1. Business Understanding
2. Define the problem

Fake product & service reviews on e-commerce platforms are a long-standing problem to both the platforms and the users. Because reviews are typically a significant factor in search rank algorithms, it becomes an incentive for sellers on platforms to exploit fake reviews to increase the visibility and sales of their products. Fake reviews are a problem because there are situations such that customers discover the products they bought are of lower quality because of the inflated ratings and spurious reviews. This would discourage customers to trust the ratings and reviews again and push them to leave the platforms ultimately. This would impede the development of platforms, so those businesses spend a great amount of money and efforts to reduce fake reviews. However, a study from UCLA found that Amazon, for example, was only able to delete around 40% of the fake reviews, and it took them an average of more than 100 days after one fake review was posted to remove it.

1. Define success

The negative impact of fake reviews could be 1). freight 2). customer service and relationship 3). the company’s reputation 4). retention. Fake reviews would create larger expectation gaps on the customers and thus drive them to return the product they bought. The customers might feel deceived and send their complaints to the customer service department, which likely requires the company to hire more agents and thus increase the operation cost. In any cases of failing to treat the customers as the way they expected, the company’s reputation stains. Eventually, the company would sink as retention rate lowers.

The target of the study is to maximumly limit the negative impact of fake reviews listed above and, in an effort, to help customers better gauge the ratings and reviews of a product.

Freight, customer satisfaction, and retention rate are quantitative metrics for a before-and-after comparison. A lower return ratio due to a quality issue, higher customer satisfaction, and higher retention rate are success targets to achieve.

1. Identify factors

A product that is available online usually is sold on more than one platform. I would collect the ratings and reviews of the same product/similar product (eg: an updated version of that product) from multiple e-commerce platform to compare and contrast. Consider various methods to filter the ratings and reviews of a product and illustrate more consistent data across different platform. If this doesn’t work out, the platform could still take advantage of the model developed to predict the damage before it happens and thus get better preparations.

1. Data Understanding
2. Gather data

Collect the ratings and reviews of the same product from multiple e-commerce platforms. In addition, explore key words of the product on social media platforms. Merging data is required for analysis. Noisy data is expected because ratings and reviews tend to be subjective. Labeled rating data is expected. Unstructured data like text and images requires manual labeling.

1. Validate data

Since ratings and reviews tend to be subjective, some kinds of products like facial products should be excluded for now. Outliers might appear because of fake reviews and emotional reviews.

1. Explore the data

Statistical analysis like OLS using the latest reviews might help explain how much impact does selected factors on the dependent variable. Visualization could be used to illustrate segmentation of reviews by geographics, sex, time, and etc.