1. **Introduction**

When a user visits an e-commerce site, there are numerous qualitative and quantitative factors that influence their decision to make a purchase. Qualitative factors being their spent budget, user needs at the moment and so on. Quantitative factors include product ratings, number of reviews, presence or absence of a product video, number of photos for the product, length of product description, price of a product – to name a few.

This capstone aims to understand which quantitative variables affects a user’s purchase decision. Consequent application of identifying the contributing factors can enable an e-commerce site to maximize profits by utilizing their user behavior data and displaying the most relevant information at all times.

1. **Data Acquisition**

Two primary data sources will be used throughout the project. The time range will be year-to-date (YTD), in order to focus on the most recent behavior.

1. Web analytics – this dataset contains conversion metrics per a unique product
2. Content management system – this dataset contains per product ID, user ratings, number of reviews, word count of content description, number of photos, videos on the site for a product, product classification
3. **Data Wrangling and Cleaning**
4. **Preliminary Exploration**
5. **Exploration and Analysis Approach**

One-variable linear regression

Segmentation by category

1. **Limitations**

This analysis limits the view of product conversion to measureable variables. There are qualitative signals that this approach does not take into consideration which in fact play a very important role in purchase decision making.

Even if all the product information is adequately laid-out on a website, a user may not buy it because their plans changed or the product isn’t a right fit for them at the given time. Such conditions are difficult to measure and this limits our understanding of why a user abandons buying the product as well as the ability to account for such conditions in the analysis.