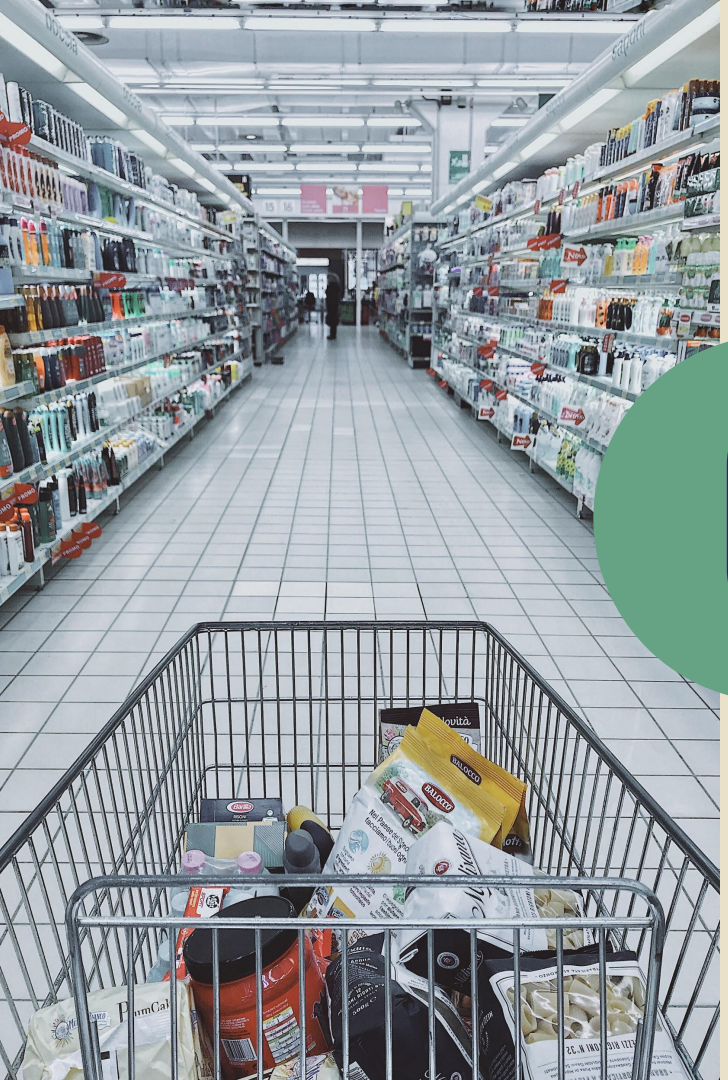
In Search of Spending

Jim Fay

Table of Contents



01

The Problem

What is the problem? How can we solve it with data science?

The Market

$602B

Online Spending

Consumer online spending with U.S. merchants in 2019.*

16%

% of Total Retail

Percent of total U.S. retail sales conducted online in 2019.*

75%

Frequent Shoppers

% of shoppers who make a purchase at least once per month.*

Our Goal

- Offer customers personalized incentives buy.
  - **Prioritize** high spending customers.
- Determine **which factors** influence consumer spending.

02

The Data

Exploring the data and its limits

Limitations

The Data

- Missing data points.
- Unavailable data.
- Not generalizable to all stores.

Sampling

- Was there a **pattern** to how this data was collected?
- Is it **exhaustive**?

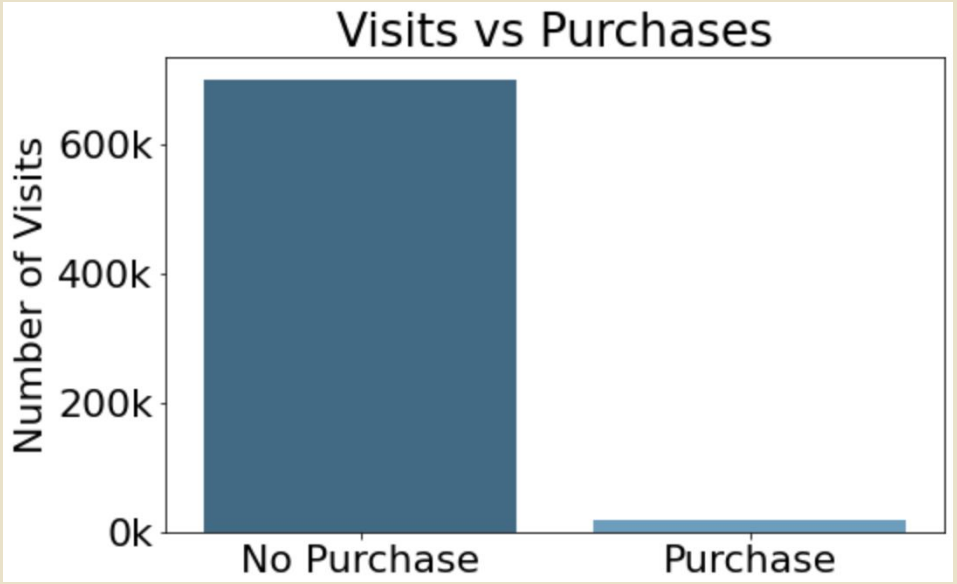
The Model

- Limited predictive power.
- Limited computer power.

The Data

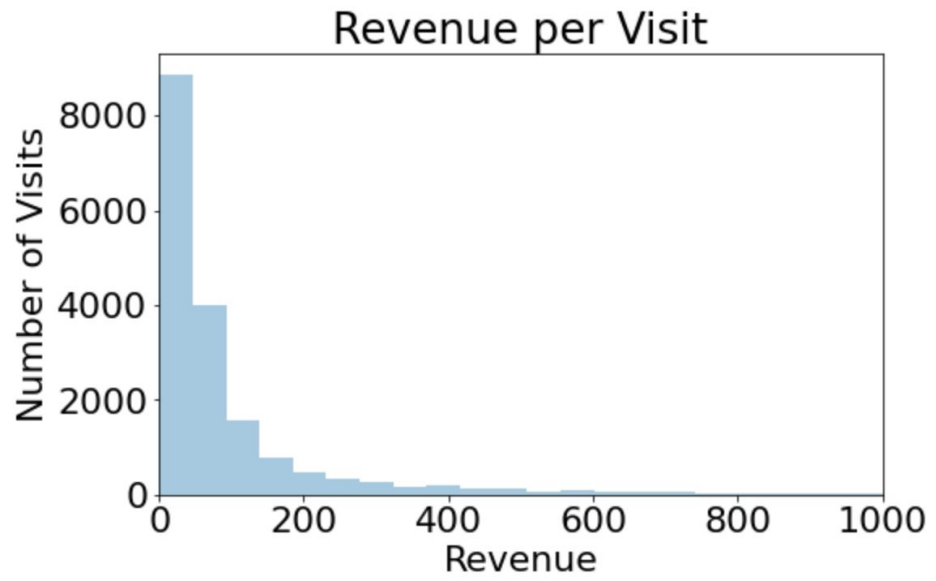
Breakdown

- 717k visits to Google's online store
- Data recorded between 2016-2018
- Types of factors used in the model:
  - Geographical
  - Device
  - Traffic Source
  - Page Views
  - Time
  - Price
- Made available through Kaggle.*



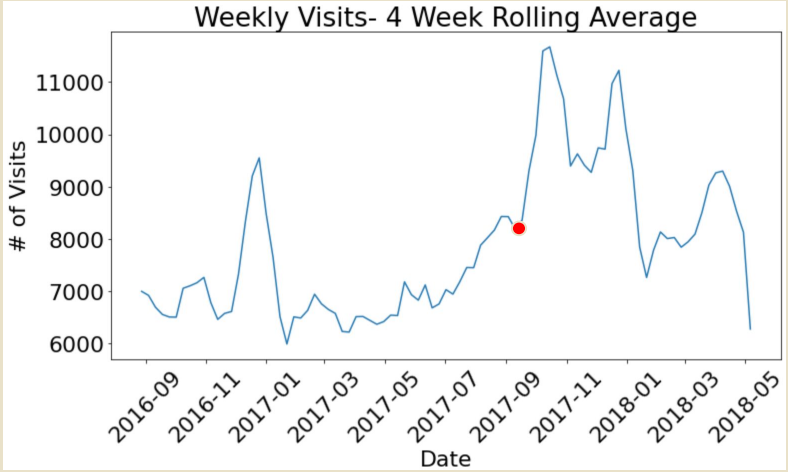
Breakdown: Purchases

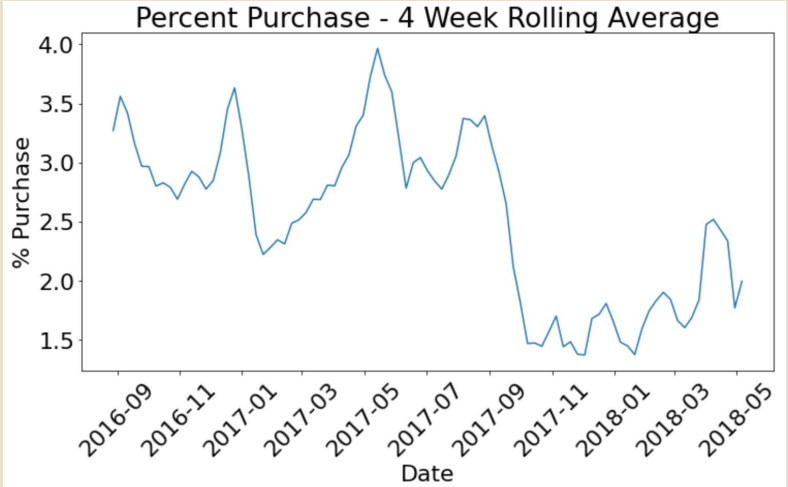
- 2.46% of visits result in a purchase.
- Most purchases are small.
- Average Purchase: $124



Purchases Over Time

- Visits increase with launch of new products.*
- New visitors don't make purchases at the same rate.





03

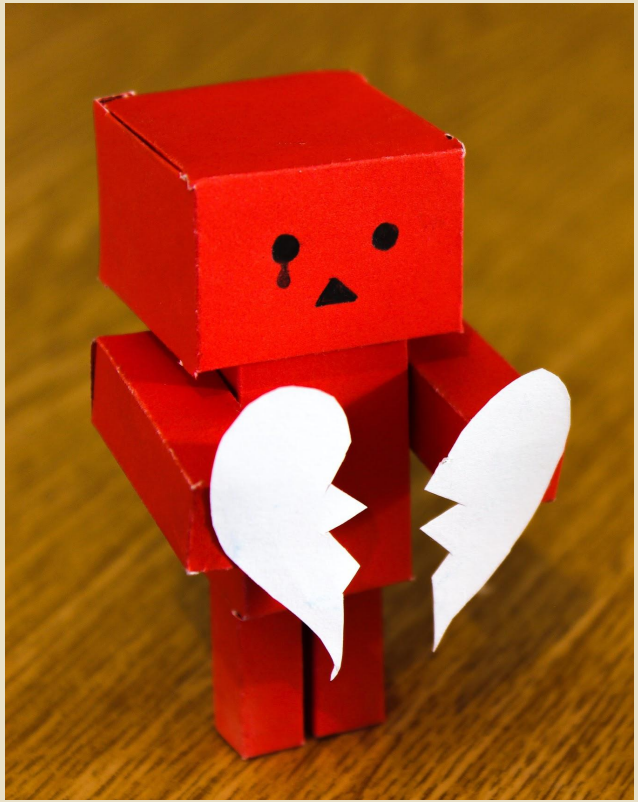
Data Analysis

Our model, results, and recommendations

The Model

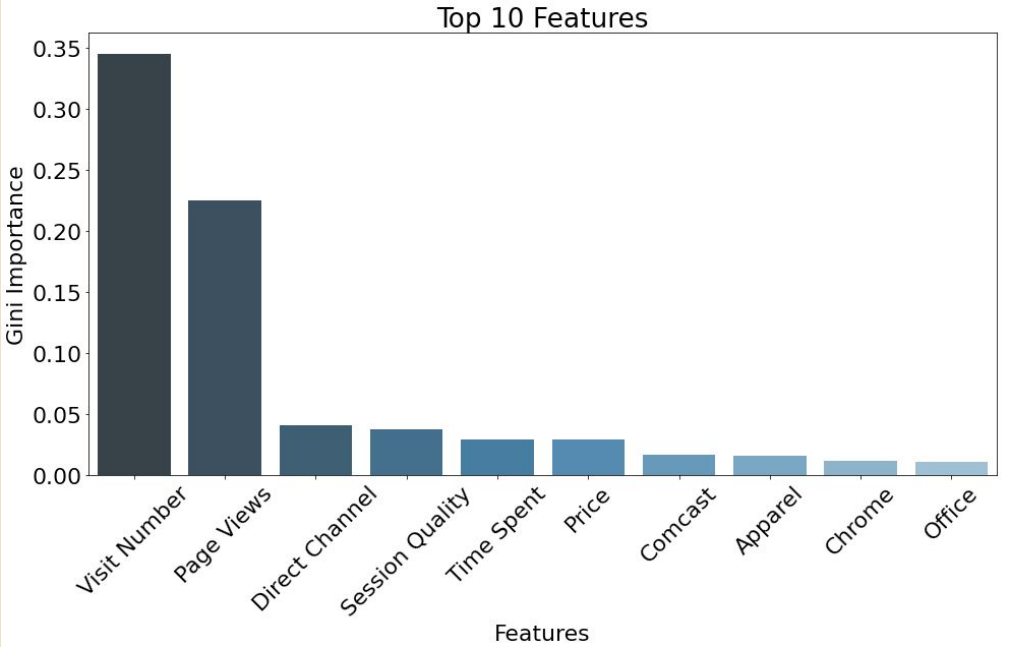
Random Forest Performance

- Cross Validation:
  - R-Squared = 8.5%
  - RMSE = 68
- Test Data Evaluation:
  - R-Squared = 0.9%



Predicting Spending

Model Results



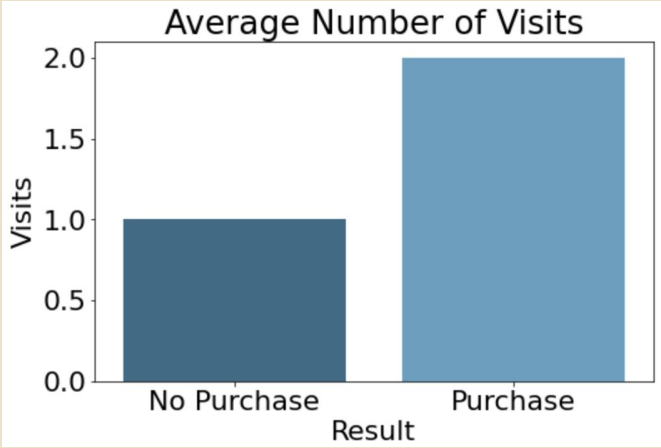
Interpretation:

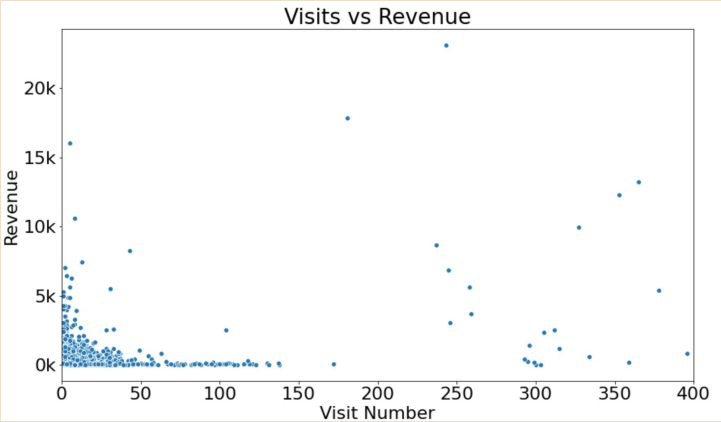
- **Activity** and **time** spent on the site are key factors.
- Which **products** were viewed is also important.

Number of Visits

Influence on Spending

- Customers that make a purchase have visited the site more.
- Most customers make a purchase within their first 25 visits.

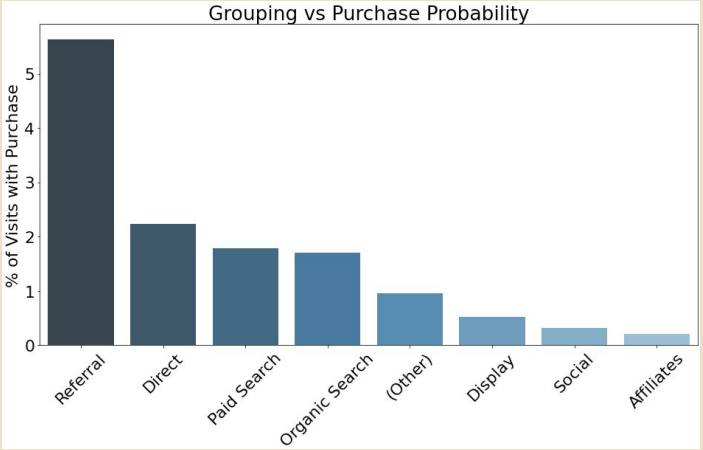


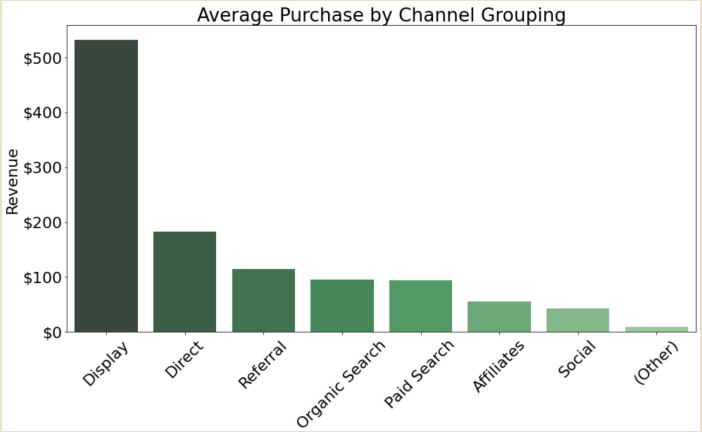


Channel Grouping

How Did the User Find the Store?

- Referrals lead to purchases.
- Banner ads lead to fewer, but larger purchases.





Recommendations

Tailor Marketing Efforts for High vs Low Spenders

- Offer better prices for bulk purchases.
- Incentivise purchases of any size with loyalty programs or one-time discounts.
- Continue to update model as new data is gathered.

04

Future Improvements

How can our solution be improved?

Improvements

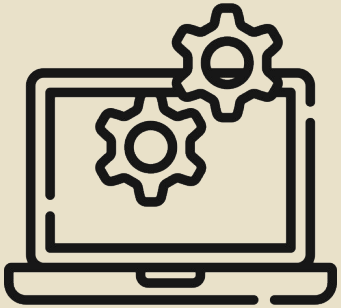
Additional Data

Incorporate **economic** data

Gain access to additional data from Google

Deployment

Deploy the model as a web application



Improve Model

Use different ML models.

Generalize model to work with other online stores.

Questions?

jrf6xh@virginia.edu

* All images in this slideshow are available royalty free from pexels