

Commerce data exercise

Part 1

Consider the following schema for a table named **Products**, which stores the category and app IDs of a product. Note that a product is not guaranteed to have both an iOS and an Android app, but assume that each product has at least one of the two.

field	type	description
product_id	STRING	unique ID of a product
category	STRING	the product's category (e.g., "Finance")
ios_app_id	STRING	unique ID of the product's iOS app (e.g., "123456789")
android_app_id	STRING	unique ID of the product's Android app (e.g., "com.app")

Consider also the following schema for a table named **Downloads**, which contains daily download data for apps (not products).

field	type	description
app_id	STRING	an app ID (could be either iOS or Android)
date	STRING	in yyyy-mm-dd format
downloads	INTEGER	the number of downloads this app got on this date

Helpful definitions

- A trailing n-day average of a metric, as of day D, is the average of the metric's value over days (D – n+1) through D (i.e., the past n days)
- Year-on-year (YoY) growth of a metric, as of day D, is the metric's value on day D divided by the metric's value on day D – 365 (i.e., a year ago)
- The product-level downloads of a product on day D is the sum of the product's iOS downloads and Android downloads on that day (if they exist)

Write a SQL query that calculates the YoY growth of the trailing 28-day average of product-level downloads, for each product and each day.

Part 2

Commerce is an age-old industry that has seen a lot of disruption over the past couple of decades. From Amazon and eBay to more recent examples like Stitch Fix and Dollar Shave Club, many companies have leveraged the internet to create successful e-commerce businesses.

We hypothesize that these e-commerce “winners” exhibit similar patterns or characteristics. If we can identify these patterns, we may be able to use them to pick out future winners as well. To test this hypothesis, we have collected sales data for several recent successful e-commerce companies.

Using this data, please:

- Propose a set of common patterns shared by these successful companies and formalize them as a precise rule that can be applied to other companies (a simple but useless example is “they all had at least \$1 of sales”)
- Interpret these patterns and evaluate whether they make intuitive sense in the context of e-commerce businesses and whether they would be useful for identifying future winners

Brandless is an e-commerce company that sells everyday essentials (e.g., snacks, tissues, soap) online. We also have data on its sales.

- Using the full dataset, train a model that predicts future monthly sales and use this model to forecast Brandless's monthly sales over 12 months (i.e., through September 2019)
- Based on the data, the success patterns you identified, the forecasted sales, and any other information you think is relevant, how likely is Brandless to become a winner? Why?
- What additional pieces of data/information would you want and what other analyses would you perform to make a higher conviction assessment?

Please prepare a presentation of your analysis that addresses the 5 questions above. Assume that the audience is data-savvy investors. Please also include the code used to build the predictive model, preferably in Python.

Link to data:

<https://www.dropbox.com/s/ibt9wcck1zfqs79/Commerce%20analysis%20-%20full%20sales%20data.xlsx?dl=0>

The following table describes what each column represents in the “Sales” tab:

Column	Description
Month	The month that this row's data represents
Sales	Sales (revenue) generated during this month
Customers	# of customers who transacted during this month
Transactions	# of transactions during this month

The other tabs contain cohort-based data for each of the companies:

Column	Description
Cohort month	The month that this row's data represents
Cohort size	# of customers who first transacted during this month
1	% of this cohort of customers who transacted again 1 month later
2	% of this cohort of customers who transacted again 2 months later
...	...
n	% of this cohort of customers who transacted again n months later

For example: In January 2013, 25 customers transacted with Blue Apron for the first time. In April 2013 (3 months later), 60% of these same customers transacted with Blue Apron again.

Notes:

- Sales, customers, and transactions are scaled down
- New cohorts will have less data available because they have not had sufficient time to age