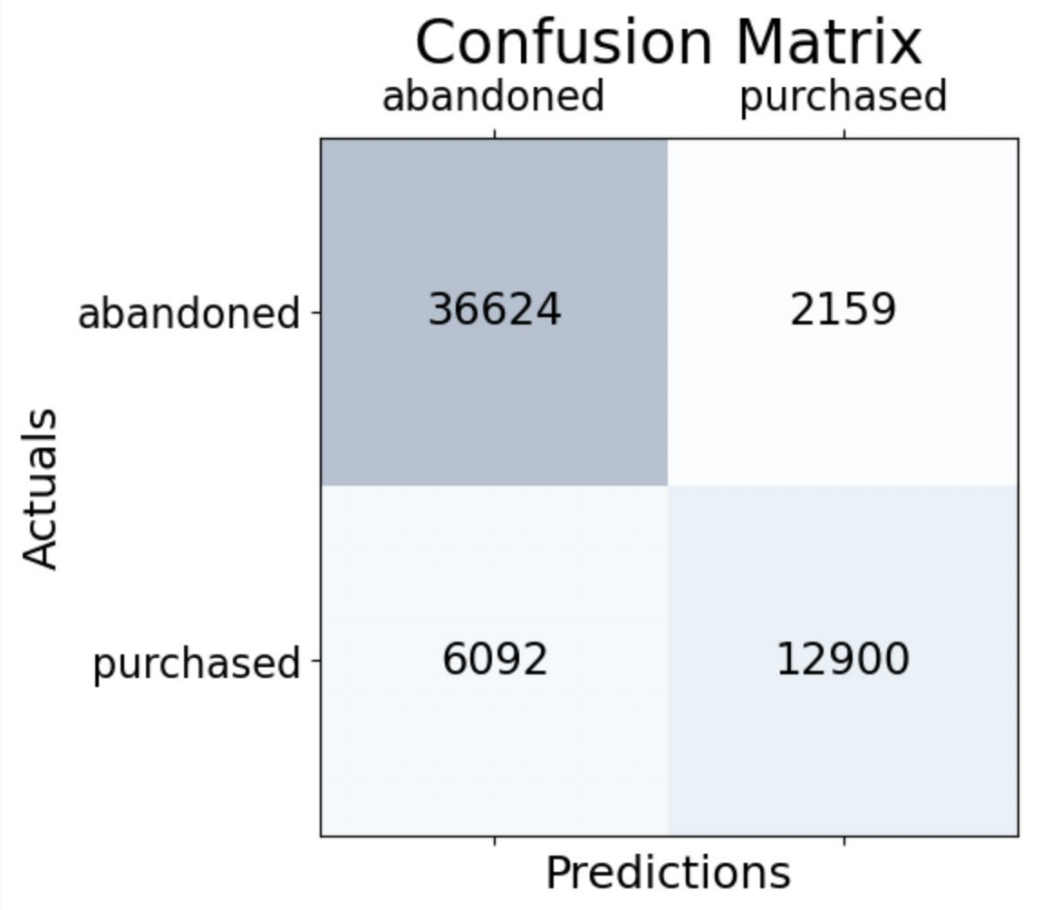


Notable Variables

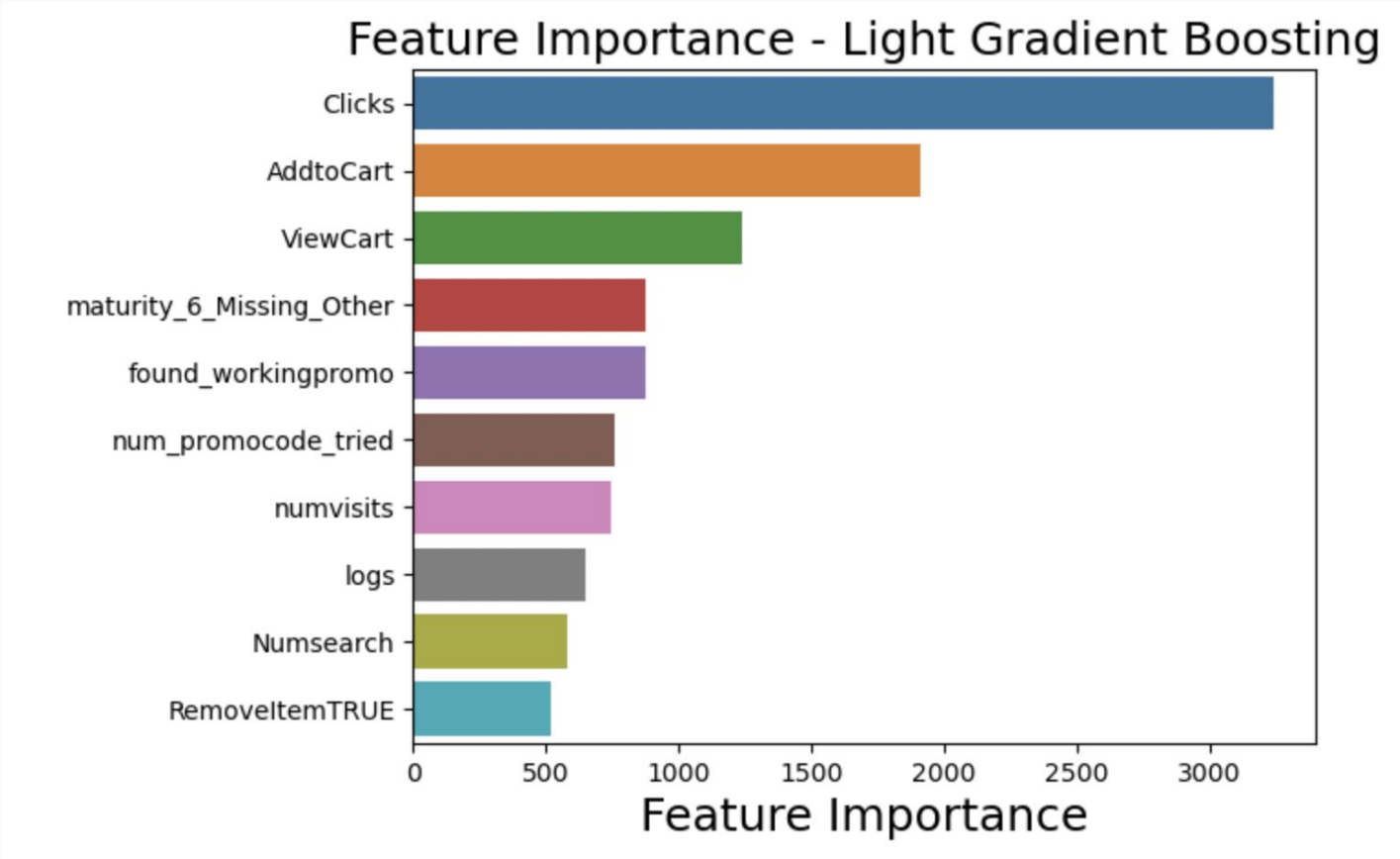


Light Gradient Boosting

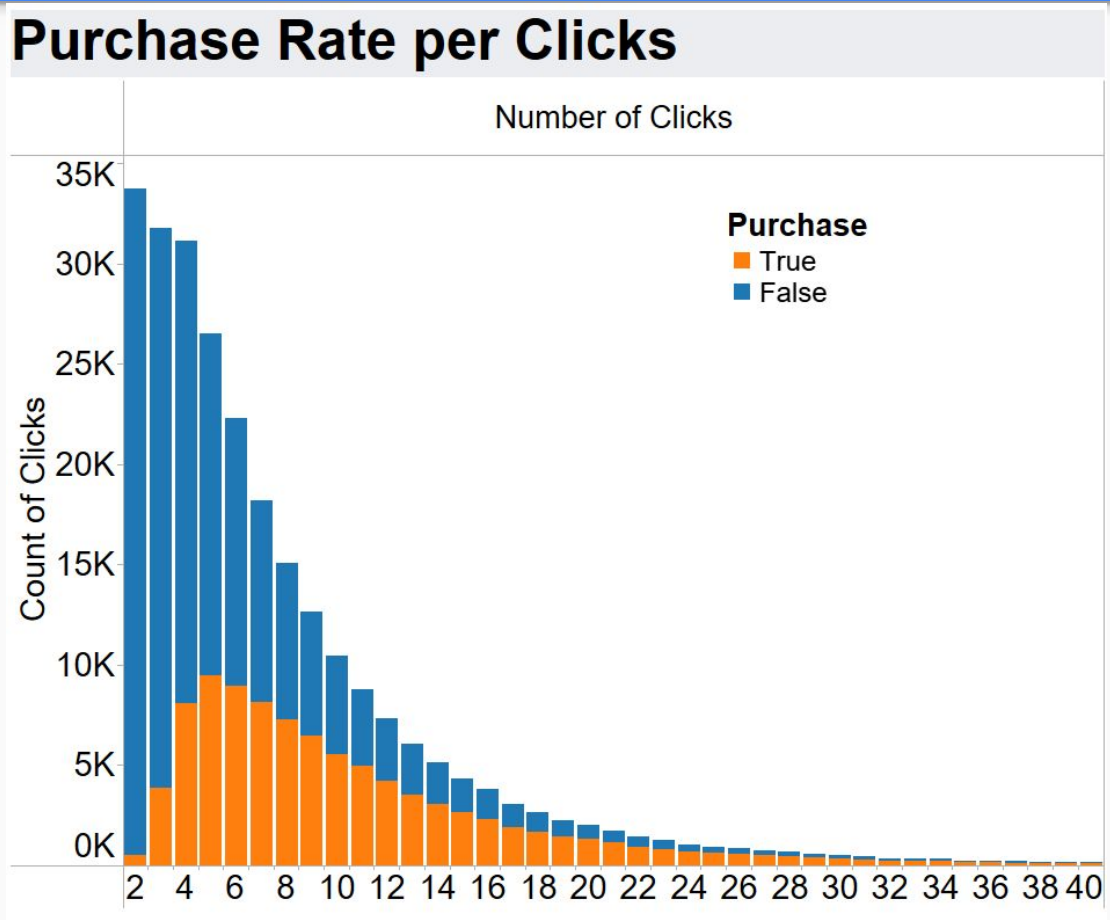
Test Accuracy: 85.70



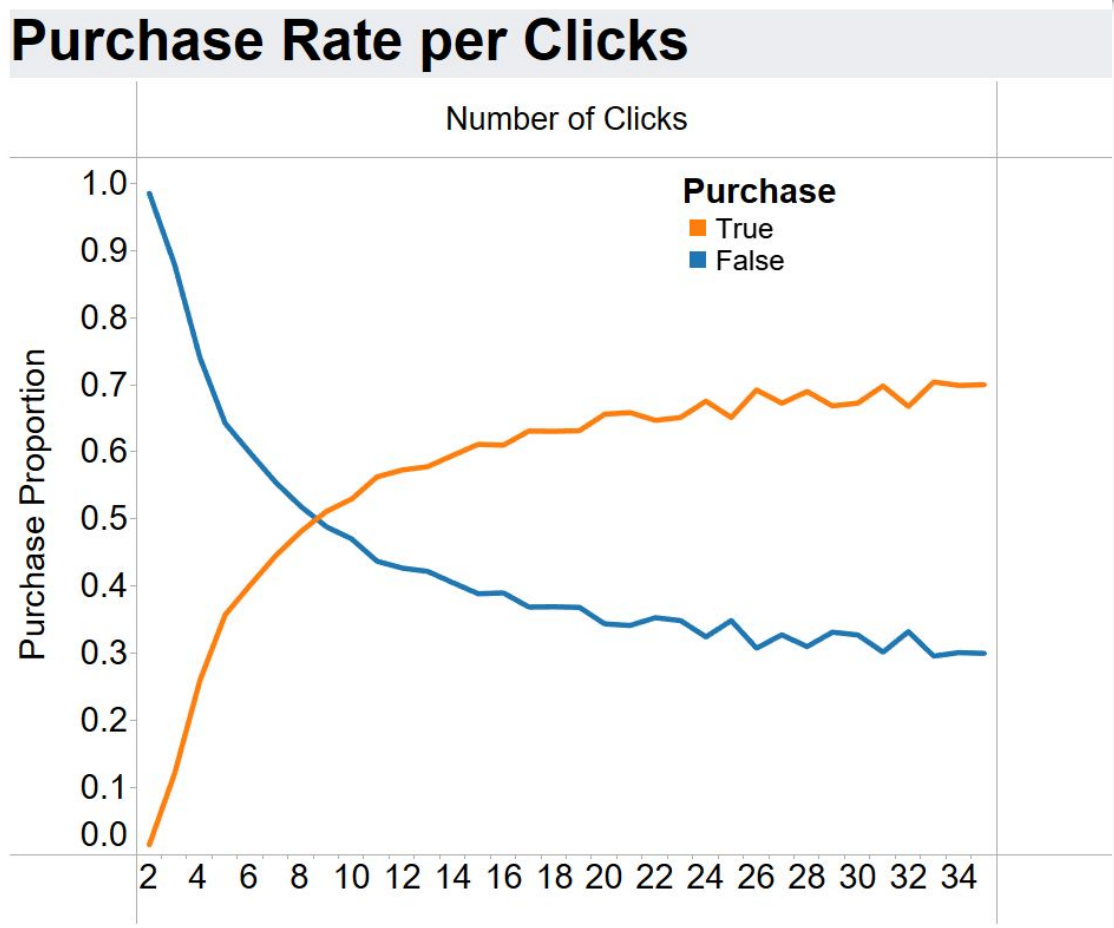
Notable Variables



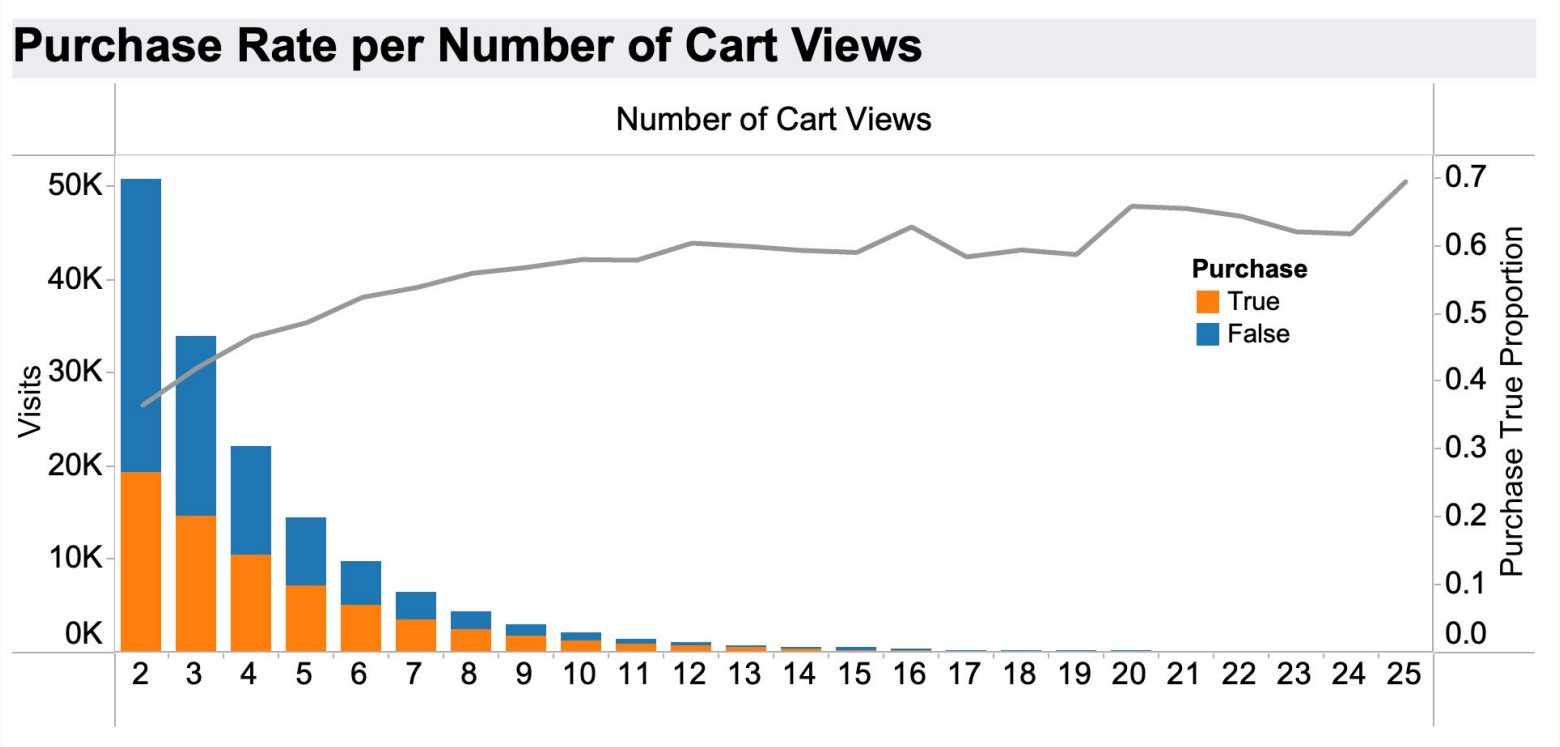
Clicks


Purchase Rate per Clicks
Number of Clicks
Count of Clicks
35K
30K
25K
20K
15K
10K
5K
0K
2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
Purchase
True
False

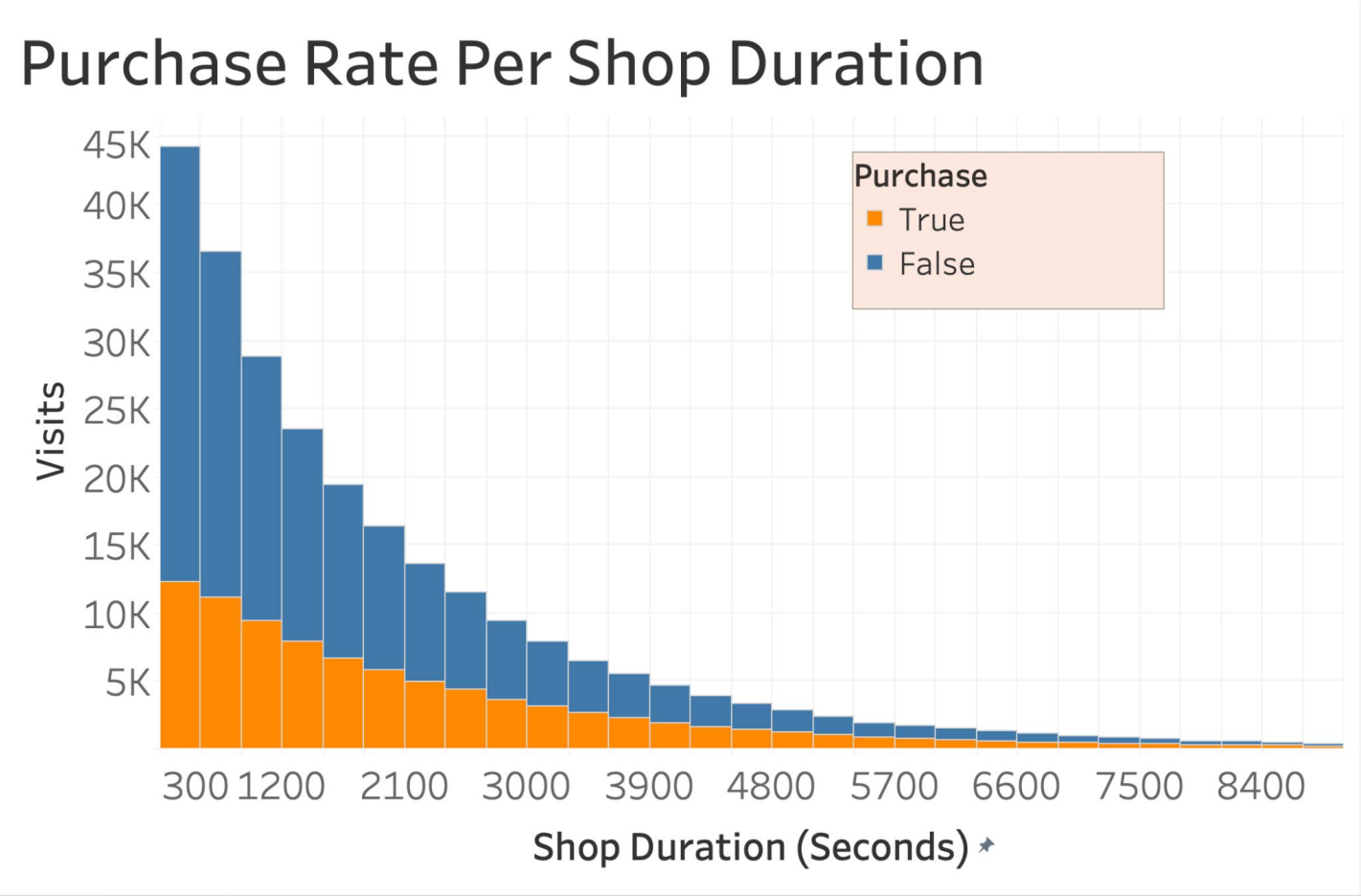
Proportion of Purchases per Clicks



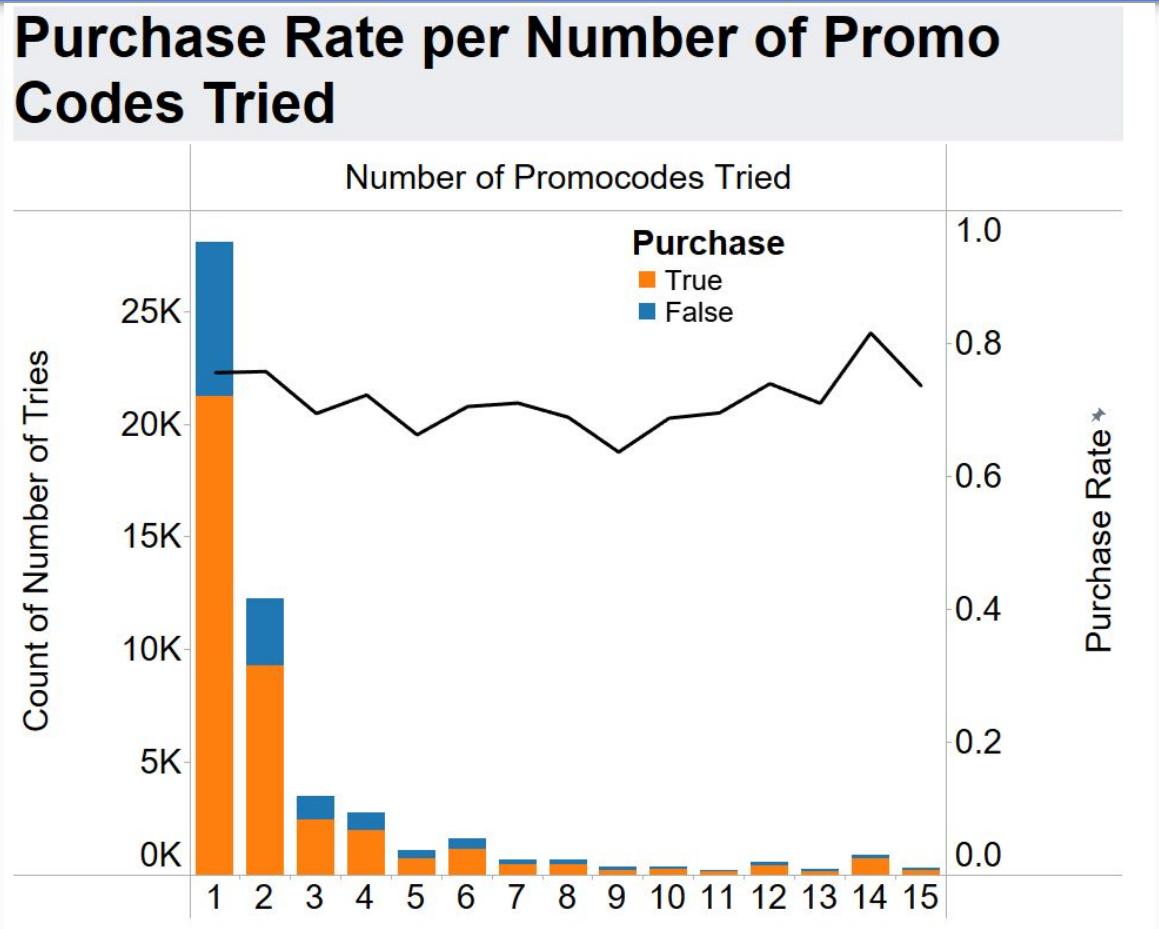
Purchase Rate by Cart View



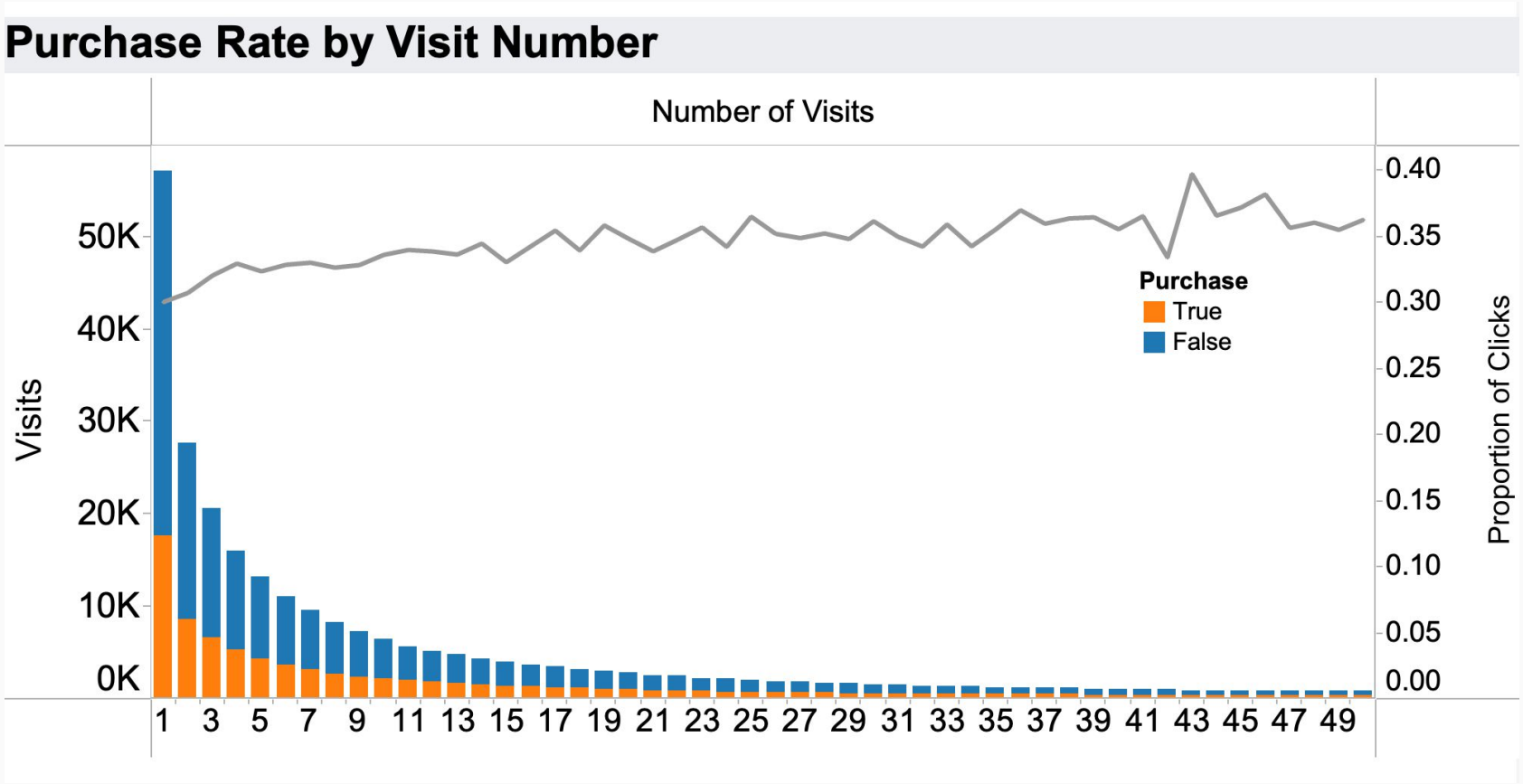
Shopping Duration



Number of Promo Codes Tried



Purchase Rate by Visit Number



Conclusion

What behavior will increase a visitor’s chances of purchasing?

- Having **more clicks** on the website
- **Longer** shop durations
- Finding a **working promo code**

Limitations/Future Plans

Imbalanced Data Set

- Number of users abandoning cart outweighs purchasers

  → Potential use of undersampling to improve models

Potential Confounding Variables

- Shop duration and clicks naturally increase because it takes time and clicks to make a purchase on Fingerhut.com.
  - In our paper we must investigate statistical significance of the increasing trends we identified in our models and plots

Future Work

Implementation of Undersampling

→ Address imbalanced data set limitation

Potential inclusion of price and review data

→ Expand variable pool to increase model accuracy

Deeper dive into impact of confounding variables

→ Address confounding variable limitation