

Abstract

The purpose of this project is to help a client competitively price their skis. My client who is trying to enter the ski market does not know how to price his skis and would like to price them competitively and receive a good return. He is having a hard time competitively pricing skis. He reached out to see if I can provide him with a data science solution that will allow him to offer the best price for skis to his customers. I approached this with first web scraping of evo.com and then conducting exploratory data analysis in google sheets. My next steps would be to create a linear regression.

Design

The desired impact is to competitively price skis based on their features.

Impact hypothesis: By identifying which features cause ski prices to increase, my client can price skis at the most competitive price to increase sales.

Solutions Path: Prediction- Based on prices of current skis and their features you can predict the price of any ski.

Assumptions

- How much does brand name impact pricing?
- Exploring customer reviews impact on quality and pricing.
- Will pricing competitively actually result in more revenue?

Data

The project used data from an online ski distribution site EVO. My python program scraped over a thousand skis from EVO.com. The data includes information on various ski features including terrain, with/without boots, with/without bindings, skiing ability, rocker type, turning radius, core laminates, full name, price, brand, year, gender, age, and condition.

[Excel Workbook](#)

Algorithms

I conducted some initial data analysis and cleaning in python. Afterwards, I exported my dataframe into Google sheets for additional exploratory analysis .

- Feature Engineering & Selection in python
 - Using Imputation to drop data points that were missing significant information
- Exploratory Data analysis in google sheets and tableau
 - Limiting Outliers by reducing my range of certain features
 - Examined graphical discontinuity
 - Categorical column grouping to identify trends
 - Vlookup, splitting columns, and other methods to clean my data in excel

Tools

- NumPy and Pandas for preliminary data manipulation
- Tableau for plotting and modeling
- Beautiful soup for web scraping and data ingestion
- Powerpoint to deliver the pitch

Communication

My project was completed in line with time requirements and can be viewed [here](#) and also contains Tableau workbooks(local and public), which are attached in the Github repository and will be presented.

Link to tableau [dashboard 1](#) and [2](#)