

**Adidas\_USA\_Retail\_Analysis**

aJANYA SHARMA

Table of Contents

[1 Introduction 1](#_Toc156942085)

[1.1 The problem statement 1](#_Toc156942086)

[1.2 Source of the data 2](#_Toc156942087)

[1.3 data processing & cleaning 3](#_Toc156942088)

[1.4 final representation of the data 3](#_Toc156942089)

[2 exploratory data analysis 4](#_Toc156942090)

[2.1 the exploration 4](#_Toc156942091)

[2.1.1 Products in Each Category 4](#_Toc156942092)

[2.1.2 Correlation between popularity , color and category 5](#_Toc156942093)

[2.1.3 Price point analysis 6](#_Toc156942094)

[2.1.4 Rating and Review Analysis 7](#_Toc156942095)

[2.1.5 Product Affinity Analysis( path to the product page) 8](#_Toc156942096)

[2.1.6 Products in various price bands 9](#_Toc156942097)

[2.2 Graphical representations 10](#_Toc156942098)

[2.2.1 Product Performance and Customer Preferences 10](#_Toc156942099)

[2.2.2 Customer Insights and Market Trends 11](#_Toc156942100)

[3 LIMITATIONS & RECOMMENDATIONS 12](#_Toc156942101)

[3.1 LiMITATIONS 12](#_Toc156942102)

[3.2 RECOMMENDATIONS 12](#_Toc156942103)

# Introduction

## The problem statement

In an effort to optimize Adidas fashion products' performance and refine marketing strategies, we aim to harness the vast array of data captured in our dataset, encompassing over 1500+ products. Our goal is to analyze this dataset to uncover actionable insights that can inform product development and enhance customer satisfaction. Specifically, we will:

1. Investigate color and category preferences across different demographics to inform product design and inventory decisions.
2. Examine the relationship between discount strategies and customer purchasing behavior to validate the impact of price sensitivity on sales volume.
3. Determine the correlation between product ratings, the number of reviews, and sales success to establish which factors most significantly influence consumer choices and satisfaction.
4. Develop a weighted rating score that combines average ratings with review counts to provide a more nuanced understanding of product popularity and quality perception.
5. Additionally, by extracting and categorizing data from the 'breadcrumbs' field, we will perform affinity analysis and price band segmentation, enabling us to visualize customer journey patterns and refine pricing strategies across various product categories

## Source of the data

This data is sourced from Kaggle. The file contains data on Adidas products available for purchase in the United States. Data description is as below:

* url: The URL of the product page on the source website. (String)
* url: The URL of the product page on the source website. (String)
* name: The name of the product. (String)
* sku: The SKU or unique identifier for the product. (String)
* selling\_price: The selling price of the product in USD or Euros. (Float)
* original\_price: The original price of the product in USD or Euros. (Float)
* currency: The currency type for the selling price and original price. (String)
* availability: The availability of the product. (String)
* color: The color of the product. (String)
* category: The category of the product. (String)
* source\_website: The source website from where the data was collected. (String)
* breadcrumbs: The breadcrumbs or path to the product page on the source website. (String)
* description: A brief description of the product provided by Adidas. (String)
* brand: The brand of the product. (String)
* images: Multiple product images provided by Adidas. (String)
* country: The country of origin/destination for the product. (String)
* language: The language in which the product page was displayed on the source website. (String)
* average\_rating: The average customer rating out of 5 stars. (Float)
* reviews\_count: The number of customer reviews for the product. (Integer)
* crawled\_at: The date and time when the data was collected. (String)

Below is screenshot of the Adidas data in the spreadsheet:

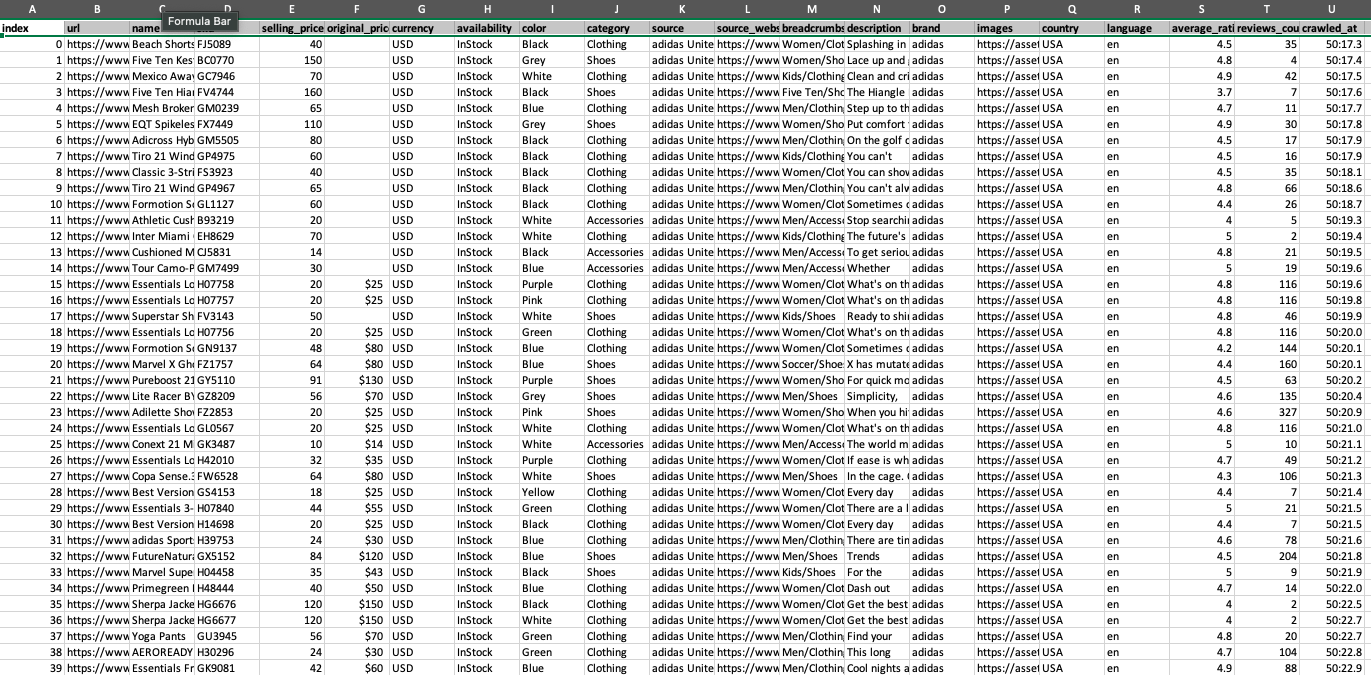


Figure 1: General Overview of the Dataset

## data processing & cleaning

Cleaning and converting raw data before processing and analysis is known as data preparation. It's a crucial stage before processing that often include reformatting data, making data changes, and integrating data sets to improve data quality.

For the Adidas data set, we have tried and looked for various inconsistencies in our data set and removed them so that we can do analysis on good quality data. We have performed the following operations on the data in excel:

1. Duplicates: No duplicates were found in the data set. This operation was performed using excel inbuilt functionality of “removing” duplicates from data tab.
2. Missing values or null values: The data was very consistent and we found around 10 missing vakues in Orignal\_Pricce column which accounted for less that 0.01% (16 out of 845) of the entire dataset.We proceeded with remove the rows which had missing original price.

## final representation of the data

The following is the final representation of the data post cleansing and pre-processing and is loaded in tableau ready for analysis.

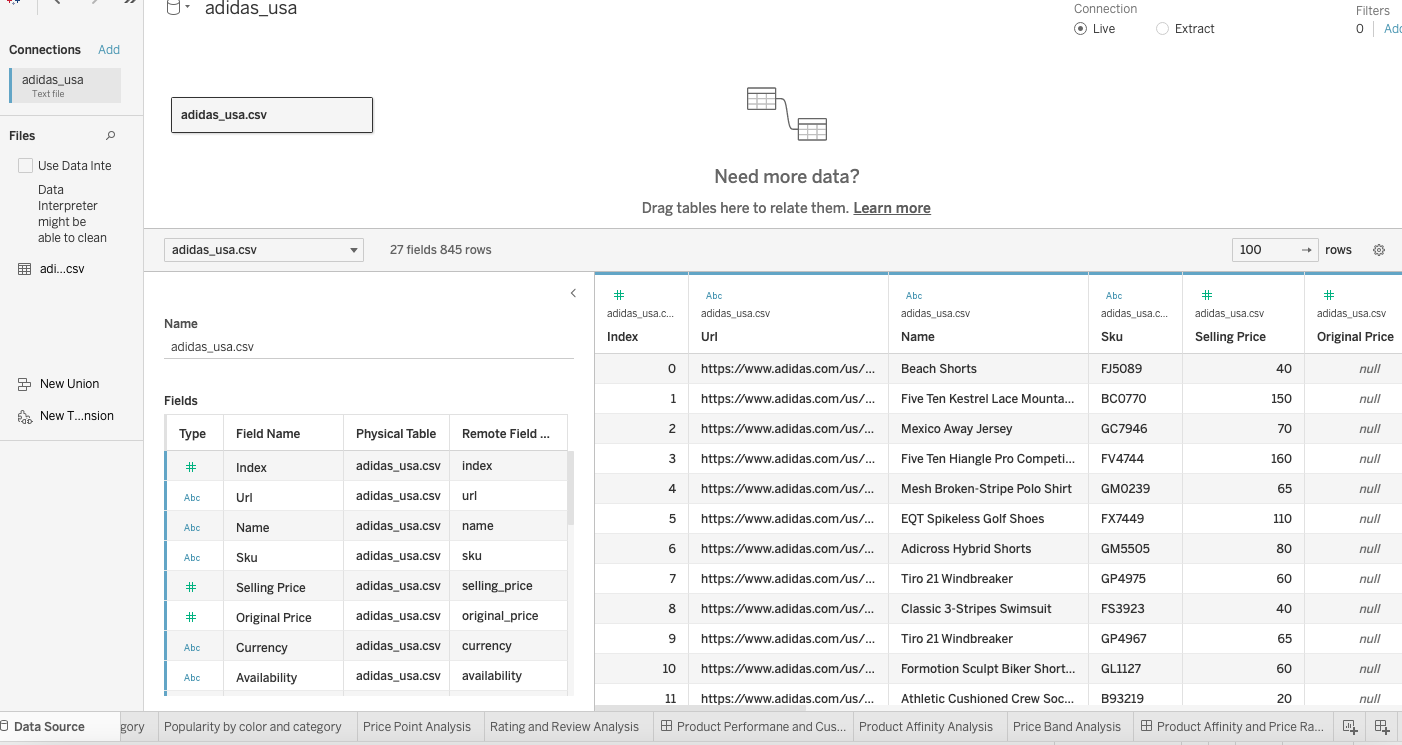


Figure 2: Processed Overview of the Dataset

# exploratory data analysis

## the exploration

Data exploration is an initial step in data analysis where we will try to get an overview about the data set and try to identify the properties of data and look for some patterns in the data.

When I did exploratory analysis, it was evident that the Adidas data set had categorical (non-metric), double integer (Metric) and date field which made up the entire data set. The findings

Once the data has been loaded, we have used tableau to do visualization and try to identify the patterns in the data set. The following findings have been accumulated and listed below in this section.

### Products in Each Category

The bar graph shows the number of products in each category with maximum items in shoes and minimum items in Accessories category.

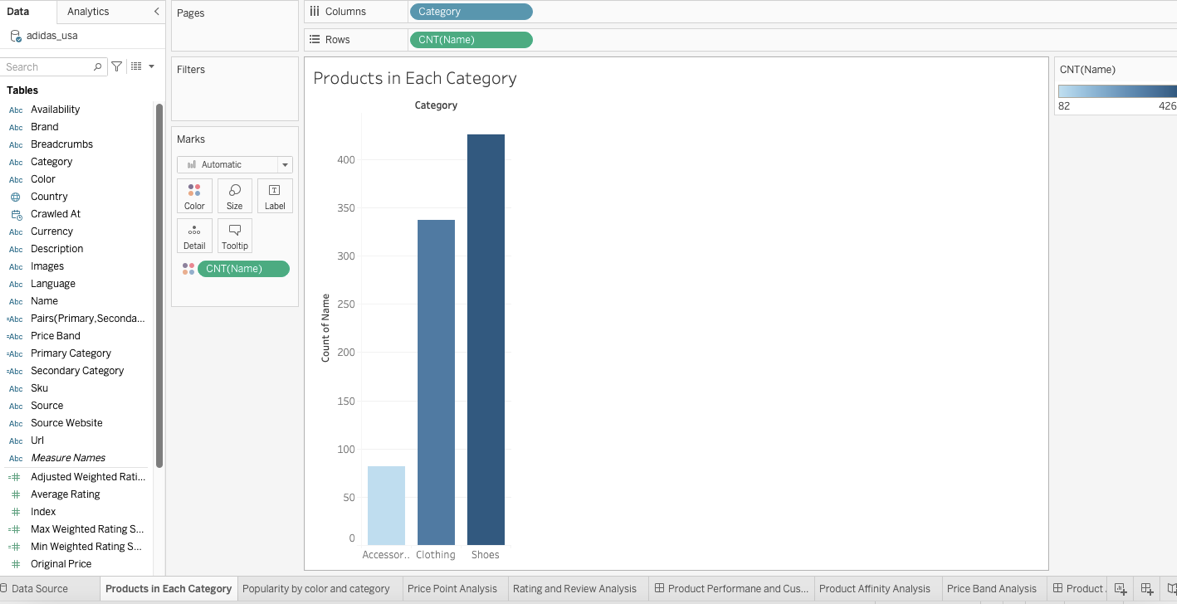


Figure 3: Products in each category

### Correlation between popularity , color and category

This visualization helps us to find what is the most popular products based on color and category. There is no sales data in the dataset so we have considered review counts as measure of popularity for this dataset. Some of the insights from this dataset are:

1)White shoes is the most popular product on on the platform followed by black and blue shoes.

2)For accessories, multicolor and white are the most preferred colors.

3)Clothing with gold color accounts for the least popularity.

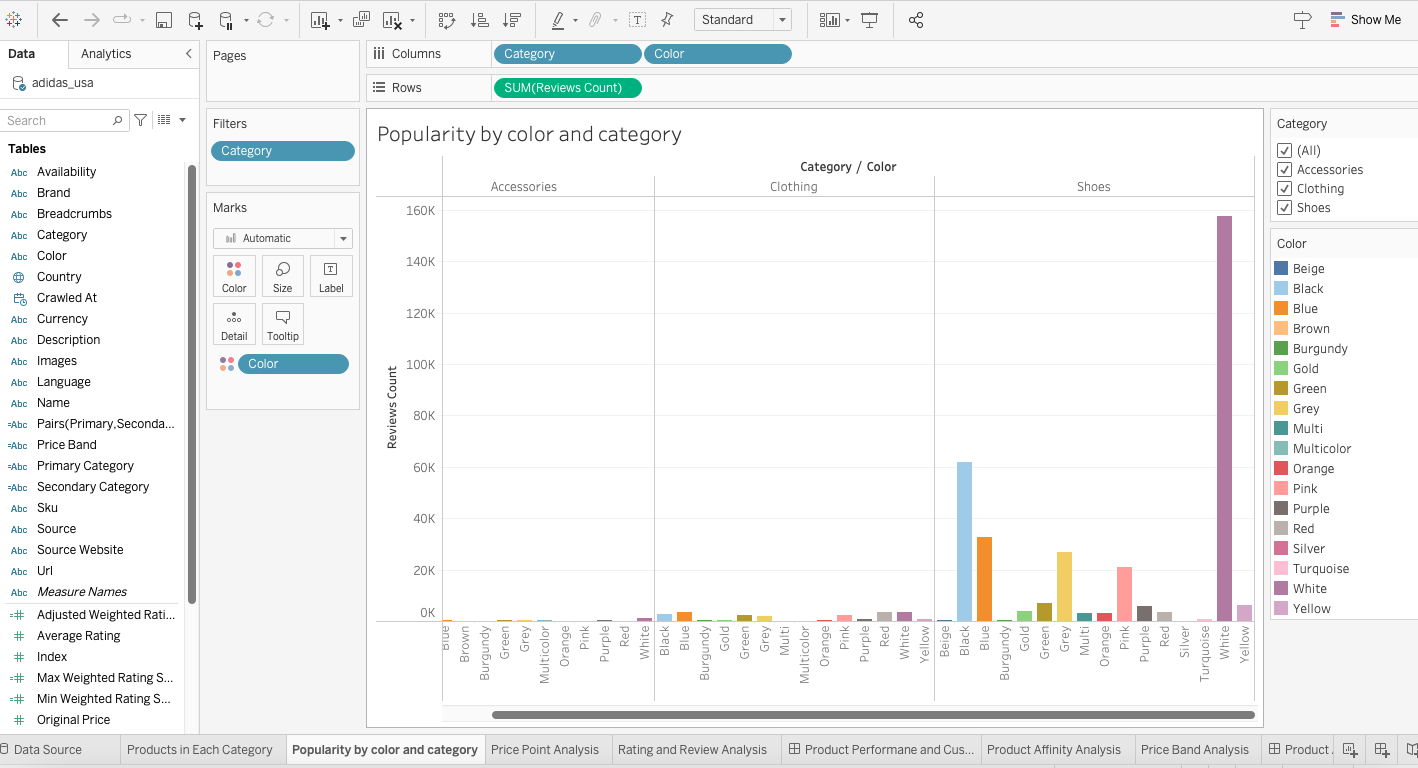


Figure 4: Findings for products popularity based on color and catrgory

### Price point analysis

A new metric “Profit “ has been introduced using calculated filed feature of tableau and is used to study profit patterns for products based on original price and sellig price difference.

This report reveals:

1. **Profit Variability**: Products at lower original price points show a high variability in profit, indicating a diverse range of discounting strategies and cost structures.
2. **Category-Specific Profit Trends**: The Shoes category exhibits a broader profit margin range compared to Accessories and Clothing, suggesting different pricing strategies across categories.

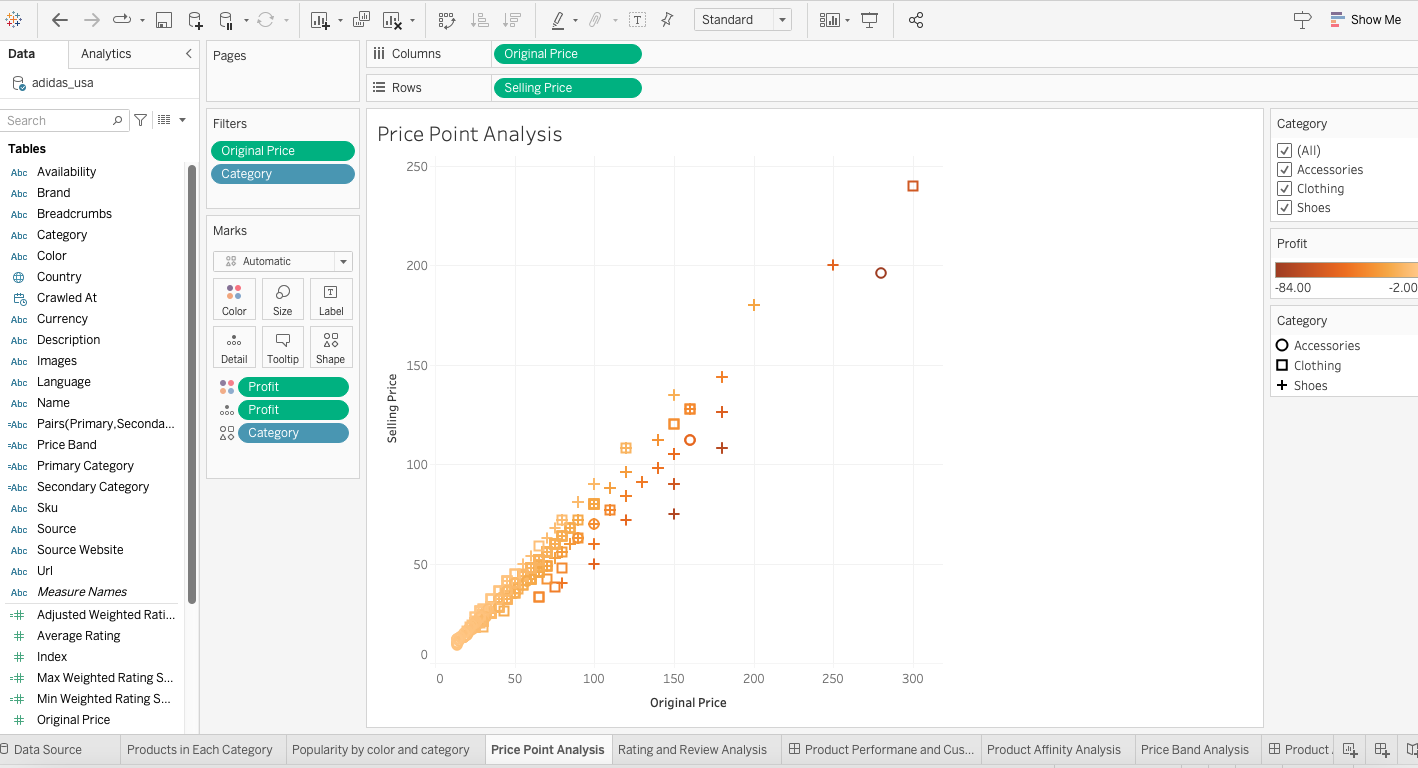


Figure 5: Profit across different categories

This visualization can help define and apply different pricing strategies across different categories to improve profits.

### Rating and Review Analysis

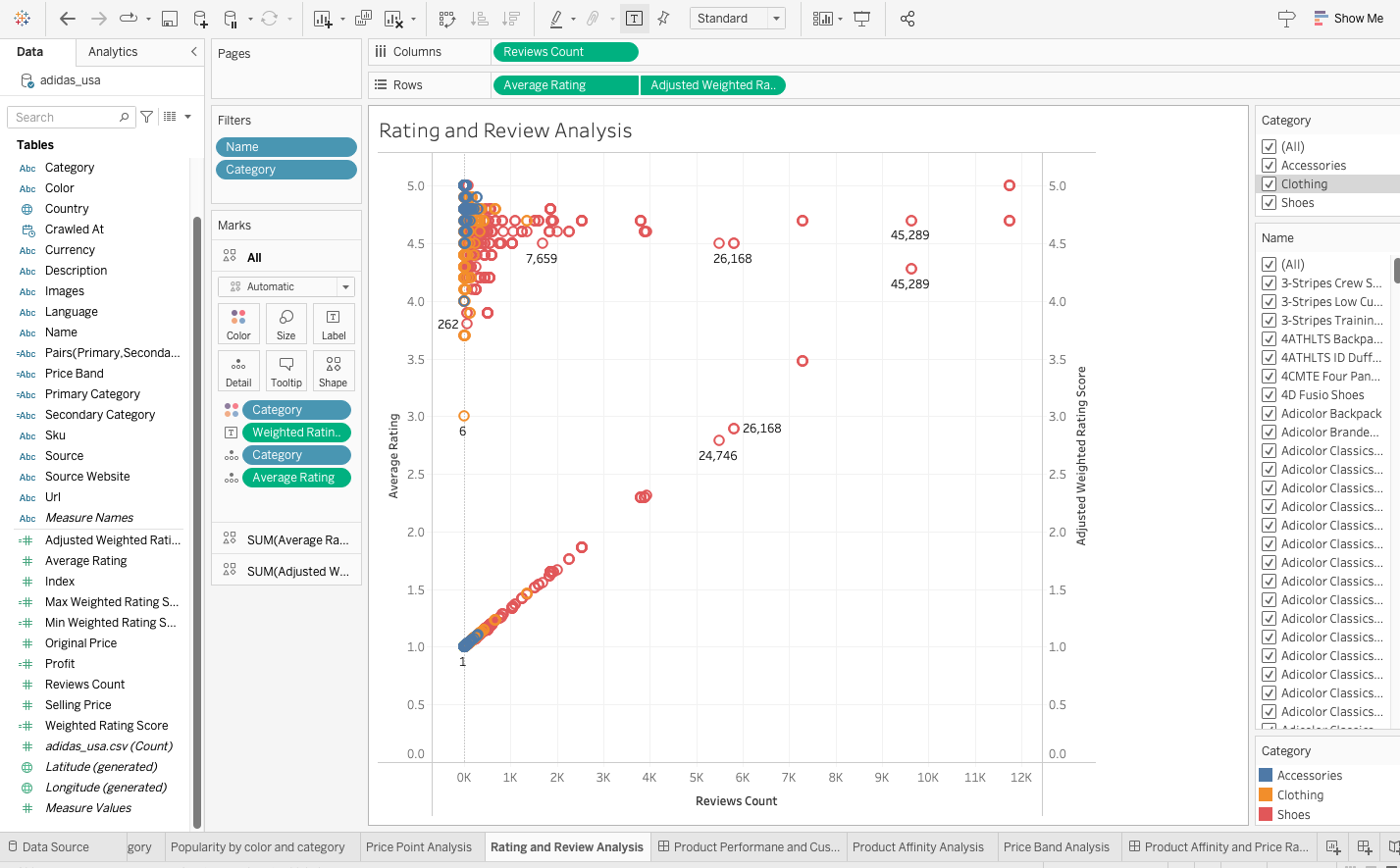


Figure 6: Adjusted weighted score and reviews

The dataset already has **Average Ratings and Review Counts.** The average rating typically represents the mean score of a product based on customer reviews.It gives a quick snapshot of overall customer satisfaction.It treats each review equally, regardless of the number of reviews. The limitation with Average rating is that product with a high average rating based on a small number of reviews is treated the same as a product with the same rating based on hundreds of reviews.

This has prompted me to introduce a new metric called “**Weighted Rating Score**” which takes into account not only the average rating but also the number of reviews.I have further calculated “Adjsuted Weighted Rating Socre” which has the same scale as Average rating which make visualizations easier to interpret.

The significance of having both an average rating and a weighted rating lies in the different aspects of product popularity and customer satisfaction they each illuminate:

* 1. **Weighted Rating Score** can be a more reliable indicator of a product's overall reputation because it reflects both the quality (via the average rating) and the quantity (via the number of reviews) of feedback.
  2. **A Weighted Rating Score** helps to balance out the score by considering the popularity of the item. For example, an average rating of 5 based on 2 reviews might be less significant than an average rating of 4.5 based on 200 reviews. The weighted rating would be higher for the latter, indicating it is likely a more trustworthy indicator of customer satisfaction.

By using both these metrics, we can distinguish between products that have high ratings due to a few positive reviews (which might not be statistically significant) and those that have slightly lower ratings but are based on a large number of reviews (which are statistically more reliable).

In summary, while the average rating is useful for understanding how well-liked a product is on average, the weighted rating is crucial for evaluating the reliability and credibility of the average rating itself. This distinction can be especially important for businesses as they assess which products are truly favored by customers and to mitigate the impact of outliers or potential review manipulation.

The visualizations shows a **Non-Linear Relationship** between the average rating and the adjusted weighted rating score. This indicates that the adjusted score does not increase proportionally with the average rating. Instead, it seems to factor in the number of reviews to moderate the score.

The Adjusted weighted rating score puts weighted rating score in the same scale of 1-5 as is average rating making it more easier to interpret.

### Product Affinity Analysis( path to the product page)

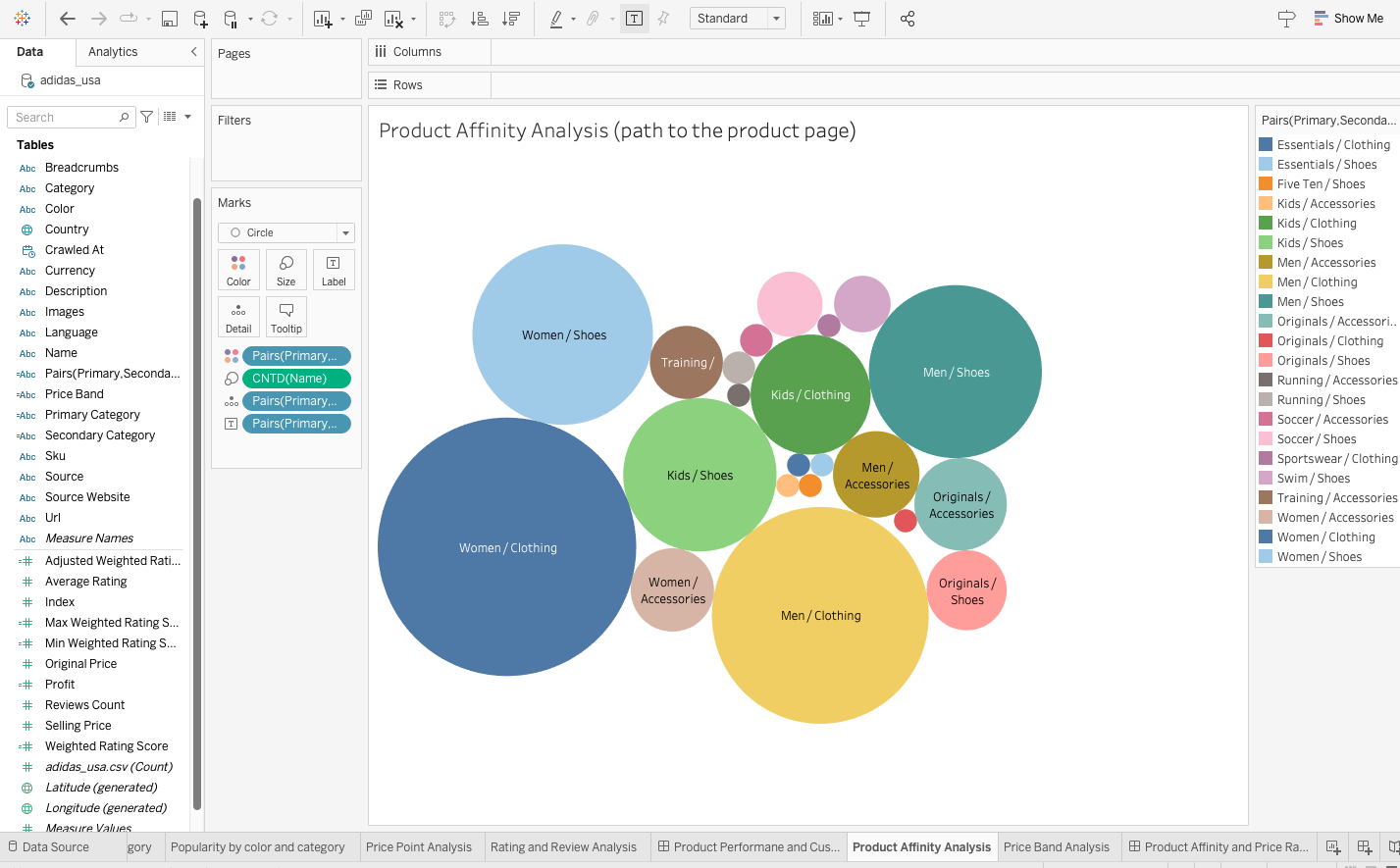


Figure 7: Path to the product page

Product Affinity Analysis is done using the 'breadcrumbs' field. The aim to find what path a user takes to reach a product.Primary category and secondary category fields are created to split the breadcrumbs. We can see hat maximum products are in women/clothing category followed by Men/Shoes. These categories can be used to build a network graph or a Sankey diagram to visualize how customers move between product categories in the future.

### Products in various price bands

Here , I am trying to identify number of product across different price ranges which may be used to discover relationship between price of product and customer preferences if that has any further relationship with customer demographics. Group products into price bands based on 'Selling Price'.Create a histogram or bar chart to show the number of products in each price band, which can highlight your pricing strategy across different categories.

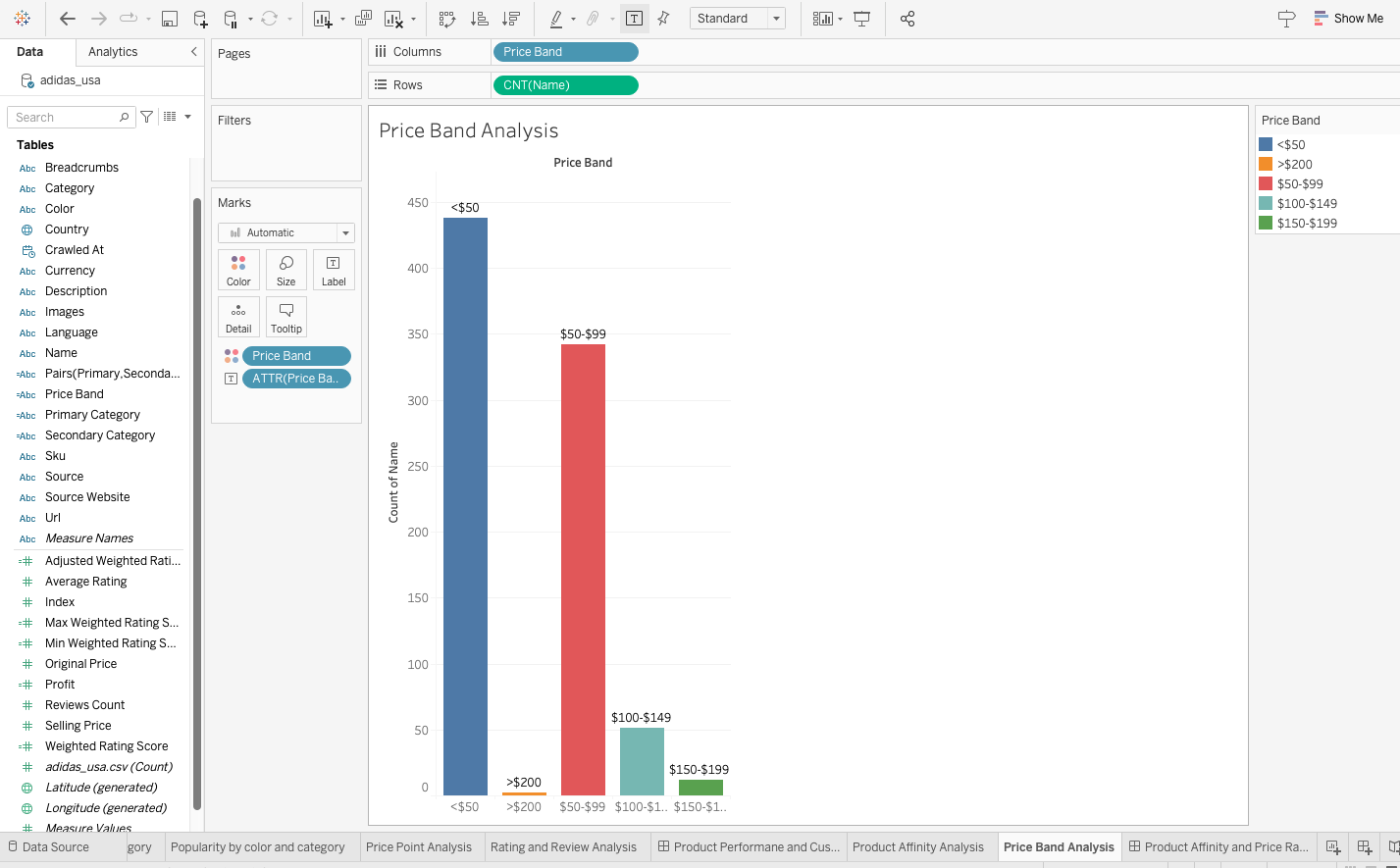


Figure 8:Price Band Analysis

Top 50 Percent of the products fall in the price band of USD $0-$50.

## Graphical representations

The final graphical representations for each of the four sections mentioned in the findings may be found here. We have created 2 dashboards to link the insights analyzed in the visualization done in the above section.

### Product Performance and Customer Preferences

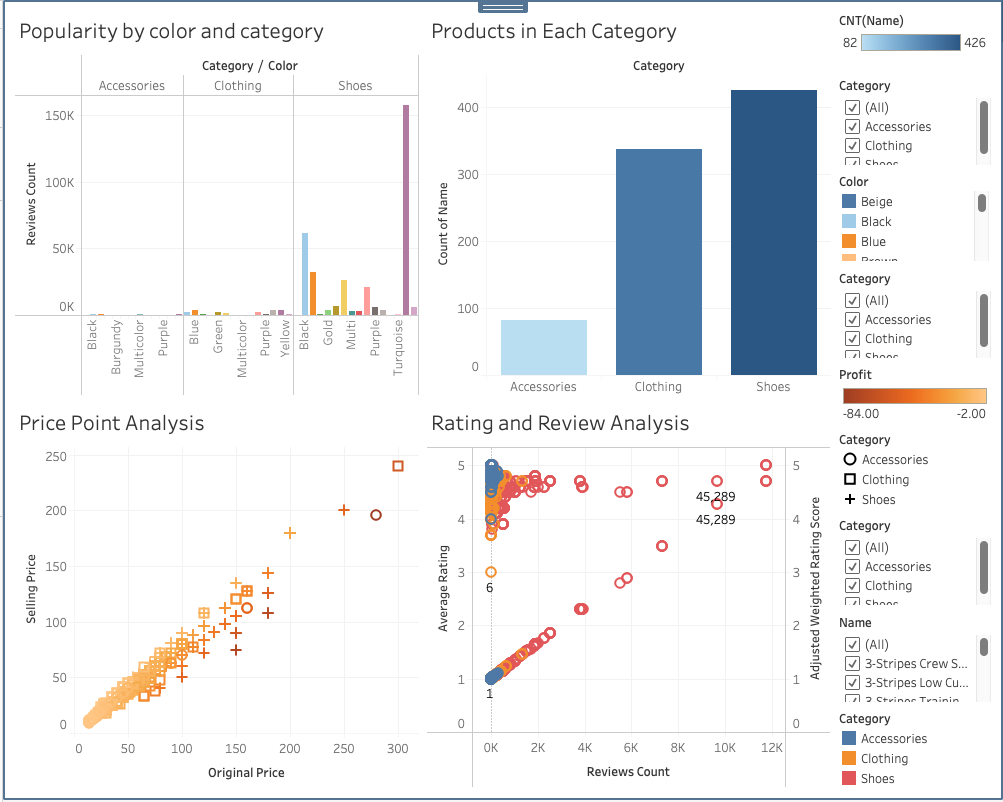


Figure 9: Analysis of product performance, and customer preferences

This dashboard is intended to be interactive. It gives a view of customer preferences such as white shoes , product performance such as profitability and a common metric “Adjsuted Weighted Score” to calculate relationship between customer reviews and rating.

1. **Rating and Review Correlation for Shoes**: The shoes category shows a cluster of products with a high number of reviews and generally high average ratings. However, there are outliers with extremely high review counts (e.g., 45,289 reviews) that do not have the highest average rating, suggesting that while some shoes are very popular, they may not be the most highly rated in terms of customer satisfaction.
2. **High-Review Products in Clothing**: The clothing category has a product with a single review (count of 1) that received a perfect average rating of 5, which is adjusted to a lower weighted score due to the low number of reviews, indicating a potential for review quantity to skew perceived quality.
3. **Weighted Rating Adjustments**: In both categories, products with a larger number of reviews tend to have their adjusted weighted rating score converge towards a median range, irrespective of the average rating. This suggests that the weighted score normalizes the effect of high review counts, providing a score that reflects both the quantity and quality of feedback.

The visualizations provide a comprehensive overview of how products are perceived across different categories, factoring in both the popularity (review count) and perceived quality (average rating) to arrive at a more balanced evaluation of customer feedback (weighted rating score).

### Customer Insights and Market Trends

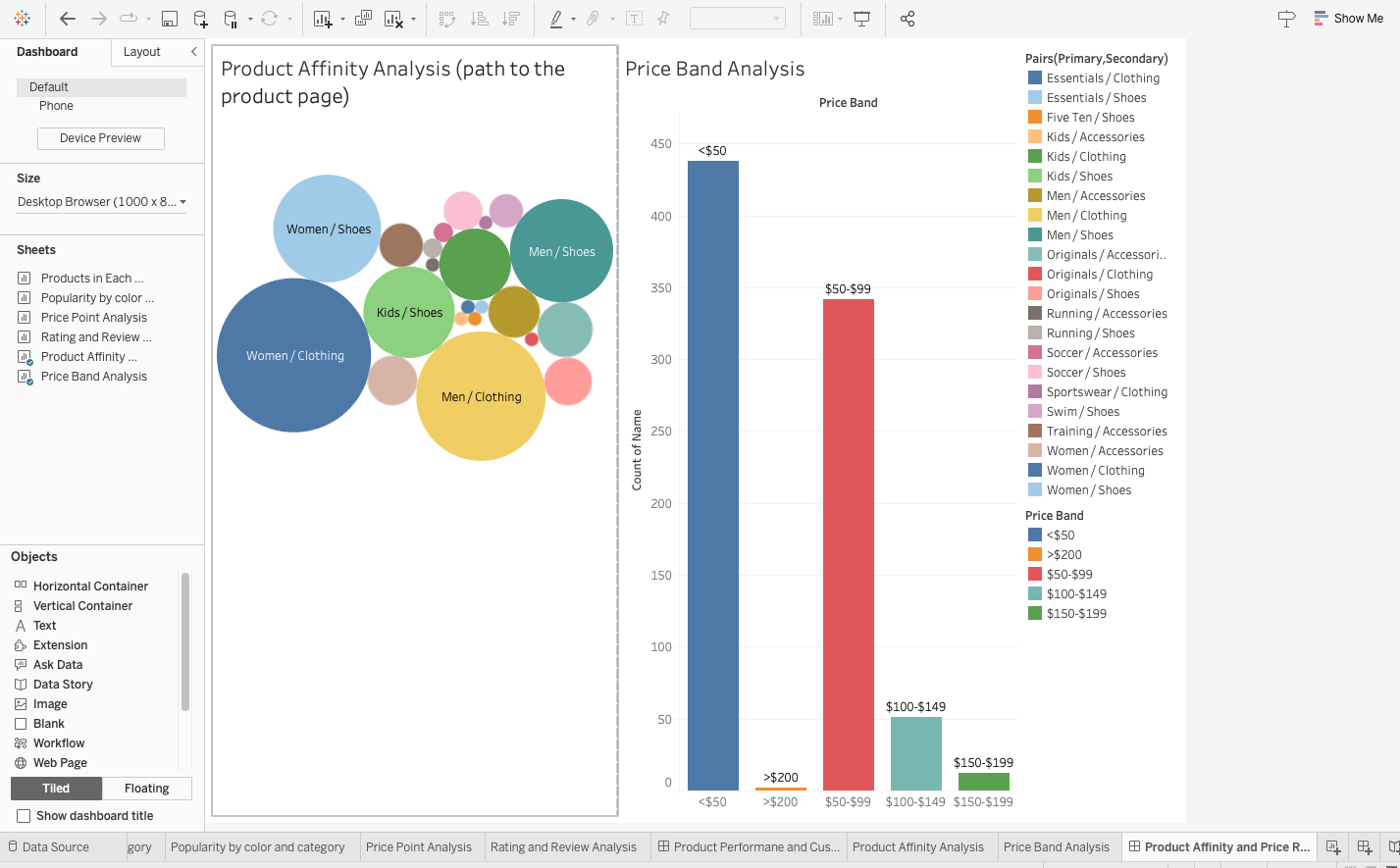


Figure 10: Product Affinity and Price Range

This dashboard is again designed for interaction with category and price band filters. This one is intended to find products based on price and breadcrumbs(primary and secondary category). This will help companies evaluate users price sentivity and navigation patterns.

# LIMITATIONS & RECOMMENDATIONS

## LiMITATIONS

The dataset and the subsequent analysis have the following limitations:

1. The approach assumes that products appearing under similar categories have a relationship which may not reflect actual customer behavior, as the true sequence of product views or purchases is unknown.
2. The dataset can have information like sales data and customer demographics which can be utilized to have more targeting marketing and product strategies.
3. Average rating is given. A product with a high average rating based on a small number of reviews is treated the same as a product with the same rating based on hundreds of reviews.

## RECOMMENDATIONS

The following are the recommendations for the dataset and the subsequent analysis:

1. This dataset needs to be combined with official adidas customer data to understand the full ecommerce and retailer data under consideration.
2. Breadcrumbs given here are how products are didvided as per application flow. If we capture breadcrumbs based on how user moves across differenct cataegories and break down these categories into primary and secondary category.can be used to build a network graph or a Sankey diagram to visualize how customers move between product categories in the future.