

Case Study

Problem Statement:

This case study uses a dataset of vendors from the Amazon marketplace to determine which Amazon merchants in the Garden category are the best or most promising. The objective is to use data-driven insights to support the acquisitions team in identifying top-performing merchants. I did this by using MySQL to parse and Power Bi to visualise the data in order to identify the important criteria that set top sellers apart.

Data Parsing:

To clean and improve the data, the dataset was parsed using SQL techniques. Below is a breakdown of the steps that make up this process:

Step 1: Eliminating Duplicates

To guarantee that only unique seller inputs are taken into account, all duplicate records were eliminated.

Step 2: Standardizing the Data

To ensure similar representation across metrics such as rating percentages, nation names, and seller links, data fields were standardised for consistent formatting.

Step 3: Handling Null or Zero Values

Based on business rules, null numbers and zeroes were either excluded or replaced with average values. Sellers with insufficient information were excluded.

Step 4: Dropping Unnecessary Columns

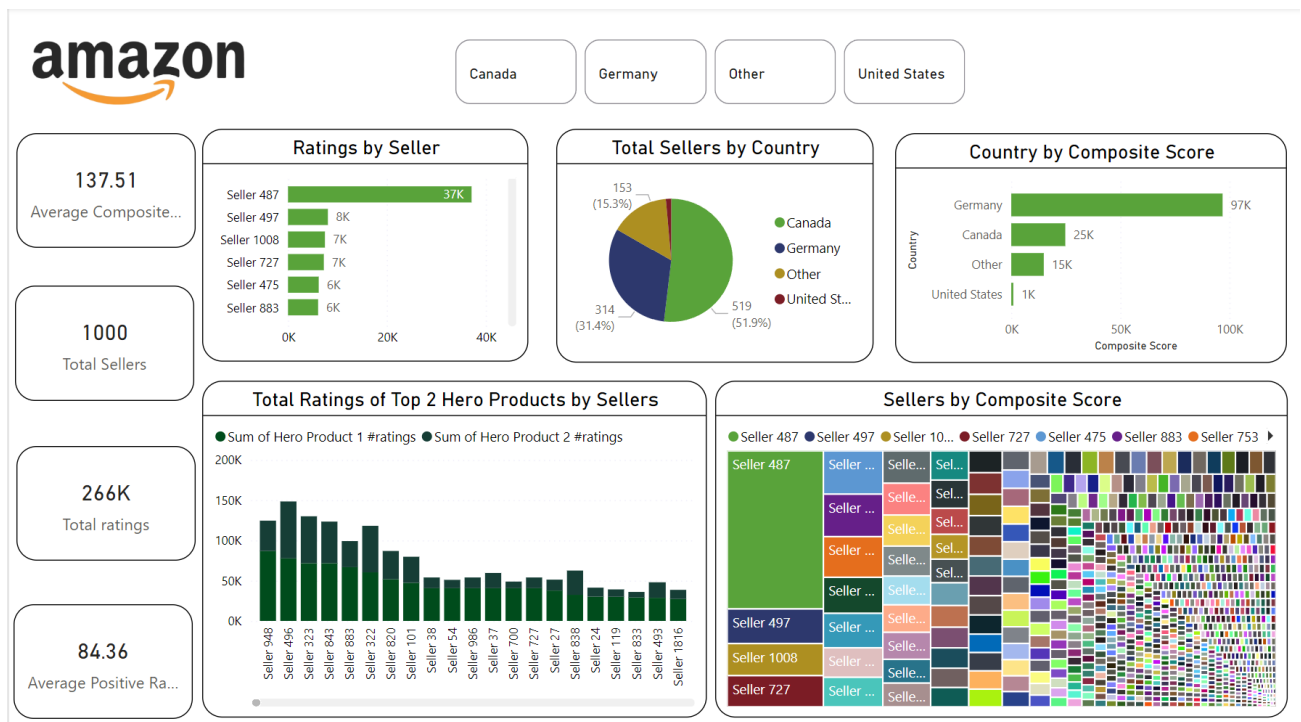
Columns that had no bearing on the analysis were eliminated from the staging table, focusing only on key metrics like rating counts, product counts, and seller performance indicators.

MySQL Code: [Click Here](#)

Cleaned Data: [Click Here](#)

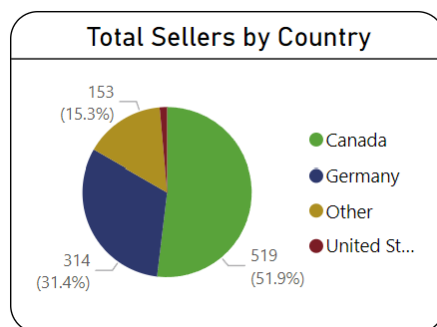
Data Visualization in Power BI:

The top-performing sellers were found and highlighted using Power BI visualisations that were created after the dataset had been cleaned and analysed using various metrics. Below is a discussion of the employed visualisations and their findings:



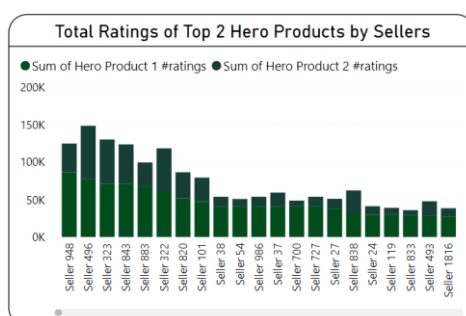
1. Pie Chart for Seller Distribution by Country:

- A pie chart was used to visualize the geographical distribution of sellers. Each seller's region was color-coded based on the count of sellers, allowing the team to easily spot high-performing countries.



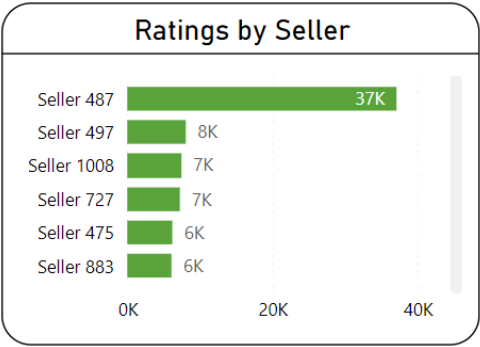
2. Stacked Column Chart for Sellers based on Hero Product Ratings:

- A stacked column chart was implemented to show the top sellers based on sum of top 2 hero product ratings, provided insight into which sellers' key products are receiving the most engagement.



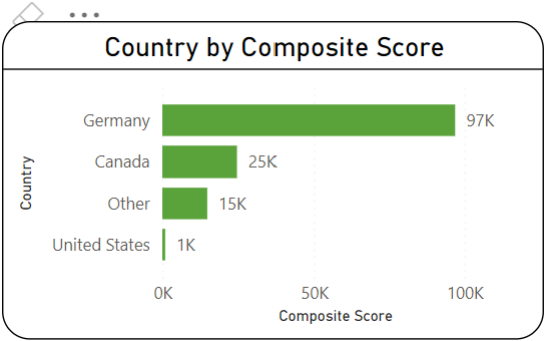
3. **Bar Chart for Sellers by Number of Ratings:**

- The Bar Chart was used to display the top sellers by number of ratings.



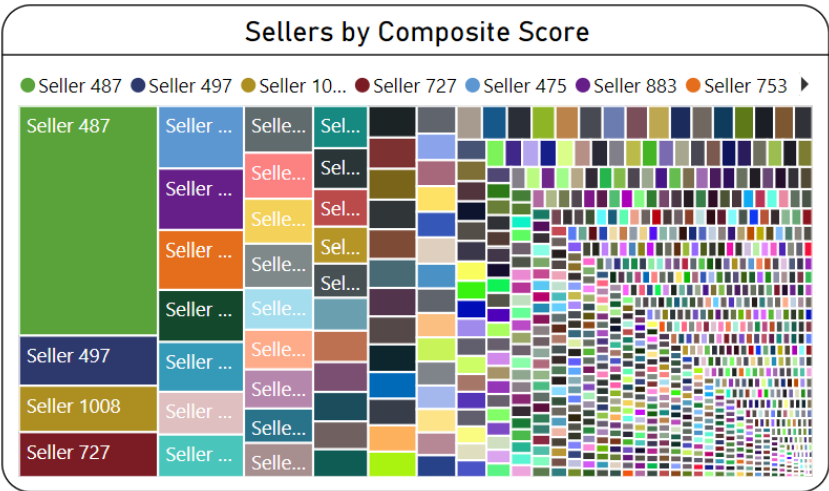
4. **Bar Chart of Countries by Composite Score:**

- A bar chart tracked to show the composite score of various countries.



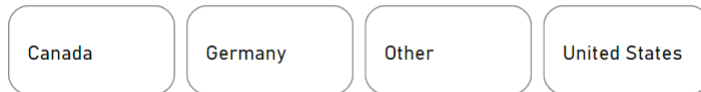
5. **Treemap for Top Sellers based on Composite Score:**

- The treemap visualized the top sellers based on composite score calculate previously, providing a clear view of market dominance.

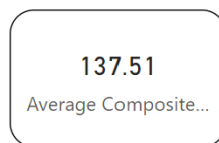


Slicer for Interactive Analysis:

To allow for tailored analysis, slicer was added to the dashboard for selecting the country. This slicer can help the acquisitions team to refine their search and focus on a particular country.



Cards Used:



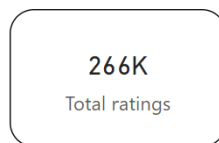
1. **Total Sellers by Growth in Rating Count:** Sum of count of all the sellers present in the dataset.

2. **Total Ratings:** Sum of count of ratings for all the sellers.



3. **Average Composite Score:** A more robust metric that considers positive rating percentage, negative ratings in last 12 months percentage and the number of ratings, giving a balanced view of seller performance:

Used the below DAX formula while creating a new column for composite score

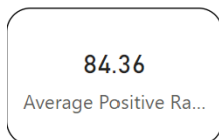


Composite_Score =

'cleaned_data'[positive_rating_percentage] * 0.4 +

(1 - 'cleaned_data'[Max % of negative seller ratings - last 12 months]) * 0.2 +

'cleaned_data'[rating_count] * 0.4



4. **Average Positive Rating Percentage:** Average of all the positive ratings.

Conclusions:

- Key metrics include an average composite score of 137.51, with 1000 sellers contributing to 266K total ratings and an average positive rating of 84.36%. The analysis highlights Germany as the top performer, leading with a composite score of 97K, followed by Canada with 25K. Germany and Canada dominate in terms of seller numbers, with Germany hosting 314 sellers (31.4%) and Canada 519 sellers (51.9%).
- Seller 487 emerges as the leader in ratings, with 37K ratings, indicating strong customer engagement. In terms of top product performance, Seller 498 and Seller 496 are noteworthy, reflecting high customer interest in their hero products.
- The Sellers by Composite Score chart showcases the distribution of scores among sellers, with Seller 487 and Seller 497 standing out. This data underscores the competitive landscape, with significant variation in seller performance.
- Overall, Germany and Canada lead in seller strength, with top sellers like Seller 487 and Seller 498 excelling in both composite scores and customer ratings, providing valuable insights for strategic decision-making.