

Technical Report

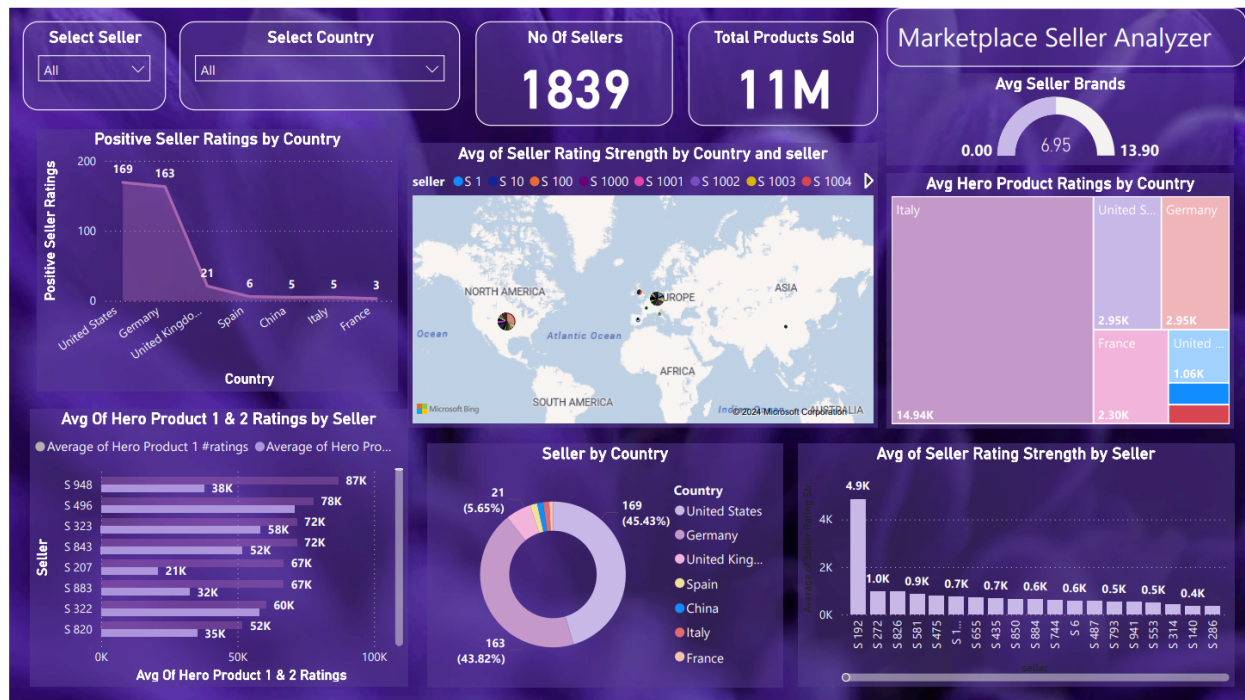
Case Study - Business Analyst - 1

Name: Kaustubh Raykar

PRN: 21070126048

Batch: 2021-25

Project: Marketplace Seller Analyzer Dashboard



PREPROCESSING REPORT

Introduction

This section of the report outlines the detailed data preprocessing steps undertaken to clean and prepare the dataset for analysis. These steps are crucial

for ensuring the accuracy and reliability of the insights generated by the dashboard.

Data Preprocessing Steps

1. Removing Unnecessary Columns

- **Columns Removed:**

- `sellerlink-url`
- `sellerstorefront-url`
- Last three empty columns in the dataset.
- **Purpose:** To streamline the dataset by eliminating irrelevant information that does not contribute to the analysis.

2. Cleaning 'Seller Product Count' Column

- **Original Format:** Mixed text strings such as "1-16 of over 2,000 results".
- **Steps Applied:**
 - Extract numeric values after "of".
 - Replace "over" and "results" with spaces to isolate numbers.
 - Convert cleaned text to whole numbers.
 - Formula Applied -

```
= Table.TransformColumns("#Previous Step", {"sellerproductcount", each Text.AfterDelimiter(_, "of"), type text})
= Table.ReplaceValue("#Previous Step", "over", "", Replacer.ReplaceText, {"sellerproductcount"})
= Table.ReplaceValue("#Previous Step", "results", "", Replacer.ReplaceText, {"sellerproductcount"})
= Table.TransformColumnTypes("#Previous Step", {"sellerproductcount", Int64.Type})
```

- **Result:** Transformed the column to display clean numeric data, facilitating accurate calculations and comparisons.
- Before Vs After

AB _C sellerproductcount - Copy	AB _C sellerproductcount
Valid 82%	Valid 81%
Error 0%	Error 0%
Empty 18%	Empty 19%
1-16 of over 2,000 results	2,000
1-16 of 123 results	123
1-16 of over 1,000 results	1,000
1-16 of 28 results	28
1-16 of over 4,000 results	4,000

3. Splitting and Cleaning 'Seller Ratings' Column

- **Original Format:** Composite strings like "88% positive in the last 12 months (118 ratings)".
- **Steps Applied:**
 - Split the column into **Positive Seller Ratings** and **Total Ratings**.
 - Extracted percentage values and converted them to a numerical format.
 - Cleaned textual descriptions to isolate numerical data.
- **Transformation Steps for 'Positive Seller Ratings':**

```
= Table.SplitColumn("#Previous Step", "sellerratings", Splitter.SplitTextByDelimiter("(", QuoteStyle.Csv), {"Positive Seller Ratings", "Total Ratings"})
= Table.TransformColumns("#Previous Step", {"Positive Seller Ratings", each Text.BeforeDelimiter(_, "%"), type text})
= Table.TransformColumnTypes("#Previous Step", {"Positive Seller Ratings", Percentage.Type})
```

- **Transformation Steps for 'Total Ratings':**

```
= Table.ReplaceValue("#Previous Step","in the last 12 months (", "", Replacer.ReplaceText,{"Total Ratings"})
= Table.ReplaceValue("#Previous Step","ratings)", "", Replacer.ReplaceText,{"Total Ratings"})
= Table.TransformColumnTypes("#Previous Step", {"Total Ratings", Int64.Type})
```

- **Result:** Enabled distinct analytical views for percentages of positive ratings and the total number of ratings, improving metric accuracy.
- Before Vs After

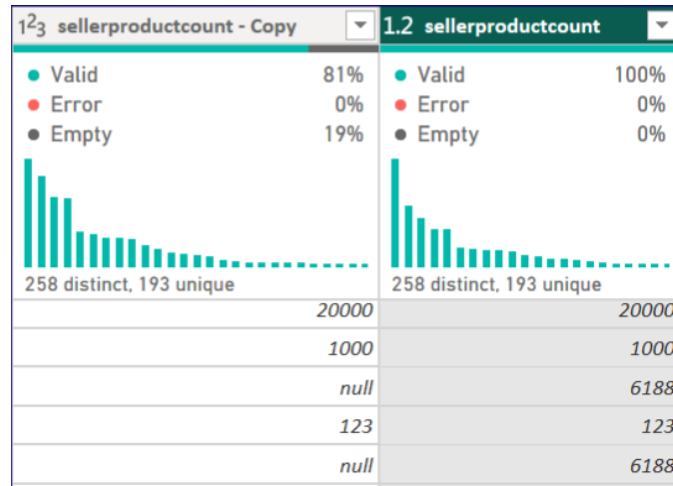
A8C sellerratings - Copy	% Positive Seller Ratings	123 Total Ratings
<ul style="list-style-type: none"> Valid 86% Error 0% Empty 14% 	<ul style="list-style-type: none"> Valid 86% Error 0% Empty 14% 	<ul style="list-style-type: none"> Valid 86% Error 0% Empty 14%
88% positive in the last 12 months (118 ratings)	88.00%	118
90% positive in the last 12 months (566 ratings)	90.00%	566
85% positive in the last 12 months (75 ratings)	85.00%	75
	null	null
81% positive in the last 12 months (52 ratings)	81.00%	52

4. Mean Imputation

- **Columns Applied:**

- Seller Product Count
- Positive Seller Ratings
- Total Ratings

- **Purpose:** To handle missing values by replacing them with the mean of their respective columns, ensuring no data point is left unanalyzed due to missing data.
- Before Vs After



5. Replacing Empty Cells with 'Unknown'

- **Columns Applied:**
 - Seller Details
 - Seller Business Name
 - Address
- **Steps Applied:** Replace all empty cells with the term "Unknown" to maintain consistency and clarity in the dataset.

6. Extracting Values from 'Seller Details' Column

- **Steps Applied:**
 - **Extract Country:** Implemented conditional checks to identify country names or codes within the text, standardizing country identification.
 - **Extract Emails:** Used text splitting and selection techniques to isolate email addresses.
 - **Extract Phone Numbers:** Identified and combined numerical strings that represent phone numbers.
- **Formulas Applied -**
- **Extract Country:**

```

= Table.AddColumn("#Previous Step", "Country", each if Text.Contains([sellerdetails], "US") or Text.Contains([sellerdetails], "United States") then "United States"
else if Text.Contains([sellerdetails], "DE") then "Germany"
else if Text.Contains([sellerdetails], "CN") then "China"
else if Text.Contains([sellerdetails], "UK") then "United Kingdom"
else if Text.Contains([sellerdetails], "FR") then "France"
else if Text.Contains([sellerdetails], "IT") then "Italy"
else if Text.Contains([sellerdetails], "ES") then "Spain"
else "Unknown")

```

- Before Vs After

ABC sellerdetails - Copy	ABC 123 Country
<div> <div>Valid 100%</div> <div>Error 0%</div> <div>Empty 0%</div> </div> <div> <div>975 distinct, 973 unique</div> </div>	<div> <div>Valid 100%</div> <div>Error 0%</div> <div>Empty 0%</div> </div>
Lohas Living Inc James Mazzello US 845 3RD Ave Fl 6 New York NY 100...	United States
Herzlich Willkommen im Amazon Shop von 1a-Handelsagentur. Für Fr...	Germany
67878u6i7 is committed to providing each customer with the highest s...	Unknown
7 days home life supermarket is committed to providing each custome...	Unknown

- **Extract Emails:**

```

= Table.AddColumn("#Previous Step", "ExtractedEmails", each let
    TextList = Text.Split([sellerdetails], " "),
    ExtractedEmails = List.Select(TextList, each Text.Contains(_, "@"))
in Text.Combine(ExtractedEmails, "; "))

```

- Before Vs After

APC sellerdetails - Copy	APC 123 ExtractedEmails
Valid 100%	Valid 28%
Error 0%	Error 0%
Empty 0%	Empty 72%
975 distinct, 973 unique	
99gaodashang is committed to providing each customer with the high...	null
Impressum: ABC-Schnäppchenmarkt GmbH Jösser Weg 10 32469 Pete...	abc-markt@web.de
AllSparesEire is committed to providing each customer with the highes...	null
Firmenwortlaut: GURU2016GmbH Geschäftsführung: Stefan Schnidar,...	Versand)support@bauguru.at ; eMail: (Geschäftsführung)office@baug...
Bestof Floral is committed to providing each customer with the highes...	null
Blumenbecker Industriebedarf GmbH Sudhoferweg 99-107 59269 Bec...	bib.beckum@blumenbecker.com
Herzlich willkommen auf unserer Internetpräsents. Wir sind ein alt ein...	null
Gesetzliche Anbieterkennung: boni-shop.com OHG Wissmannstraße 5...	amazon@boni-shop.com

- **Extract Phone Numbers:**

```
= Table.AddColumn("#Previous Step", "ph", each let
    TextList = Text.Split([sellerdetails], " "),
    ExtractedNumbers = List.Select(TextList, each
        Text.StartsWith(_, "+") or Text.StartsWith(_, "
    ("))
    in Text.Combine(ExtractedNumbers, " "))
```

- Before Vs After

APC sellerdetails	APC 123 Extracted Phone Number
Valid 100%	Valid 34%
Error 0%	Error 0%
Empty 0%	Empty 66%
975 distinct, 973 unique	
Dataanalyst is committed to providing each customer with the highest...	null
Impressum Gesetzliche Anbieterkennung: Daniel Vogeley und Christop...	022342509631 E-Mail: Ust-IdNr.:
DINAN is committed to providing each customer with the highest stan...	null
Flukey LLC is committed to providing each customer with the highest s...	null
Vor- und Nachname des Unternehmensvertreters: Rizhi Jiang Name d...	10169-0499 720-532-4411 81-3873597
HOCK Design e.K. Kölner Strasse 45 51766 Engelskirchen Germany Tel...	02262-7075160 amazon@hockdesign.de
ISE Forest & Garden is committed to providing each customer with the...	null
IMPRESSUM itenga GmbH Am Umspannwerk 2 66557 Illingen Deutsch...	06825 89697-0 06825 89697-20 E-Mail: EU-Kommission
IMPRESSUM Johannes Zierleyn KG Am Schweinemarkt 3 48529 Nord...	+49 +49 E-Mail:

- **Purpose:** These extractions and transformations make detailed contact information and geographical data readily available for analysis.

7. Data Type Standardization

- **Purpose:** Logically adjusted the data types of all columns to align with their content (e.g., numeric data types for ratings and counts, text for descriptive

fields), ensuring that data operations performed during analysis are appropriate and efficient.

Conclusion

The preprocessing steps outlined above were critical in transforming the raw dataset into a structured, clean format ready for advanced analysis. This meticulous preparation not only enhances the performance of the data analysis tools but also ensures that the insights derived are based on accurate and relevant data.

By documenting these steps, I provide transparency in my methodology and reinforce the reliability of the data-driven insights provided by the Marketplace Seller Analyzer Dashboard.

DASHBOARD REPORT

Introduction

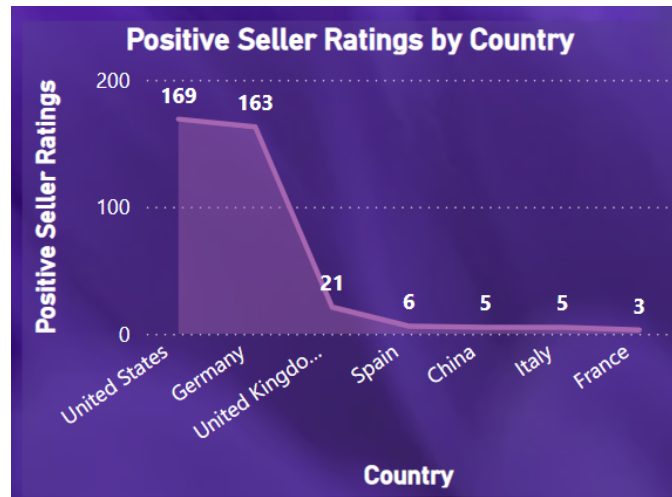
This comprehensive dashboard aims to analyze Amazon merchant data to identify promising e-commerce acquisition targets, utilizing advanced DAX formulas and dynamic slicers to offer deep insights into seller performance and market dynamics.

Detailed Visualization and Analysis

1. Positive Seller Ratings by Country

- **Visual Type:** Line Chart
- **Data Source & DAX:**
 - Column: `Positive Seller Ratings` extracted directly from the data.
 - Insight: The chart reveals that the United States and Germany have the highest positive seller ratings, suggesting superior customer satisfaction and seller performance.
- **Technical Detail:** Utilizes direct column data to represent positive ratings trends over countries.

-

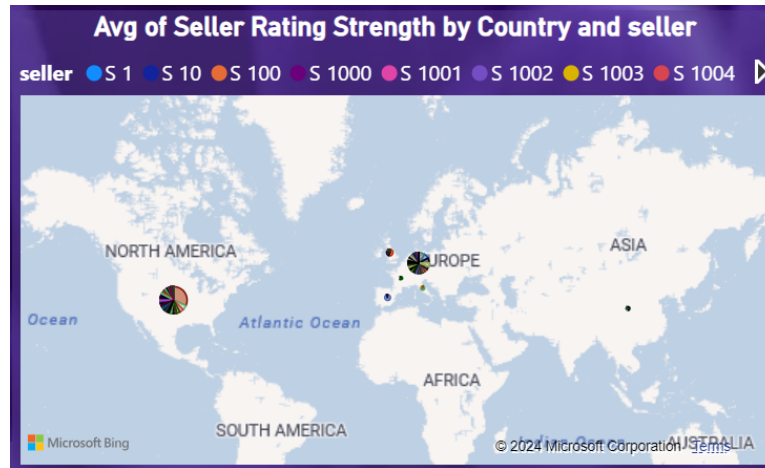


2. Avg of Seller Rating Strength by Country and Seller

- **Visual Type:** World Map
- **DAX Formula:**

```
Seller Rating Strength = ([Positive Seller Ratings] * [Total Ratings]) / [Seller Product Count]
```

- **Details:**
 - Location mapped by **Country**
 - Bubbles sized by **Avg of Seller Rating Strength**
 - Insight: This map visualization highlights the geographic distribution of seller rating strength, with significant clusters in North America and Europe, indicating robust markets.
 -

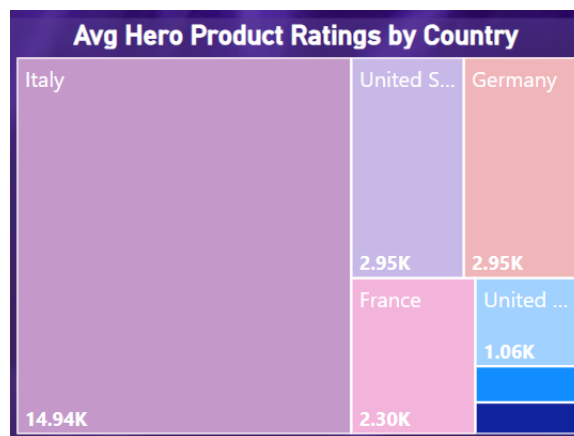


3. Avg Hero Product Ratings by Country

- **Visual Type:** Treemap
- **DAX Formula:**

Avg Hero Product Ratings = ([Hero Product 1 #ratings] + [Hero Product 2 #ratings]) / 2

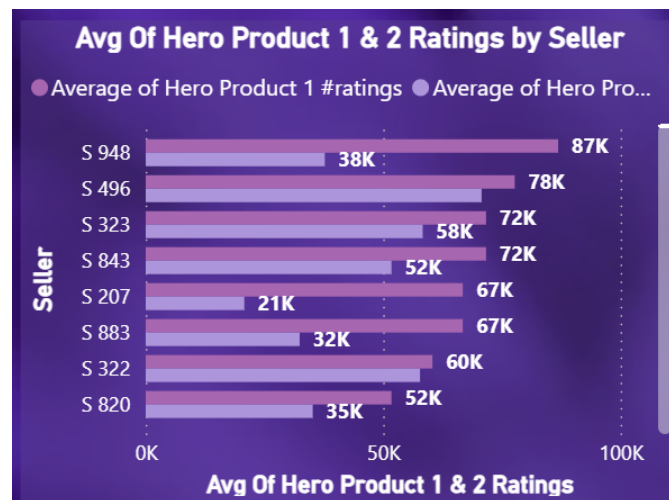
- **Details:**
 - Grouped by **Country**
 - Insight: Indicates the popularity and customer reception of hero products in various countries.
 -



4. Avg of Hero Product 1 & 2 Ratings by Seller

- **Visual Type:** Bar Chart
- **Details:** Represents the average ratings for the top two products offered by each seller, showcasing which sellers have the strongest product offerings.
- **Technical Detail:** Averages taken directly from product ratings columns to assess product performance.

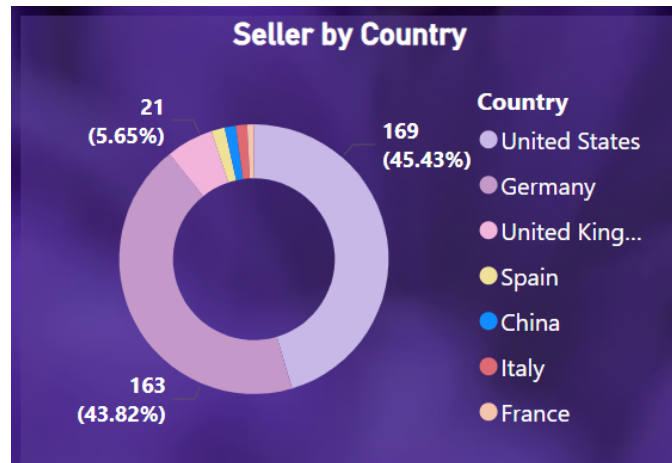
-



5. Seller by Country

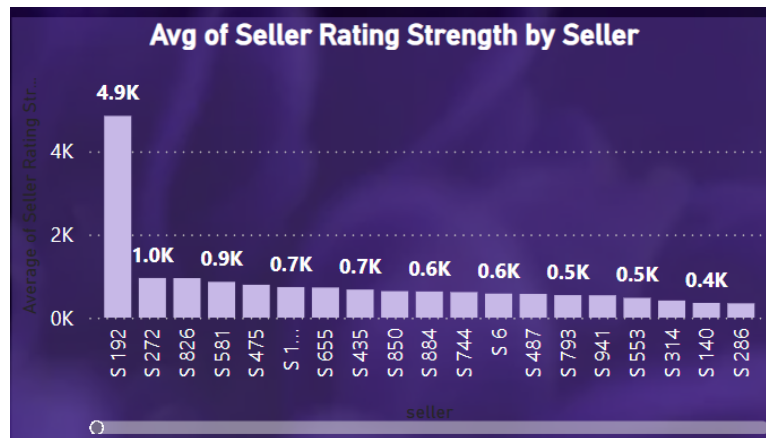
- **Visual Type:** Pie Chart
- **Details:** Shows the distribution of sellers across different countries, using direct count from the data.
- **Insight:** The majority of sellers are based in the United States, followed by significant numbers in Germany and Italy, indicating a concentration of potential acquisition targets in these areas.

-



6. Avg of Seller Rating Strength by Seller

- **Visual Type:** Bar Chart
- **Details:** Utilizes the same measure as the world map, focusing on individual seller performance.



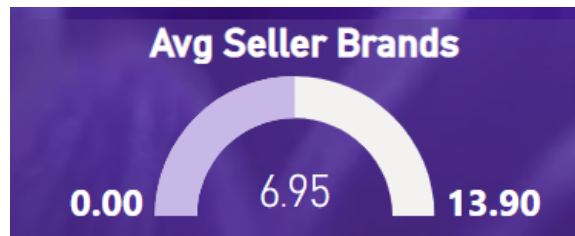
7. Avg Seller Brands Gauge

- **Visual Type:** Gauge
- **DAX Formula:**

```
Avg Seller Brands = AVERAGE([Count of Seller Brands])
```

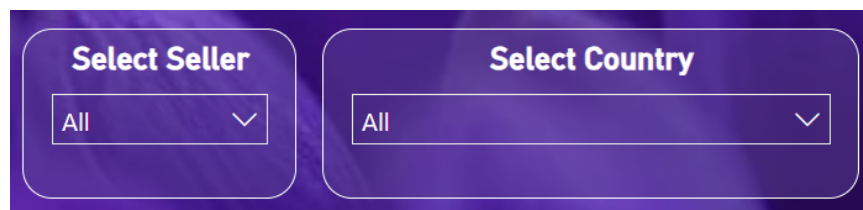
- **Details:**
 - Measures average brand count per seller.

- Insight: Provides a gauge of how diverse each seller's product portfolio is.
-



8. Slicers: Select Seller & Select Country

- **Functionality:**
 - **Select Seller:** Allows users to filter the dashboard based on individual seller data.
 - **Select Country:** Extracted from `Seller Details`, enabling geographic filtering.
- **Details:** Enhances user interaction, allowing for dynamic exploration of the data.
-



9. Total Products Sold

- **Visual Type:** Numeric Indicator
- **DAX Function:**

```
Total Products Sold = SUM([Seller Product Count])
```

- **Details:** Summarizes the total volume of products sold by all sellers, giving a sense of market size.



Overall Assessment

This dashboard effectively uses data visualization to provide strategic insights into seller performance across different metrics and geographies. It uncovers areas of high performance and potential concerns, guiding strategic decisions in the acquisition process. It not only demonstrates the power of data visualization but also utilizes advanced DAX measures and dynamic slicers to provide a deep, interactive analysis. The detailed technical setup behind each visual ensures that stakeholders can trace the logic used, reinforcing the reliability of the insights provided.

Conclusion

The insights derived from the Marketplace Seller Analyzer Dashboard highlight the crucial areas for potential investment and acquisition, underscoring the importance of data-driven decision-making in e-commerce. By effectively combining data extraction techniques, DAX calculations, and dynamic visualizations, this dashboard serves as a crucial tool for strategic decision-making in e-commerce acquisitions. It allows stakeholders to pinpoint high-potential targets, understand market dynamics, and strategize effectively based on comprehensive data analysis.