

Ask

The questions being asked in this analysis are:

- What are the most common words in the object columns
- What's the most common Category
- The difference in pricing between Max Retail Pricing(MRP) and Selling Price
- The most common type of discount

Prepare

The dataset is from [Kaggle](#). It contained a CSV file containing information on approximately 30,000 items.

Process

The first thing I did was inspect the CSV file in Excel for spelling and duplicates. I deleted the duplicates and corrected the spelling mistakes. Next, I submitted the CSV file to the notebook to start the data cleaning process. I started by checking for nulls and there weren't any. Following that, I dropped the 'Unnamed: 0' column because there was no helpful information in that column. After that, I deleted 'Rs\n' before all MRP prices and converted MRP and Selling Price to integers. I also deleted 'Size' before all listed sizes from the 'Sizes' column.

Analyze

I started my analysis by finding the most common words in the 'Sizes', 'BrandName', 'Category', and 'Detail' Columns. In the 'Size' word cloud the most commonly used word is 'Large'



In the 'BrandName' word cloud the most commonly used word was 'Zink'



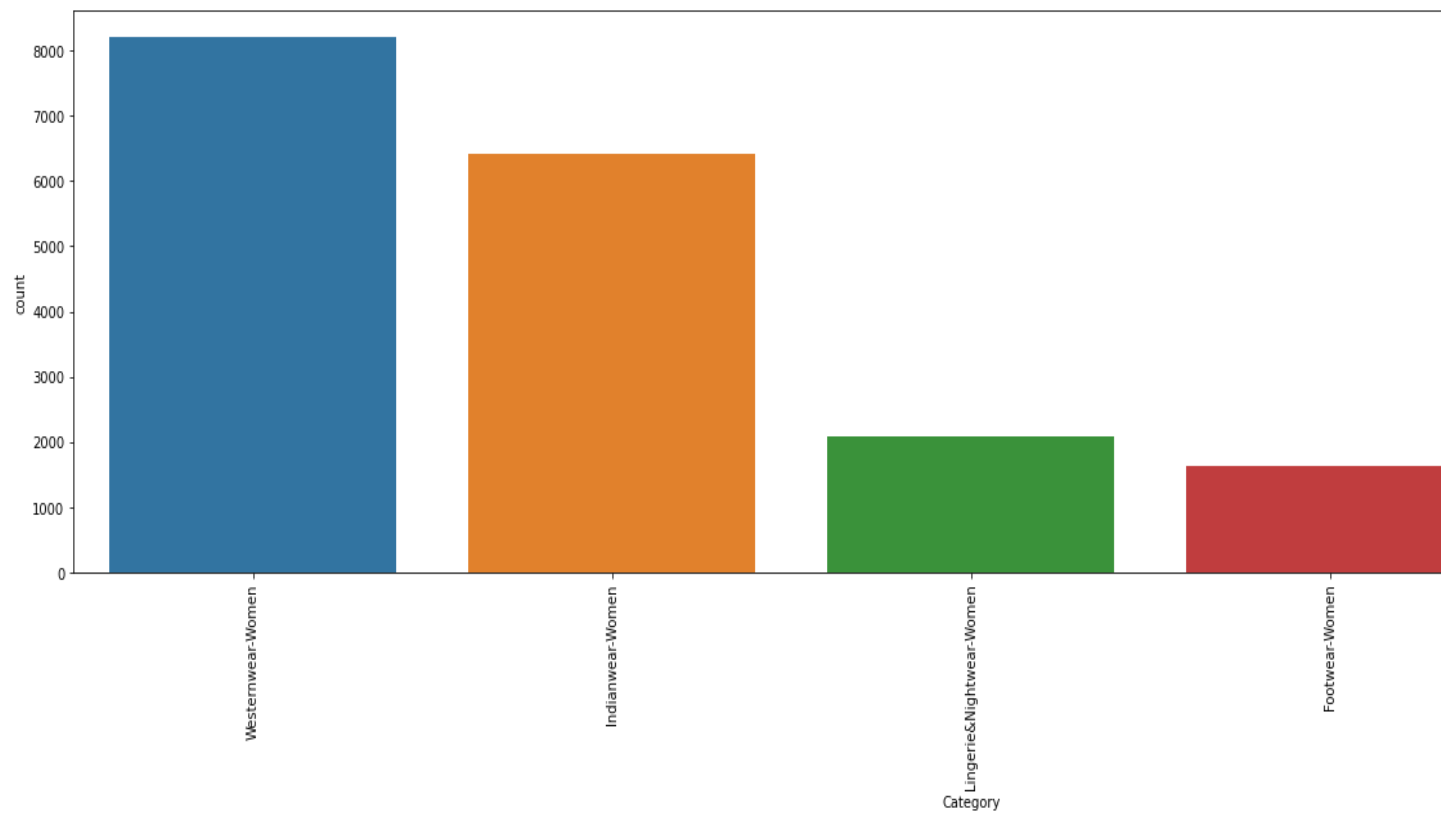
The most commonly used word in the 'Category' word cloud was 'Women'



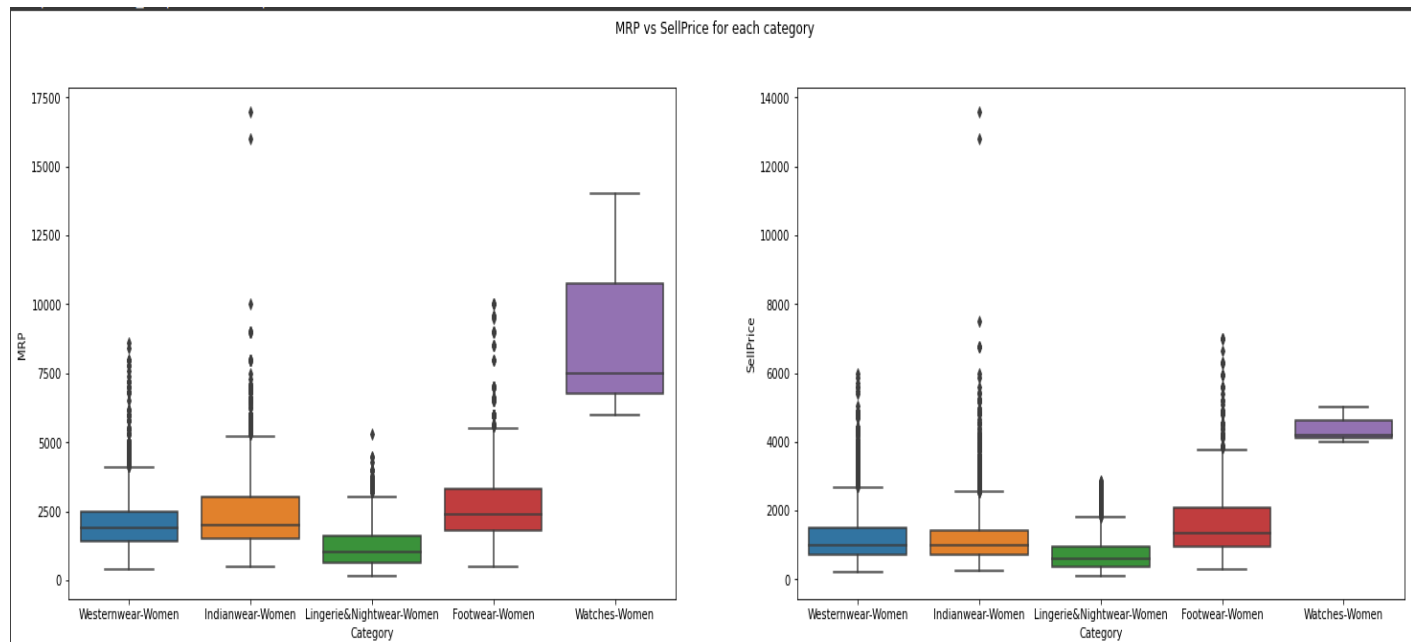
Lastly, the most commonly used word in the 'Details' word cloud was also 'Women'



To answer the second question, I used a count plot to see what was the most common category and It was Western women. It was used over 8000 times.



In my next analysis, I wanted to compare the price difference between MRP and the selling price. What I found was that MRP's prices vary more than the selling prices with bigger box plots. While the Selling price has less varying prices.



Lastly, I found that the most common discount was the 50% off discount

