Northwind Traders – Full Project

By Karl Toto Same

INTRODUCTION :

With relentless dedication to mastering Data Analytics, I am now embarking on a significant milestone—my first major project. This journey takes us into the data of Northwind Traders, a fictitious gourmet food supplier, where I am poised to transform my knowledge into visual insights.

The objective of this project is to leverage my newly acquired skills in EXCEL, SQL and Tableau to conduct a comprehensive data analysis of Northwind Traders—a fictitious food company. Through this analysis, I aim to uncover valuable insights into sales trends, product performance, employee performance, customer behavior and shipping cost. By synthesizing data and employing data visualization techniques, I intend to provide actionable recommendations that can drive informed decision-making for the company.

I downloaded the dataset on the website Maven Analytics as a folder containing CSV files. The data includes information on customers, products, orders, shippers, and employees.

Table	Field	Description
orders	orderID	Unique identifier for each order
orders	customerID	The customer who placed the order
orders	employeeID	The employee who processed the order
orders	orderDate	The date when the order was placed
orders	requiredDate	The date when the customer requested the order to be delivered
orders	shippedDate	The date when the order was shipped
orders	shipperID	The ID of the shipping company used for the order
orders	freight	The shipping cost for the order (USD)
order_details	orderID	The ID of the order this detail belongs to
order_details	productID	The ID of the product being ordered
order_details	unitPrice	The price per unit of the product at the time the order was placed (USD - discount not included)
order_details	quantity	The number of units being ordered
order_details	discount	The discount percentage applied to the price per unit
customers	customerID	Unique identifier for each customer
customers	companyName	The name of the customer's company
customers	contactName	The name of the primary contact for the customer
customers	contactTitle	The job title of the primary contact for the customer
customers	city	The city where the customer is located
customers	country	The country where the customer is located
products	productID	Unique identifier for each product
products	productName	The name of the product
products	quantityPerUnit	The quantity of the product per package
products	unitPrice	The current price per unit of the product (USD)
products	discontinued	Indicates with a 1 if the product has been discontinued
products	categoryID	The ID of the category the product belongs to
categories	categoryID	Unique identifier for each product category
categories	categoryName	The name of the category
categories	description	A description of the category and its products
employees	employeeID	Unique identifier for each employee
employees	employeeName	Full name of the employee
employees	title	The employee's job title
employees	city	The city where the employee works
employees	country	The country where the employee works
employees	reportsTo	The ID of the employee's manager
shippers	shipperID	Unique identifier for each shipper
shippers	companyName	The name of the company that provides shipping services

DATA PREPROCESSING AND CLEANING:

In the initial stage of data preparation, I imported the dataset into Excel using Power Query. To ensure clarity and consistency, I set the first row as column headers for the 'Customers' sheet.

One significant challenge encountered during this phase was the presence of invalid characters, particularly accents, in both the 'Customers' and 'Products' sheets. These characters posed compatibility issues when importing the data into MySQL.

Sir Rodney's Scones
Gustaf's Knackebr🞋d
Tunnbr🞋d
Guarana Fantastica
NuNuCa🞋Nu🞋-Noug...
Gumb🞋r Gummib🞋rc...
Schoggi Schokolade
R🞋ssle Sauerkraut
Th🞋ringer Rostbratwu...
Nord-Ost Matjeshering

To address this, I meticulously identified and replaced each problematic '?' character with a 'e,' choosing 'e' using the Search & Replace function on EXCEL due to its prevalence in various languages, including English, French, Spanish, and German.

These data cleaning and preprocessing steps in Excel were essential to ensure data consistency and quality, setting the stage for a robust analysis.

In the second stage of data cleaning, I employed the import wizard in MySQL Workbench to bring in the preprocessed datasets. At this point, my focus remained on processing and cleaning the data.

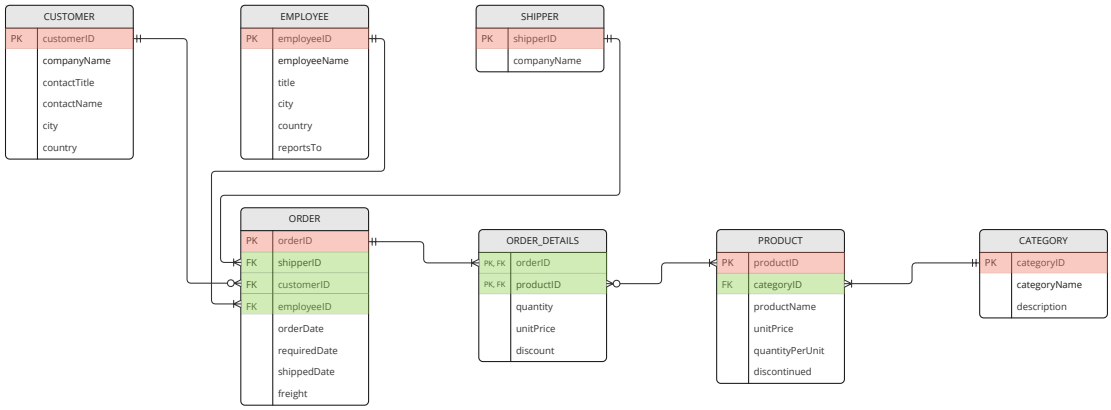
I initiated this stage by changing certain field types such as text to timedate for ‘orderdate’, ‘requiredDate’ and ‘shippedDate’. Additionally, I standardized the date format within the 'Orders' table to streamline further analysis.

Then I addressed missing values in the 'Orders' table, to ensure uniformity and consistency by setting them to 'NULL.'

Furthermore, I paid attention to enhancing data clarity and readability by refining column names in the 'Products' table, a crucial step to facilitate a more structured and effective analysis process.

DATA EXPLORATION:

To have a better understanding of the relationships between the different data tables, I have created an Entity Relationship Diagram using Miro.



This diagram serves as a visual roadmap, outlining the structure of our dataset and the relationships between different entities. It provides a starting point for our 'Data exploration' phase, helping us formulate targeted questions to ask as we dive deeper into the data. By understanding how various elements are connected, we can uncover meaningful insights and direct our analysis effectively.

During the 'Data Exploration' phase, I made extensive use of MySQL Workbench. Through the thoughtful crafting of SQL queries and judicious table joins, I embarked on a general examination of the dataset. This effort was aimed at uncovering primary insights that shed light on various aspects of the business.

My queries and data manipulations were carefully designed to reveal general KPIs, evaluate product performance by country and customer, understand

employee contributions, gain insights into customer behavior, and scrutinize freight costs. These analytical endeavors yielded valuable findings that would serve as the basis for a more thorough analysis later.

After gaining general insights during the exploration phase, I have come up with 6 questions that I thought would be relevant to answer for Northwind Traders stakeholders:

1. ***What are the notable trends in sales over time, and how have they evolved?***
2. ***Which products have performed the best and worst in terms of sales?***
3. ***Which product categories contribute the most and least to overall sales revenue?***
4. ***Can we identify any specific customers who have made substantial purchases or exhibit unique buying behavior?***
5. ***Are shipping costs consistent among different shipping providers, and if not, what variations exist?***
6. ***Which employees have demonstrated superior and inferior performance in sales?***

<u>Interlude: Reflecting on Progress</u>

In this section, I'd like to pause and reflect on the progress we've made in our project so far. It provides an opportunity to share some initial observations, the challenges we've encountered, and the insights we've gained along the way.

Our journey began with the crucial task of cleaning and preprocessing the data. This involved removing special characters, handling missing values, and ensuring data consistency. These initial steps laid the foundation for meaningful analysis.

One significant challenge I faced at this stage was dealing with special characters. Looking back, it was quite frustrating that my only solution at the time was to replace every '?' character with an 'e'. This resulted in some inaccuracies in product names, but it seemed like the only viable option to my knowledge at that point.

Once the data was prepared, the exploration phase allowed us to gain a comprehensive view of the company's operations. It has been truly enlightening to see the data come to life and unveil its underlying stories.

However, as I delved deeper into this dataset with the primary goal of gaining insights into the company's performance, I came to the conclusion that we could streamline our approach. It became apparent that we could accomplish data exploration, analysis, and visualization all within a single tool: Tableau. Additionally, I realized that, for this straightforward objective, the analysis could follow the visualization. I'll provide more details on in the next section.

This realization was somewhat frustrating, considering the considerable amount of time I had already invested in using MySQL during the exploratory phase.

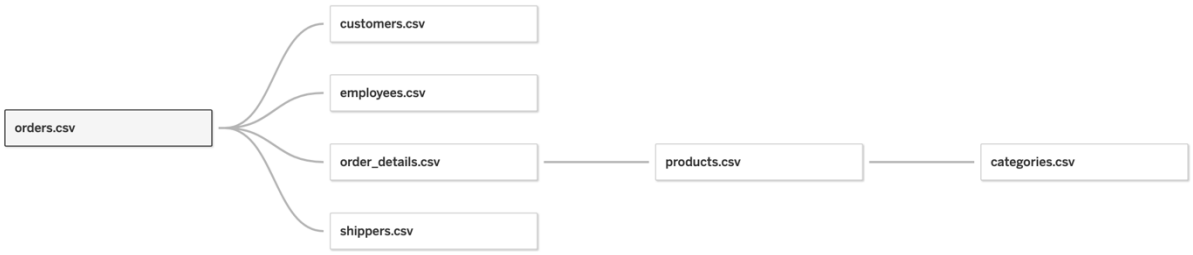
From this reflection exercise, one significant lesson I've learned as a Data Analyst is the importance of discerning when a particular tool is relevant for a given task and when it's not.

DATA VISUALIZATION AND ANALYSIS:

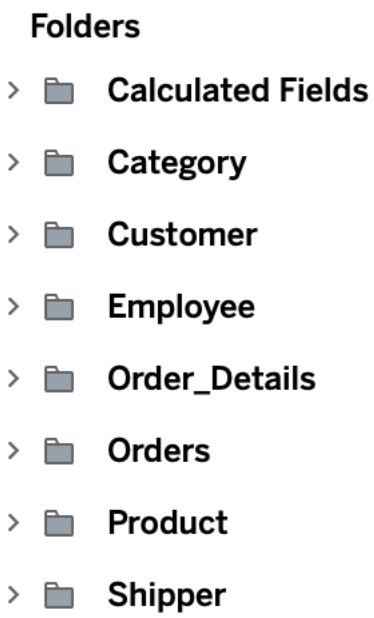
Our main goal is to understand how well Northwind Traders is doing in various aspects like sales, products, employees, and shipping costs. In a typical data project, we would first analyze the data before creating visualizations. However, in this case, I've chosen to make an interactive Tableau Dashboard first. This dashboard will allow anyone to quickly find answers to questions about these areas. For example, someone at Northwind Traders could easily explore how a specific product has been performing over a certain time period with a particular customer.

In my opinion, it is the most suited solution to meet this project's objective.

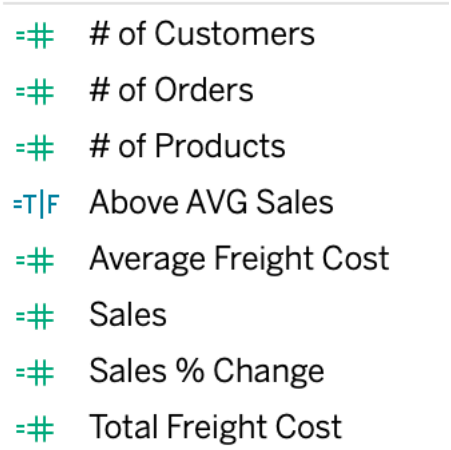
After making that decision, I proceeded to import the various CSV files into Tableau. Utilizing my prior understanding of the database schema for this dataset, I was able to efficiently establish connections between the different tables within the Data Source Menu of Tableau.



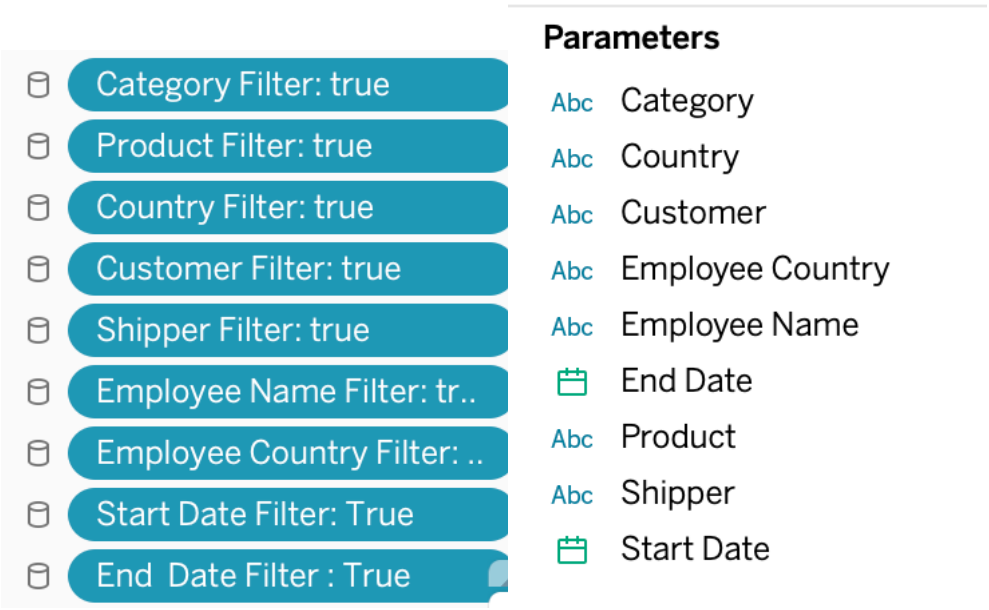
From there I organized the datasets and different measures by folder to have increased clarity on the software 'Data sidebar'.



I then created multiple ‘Calculated fields’ using the Tableau coding language which is similar to SQL. These fields are used in the different sheets.



Next, I began crafting individual Tableau sheets, each focused on a specific insight. While constructing these sheets, I also designed "Filters" linked to "Parameters" to enhance interactivity within the final dashboard. This approach ensures that users can dynamically explore the data and gain valuable insights.



Finally, I merged all of these sheets into one fully interactive dashboard you can find HERE. (Tip: Once on dashboard click on the bottom right button for full display)

We are now well-equipped to address the questions posed:

1. ***What are the notable trends in sales over time, and how have they evolved?***

Over the last three years, our sales have exhibited a consistent upward trend, marking an impressive 168% surge from January 2013 to April 2015. Notably, the first quarter of 2015 witnessed the most remarkable quarterly growth of the entire three-year period, boasting a substantial increase of 64.3%. Although there is a noticeable dip in sales for May 2015, it's important to note that the month's activity isn't yet concluded, rendering it premature to draw any significant conclusions at this juncture. Nonetheless, our sales data strongly indicates an expectation of sustained growth in the forthcoming months.

2. Which products have performed the best and worst in terms of sales?

In terms of product performance, Côte de Blaye from the Beverages Category stands out as the top performer, generating a remarkable $141,397 in sales over the three-year period. This impressive figure accounts for 11.17% of the total sales. Following closely in second place is Theringer Rostbratwurst from the Meat & Poultry Category, with sales totaling $80,369, representing 6.35% of the overall sales. Taking the third spot is Raclette Courdavault, which contributed $71,156 in sales, making up 5.62% of the total sales.

Conversely, the three lowest-performing products include Chocolade from the Confections category, which generated only $1,369, accounting for a mere 0.11% of total sales. Geitost from the Dairy Products category generated $1,648, representing 0.13% of the total sales. Lastly, Genen Shouyu from the Condiments category contributed $1,785 in sales, making up just 0.14% of the total sales.

It's noteworthy to mention that the second most-selling product has been discontinued, although specific details regarding the reasons for this discontinuation are currently unavailable.

3. ***Which product categories contribute the most and least to overall sales revenue?***

The top three categories contributing the most to the overall sales revenue are as follows:

1. Beverages: $267,868, representing 21.16% of total sales.
2. Dairy Products: $234,507, comprising 18.53% of total sales.
3. Confections: $167,357, accounting for 13.22% of total sales.

Conversely, the three categories with the least contribution to overall sales revenue are:

1. Grain & Cereals: $95,745, making up 7.56% of total sales.
2. Produce: $99,984, constituting 7.90% of total sales.
3. Condiments: $106,047, representing 8.38% of total sales.

This breakdown provides a clear view of which product categories are the major drivers of sales revenue and which have a relatively smaller impact on the overall sales picture.

4. ***Can we identify any specific customers who have made substantial purchases or exhibit unique buying behavior?***

Among our customers, QUICK-Stop from Germany emerges as the top contributor to our sales, providing a significant total of $110,277. Notably, Côte

de Blaye appears to be a particular favorite among their purchases, accounting for an impressive 20.8% of their total acquisitions.

Taking the second spot on our list is Ernest Handel from Austria, who has contributed $104,875 to our sales. Their purchasing pattern predominantly leans towards Dairy products, which make up a substantial 23.4% of their total purchases.

In the third position, we find Save-a-lot Markets from the USA, making a substantial contribution of $104,362 to our sales figures. Within their purchases, the Meat & Poultry category holds significant sway, constituting 26.5% of their total sales. Notably, Theringer Rostbratwurst stands out within this category, representing 11.54% of their total sales on its own.

5. ***Are shipping costs consistent among different shipping providers, and if not, what variations exist?***

When comparing shipping costs among different providers, we notice some variations. United Package appears to have the highest average freight cost at $86.64, but it's important to consider that they have shipped the largest number of orders, totaling 326. Federal Shipping and Speedy Express, on the other hand, have shipped a similar number of orders, with 255 and 249 respectively. However, Federal Shipping has a higher average freight cost, averaging $80.44 per order, while Speedy Express is more cost-effective, with an average cost of $65 per order.

Furthermore, when examining shipping costs by category, United Package consistently charges the highest on average across all categories. Federal Shipping follows closely behind, with relatively higher costs as well. In contrast, Speedy Express tends to be the most economical option overall.

These variations in shipping costs among providers highlight the importance of carefully considering the choice of shipping partner depending on factors like order volume, budget constraints, customer location and delivery requirements.

6. ***Which employees have demonstrated superior and inferior performance in sales?***

The USA team has outperformed the UK team in terms of both the number of orders and total sales. Specifically:

- The USA team completed a total of 606 orders, contributing a substantial $794,349 to total sales, which represents 62.76% of the overall sales.

- In contrast, the UK team managed 224 orders, resulting in $471,444 in sales, constituting 37.24% of the total sales.

This comparison clearly highlights the superior performance of the USA team in terms of both order volume and total sales revenue. Now, we can further analyze individual employee performance within these teams:

USA Team:

- Margaret Peacock leads the pack with an impressive 156 orders and $232,891 in sales.
- Janet Leverling follows closely in second place, having secured 127 orders and $202,813 in sales.
- Nancy Davolio comes in third place, completing 123 orders and $192,108 in sales.
- Andrew Fuller occupies the fourth place with 96 orders and $166,538 in sales.
- Laura Callahan, although managing the team, had 104 orders but only achieved $126,862 in sales, she still holds sixth place overall.

UK Team:

- Robert King emerges as the top performer in the UK Team and fith overall with 72 orders and $124,568 in sales.
- Anne Dod is seventh on the list with 43 orders and $77,308 in sales.
- Michael Suyama, with a respectable total of 67 orders, has generated $73,913 in sales, placing him in the eighth position among the sales team members.
- Steven Buchanan, unfortunately, performed less optimally within the UK Team as well as overall, with 42 orders and $68,792 in sales.

It's noteworthy that Laura Callahan and Steven Buchanan, despite being team managers, have the lowest sales figures within their branches. This might be attributed to their managerial responsibilities, which could potentially divert their focus away from individual sales efforts.

FINAL REPORT & RECOMMENDATIONS

This report provides insights into the overall performance of Northwind Traders, analyzing various aspects such as sales trends over time, product performance, customer contributions, shipping costs, and employee sales performance.

Over the last three years, sales have shown consistent growth, with an impressive 168% increase from January 2013 to April 2015. The first quarter of 2015 marked the highest quarterly growth, with a 64.3% increase. While there was a slight dip in May 2015, it's too early to draw conclusions. Overall, sales data indicates expectations of continued growth.

The best-performing products include Côte de Blaye from the Beverages Category, Theringer Rostbratwurst from the Meat & Poultry Category, and Raclette Courdavault. In contrast, the lowest-performing products are Chocolade, Geitost, and Genen Shouyu.

Beverages, Dairy Products, and Confections are the top categories contributing the most to overall sales revenue. On the other hand, Grain & Cereals, Produce, and Condiments contribute the least.

Notable customer contributions come from QUICK-Stop, Ernest Handel, and Save-a-lot Markets, with each exhibiting unique buying behaviors and preferences.

Shipping costs vary among different providers. United Package has the highest average freight cost, followed by Federal Shipping, while Speedy Express is the

most economical choice. The choice of provider should consider factors like order volume and budget constraints.

The USA team outperforms the UK team in terms of both order volume and total sales revenue. Among individual team members, Margaret Peacock, Janet Leverling, and Nancy Davolio lead the USA team. Robert King stands out in the UK team. Notably, Laura Callahan and Steven Buchanan, who are also team managers, have lower sales figures, possibly due to managerial responsibilities.

Northwind Traders has seen consistent sales growth over the past three years, with certain products and product categories performing exceptionally well. Customer contributions and shipping costs also play significant roles in sales dynamics. Employee sales performance varies, with team managers demonstrating lower individual sales, likely due to their managerial roles.

Based on the information provided in the report, several recommendations can be made to optimize the company's performance:

- Product Focus:

Given the strong performance of products like Côte de Blaye and Theringer Rostbratwurst, consider investing more resources in marketing and promoting these high-performing products to boost sales further.

- Product Diversification:

While certain product categories contribute significantly to sales, explore opportunities to diversify our product offerings to attract a broader customer base and reduce reliance on specific categories.

- Customer Relationship Management:

Nurture relationships with top-contributing customers like QUICK-Stop, Ernest Handel, and Save-a-lot Markets. Personalize offers and promotions to cater to their preferences and incentivize repeat business.

- Shipping Cost Optimization:

Analyze shipping costs further to identify opportunities for cost reduction. Negotiate with shipping providers for better rates, especially for high-volume shipping.

- Employee Training and Support:

Provide additional training and support to team managers like Laura Callahan and Steven Buchanan, as their dual roles may affect their individual sales performance. Ensure that managerial responsibilities do not overshadow their sales efforts.

- Market Expansion:

Explore opportunities for expanding into new geographic markets to tap into untapped customer bases and increase overall sales.

- Product Lifecycle Management:

Regularly assess product performance and discontinue underperforming products while introducing new, innovative items to keep the product catalog fresh and appealing to customers.

- Customer Segmentation:

Segment customers based on their preferences and purchase behavior to tailor marketing and promotional efforts more effectively.

- Market Research:

Conduct market research to stay updated on industry trends, customer preferences, and competitive landscape to make informed decisions.

- Performance Monitoring:

Continuously monitor sales performance, track key performance indicators (KPIs), and adjust strategies as needed to adapt to changing market dynamics.

Implementing these recommendations can help Northwind Traders build on its strengths, mitigate weaknesses, and seize new opportunities for growth and profitability. Regularly assessing and adjusting strategies based on data and market insights will be essential in maintaining a competitive edge in the industry.

POST-PROJECT REFLECTION AND LESSONS LEARNED:

Upon completing this project, several key observations and reflections have come to light, shedding light on areas where improvements can be made and lessons that can be applied to future endeavors.

- Data Cleaning Challenges:

During the data cleaning phase, I encountered certain issues, particularly with handling special characters. It became evident that a deeper understanding of data cleaning techniques, especially when dealing with non-standard characters, would be beneficial. Moving forward, I recognize the need to enhance my skills in this area to streamline the cleaning process and ensure data quality.

- Tool Selection Proficiency:

In hindsight, I acknowledge the importance of rapidly identifying the most suitable tools for specific tasks within a project. In future projects, I aim to improve my ability to assess and select the most relevant tools efficiently, optimizing workflow and productivity.

- Data Analysis Challenges:

While I am content with the dashboard I created, it is worth noting that not all relevant information was readily available. To provide the most pertinent insights and answers to questions, I had to invest additional effort in refining the

data analysis process. This experience underscores the importance of thorough data analysis and the need to ensure that the information presented is the most relevant for addressing project objectives.

- Data Visualization Enhancement:

Additionally, I intend to refine my use of Tableau, exploring advanced features such as Viz in Tool Tip to make the most of available dashboard space. In this project, I faced the challenge of limited space and had to create multiple versions of each sheet to present comprehensive information. Enhancing visualization techniques will be a focus for future projects.

- Continuous Improvement:

Rather than investing additional time in this particular project, I have chosen to apply the lessons learned to future endeavors. This decision allows me to actively track and measure progress in my journey, comparing future projects with past ones. By consistently implementing improvements and best practices, I aim to deliver increasingly refined and effective project outcomes.

This post-project reflection highlights the commitment to ongoing learning and improvement, ensuring that each project becomes an opportunity to refine skills and enhance results. It underscores the value of practical experience and continuous growth in our data analysis and visualization efforts while emphasizing the importance of meticulous data analysis to ensure the relevance of the information presented.