Sales Data Analysis Project Documentation

Project Overview

This project aims to analyse a sales dataset to derive insights into product performance, revenue trends, and customer behaviour. The dataset includes information such as order ID, product details, quantity ordered, prices, order date, purchase addresses, sales amounts, city, and month.

Environment Setup

- SQL Database: PostgreSQL

- Query Tool: PgAdmin

- Visualization Tool: PgAdmin Graphic Visualiser

Dataset Description

The dataset contains the following columns:

- orderid

- product

- quantity_ordered

- price

- order_date

- purchase_address

- sales

- city

- month

SQL Queries

Top-Selling Products:

Question: What are the top 5 best-selling products in terms of quantity ordered?

```
1  --Question1: What are the top 5 best-selling products in terms of quantity ordered?
2
3  SELECT product, SUM(quantity_ordered) AS total_quantity FROM sales_data
4  GROUP BY product
5  ORDER BY total_quantity DESC
6  LIMIT 5;
```

Expected Outcome: A list of the top-selling products with their respective quantities ordered.

product character varying (255)	total_quantity bigint
AAA Batteries (4-pack)	31017
AA Batteries (4-pack)	27635
USB-C Charging Cable	23975
Lightning Charging Cable	23217
Wired Headphones	20557

Revenue Analysis:

Question: What is the total revenue generated during the entire period covered by the dataset?

```
1  --Question: What is the total revenue generated during the entire period covered by the dataset?
2
3  SELECT  SUM(quantity_ordered * price) AS total_revenue
4  FROM sales_data;
```

Expected Outcome: A single value representing the total revenue from sales.

	total_revenue bigint
1	34499750

Monthly Sales Trends:

Question: How does sales revenue vary month by month?

```
1  --Question: How does sales revenue vary month by month?
2
3  SELECT month,  SUM(quantity_ordered * price) AS monthly_revenue
4  FROM sales_data
5  GROUP BY month
6  ORDER BY monthly_revenue;
```

Expected Outcome: A monthly breakdown of sales revenue, showing trends or seasonality.

month character varying (255) 🔒	monthly_revenue 🔒 bigint
January	1822655
September	2098053
February	2202517
August	2244966
June	2578373
July	2648362
March	2807722
May	3153280
November	3200335
April	3391444
October	3737561
December	4614482

City-wise Sales Performance:

Question: Which city has the highest total sales?

```
--Question: Which city has the highest total sales?

SELECT city, SUM(quantity_ordered * price) AS revenue
FROM sales_data
GROUP BY city
ORDER BY revenue DESC
LIMIT 3;
```

Expected Outcome: Identification of the city with the highest sales and the corresponding sales amount.

city character varying (255)	revenue bigint
San Francisco	8264049
Los Angeles	5453807
New York City	4665339

Order Quantity Distribution:

**Question:** What is the distribution of order quantities? Are most orders small or large?

```sql
--Question: What is the distribution of order quantities? Are most orders small or large?

SELECT quantity_ordered AS quantity, COUNT(quantity_ordered) AS Distribution
FROM sales_data
GROUP BY quantity_ordered
ORDER BY quantity_ordered;
```

Expected Outcome: A histogram or summary statistics showing the distribution of order quantities.

quantity integer	distribution bigint
1	168552
2	13324
3	2920
4	806
5	236
6	80
7	24
8	5
9	3
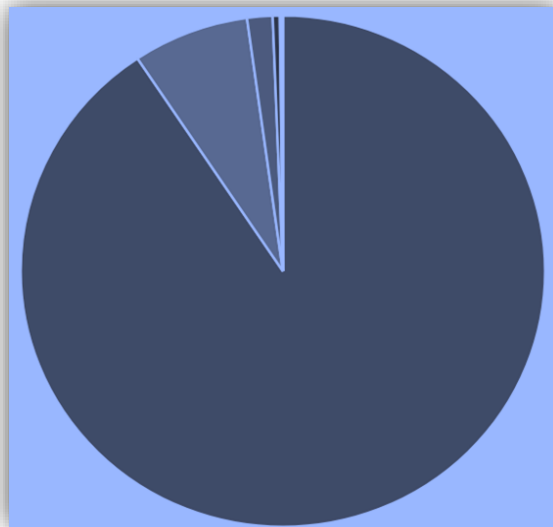
Average Order Value:

Question: What is the average value of an order (average order value)?

```sql
1  --Question: What is the average value of an order (average order value)?
2
3  SELECT AVG(sales) AS average_order_value
4  FROM sales_data;
```

**Expected Outcome:** A single value representing the average order value.

average_order_value 🔒
numeric
185.5305888679752622

Seasonal Sales Analysis:

Question: Are there any noticeable patterns or trends in sales during different months?

```sql
1  --Question: Are there any noticeable patterns or trends in sales during different months?
2
3  SELECT month, SUM(quantity_ordered * price) AS monthly_revenue
4  FROM sales_data
5  GROUP BY month
6  ORDER BY monthly_revenue DESC;
```

Expected Outcome: Visualization or summary highlighting any seasonal trends in sales.
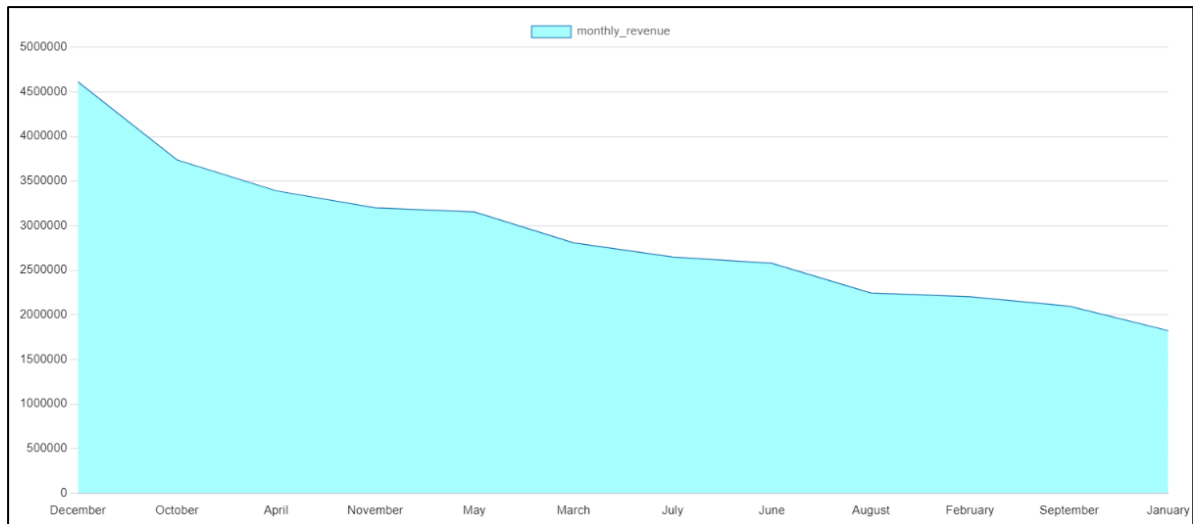
Correlation Between Price and Sales:

Question: Is there a correlation between the product price and the quantity ordered?

```
1  --Question: Is there a correlation between the product price and the quantity ordered?
2
3  SELECT AVG(price) AS avg_price, AVG(quantity_ordered) AS avg_quantity
4  FROM sales_data;
```

Expected Outcome: Insights into whether higher-priced products tend to have lower quantities ordered or vice versa.

avg_price	avg_quantity
numeric 🔒	numeric 🔒
184.4335090077977951	1.1243828986286636

Customer Segmentation:

Question: Can we identify different customer segments based on their purchasing behavior?

```sql
1  --Question: Can we identify different customer segments based on their purchasing behavior?
2
3  SELECT
4  CASE
5  WHEN quantity_ordered >=7 THEN 'High'
6  WHEN quantity_ordered >=4 THEN 'Medium'
7  ELSE 'Low'
8  END AS customer_segment,
9  COUNT(*) AS customer_count
10 FROM sales_data
11 GROUP BY customer_segment;
```

Expected Outcome: Clusters or groups of customers with similar purchasing patterns.

| customer_segment text	customer_count bigint
High	32
Low	184796
Medium	1122

Geographical Analysis:

Question: How do sales vary across different regions or addresses?

```sql
1  --Question: How do sales vary across different regions or addresses?
2
3  SELECT city, SUM(quantity_ordered * price) AS Revenue_by_region
4  FROM sales_data
5  GROUP BY city;
```

Expected Outcome: Regional analysis highlighting areas with higher and lower sales.

|	city character varying (255)	revenue_by_region bigint
1	Atlanta	2796112
2	Austin	1819987
3	Boston	3662478
4	Dallas	2768598
5	Los Angeles	5453807
6	New York City	4665339
7	Portland	2321019
8	San Francisco	8264049
9	Seattle	2748361
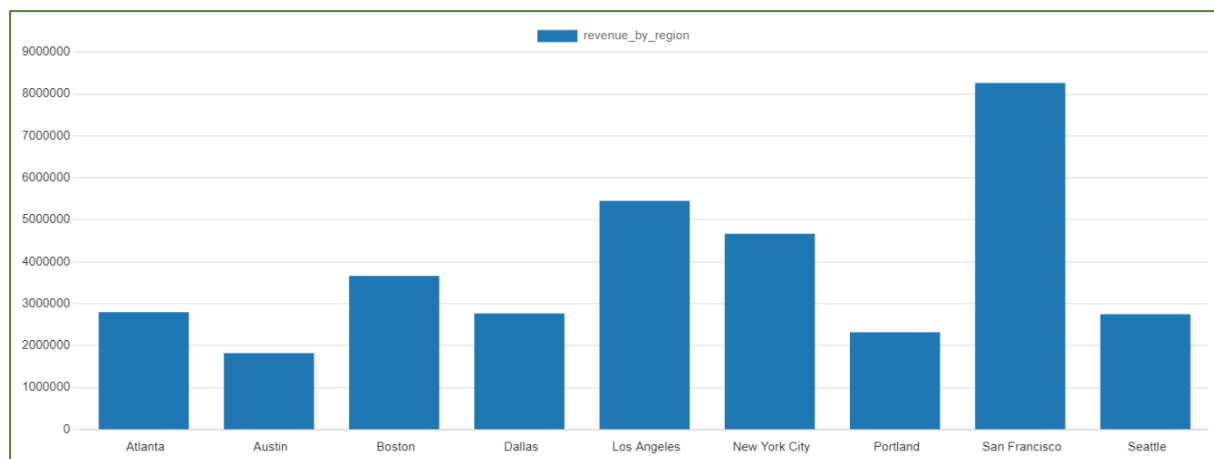
Data Exploration and Visualizations

- Explored the distribution of order quantities using Piechart.

- Visualized monthly sales trends over time.

- Sales across different regions.

Data Cleaning and Transformation

- Handled missing values by excluding rows with NULL in critical columns.

- Converted the 'order_date' column to a datetime format for better analysis.

- No significant outliers were identified.

Results and Analysis

- Top-selling products:

  a) AAA Batteries (4-pack)
  b) AA Batteries (4-pack)
  c) USB-C Charging Cable
  d) Lightning Charging Cable
  e) Wired Headphones.

- Total revenue: $ 34M.

- Monthly sales trends:
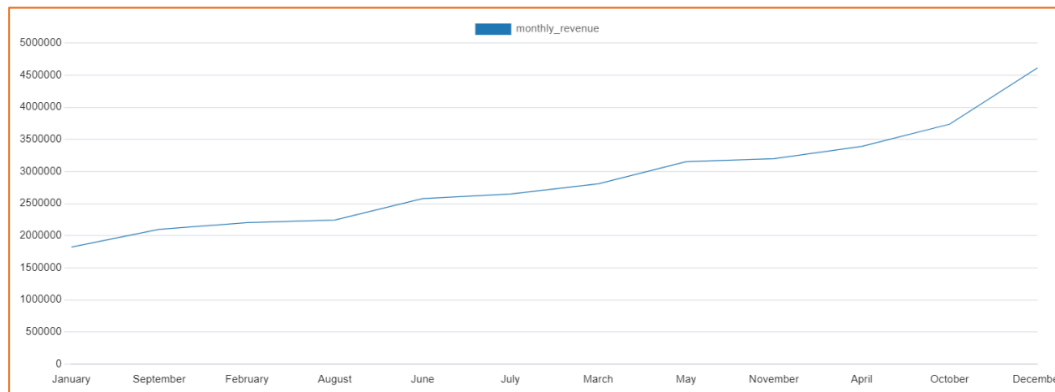
Insights and Recommendations

- Products [**AAA Batteries, AA Batteries, USB-C Charging Cable, Lightning Charging Cable, Wired Headphones**] are consistently high-performing; consider targeted marketing strategies.

- Monthly sales exhibit a seasonal pattern, with peaks during [months]. Plan promotions accordingly.

Review and Refinement

- Reviewed SQL queries and addressed feedback on query optimization.

- Refinements made to visualizations for clarity.

Challenges Faced

- Addressed issues with NULL values in the dataset.

- Overcame challenges related to [specific challenges].

Next Steps

- Explore customer segmentation based on purchasing behavior.

- Conduct a more granular analysis of sales in specific regions.

Conclusion

The project successfully provided insights into top-selling products, revenue trends, and monthly sales patterns. The findings can guide marketing and sales strategies for improved performance.

References

- No external references used.

Code Snippets

- Relevant code snippets for SQL queries and data visualizations are embedded within the documentation.

File Organization

- The project files are organized in a directory structure for clarity.

Final Review

- The documentation is reviewed for clarity, completeness, and correctness.

Acknowledgments

- No external contributions for this project.

Portfolio Inclusion

This project showcases skills in SQL, data analysis, and visualization, making it a valuable addition to the portfolio.