

Shopping Trip Classifier

What does your cart say about your shopping trip?

Sanghamitra Dutta

25th Oct 2017



Objective

Classify the type of shopping trip
based on
Products bought

Data Description

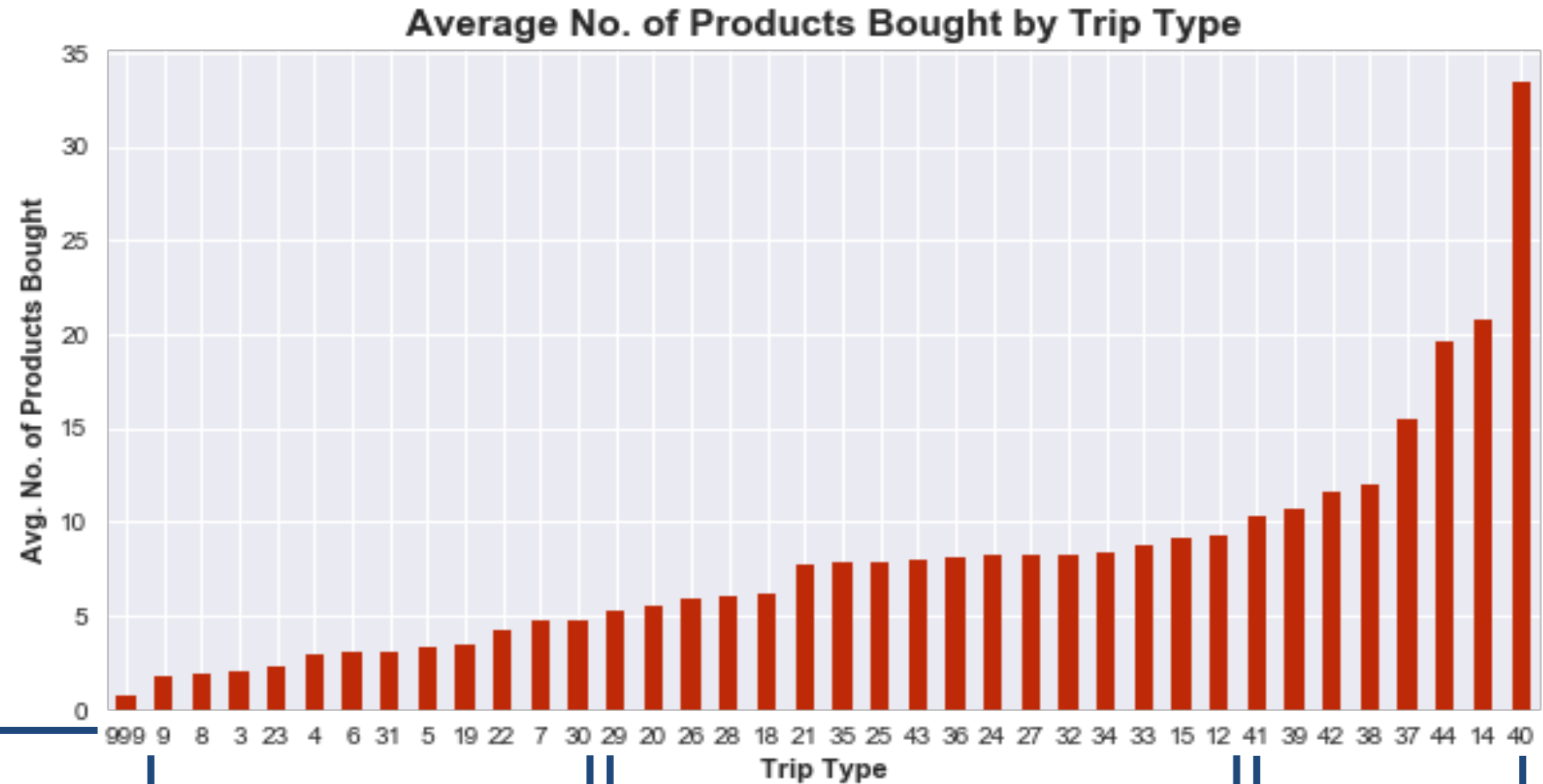
- Data acquired from Walmart Kaggle Competition
- Label : Trip Type
 - An id code to describe the trip type
 - 38 different types based on Walmart's proprietary method
- Features:
 - Visit Number: Id for an unique visit (over 95K visits)
 - Department Description - 68 Product Departments/ Product Categories
 - Scan Count – No. of products bought

Data Exploration

Deciphering Walmart's Trip Type Coding Based on:

- No. of Products Bought
- Product Categories Bought

Trip Types



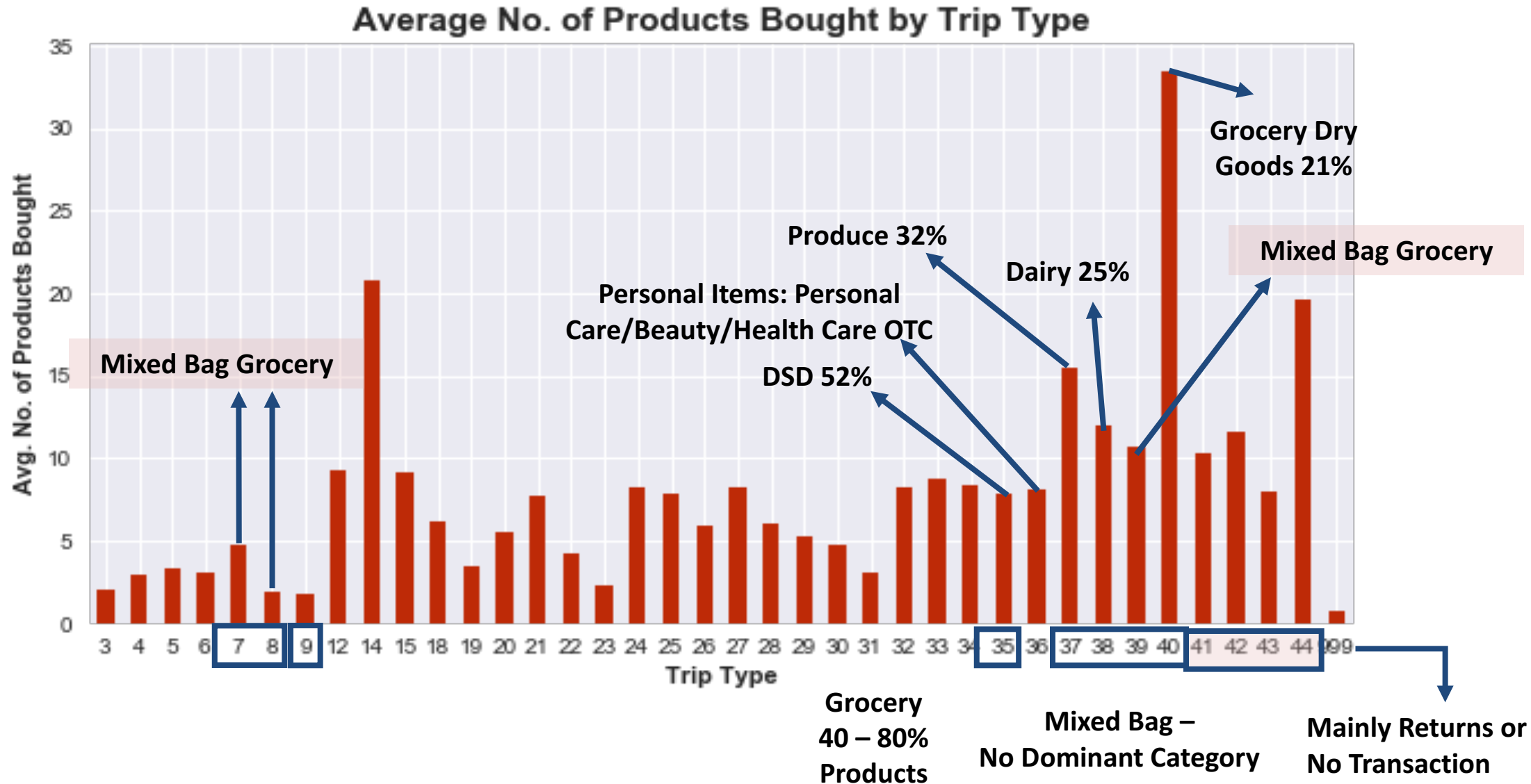
Mainly Returns or
No Transaction

Avg.: 1 to 4 Products

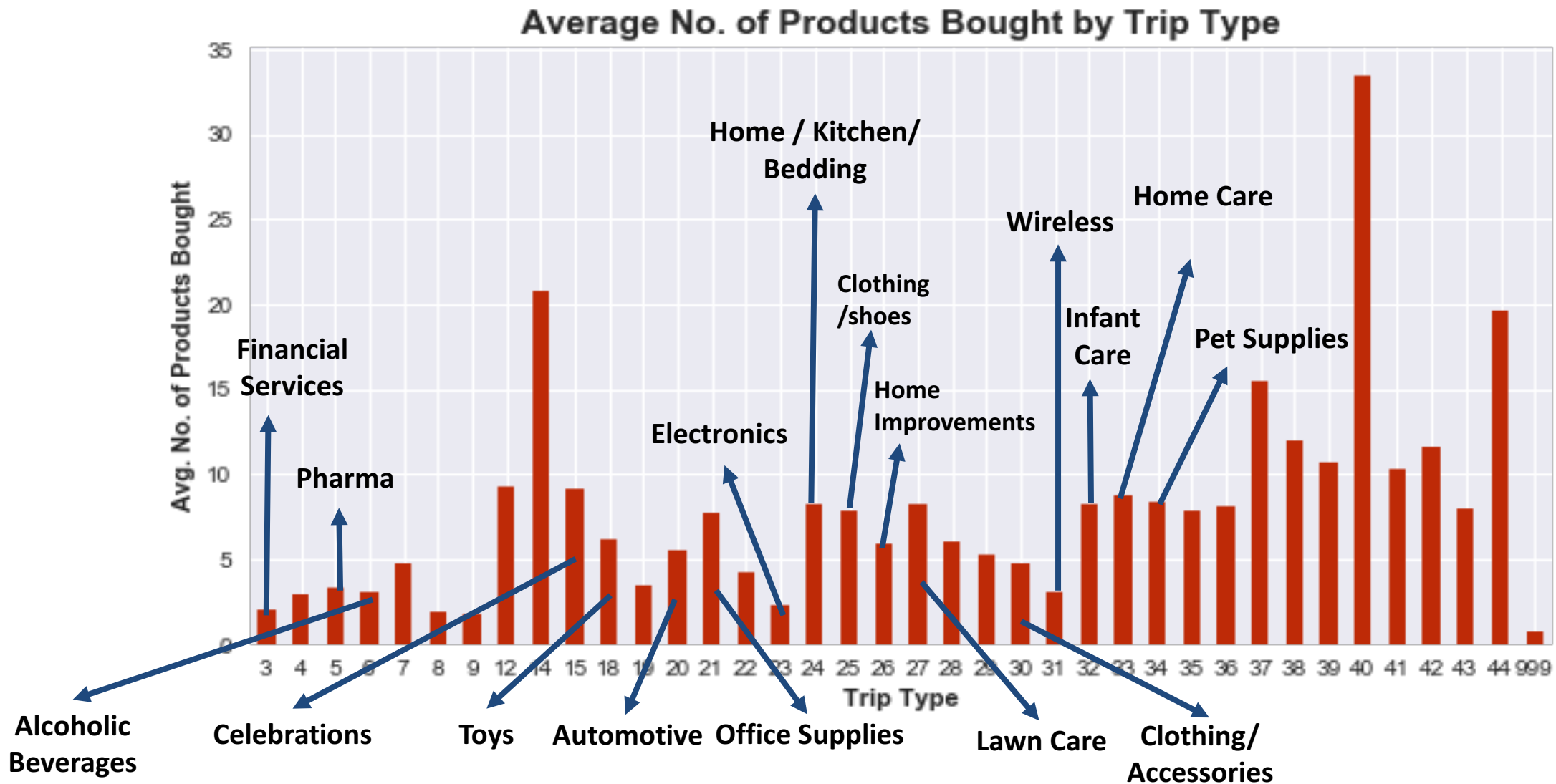
Avg.: 5 to 9 Products

Avg.: 10 or more Products

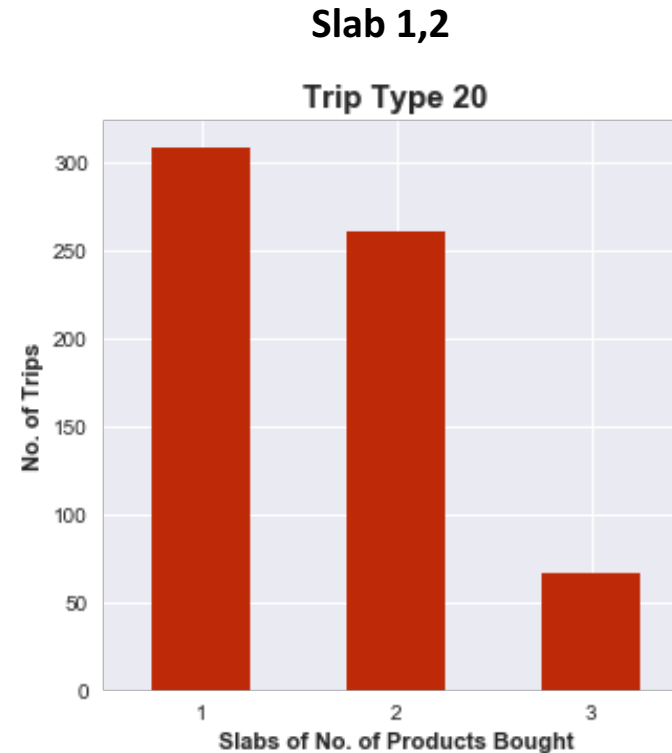
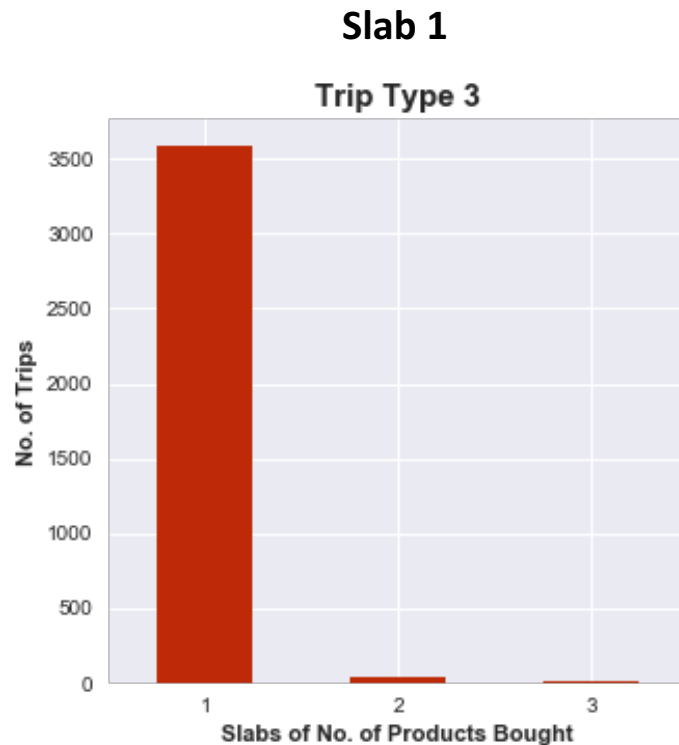
Splitting Trip Type by Categories Bought



Splitting Trip Type by Predominant Categories Bought



Binning Trip Types based on No. of Products bought



Slab Definitions

Slab0: No Transactions

Slab1: 1 to 4 Products

Slab2: 5 to 9 Products

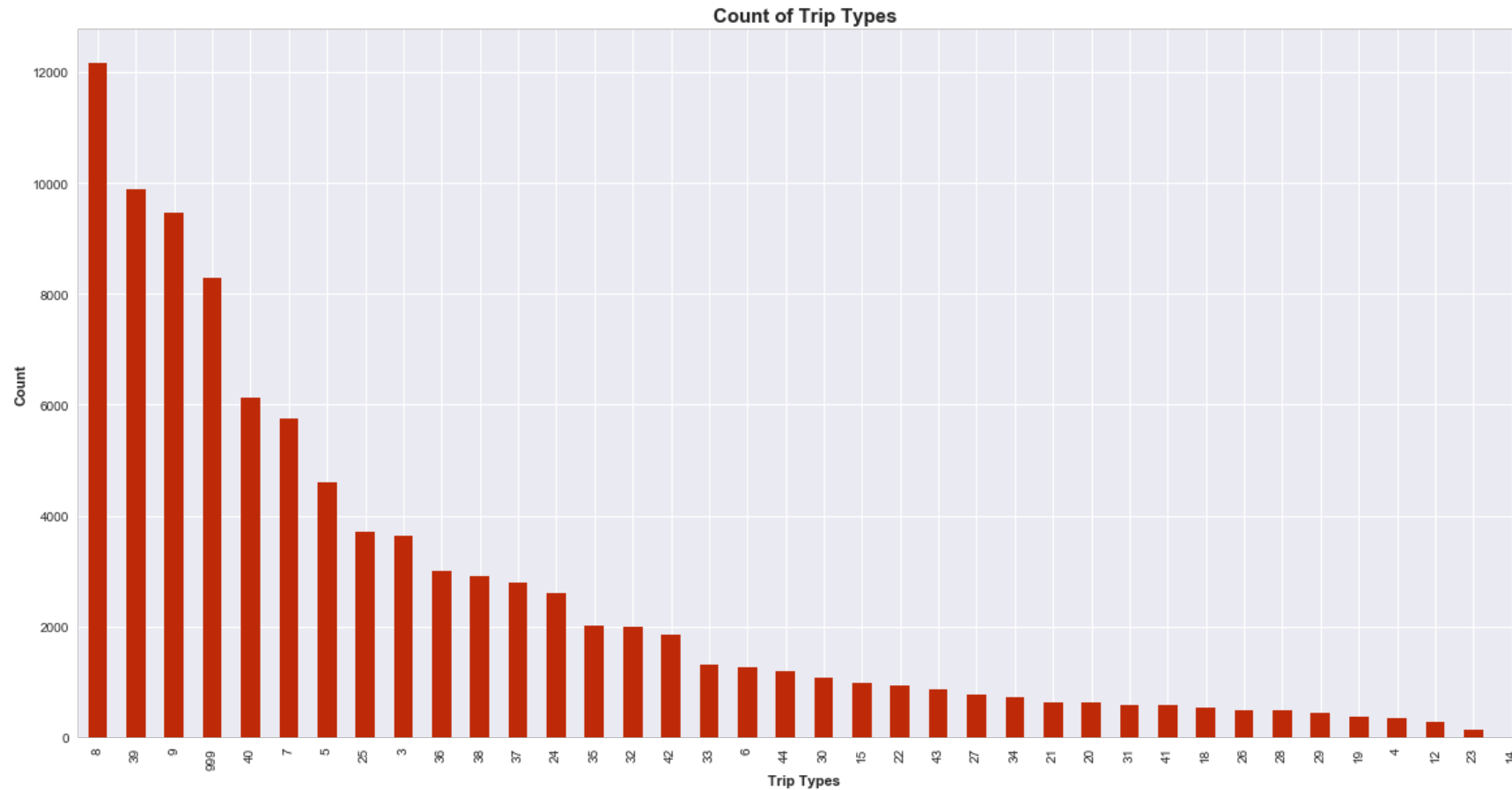
Slab3: 10 or more Products

Features Added/Transformed

- Recreated each trip with the category & count of products bought
- Dummy Variables For Weekdays
- Total No. of Categories Bought
- Total No. of Products Bought
- Returns Only

Modeling

Unbalanced Classes



Balanced Classes

Up-sampled Minority Classes and Down sampled Majority Classes to 1800 samples
Dropped 7 Classes with very low samples

Model Performance

Model	Cross Validated Accuracy Score (%)
Random Forest	74.54
SVM Linear Kernel	70.64
SVM RBF Kernel	70.01
Logistic Regression (Multinomial)	67.02
KNN (k=1)	65

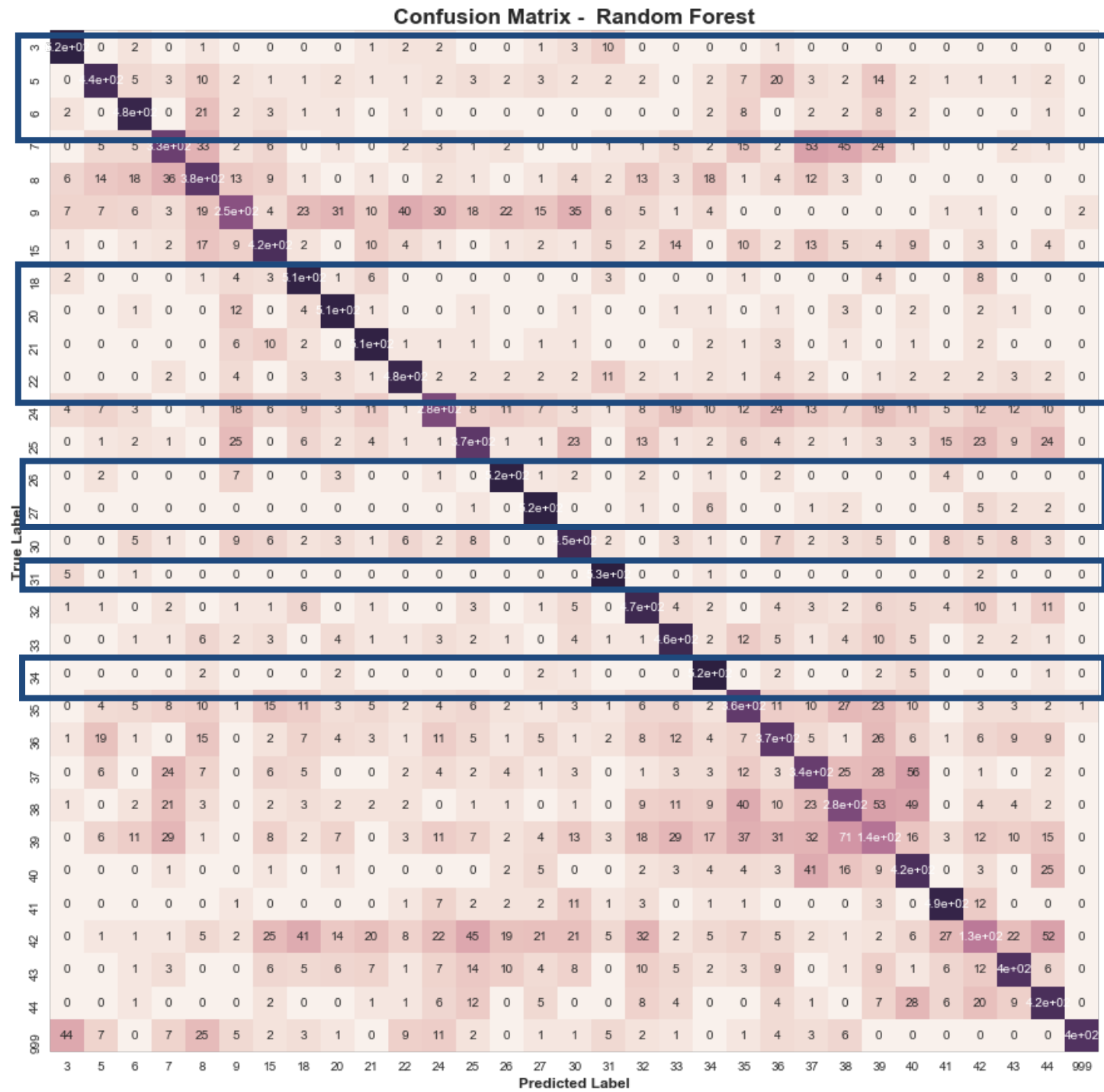
A closer look at Random Forest

Using Confusion Matrix
&
Classification Report

Random Forest Successes

Classification Report:
Success: F1 Score > 0.85
Failure: F1 Score < .65

Slab Definitions
Slab1: 1 to 4 Products
Slab2: 5 to 9 Products
Slab3: 10 or more Products



TripType:3 (Slab1)
Financial Services: 80%

Trip Type: 5 (Slab1,2)
Pharma OTC & RX: 58%

TripType:6 (Slab1)
LIQUOR,WINE,BEER- 30%
CANDY, TOBACCO,
COOKIES- 20 %

TripType:18 (Slab1 & 2)
Toys - 40.4%
Impulse Merchandise -12.4%

TripType:20 (Slab1 & 2)
Automotive- 54.44%
DSD Grocery -6.85%

TripType:21 (Slab1 & 2)
Office Supplies- 34.42%
Fabrics & Crafts -31.35%

TripType:22 (Slab1)
Electronics- 20%
Media & Gaming-12%

Trip Type:26 (Slab1,2)
Home Improvement -56%

Trip Type:27 (Slab1,2,3)
Lawn Care/ Horticulture-60%

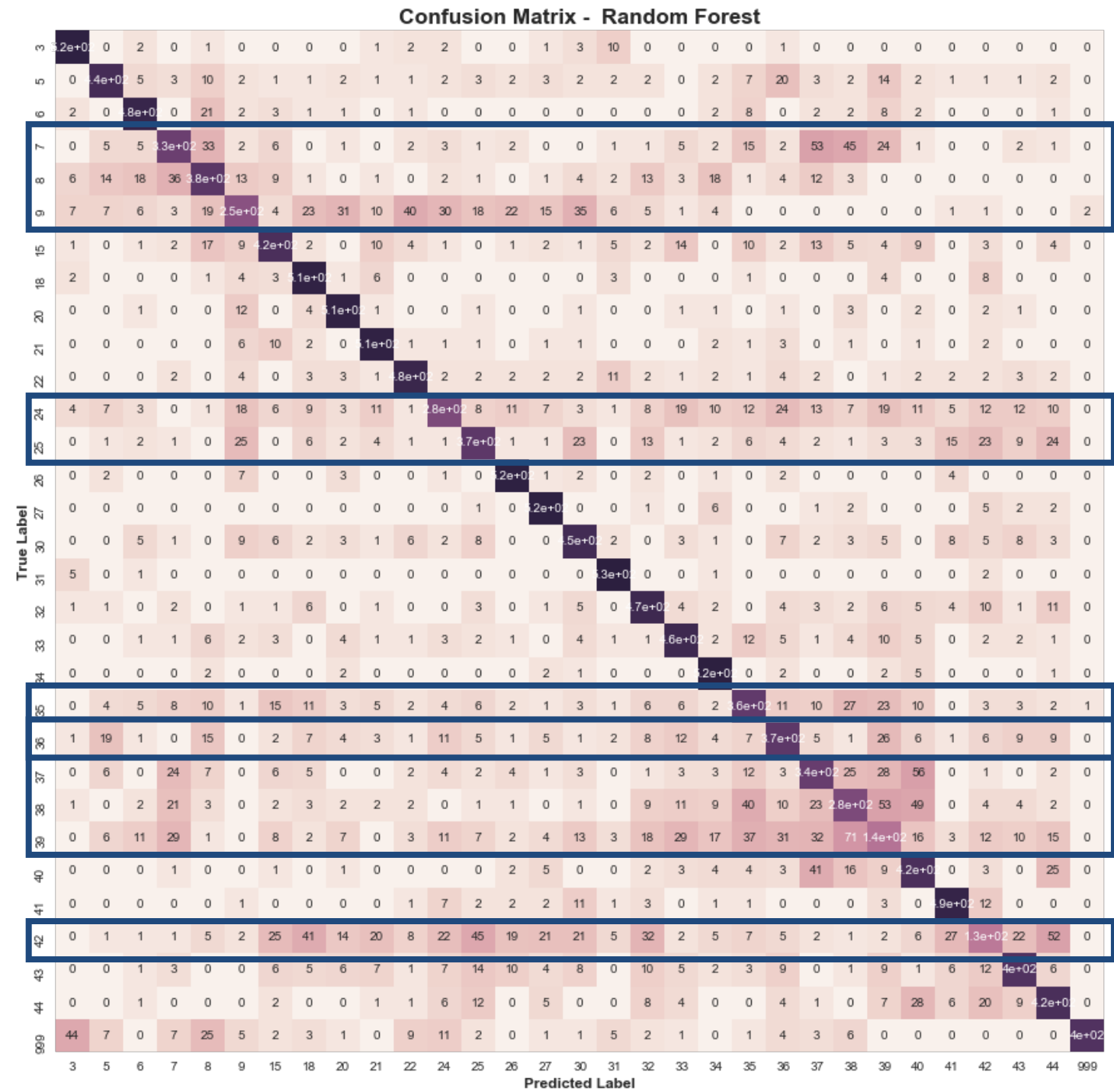
Trip Type:31 (Slab1)
Wireless -54%

Trip Type:34 (Slab1,2,3)
Pet Supplies - 61%

Random Forest Failures

Classification Report:
Success: F1 Score > 0.85
Failure: F1 Score < .65

Slab Definitions
Slab1: 1 to 4 Products
Slab2: 5 to 9 Products
Slab3: 10 or more Products



Trip 24:

- Each category has low contributions ~12%
- Lots of products bought along with grocery

Trip 25:

- None of the individual categories are high contributors
- Trip 30(shoe/accessories/clothing trip) is more accurate as it has lesser no. of products are bought

Trip 36:

- It might be confusing the model as these categories appear frequently in grocery classes

Trip Type:42 (Slab1,2,3)
Mixed Bag

Trip Type:24 (Slab1,2,3)
Home , Kitchenware & Bedding Items ~12% each

Trip Type:25 (Slab1 2,3)
Clothing /shoes – 56%

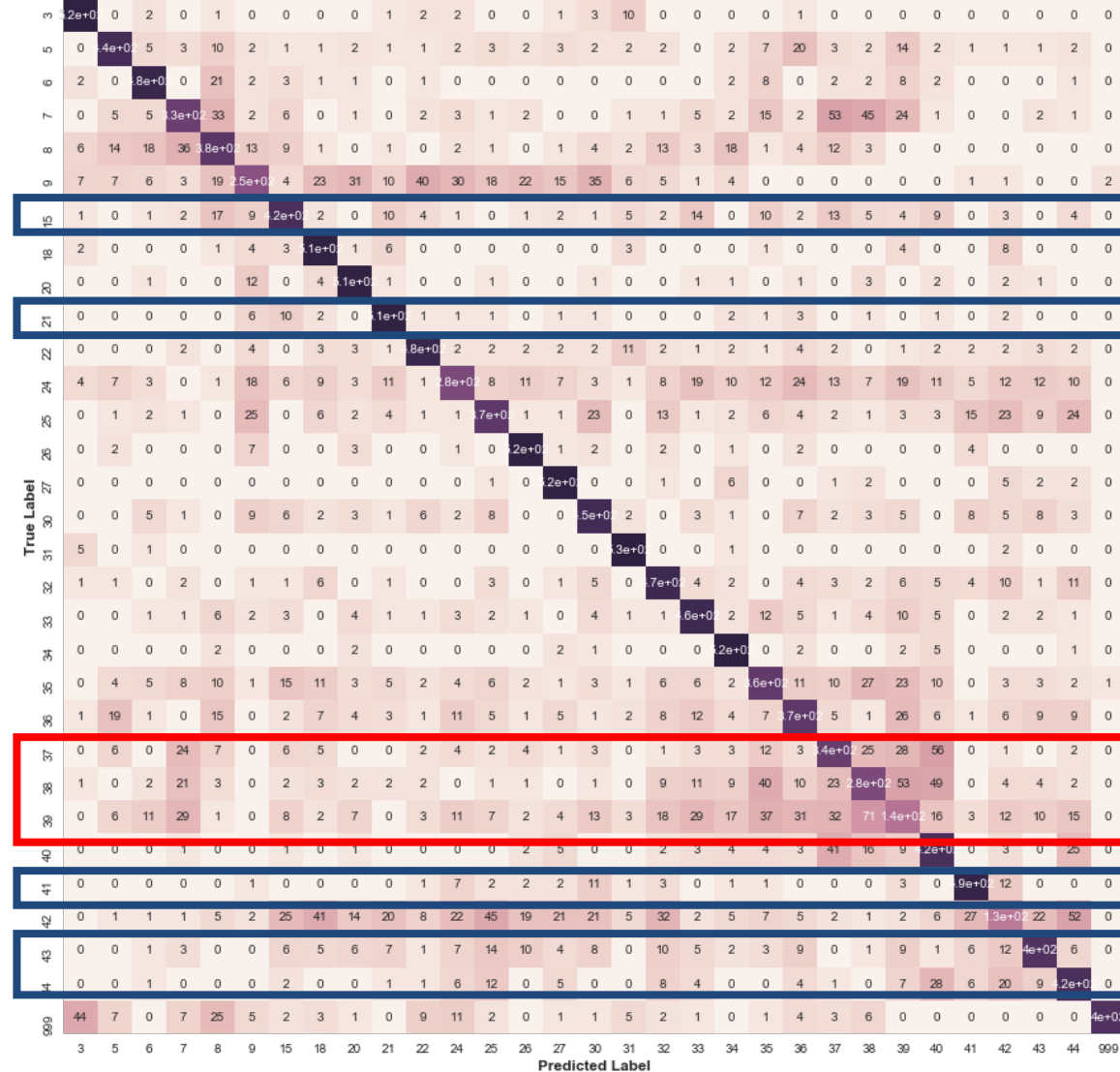
Trip Type:36 (Slab1,2,3)
Personal Items: – 70%
Personal Care - 40%

Evaluating Random Forest vis-à-vis Other Top Models

Using Confusion Matrix

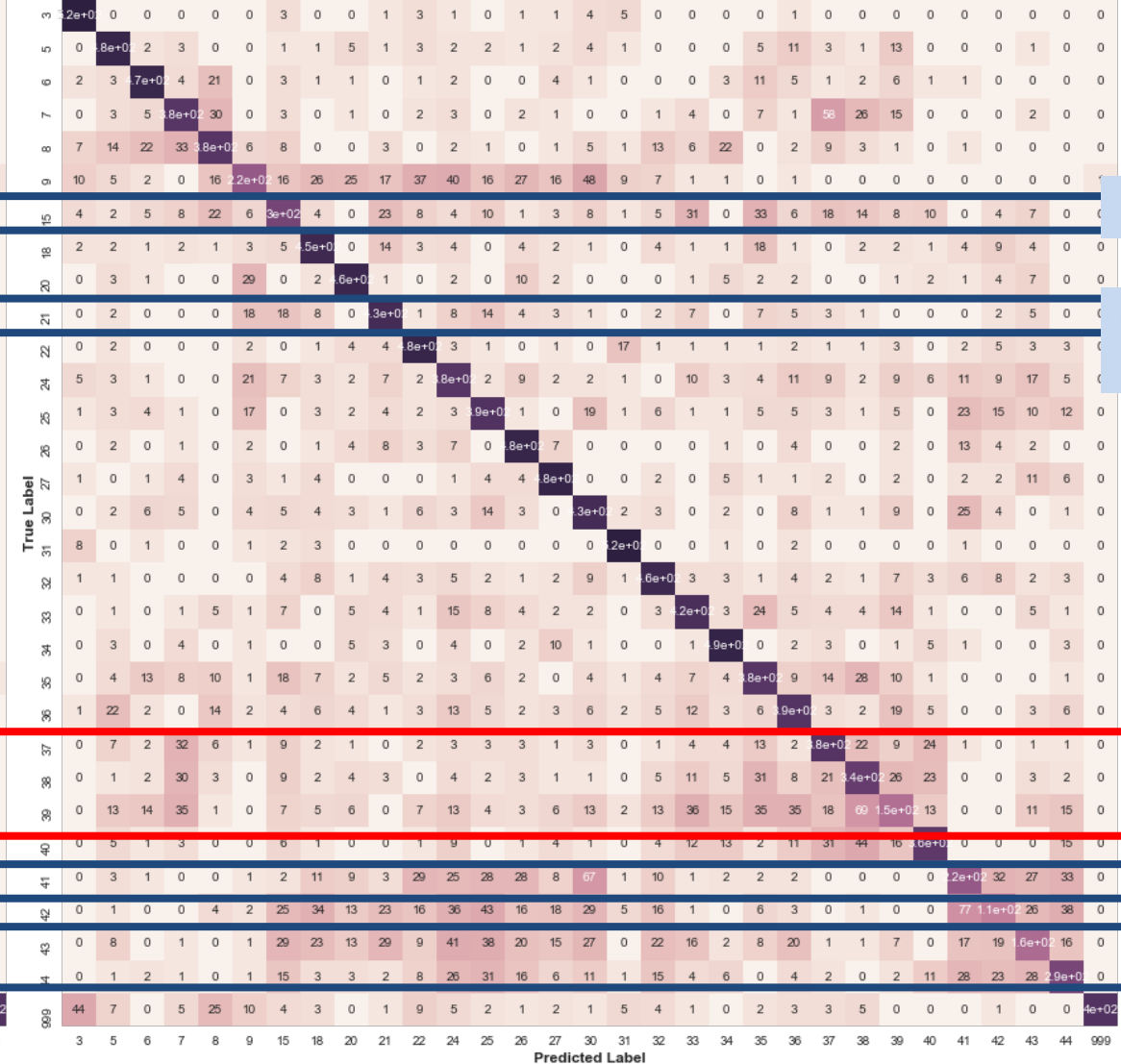
Random Forest Vs SVM Linear

Confusion Matrix - Random Forest



RFC Sores Better: 15, 21, 41, 43, 44

Confusion Matrix - SVM, kernel = Linear



SVM(Linear) Scores Better: 37, 38, 39

Celebrations

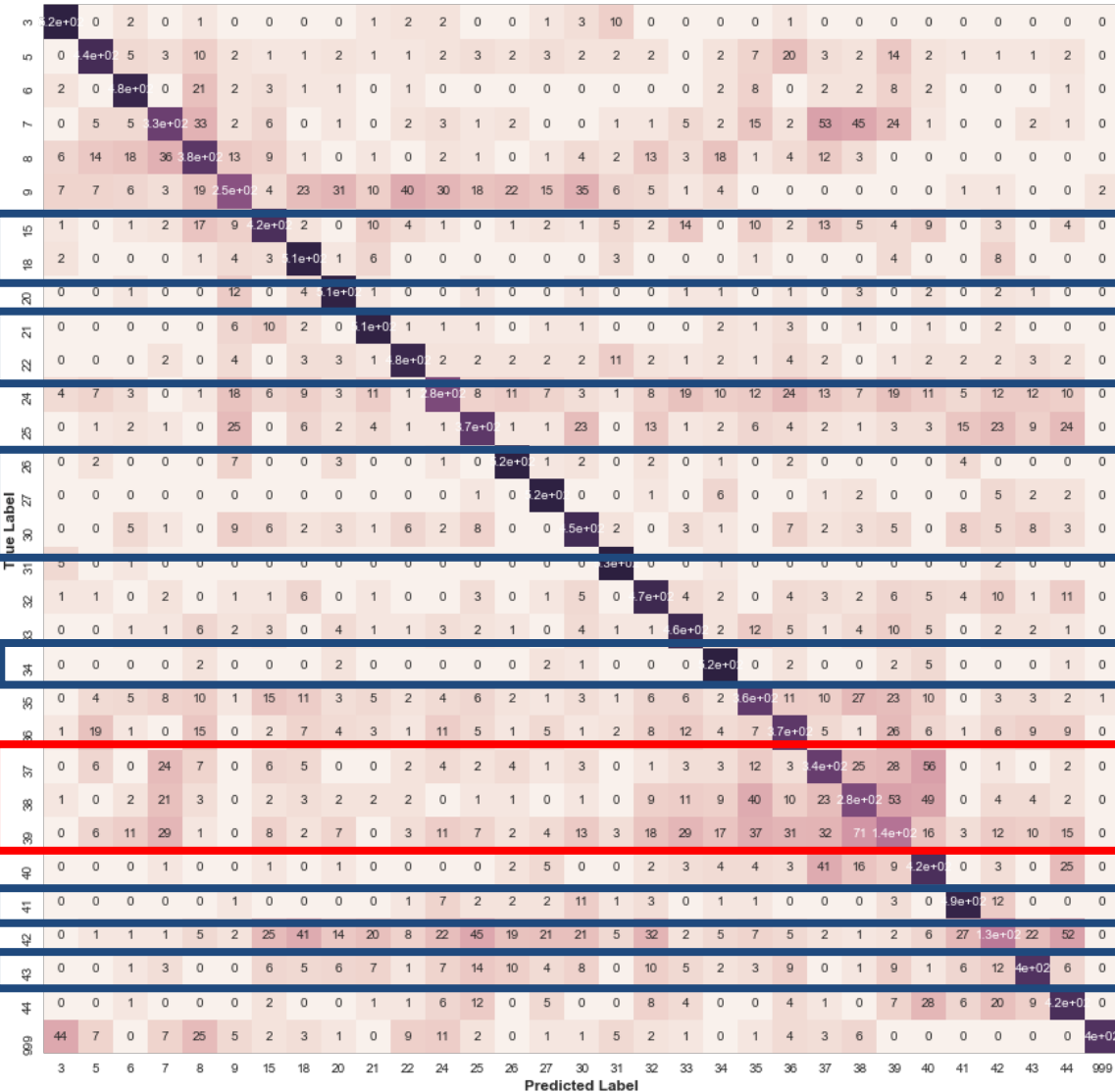
Office
Supplies

Grocery
Categories

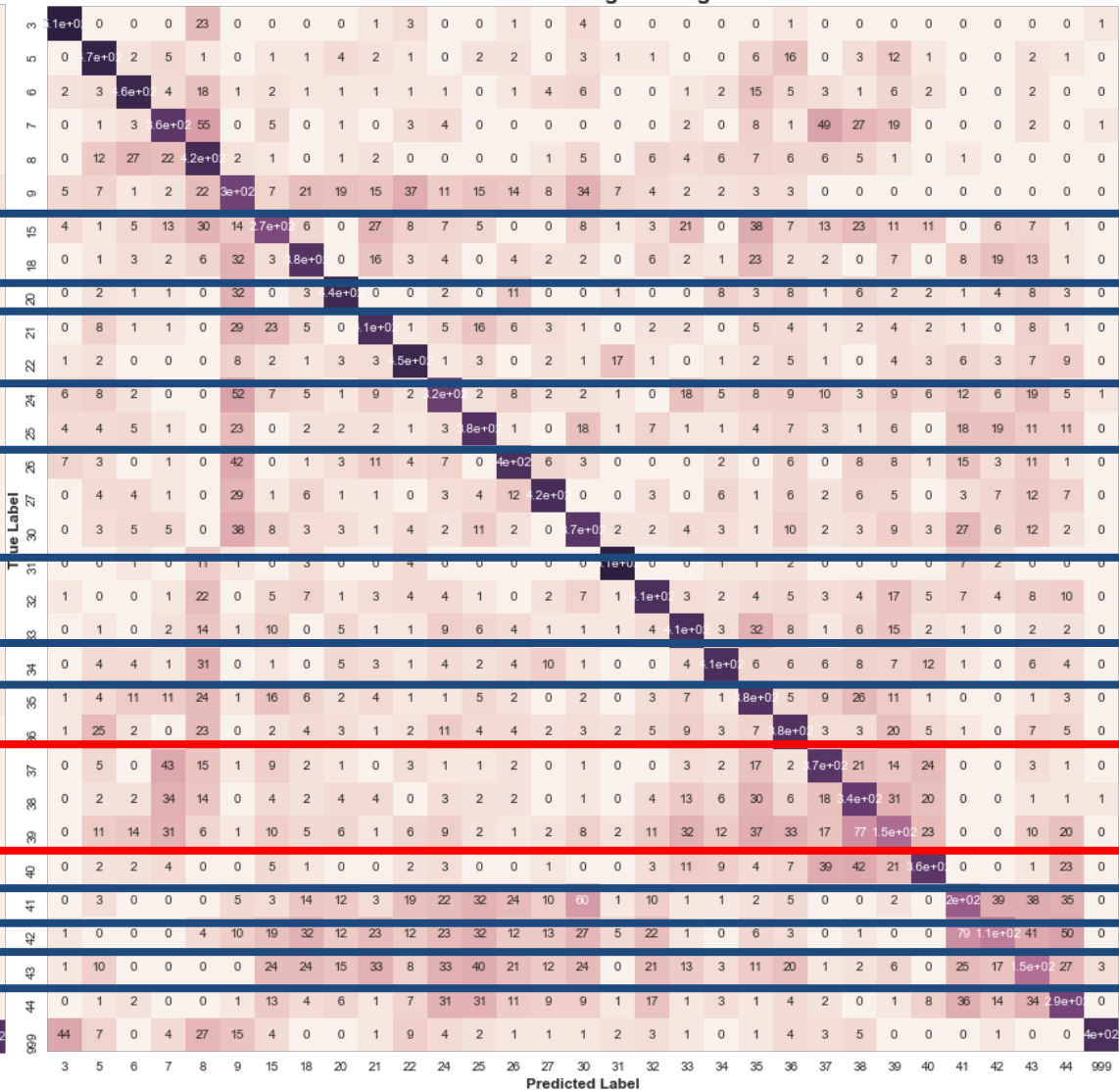
Mixed Bag
Categories

Random Forest Vs Logistic Regression

Confusion Matrix - Random Forest



Confusion Matrix - Logistic Regression



Conclusion

- Key Findings

- At an overall level Random Forest performance is superior to all models
- Random Forest does well in classifying focused trips where:
 - Specific and less frequent categories are purchased
 - Purchases are limited to fewer categories and total products

- Further Feature Development

- Within mix bag categories for both grocery and non-grocery find features to distinguish these classes

Thank You!!

Appendix

Trip Type:3 Slab1 Financial Services Impulse Merchandise	Trip Type: 5 Slab1,2 Pharma OTC & RX: 58%	Trip Type:6 Slab1 LIQUOR,WINE,BEER- 30.0% CANDY, TOBACCO, COOKIES- 20%	Trip Type:7 Slab1,2 Mix Bag Grocery- 73%	Trip Type: 8 Slab1 Mixed Bag Grocery – 50% Personal Items: -27%	Trip Type:9 Slab1 Mixed Bag Non- Food
Trip Type:15 Slab1,2,3 Celebrations-35%	Trip Type:18 Slab1,2, Toys-40%	Trip Type:20 Slab1 & 2 Automotive- 54.44% DSD Grocery -6.85%	Trip Type: 21 Slab1,2 Office Supplies - 34 Fabrics & crafts - 31	Trip Type: 22 Slab1,2 Electronics/Gaming/ Camera: 35%	Trip Type:24 Slab1,2,3 Home , Kitchenware & Bedding Items -46%
Trip Type:25 Slab1 2,3 Clothing /shoes – 56%	Trip Type:26 Slab1,2 Home Improvement - 56%	Trip Type:27 Slab1,2,3 Lawn Care/ Horticulture-60%	Trip Type:30 Slab1,2 Clothing/Accessories/ Shoes -60%	Trip Type:31 Slab1 Wireless –54%	Trip Type:32 Slab1,2,3 Infant Care –55%
Trip Type:33 Slab1,2,3 Home Care –56%	Trip Type:34 Slab1,2,3 Pet Supplies – 61%	Trip Type:35 Slab1 Grocery – 86% DSD Grocery 21%	Trip Type:36 Slab1,2,3 Personal Items: – 70% Personal Care - 40% Beauty – 17%	Trip Type:37 Slab1 Grocery – 82% Produce – 30%	Trip Type:38 Slab1 Grocery – 80% Dairy – 25%
Trip Type:39 Slab 2,3 Grocery – 60% Mix Bag of Grocery	Trip Type: 40 Slab 3 Grocery: – 70% Grocery Dry Goods - 20%	Trip Type: 41,42,43,44 Slab1,2,3 Mixed Bag	Trip Type: 999 Slab 0,1 Returns No Transactions		

Weekend Vs Weekday

