

## Steps used for data cleaning:

1. Connect to Data:
  - Open Tableau Prep and connect to your data source. You can connect to various types of data, including spreadsheets, databases, and cloud services.
2. Review Data:
  - Once your data is loaded, take some time to review it. Check for inconsistencies, missing values, duplicate entries, and data types.
3. Remove Unnecessary Columns:
  - Eliminate any columns that are not needed for your analysis. This helps streamline your dataset.
4. Rename Fields:
  - Rename columns to make them more understandable and consistent, ensuring that they follow a naming convention.
5. Handle Missing Values:
  - Address any missing values by filling them in with appropriate substitutes (mean, median, mode), removing rows, or flagging them for review.
6. Remove Duplicates:
  - Identify and remove duplicate records to ensure accuracy in analysis.
7. Clean Data Formats:
  - Standardize the formatting of your data, such as dates and text fields. For example, convert all text to lowercase or trim whitespace.
8. Split Columns:
  - If you have concatenated data (like full names), split them into separate columns (for example, first name and last name).
9. Modify Data Values:
  - Replace incorrect or inconsistent values (e.g., changing "NY" to "New York").
10. Create Calculated Fields:
  - Create any necessary calculated fields to derive new insights or metrics from your existing data.
11. Sort and Filter Data:
  - Apply sorting and filtering to focus on relevant data points.
12. Preview and Validate:
  - Continuously preview the cleaned data to ensure that the preparations align with your objectives. Validate any transformations or calculations performed.
13. Output the Cleaned Data:
  - Once the data is cleaned and prepared, output the results to Tableau Desktop or save it in the desired format for further analysis.

[illegible]

1. Removed the null values from the columns necessary to take decision such as sellerratings, sellerproductcount
2. Checked for duplicate rows if present
3. Parsed required data:
  - a. Seller percentage:  
LEFT([sellerratings],  
FIND([sellerratings], "%") - 1)  
Add percentage sign with: [sellerratings] + "%"
  - b. Seller product count: REPLACE([sellerproductcount], "1-16 of ", "") ->  
REPLACE([sellerproductcount], " results", "") -> REPLACE([sellerproductcount],  
"over ", "")
  - c. Seller ratings:  
MID([sellerratings],  
FIND([sellerratings], "(") + 1,  
FIND([sellerratings], ")") -  
FIND([sellerratings], "(") - 1)
  - d. Email:  
REGEXP\_EXTRACT([sellerdetails], "E-Mail:\s\*(.\*?)\n")  
REGEXP\_EXTRACT([sellerdetails], "Mail:\s\*(.\*?)\n")
  - e. Telephone number:

REGEXP\_EXTRACT([sellerdetails], "Tel:\s\*(.\*?)\n")

REGEXP\_EXTRACT([sellerdetails], "Telefon:\s\*(.\*?)\n")

REGEXP\_EXTRACT([sellerdetails], "Tel:.\s\*(.\*?)\n")

- f. To ensure that we have atleast one of the contact details, added a filter:

ISNULL([Calculation5]) AND ISNULL([Calculation3]) AND ISNULL([Calculation1])

AND ISNULL([email])

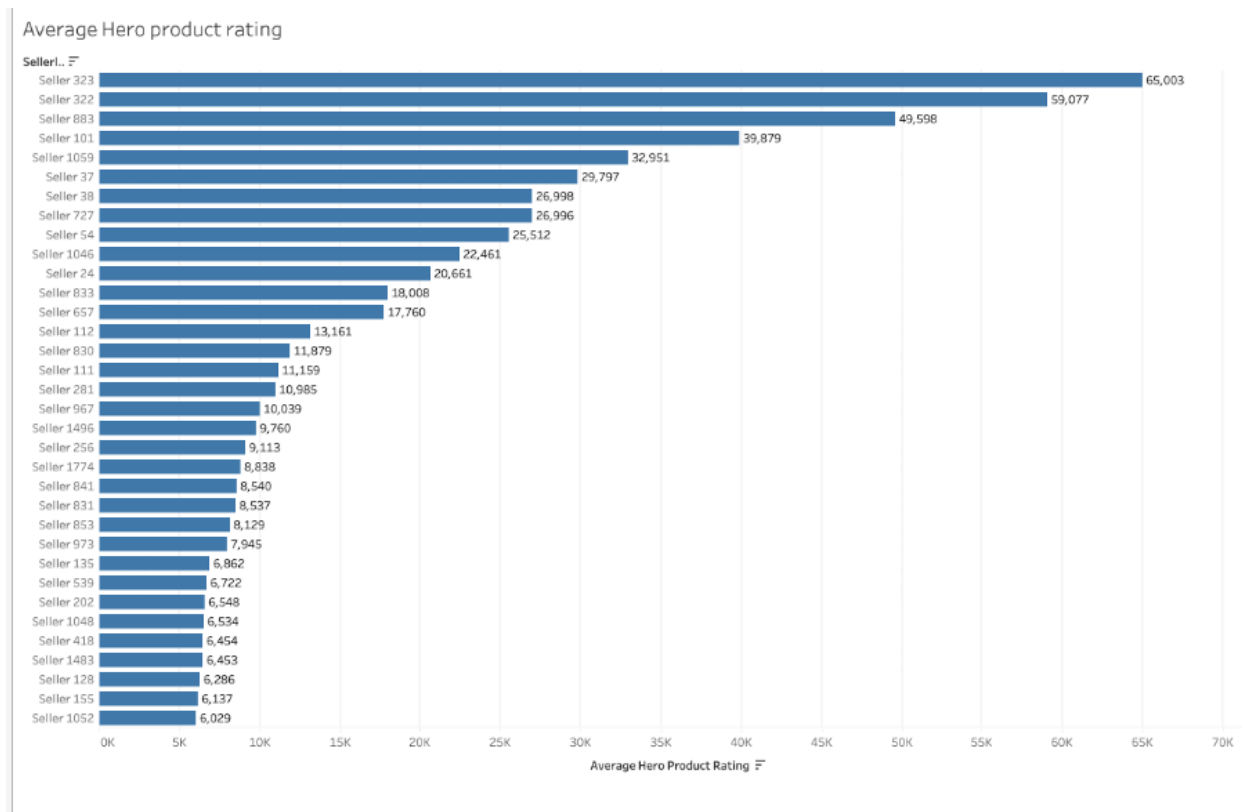
- g. To get the average of the ratings of hero product 1 and hero product 2:

$([Hero\ Product\ 1\ \#ratings] + [Hero\ Product\ 2\ \#ratings])/2$

- h. To get the business location from the address:

RIGHT([businessaddress], 2)

Considering the purpose to find the best or most promising Amazon merchants that the Acquisitions team should reach out to, and acquire, we can filter those sellers with average rating more than 6k.



### Insights from the above achieved graph:

- Top Performers:** Seller 323 has the highest average Hero product rating, significantly leading with 65,003, Seller 322 with 59,077 and Seller 883 with 49,598 ratings. These three sellers dominate the chart, indicating they have a strong market presence or customer satisfaction.
  - Mid-Tier Sellers:** There is a noticeable drop in average ratings after the top three sellers. Sellers like 101 and 1059, with ratings of 39,879 and 32,951 respectively, show a middle ground in performance, suggesting they are strong but not at the top tier.
  - Lower Performers:** The sellers at the bottom of the chart, have significantly lower ratings, ranging from around 6,000 to 7,000. This indicates either lower customer engagement, lesser product visibility, or potentially lower product satisfaction.
- These insights can help in identifying which sellers are performing well and which might need strategic improvements.

Now, we can observe the seller rating percentage for the selected sellers to get an idea about the seller's performance. We have picked up those sellers who has a rating greater than 90%.

Sellerratingspercent	Sellerlink	
100%	Seller 111	11,159
99%	Seller 153	5,979
	Seller 539	6,722
98%	Seller 128	6,286
	Seller 418	6,454
	Seller 1048	6,534
	Seller 135	6,862
	Seller 853	8,129
97%	Seller 724	5,110
	Seller 284	5,569
	Seller 1724	5,991
	Seller 1052	6,029
	Seller 202	6,548
	Seller 831	8,537
	Seller 101	39,879
96%	Seller 1503	5,064
	Seller 973	7,945
	Seller 1774	8,838
	Seller 657	17,760
95%	Seller 1054	5,817
	Seller 833	18,008
	Seller 54	25,512
94%	Seller 155	6,137
	Seller 841	8,540
	Seller 883	49,598
	Seller 322	59,077
93%	Seller 1496	9,760
	Seller 830	11,879
	Seller 38	26,998
92%	Seller 140	5,897
	Seller 256	9,113
	Seller 1059	32,951
	Seller 323	65,003
91%	Seller 1483	6,453
	Seller 37	29,797
90%	Seller 956	5,188
	Seller 967	10,039

## Insights from the above achieved graph:

The graph you provided shows the seller rating percentage for a selected group of sellers who have a rating greater than 90%. Here are some key insights based on the graph:

1. **Top Performers:** The sellers at the top of the graph (Seller 1119, Seller 153, Seller 319) have achieved a perfect 100% seller rating. This indicates exceptional performance and customer satisfaction.
2. **Variability in Performance:** As the ratings slightly decrease, there is a noticeable increase in the number of sellers within each percentage bracket. This could indicate more competition and variability in customer satisfaction levels as ratings decrease.
3. **Seller Rating vs. Transaction Volume:** Alongside the percentage, the number of transactions or interactions (not explicitly shown but often correlated with the size of the rating bar in such graphs) could be an important factor. Sellers with high ratings but also high transaction volumes are likely more reliable than those with fewer transactions, as their high rating is maintained over a larger sample size.
4. **Strategic Importance:** Sellers with ratings closer to 90% and below (which have not been considered in order to extract only the top sellers) may need strategic improvements to enhance their customer satisfaction and reach the top tier. These sellers represent potential targets for further analysis to understand what specific factors are limiting their rating increase.

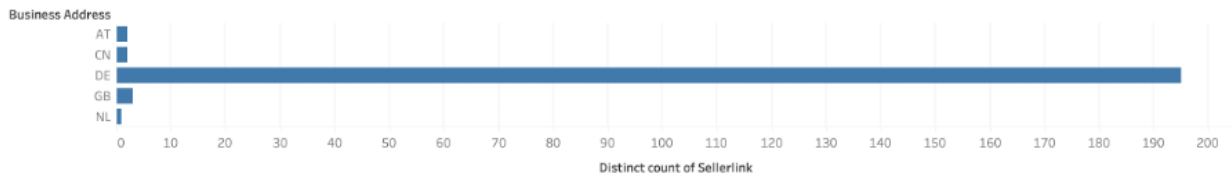
This graph can help in making strategic decisions regarding which sellers to partner with, offer incentives for improvement, or prioritize for developmental help to maintain high customer satisfaction standards.

Upon parsing out the phone number or email whichever is available in the data set we can connect to the top sellers selected for acquisition.

Sellerlink	Phn Number2	Email
Seller 323	Null	info@limuno.com
Seller 322	0371 400 7682	shop@varia-store.com
Seller 883	+4962369329325	service@vitamed-quinger..
Seller 101	+ 49 (0)2273 / 991-9325 (..	Null
Seller 1059	04942203673	amazon@wtnord.de
Seller 38	+49 5921 834011	nils.holm@zierleyn.de
Seller 727	06206 - 951326 - 0	support@trendmile.com
Seller 54	02951/9849790 (Montag ..	info@kaleidoshop.de
Seller 1046	02431 9569550	amazon@wohn-idee24.de
Seller 24	02521/8406-0 E-Mail: bib...	bib.beckum@blumenbeck..
Seller 657	06772 9669046 Telefax: 0..	amazon@tmf-racing.com ..
Seller 112	Null	info@kiddy-hits.de
Seller 830	030/86204408	amazon@verpacking.com
Seller 111	+49(0)3386 280979	Null
Seller 967	+49 3331 - 3010560	Null
Seller 1496	+49 30 436 07 330 (Mo. bi..	support@yovivo.de
Seller 256	01789819900	Null
Seller 1774	03475 714957	info@zooheld.de
Seller 841	06831-9458970Telefax: 0..	info@lenkgetriebe.netGe..
Seller 831	071312798611	info@fairmarkt24.de
Seller 853	07841628433	shop@violadirekt.eu
Seller 973	07191-3272-25	amazon@welovebags.de
Seller 135	03054619947	amazon@kjells.de
Seller 539	0314 / 25 41 71 50	Null
Seller 202	05173.2409944	Null
Seller 1048	49 (0) 69 15625967	shop@wohnraum-berger...
Seller 418	Null	info@lubalo-ug.de
Seller 1483	0151 23758639UST-ID: D..	info@xylabis.deTelefon: 0..
Seller 128	0938297910	info@kirchner24.de
Seller 155	Null	info@kobalz.de
Seller 1052	+49 (0)36338 59889-00	info@highlight-led.de
Seller 1724	09834/975133Telefax: 09..	zierfischtreff@aol.comU..
Seller 153	04084605742	mail@knobelbox.com
Seller 140	05141 5934480	Null
Seller 1054	+49 (0) 8191/9406800	info@wortek.de
Seller 284	0177 33 709 77	distribute@leucht-fachha..
Seller 956	02405 / 4068714	info@wasserflora.de
Seller 724	04022853663	info@trekvogel.de
Seller 1503	02162/8154770	info@kiosk-osman.de
Seller 805	0423166811	amazon@vbs-versand.de
Seller 486	08623 - 98728015	amazon
Seller 1023	03471621561	amazon2015@werkzeugl..
Seller 494	04263 9380-500	support@roehrs-baumark..
Seller 573	0241/51963622	info@teppichmax.de
Seller 697	Null	info@tortendeko-shop.de
Seller 684	Null	amazon@werkzeug-schul..
Seller 178	Null	info@kotte-zeller.de
Seller 1047	08721916360	info@wohnkult.net
Seller 156	+49 (0) 9721 38895-15	Null
Seller 72	01742483335	katerrot27@aol.com
Seller 572	0202/90509020 E-Mail: s..	service@teppichcenter24..
Seller 96	05115455911	kerima.tools@gmx.de

Using different extraction and cleaning methods I have achieved a clean table of the sellers with their phone numbers and their email addresses. This will help me to gain a clear knowledge regarding their contact details for future reference.

For further information based on after data cleansing, with this graph we can say that most of the sellers are from DE, the business locations we extracted for the top sellers from the business address provided in the data sample.



### Insights observed based on overall data set:

The distribution of business addresses across different countries or regions, denoted by their two-letter country codes. Here's a breakdown of the insights:

1. **DE (Germany):** Dominates the chart with the highest number of entities, significantly outnumbering the other countries. There are approximately 170-180 entities associated with this address.
2. **AT (Austria), CN (China), GB (United Kingdom), and NL (Netherlands):** These countries have far fewer entities, each with only a handful of records, likely fewer than 10 in each case.

Interpretation:

- **Concentration in DE:** The overwhelming majority of the business addresses are in Germany. This could suggest that the data set is skewed towards German businesses, or it may reflect a particular business focus or data collection in this region.
- **Minimal Distribution Elsewhere:** The other countries have minimal representation, indicating a possible lack of business presence or data collection from these regions in the data set.

This distribution suggests that any analysis or conclusions drawn from this data set will heavily reflect the situation in Germany rather than a more balanced, multinational perspective.



# **Conclusion:**

## **1. Dominance by a Few Sellers:**

- A small group of sellers, particularly Sellers 7, 323, and 322, dominate the marketplace in terms of average product ratings and overall performance. These sellers consistently receive high ratings, indicating strong customer satisfaction and market presence.

## **2. Wide Performance Disparity:**

- There is a significant disparity between the top-performing sellers and the rest. While the leading sellers have very high ratings, many other sellers fall into mid or lower tiers, with ratings significantly lower. This suggests that while some sellers excel, many others struggle to reach the same level of performance.

## **3. Potential for Improvement Among Lower-Tier Sellers:**

- A considerable number of sellers have ratings in the lower range (90%-95%), indicating room for improvement. These sellers could benefit from targeted support, such as better visibility, enhanced customer service, or product quality improvements, to boost their ratings and overall performance.

## **4. Opportunity for Broader Seller Success:**

- The marketplace is currently skewed towards a few high-performing sellers, but there is an opportunity to elevate mid and lower-tier sellers. With appropriate strategies, such as incentivizing quality improvements or providing more resources, the marketplace could see a more balanced representation of successful sellers.

## **5. Importance of Customer Feedback:**

- The ratings and performance data highlight the critical role of customer feedback in determining seller success. Sellers who actively engage with customer feedback and make improvements are more likely to achieve higher ratings and stand out in the marketplace.

Hence, the marketplace is characterized by a few dominant sellers with very high ratings, but there is significant potential for growth and improvement among many other sellers. By fostering a more competitive and supportive environment, Amazon can help a broader range of sellers succeed, ultimately leading to a more diverse and robust marketplace.