Enhancing Customer Satisfaction at Shopzilla Through Data Analysis

(Project 1)

By: Shantal Rivas

Tools Used: BigQuery (SQL), Kaggle (Dataset) & Tableau

Dataset Owner: Dee Dee

Introduction

In today's highly competitive e-commerce landscape, where customers have endless choices, maintaining high customer satisfaction is essential. A strong **Customer Satisfaction Score (CSAT)** not only fosters brand loyalty but also plays a key role in overall business success.

For this project, I analysed customer satisfaction trends using a **fabricated dataset from Kaggle** (created with the Faker library to simulate real-world data). To ensure the analysis is both relevant and practical, I developed a business case in which **Shopzilla, a leading e-commerce platform, is experiencing a decline in CSAT scores due to common customer complaints**.

My goal is to **identify key factors contributing to lower CSAT scores and provide data-driven recommendations to improve customer satisfaction**. This project also showcases the **data analytics skills I acquired through the Google Data Analytics Certificate**, including **SQL (BigQuery)** for data extraction, **Tableau** for visualization, and structured problem-solving using the **Google Data Analytics Process** (**Ask, Prepare, Process, Analyse, Share, Act**).

Through this analysis, I explore **customer complaints, evaluate agent performance, and assess how response time impacts CSAT**. Ultimately, my recommendations aim to **reduce complaints, enhance agent efficiency, and improve overall customer satisfaction**, ensuring Shopzilla remains competitive in the e-commerce industry.

ASK: Defining the Business Problem

Background:

Shopzilla, a prominent e-commerce platform, has recently been facing a rise in negative customer feedback, leading to a decline in Customer Satisfaction (CSAT) scores. Customers have reported a variety of issues related to orders, returns, and overall customer support.

To address this, I defined the following business objectives:

- **Identify key factors influencing customer satisfaction**: What are the most frequent customer complaints?
- **Analysing Customer Feedback**: What Do Customers Complain About the Most?
- **Assess the impact of agent performance on CSAT scores**: Is response time too long and affecting CSAT? How are customers rating the average service provided by agents?
- **Recommend strategies to improve the customer experience and reduce complaints**.

In addition, I identified key stakeholders who would benefit from these insights:

- **Customer Support Managers**: Optimize agent response times and enhance training programs.
- **Operations Team**: Improve logistics and streamline the issue resolution process.
- **Executives & Strategy Team**: Leverage insights to drive business improvements.

PREPARE: Load & Understand the Data

Dataset Overview:

I uploaded my dataset to **BigQuery (Google Cloud)** for analysis. To ensure the data was properly uploaded and structured, I first examined its contents by running the following SQL query:

SQL Query:

```
SELECT *

FROM `project-1-449815.shopzilla_analysis.customer_satisfaction_data`

LIMIT 10;
```

💡 **Key Insights:**

- The dataset contains **85,907 rows and 20 columns**, with a mix of text and numerical data.
- **CSAT scores range from 1 to 5**, providing a clear metric for customer satisfaction.

Key Columns for Analysis:

I identified the most relevant columns for this analysis:

- **CSAT Score** → Customer satisfaction rating (1-5).
- **Issue Category** → Type of complaint (e.g., delivery, returns, product issues).
- **Order Date Time** → Timestamp of when the order was placed (important for time-based trends).
- **Product Category** → Category of the purchased product (helps identify high-compliant products).
- **Issue Reported Time** → Timestamp of when the issue was reported.
- **Issue Resolution Time** → Duration taken to resolve the issue.
- **Order ID** → Unique identifier for each order (useful for tracking duplicates).
- **Customer Remarks** → Customer feedback or comments categorized by issue type.

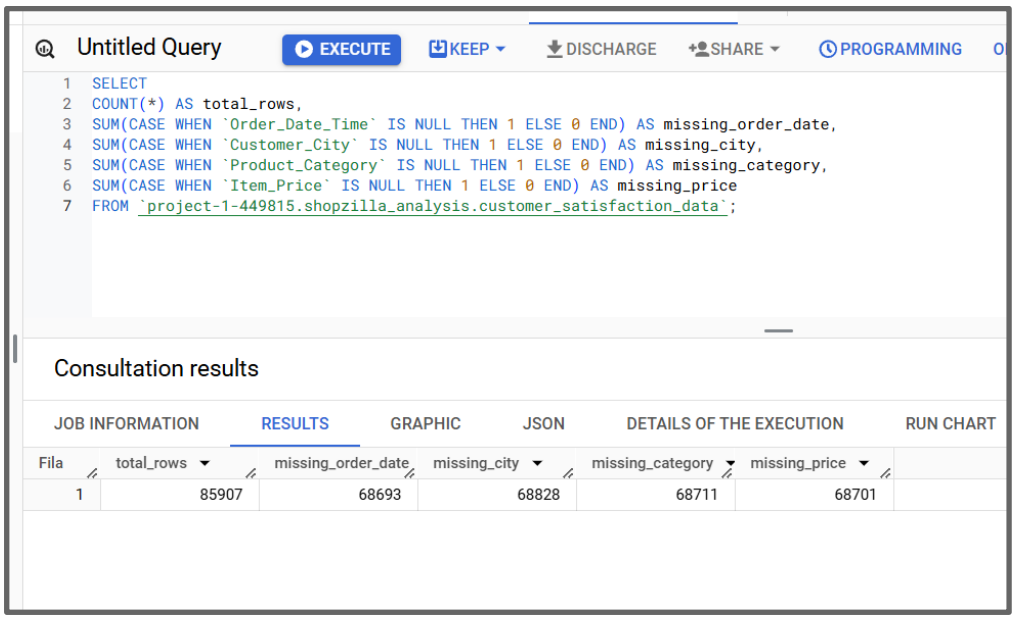
PROCESS: Data Transformation & Preprocessing

Before diving into analysis, I cleaned the data to ensure **accuracy, consistency, and completeness**. Here's how I approached the preprocessing phase:

Handling Missing or Null Values:

To identify missing values in key columns, I ran the following **SQL query**:

```
SELECT
COUNT(*) AS total_rows,
SUM(CASE WHEN `Order_Date_Time` IS NULL THEN 1 ELSE 0 END) AS missing_order_date,
SUM(CASE WHEN `Customer_City` IS NULL THEN 1 ELSE 0 END) AS missing_city,
SUM(CASE WHEN `Product_Category` IS NULL THEN 1 ELSE 0 END) AS missing_category,
SUM(CASE WHEN `Item_Price` IS NULL THEN 1 ELSE 0 END) AS missing_price
FROM `your_project_id.shopzilla_analysis.your_table_name`;
```



💬 Key Consideration:

Since I used the **free BigQuery tier (Sandbox)** for this project, I did not have access to certain operations like **UPDATE** or **DELETE**. To work around this limitation, I used the **CREATE OR REPLACE TABLE** operation to clean and transform the data as needed.

To ensure accurate analysis, I performed several key data cleaning steps, including handling missing values, checking for duplicates, converting time formats, and correcting invalid data.

1️⃣ Handling Missing Values:

Since **accurate timestamps** are critical for time-based analysis, I **removed rows with missing Order_Date_Time** using the following query:

```
CREATE OR REPLACE TABLE `project-1-449815.shopzilla_analysis.new_table_name` AS
SELECT *  FROM `your_project_id.shopzilla_analysis.your_table_name`
WHERE `Order_Date_Time` IS NOT NULL;
```

Next, I addressed missing values in **categorical columns** (Customer_City and Product_Category). Instead of removing these rows, I replaced null values with "Unknown" or "Unspecified" to maintain data integrity:

```
CREATE OR REPLACE TABLE `project-1-449815.shopzilla_analysis.new_table_name` AS
SELECT  COALESCE(`Customer_City`, 'Unknown') AS `Customer City`,
COALESCE(`Product_Category`, 'Unspecified') AS `Product Category`  FROM
`your_project_id.shopzilla_analysis.your_table_name`;
```

This ensures that the dataset remains complete and usable for analysis without unnecessary data loss.

2️⃣ Checking for Duplicates

To verify data consistency, I checked for duplicate entries in key fields (Order ID and Customer ID) using the following query:

```
SELECT `Order ID`, COUNT(*)
FROM `your_project_id.shopzilla_analysis.your_table_name`
GROUP BY `Order ID`
HAVING COUNT(*) > 1;
```

✅ **No duplicates were found.** However, if any had been present, I would have removed them using:

```
CREATE OR REPLACE TABLE `your_project_id.shopzilla_analysis.new_table_name` AS
SELECT DISTINCT *
FROM `your_project_id.shopzilla_analysis.your_table_name`;
```

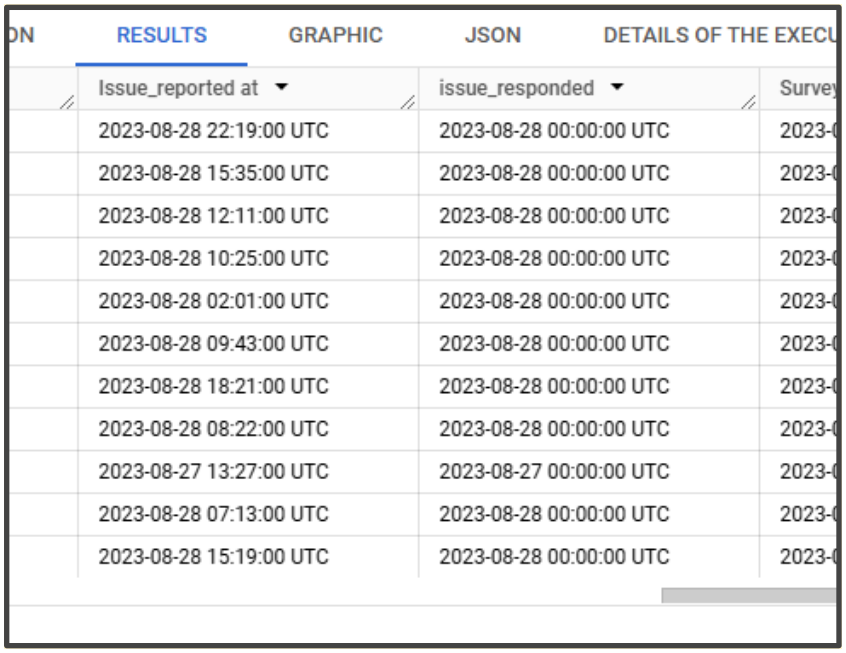
3️⃣ Converting Response Time from Minutes to Hours

To make response time easier to interpret, I converted it from **minutes to hours**:

```
CREATE OR REPLACE TABLE `your_project_id.shopzilla_analysis.new_table_name` AS
SELECT *, TIMESTAMP_DIFF(`issue_responded`, `Issue_reported_at`, HOUR) AS
`Response_Time_Hours` FROM `your_project_id.shopzilla_analysis.your_table_name`;
```

4️⃣ Fixing Negative Timestamp Values

During data validation, I identified **negative values** in the issue_responded and issue_reported_at columns. These cases represented instances where the **response timestamp was recorded before the issue was reported**, which is logically incorrect and would skew the analysis. See below:



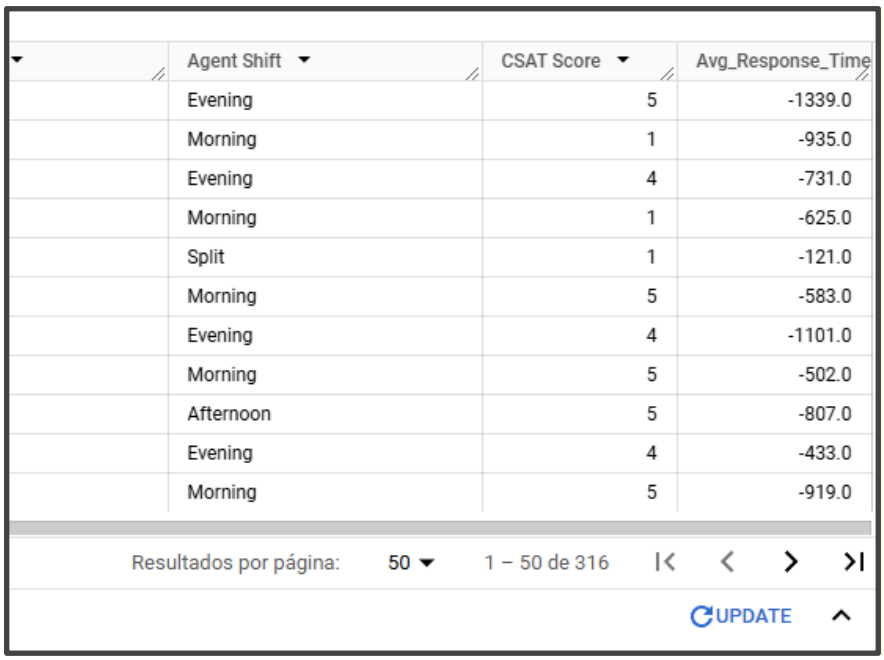
ON | RESULTS | GRAPHIC | JSON | DETAILS OF THE EXECU

	Issue_reported at	issue_responded	Survey
	2023-08-28 22:19:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 15:35:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 12:11:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 10:25:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 02:01:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 09:43:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 18:21:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 08:22:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-27 13:27:00 UTC	2023-08-27 00:00:00 UTC	2023-0
	2023-08-28 07:13:00 UTC	2023-08-28 00:00:00 UTC	2023-0
	2023-08-28 15:19:00 UTC	2023-08-28 00:00:00 UTC	2023-0

Identifying Invalid Rows:

To determine how many rows contained negative values, I ran the following query:

*SELECT * FROM `project-1-449815.shopzilla_analysis.your_table_name` WHERE TIMESTAMP_DIFF(issue_responded, issue_reported_at, MINUTE) < 0;*



	Agent Shift	CSAT Score	Avg_Response_Time
	Evening	5	-1339.0
	Morning	1	-935.0
	Evening	4	-731.0
	Morning	1	-625.0
	Split	1	-121.0
	Morning	5	-583.0
	Evening	4	-1101.0
	Morning	5	-502.0
	Afternoon	5	-807.0
	Evening	4	-433.0
	Morning	5	-919.0

Removing Invalid Data:

Since **BigQuery does not support direct row deletion**, I created a **new table** that excluded these erroneous rows:

```
CREATE OR REPLACE TABLE `project-1-449815.shopzilla_analysis.new_table_name` AS
SELECT *
FROM `project-1-449815.shopzilla_analysis.your_table_name`
WHERE TIMESTAMP_DIFF(issue_responded, issue_reported_at, MINUTE) >= 0;
```

By filtering out these records, I ensured that the dataset remained **accurate and reliable** for further analysis.

Key Insights from the Data Cleaning Phase

1. **Missing Values Identified:**
   I discovered missing values in key columns such as Order_Date_Time, Customer_City, Product_Category, and Item_Price, indicating a need for improved data completeness.
2. **Rows Removed for Accuracy:**
   To ensure accurate analysis, I removed any rows with missing Order_Date_Time, as complete timestamps are crucial for tracking customer behavior effectively.
3. **Filling in Gaps:**
   To maintain data integrity, I filled in missing values for Customer_City and Product_Category with "Unknown" and "Unspecified." This allowed the analysis to continue without interruption and prevented gaps in the dataset.

4. **No Duplicates Found:**
   A thorough check revealed that there were no duplicate entries in key fields like Order ID and Customer ID, ensuring each record was unique and reliable.
5. **Time Data Converted:**
   I converted average response times from minutes to hours, which simplified interpretation and made the data more user-friendly.
6. **Invalid Data Removed:**
   Rows with negative values in the issue_responded and issue_reported_at columns were identified and removed, as they represented logically inconsistent timestamps.
7. **Clean Tables Created:**
   Since I was using the free tier of BigQuery (which doesn't support direct row deletion), I created new tables to ensure that the dataset was thoroughly cleaned and ready for accurate analysis.

ANALYZE: Data Exploration & Insights

In this analysis, I focused on identifying the key factors impacting customer satisfaction for Shopzilla, using SQL queries in BigQuery and visualizations in Tableau. Here's a concise breakdown of the findings and recommendations:

1. Identifying Most Common Customer Complaints (What are the most frequent customer complaints?)

To identify the most frequent complaints, I ran the following SQL query:

```
SELECT category AS Issue_Category, COUNT(*) AS Complaint_Count
FROM `project-1-449815.shopzillaa_analysis.your_table_name`
GROUP BY category
ORDER BY Complaint_Count DESC;
```

Consultation results

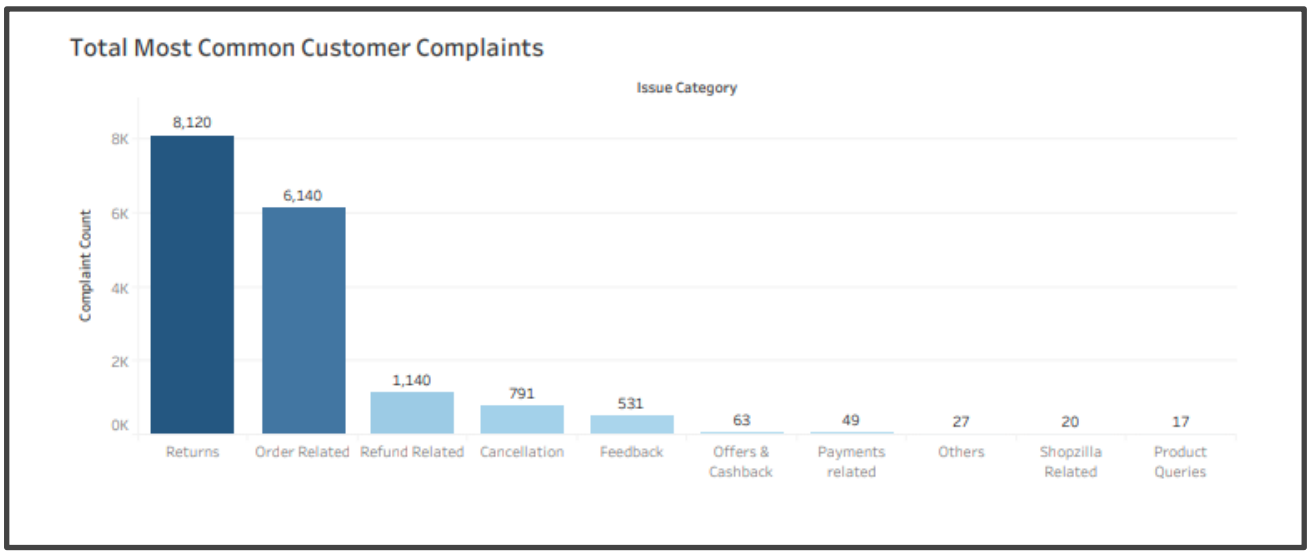
JOB INFORMATION | RESULTS | GRAPHIC | JSON | DETAILS OF THE

Fila	Issue_Category	Complaint_Count
1	Returns	8242
2	Order Related	6284
3	Refund Related	1172
4	Cancellation	802
5	Feedback	538
6	Offers & Cashback	63
7	Payments related	49
8	Others	27
9	Shopzilla Related	20
10	Product Queries	17

This query counts the number of complaints per issue category and sorts the results by frequency, allowing us to prioritize improvements in the most common areas.

To visualize these results, I created a bar chart in Tableau, which helped translate the raw data into a clearer format. Although the free version of Tableau did not allow me to create a pie chart (which would have been more ideal), the bar chart still effectively communicated the key insights. I also calculated the percentage distribution of complaints for each category using the following formula:

Percentage = (Complaint Count / Total Complaints) × 100



💡 **Key Insight:**

- Returns: 48.1% of complaints
- Order-related issues: 36.4% of complaints
- 📊 The sum of the total complaints per category is: **16,898**

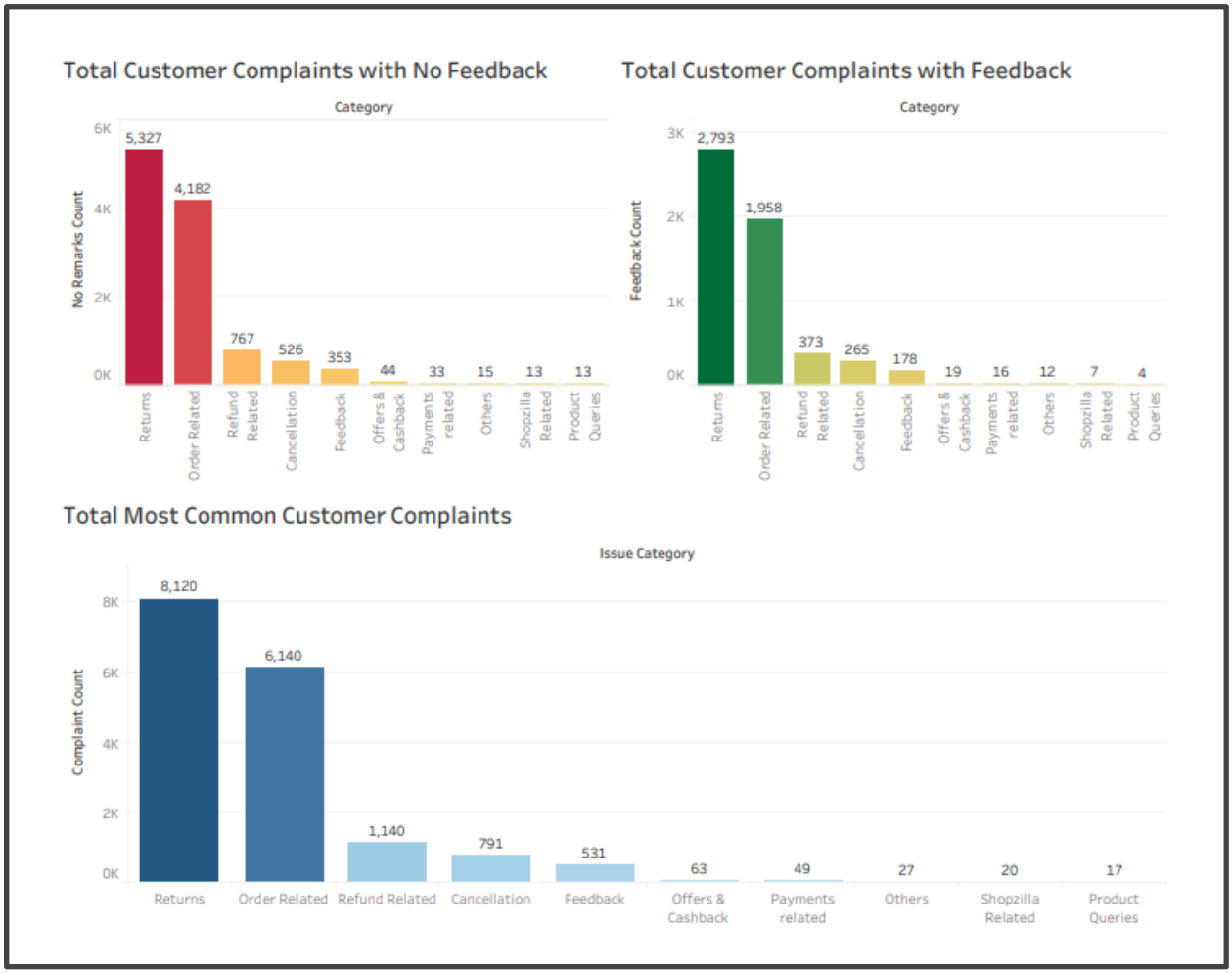
This indicates that customers are primarily struggling with returns and order-related issues, which could involve problems like product quality, incorrect orders, or a complicated return process.

2. Analyzing Customer Feedback (What do customers complain about the most?)

I focused on analyzing customer feedback across the most common complaint categories, specifically identifying instances where customers did not provide remarks (i.e., "No remarks"). Additionally, I noticed several entries with non-informative symbols (e.g., ??, ..), which did not offer meaningful insights.

To assess this, I ran the following query to determine how many complaints lacked detailed feedback or did not offer meaningful insights:

After this, I analysed the total customer complaints with feedback in the categories of Returns, Order Related, Cancellation, and Refund Related in BigQuery. Then, I transformed the raw data into visuals using Tableau and created a Dashboard to share my insights, as shown below:



- Out of 8,120 complaints related to returns, 5,327 had no feedback.
- Out of 6,140 complaints related to order issues, 4,182 had no feedback.

These findings highlight the importance of improving data collection methods and encouraging more detailed feedback from customers. More comprehensive feedback would allow Shopzilla to take targeted actions to address specific problems and enhance customer satisfaction.

📝 Recommendations for Improving Data Collection and Customer Feedback:

Enhance Feedback Collection Methods:

- Encourage customers to provide more detailed feedback by adding open-ended text fields or multiple-choice questions to the complaint submission form.
- **Why?** More detailed feedback helps identify specific pain points and allows for targeted improvements.

Implement Follow-up on No-Remark Complaints:

- Introduce an automated follow-up system to reach out to customers who submit complaints without feedback.
- **Why?** Proactively requesting more details will help close the feedback loop and provide actionable insights.

Incentivize Detailed Feedback for Specific Categories:

- Focus on complaints related to returns and order issues. Offer incentives (e.g., discounts or loyalty points) for providing detailed feedback.
- **Why?** Incentivizing feedback will encourage customers to provide insights into the underlying causes of common issues, helping to improve services and reduce complaints.

📝 Recommendations to Improve Customer Satisfaction (CSAT)

Shopzilla's customer satisfaction (CSAT) has declined primarily due to complaints about incorrect or damaged items and late or missing shipments. Based on the data analyzed, the main areas of concern are Returns, Order-Related, Cancellation, and Refund-Related issues. Below are actionable recommendations:

1. **Fix Quality Control for Products**
   - **Problem:** Incorrect or damaged items are frequently reported in returns and order-related complaints.
   - **Solution:** Improve quality checks at various stages of fulfillment, including pre-shipping inspections and better packaging.
2. **Speed Up and Improve Shipping**
   - **Problem:** Late or missing shipments are causing frustration.
   - **Solution:** Collaborate with reliable shipping carriers, offer multiple delivery options, and provide real-time tracking updates.
3. **Simplify the Returns Process**
   - **Problem:** Customers struggle with the returns process.
   - **Solution:** Offer prepaid return labels, clear instructions, and faster refund or store credit processing.
4. **Better Communication with Customers**
   - **Problem:** Lack of transparency regarding order status and delays.
   - **Solution:** Automate notifications about order progress and delays, and make it easy for customers to contact customer support.
5. **Streamline Refunds**
   - **Problem:** Refunds are slow and complicated.
   - **Solution:** Implement faster refund processes, offering full or partial refunds or store credits quickly.

6. **Offer Satisfaction Guarantees**
    - **Problem:** Customers lack confidence in product quality or delivery service.
    - **Solution:** Offer a "no-questions-asked" satisfaction guarantee with a clear return policy.
7. **Incentivize Feedback**
    - **Problem:** Feedback collection is inconsistent.
    - **Solution:** Offer incentives like discounts or loyalty points to encourage feedback from customers, helping build trust and improve services.
8. **Track and Analyze Complaints Regularly**
    - **Problem:** Complaints are not tracked systematically.
    - **Solution:** Set up an automated system to categorize and prioritize complaints, allowing the team to address the most impactful issues promptly.
9. **Enhance Customer Support**
    - **Problem:** Customer support response times are slow and unhelpful.
    - **Solution:** Provide training and resources for support agents, and leverage tools like AI to speed up issue resolution.
10. **Improve the Ordering Process**
    - **Problem:** Incorrect items are frequently shipped due to unclear product descriptions.
    - **Solution:** Improve product descriptions, images, and size guides, and allow customers to review their order before confirming.

3. Assessing the Impact of Agent Performance on CSAT Scores (Is Response Time Too Long and Affecting CSAT? How Are Customers Rating the Average Service from Agents?)

To understand the impact of response time on customer satisfaction (CSAT), I analyzed the average CSAT scores across all agents and calculated the average response time—from the moment the issue was reported until it was resolved. The following query was used:

```
SELECT ROUND(AVG(`CSAT Score`), 2) AS Avg_CSAT,
ROUND(AVG(TIMESTAMP_DIFF(`issue_responded`, `Issue_reported at`, HOUR)), 2) AS Avg_Response_Time_Minutes
FROM `project-1-449815.shopzilla_analysis.your_table_name`;
```

This query calculates the average CSAT score and the average response time in hours, by measuring the time difference between when an issue is reported and when it is responded to. The results are rounded to two decimal places for clarity.

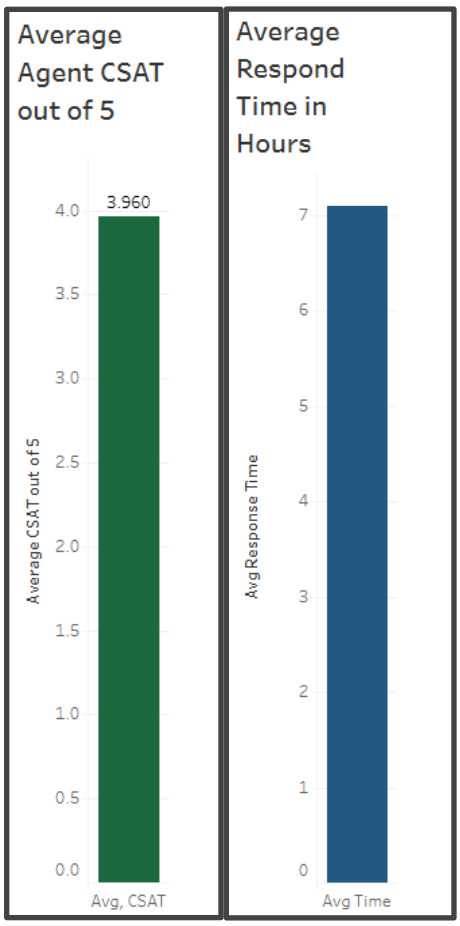
Consultation results

JOB INFORMATION | RESULTS | GRAPHIC | JS

Fila	Avg_CSAT	Avg_Response_Time
1	3.96	7.09

Key Insights:

- **Avg CSAT Score**: 3.96
- **Avg Response Time**: 7.09 hours



The average CSAT score of 3.96 is a strong result, indicating a generally positive customer experience. According to research from Flinder, a smart finance functions company, a CSAT score of 3.96 out of 5 is considered excellent in the e-commerce sector.

The average response time of 7.09 hours is within an acceptable range, as Zendesk's research suggests that response times up to 8 hours are often acceptable, particularly when issues require detailed investigation or coordination. While this is good, there is always room for improvement, and faster response times could contribute to even higher

levels of customer satisfaction. Below, I've outlined some recommendations to further enhance performance.

Recommendations:

1. **Monitor Agent Response Times**: While the current response time of 7.09 hours is within an acceptable range, reducing it—especially for simpler or more urgent issues—could have a positive impact on CSAT scores.
2. **Equip Agents with the Right Tools**: Providing training and resources that enable agents to resolve issues more efficiently—without sacrificing quality—can make a significant difference. Tools like preset responses or AI-driven assistance can help agents work more quickly and effectively.
3. **Embrace Technology**: Implement automated responses for common or straightforward inquiries. This will allow agents to focus on more complex issues. Additionally, AI-powered tools can assist agents in providing faster resolutions while maintaining a personalized experience for customers.
4. **Set Clear Response Time Goals**: Strive to reduce the average response time to below 7 hours, as recommended by Zendesk. This could involve optimizing workflows, automating tasks, or even adding more agents during peak times to ensure quicker responses.

SHARE: Sharing Visual Insights

To make the data more accessible and actionable for stakeholders, I used Tableau to create clear, visual charts that transform complex data into meaningful insights. A detailed presentation of these insights, along with the recommendations, is available here:

https://docs.google.com/presentation/d/1WNz4MAAoenA1kzr79OufFHFMAW9AQytP/edit?usp=sharing&ouid=100623317356432856394&rtpof=true&sd=true

ACT: Implementing Recommendations & Tracking Progress

In this phase, the focus is on executing the recommendations to enhance customer satisfaction and streamline operations, while closely monitoring the impact of these changes to ensure they are driving positive results.