Jackie Kirschner
Capstone Proposal

Google Merchandise Store Churn

Improving conversion rate is key for any e-commerce store. Conversion rate can be defined as the percentage of website visitors who buy something on the site.

Potential customers are driven to the site in various ways such as search engines, social media, paid ads, or referral sites. They browse the site from all over the world, and shopping can be done on desktop, mobile, or tablet with various browser types and text translated into their natural language.

Google Analytics tracks several metrics including how visitors reached the site, landing page, IP address location, device type, number of times on site, number of web pages visited, time spent on site, and conversions.

First time visitors are especially valuable for e-commerce stores. If users convert the first time they visit, marketing dollars are saved. However, it's fairly unusual for a first time visitor to purchase.

Question:
Can I predict whether or not a first-time visitor will convert? If so, what are the most important features leading to a purchase?

What are the data and do you have them, and have you looked at them?
Google published their merch store analytics from 8/1/2016-8/1/2017. I queried and downloaded one week of data from Google BigQuery. There are ~12,000 rows and 8 columns (mostly categorical to be dummied) of web session data.

What is the MVP (minimum viable product)?
Train a logistic regression model to properly classify first-time visitors who convert. Discover most important features leading to a purchase.

MVP +, MVP ++, MVP +++, etc:
- Improve the model (hyper-parameters, dimensionality reduction, clustering).
- Compare with other classification models.
- Find thresholds for numeric features (page views, time on site) that lead to max conversions.
- Run the model on a group of returning visitors - are most important features the same? How does this translate to marketing strategy?
- Use big data tools to run the model on an entire year of website sessions.