

PARCH AND POSEY DATA ANALYSIS

PROJECT WORK



February 16, 2024

hertechtrailacademy

hertechtrailacademy

**TABLE OF CONTENT**

1. Introduction
2. Dataset Description
3. Analytical Questions
4. Methodology
5. Visualisation and Insights
6. Conclusion
7. **INTRODUCTION**

Navigating through the intricacies of Parch and Posey's business data, we embark on a comprehensive analysis. Our exploration begins with understanding the total revenue generated, examining its fluctuations across different quarters or months. We further delve into the distribution of average sales for standard quantities across various regions. Identifying the top 5 customers based on their total purchase amounts and listing customers with the second letter 'a' in their names add depth to our analysis. Calculating total sales and average order values by region provides insights, while exploring the frequency of customer orders adds a dynamic aspect. Additionally, we limit our focus to the top 10 customers with total revenue below $50,000. We then pinpoint the month with the highest sales and conclude by finding the minimum and maximum gloss\_qty ordered by each client, rounding out our exploration of the Parch and Posey dataset.

1. **DATASET DESCRIPTION**

The Parch and Posey dataset serves as the primary source of data for this project, providing insights into the sales and customer data of a fictional company. This dataset is publicly available, created for educational data analysis projects and to mimic real-world sales transactions and customer interactions.

The dataset consists of five tables-**accounts**,**orders**, **sales\_reps**, **region** and **web\_events** consisting of approximately 10,000 records and about 25 columns with each column detailing sales transactions and capturing different attributes of the data.

1. **ANALYTICAL QUESTIONS**

Below are the ten analytical questions and its corresponding queries for insights

1. What is the total revenue generated by Parch and Posey?

**SELECT sum(total\_amt\_usd) total\_revenue**

**FROM orders**

1. How does the total revenue vary over different quarters or months?

**SELECT DATE\_TRUNC('month', occurred\_at) AS month, SUM(total) AS total\_revenue**

**FROM orders**

**GROUP BY 1**

**ORDER BY 1**

1. What is the distribution of average sales of standard qty in different regions?

**SELECT r.name, AVG(standard\_qty) Average\_sales, SUM(standard\_qty) Total\_average\_sales**

**FROM orders o**

**JOIN accounts a**

**ON o.account\_id = a.id**

**JOIN sales\_reps s**

**ON a.sales\_rep\_id = s.id**

**JOIN region r**

**ON s.region\_id = r.id**

**GROUP BY r.name**

**ORDER BY Average\_sales desc**

1. Who are the top 5 customers based on their total purchase amount?

**SELECT a.name, o.account\_id, SUM(total\_amt\_usd) AS total\_purchase\_amount**

**FROM orders o**

**JOIN accounts a**

**ON a.id = o.account\_id**

**GROUP BY a.name, o.account\_id**

**ORDER BY total\_purchase\_amount desc**

**LIMIT 5**

1. List of customers that have the second letter as ‘a’ in their name.

**SELECT name**

**FROM accounts**

**WHERE name LIKE '\_a%'**

**ORDER BY name desc**

**LIMIT 10**

1. Calculate the total sales and average order value by region of the Parch and Posey dataset.

**SELECT r.name, AVG(total\_amt\_usd) Average\_order\_value, SUM(total\_amt\_usd) Total\_order\_value**

**FROM orders o**

**JOIN accounts a**

**ON o.account\_id = a.id**

**JOIN sales\_reps s**

**ON a.sales\_rep\_id = s.id**

**JOIN region r**

**ON s.region\_id = r.id**

**GROUP BY r.name**

**ORDER BY Average\_order\_value asc**

1. How frequently do customers place orders.

**SELECT a.name, COUNT(\*) AS order\_count**

**FROM orders o**

**JOIN accounts a**

**ON a.id = o.account\_id**

**GROUP BY a.name**

**ORDER BY order\_count desc**

**LIMIT 10**

1. With a limit of 10, list the customers with a total revenue less than $50,000.

**SELECT o.account\_id, a.name, total\_amt\_usd**

**FROM orders o**

**JOIN accounts a**

**On a.id = o.account\_id**

**WHERE total\_amt\_usd < 50000**

**ORDER BY total\_amt\_usd desc**

**LIMIT 10**

1. Which month has the highest sales in the Parch and Posey dataset.

**SELECT DATE\_PART ('month',occurred\_at) AS month, SUM(total\_amt\_usd) AS highest\_sale**

**FROM orders**

**GROUP BY 1**

**ORDER BY 2 desc**

1. Find the minimum and maximum number of gloss\_qty ordered by each client.

**SELECT a.name, o.account\_id, MIN(gloss\_qty) MINIMUM, MAX(gloss\_qty) MAXIMUM**

**FROM orders o**

**JOIN accounts a**

**ON a.id = o.account\_id**

**GROUP BY a.name,o.account\_id**

**ORDER BY account\_id desc**

**4. METHODOLOGY**

The approach used in this analysis primarily involved querying the dataset using SQL(Postgres) to extract relevant information based on predefined analytical questions as identified above. The SQL queries were used to aggregate, filter and manipulate the data to generate the insights. Examples of data manipulation techniques used include JOINs, GROUP BY and aggregate functions were employed to transform and aggregate the data for analysis.

For visualisation, Tableau software was employed to create it. Making use of different charts and graphs, these were utilised to present the findings visually incorporating filters, parameters, reference line, average line, constant line and more. Finally, some selected charts came together to create a dashboard.

**5. VISUALISATION AND INSIGHTS**

Question 2 of our analytics compares the total revenue over the months where February recorded the lowest revenue sales with an amount of $1.31m and the highest revenue recorded in December for $3.13m. In all, we can say the company had a sturdy rise in its revenue over the year.

In question 3, we find the average sales of standard\_qty amongst the four regions as 484,586, with midwest and west regions falling below the line. Northeast and Southeast recorded the highest sales of 647,000 and 566,000 respectively. However, in each region the average sales is the same as the sales recorded.

In question 4, we identified the top customers based on their total purchase amount and identified the following customers as leading the chart with EOG resources amassing an amount of $382,873, followed by Mosaic with $345,619 and lastly Republic Services with $293,861 amongst the top five. The chart is interactive, making use of parameters and filters to engage users who would like to view more customers within the top ten performing.

In question 6, we seek to identify the average order value and total sales by region. Averagely, the least orders received by a region is 3,191 from the Southeast and 3,626 being the highest orders from the West region. In terms of total sales, Northeast came on top with total sales of 7,744,405 followed by Southeast with 6,458,497.

With question 7, we identify the frequencies by which each customer places an order. Using a histogram chart, we limit our results to the top 10 orders allowing the user the opportunity to switch view within the specified number of customer orders, and the highest being 71 orders by Leucadia National.

Furthermore, question 8 lists the customers with total revenue less than $50,000. Between the years 2013 to 2017 like Bank of New York Mellon Corporation, Nike, Delta Airlines among others have a revenue less than $50,000. Amongst the least 10, the average total was $1,285 with an option to view other similar customers using the parameter created.

Question 9, uses a line chart to identify the highest sales by month, December recording $3,129,412(highest) and January with $1,337,661 being the lowest.

Lastly, question 10 highlights the minimum and maximum gloss\_qty associated with each client with clients like CBRE Group, Peter Kiewit Sons’ and Publix Super Marketshaving a minimum order below 50 and clients like HCA holdings, Starbucks and Reliance Steel and Aluminium having maximum orders above 500.

Dashboard link - [here](https://public.tableau.com/views/ProjectWork_17079479783600/Dashboard2?:language=en-US&publish=yes&:sid=&:display_count=n&:origin=viz_share_link)

**6. CONCLUSION**

In conclusion, leveraging insights from the analysis of the Parch and Posey dataset can inform strategic decision-making, enhance operational efficiency, and drive sustainable business growth. Continual monitoring and adaptation to evolving market dynamics will be essential for maintaining competitiveness and achieving long-term success.