

E-Commerce Problem Statement

Background

Electronic commerce or more commonly known as e-commerce is defined as the activity or purchasing or selling good through online electronic means or over the world wide web. The advent of the internet enabled unprecedented growth in communication channels spanning the entire globe. As a result, industries emerged and markets for online purchases developed to meet demand for products not available domestically. In the United States, [reports](#) indicate that from the time period of 1998-2015, e-commerce sales have been growing nine times the rate to that of traditional in-store sales, and have accounted for roughly 7.2 percent of all retail sales in the year 2015. This statistic is only expected to get larger as a result of recently designed mobile commerce sales, and overall estimates indicate a total valuation of worldwide [B2C e-commerce](#) sales amounting to 2.3T USD in 2018.

This large expansion has seen a drastic development of novel companies offering a multitude of services within the overall e-commerce landscape. Amazon, Alibaba, e-Bay, and Shopify are a few prominent examples which operate within the vast e-commerce space. As a result of this globalization and spread of e-commerce platforms, it is imperative for traditional retail sales industries to engage in e-commerce activities to further advance their revenue streams, remain profitable and operational.

Fundamentally, e-commerce has also led to a shift in marketing and advertising strategies, enabling companies to reach users via multiple channels (i.e., mobile applications) and expanding their reach with minimal effort. This in consequence has also generated substantial amounts of consumer data, which through careful mathematical analysis offers a powerful yet insightful ability to under consumer spending behaviors, which in turn drives their sales.

In addition, specifically tailored programmatic advertising and extraction of consumer data has brought on a new wave of companies and industries. The exponential growth of data generated through the “online shopping experience” enables companies to learn and extract fundamental information from the transactions to develop an insight into customer segments, biases, pricing and spending habits. We are at a cusp whereby further customer customization and improved user experience on all e-commerce platforms is advancing as a means of fierce competition. Examples of such innovative strategies include the incorporation of augmented reality, offering of chatbots and multiple methods of payment, on-site personalization and voice searches. Analyzing how e-commerce has developed overtime may gleam insight into how such technological disruptions have shifted society from a socioeconomic standpoint and the impact it has had on various industries.

Your Task

Your goal is to use the various e-commerce data streams in order to discover and analyze patterns in online shopping and make recommendations about dynamic pricing opportunities or more effectively leveraging insights related to customer segmentation. More broadly, you should consider how e-commerce has impacted consumer spending habits and adaptation of online platforms in various industries.

We have partially pre-cleaned several datasets for your use, including daily pricing data from several US retailers, product metadata and reviews from Amazon, transaction-level data in the

UK and Brazil, and online event-based consumer behavior data from RetailRocket. Note that these datasets are from distinct time periods and have very little overlap, but we encourage you to synthesize insights and extrapolate across different types of e-commerce data.

You are asked to pose your own question and answer it using the available datasets as well as any supplementary datasets you may find. What is important is both the creativity of your question and the quality of your data analysis. **You need not be comprehensive; depth of insight is more important over breadth of the question posed.**

Submissions may be predictive, using machine learning and/or time series analysis to predict future bike share trends. Submissions may also be illuminating, through the use of data visualizations or through sound statistical tests.

Consider exploring one of the sample questions below, or creating your own variation. Creativity in formulating your own question is encouraged; **however, it should not be at the expense of analytical depth, precision, and rigor, which are far more important.**

Sample Question 1: Given the ubiquity of online price-comparison services, what are the dynamics of product pricing among competing retailers?

Sample Question 2: What aspects of consumers' online behavior is useful for businesses to better understand their customers and predict consumer profitability?

Sample Question 3: How have broader trends in e-commerce impacted consumer spending habits and adaptation of online platforms in various industries?

Datasets

The provided datasets are spread across five primary datasets. Your team should only use the datasets that are relevant to your chosen question/topic. The raw data sources are noted but we encourage you to use our data since they have been organized and cleaned.

retailer_pricing

Dynamic product pricing data from 6 competing US retailers, from August 2019 to March 2020. ~22.5 million rows & 11 columns. Size: ~1GB zipped, ~3GB unzipped. Source: not public.

amazon_reviews (2)

Reviews and product metadata of millions of Amazon products, from 1996 - 2018. Size: ~45MB zipped, ~120MB unzipped. [Source](#).

- **beauty_reviews:** ~370,000 rows & 9 columns.
- **beauty_metadata:** ~33,000 rows & 9 columns.

online_consumer_behavior (2)

Online event-based consumer behavior data from RetailRocket, over 4 to 5 months. ~2.7 million rows. Size: ~1GB zipped. [Source](#)

- **events.** ~2.7 million rows & 5 columns.
- **item_properties.** ~20 million rows & 4 columns.

UK_retail_transactions

Transactions dataset from a UK-based online retailer, from 2009 - 2011.

~1 million rows & 8 columns. Size: ~15MB zipped, ~100MB unzipped. [Source](#)

BR_retail_transactions (8)

Transactions dataset containing 100k orders made at multiple marketplaces in Brazil through the [Olist Store](#), from 2016 - 2018.

Size: 45MB zipped, 120MB unzipped. [Source](#)

Additional Datasets

Some good starting points for additional datasets are this compilation of [recommender systems datasets](#) as well as this list of [e-commerce and retail datasets](#). You may also want to draw on survey and report-level data to motivate your report. Here are a few sources you may find useful: [one](#), [two](#), [three](#), [four](#), and [five](#).