**Women’s shoe price – Exploratory data analysis**

**Data**

This dataset is from Kaggle’s competitions a sample list of 10,000 women's shoes and their product information provided by Datafiniti's Product Database. Dataset includes 33801 observations and 52 features. Some of those are shoe name, brand, price, and more.

**Question**

Selling factors of shoes brand name or price?

**Hypothesis**

Brand name attracts more customers

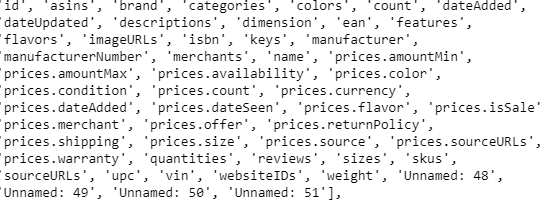
**Environment**

Python- Jupyter notebook

**Exploratory data analysis of shoe price**

In the preliminary exploration of data, our dataset contains 33801 value and 52 features. That includes categorical, numerical and Boolean values. Some of the features contains null values.

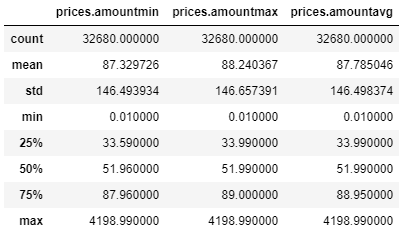
Column names in dataset:



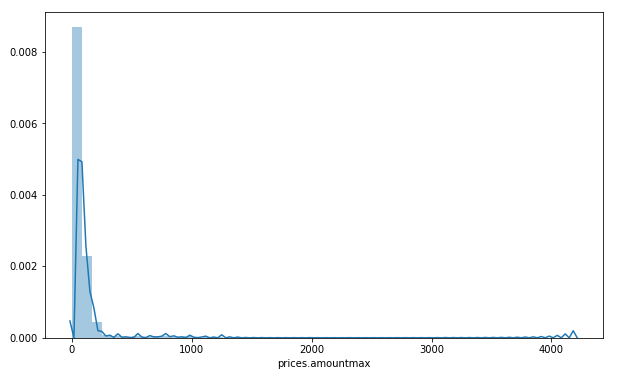
**Data cleaning and preprocessing**

Data cleaning includes changing the column names and values to lowercase, eliminating columns that contains more than 50% of null values. Also created a new column prices.amountavg using the values of minimum and maximum price values. After removing the columns with null values over 50% new dataset contains 24 columns only.

**Price features in Data**



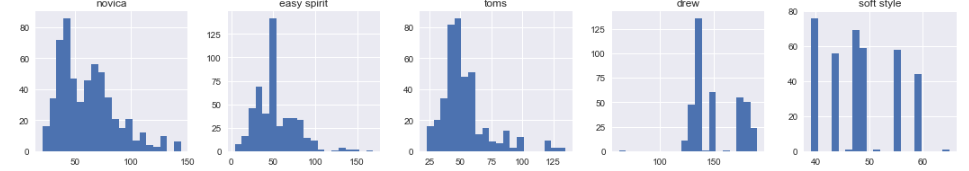
**Price distribution plot**



Avg shoe price in our data is $88 and median shoe price is $51 most expensive shoes are sold for $4198.







Prices across brand not showing normal distribution.

Outliers after removing expensive non-shoe items from data.



