# OLAJIDE ABDULLATEEF

## PROJECT REPORT

## (The Analytics Hackathon 2025)


**November 2025**

# Report Outline

## Project Overview

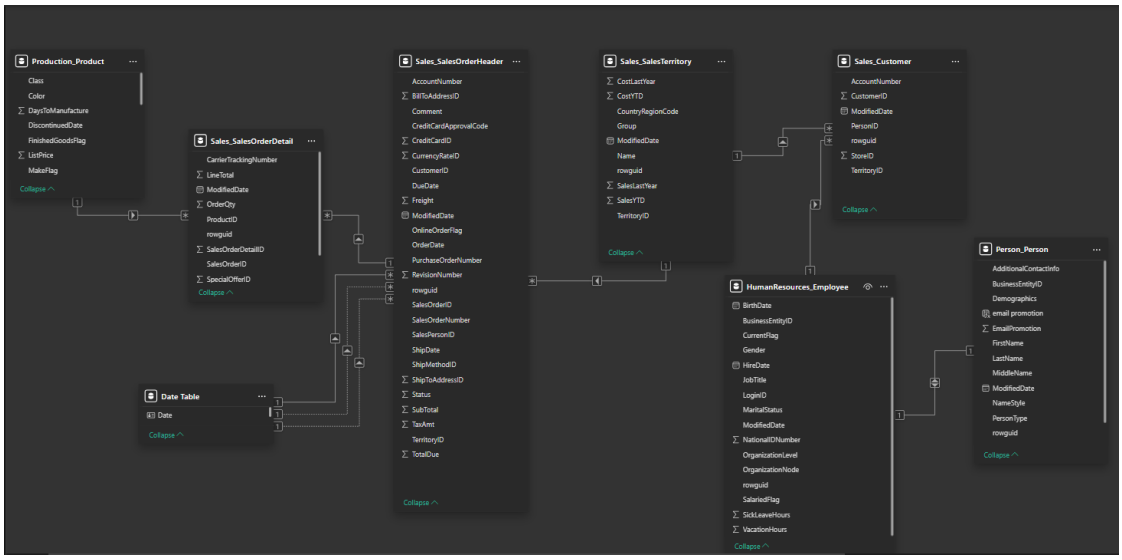
AdventureWorks is a global manufacturing company that sells products through online and reseller channels. Management needs a clear, evidence-backed understanding of business performance, customer behaviour, and product trends to inform pricing, inventory, marketing, and territory strategies.

As an Analyst at AdventureWorks**,** the objective is to explore and analyse the AdventureWorks dataset, produce five high-quality, business-relevant insights supported by queries and visualizations, and document the full analysis process, and also deliver a presentation-ready report.

## Methodology

- **Data Collection/Source:**
  AdventureWorks2019 database was downloaded from the internet and restored in Microsoft SQL Server. Connection was secured to the database and tables are queried with SQL. For visualization, given the limitations of Direct Query, data was imported into Power BI by first writing a Pandas script to connect to the Database to read the tables into a local computer and then imported manually into Power BI to create a semantic model.

- **Semantic/Data Model Design:**
  The model is designed using a star schema, with **SalesOrderHeader** table acting as the **Fact Table,** and others acting, the dimension tables. A date table was created to connect to the date columns on the Fact Table, with only one active relationship to the OrderDate column.



## Explanation of Tables

- **Sales.SalesOrderHeader:** This table contains a unique record of all sales. Each SalesOrderID is assigned a unique SalesOrderNumber for each sales record. It also contains OrderDate, DueDate and ShipDate for each Sales and the respective CustomerID (distinct count for CustomerID column shows that some customers placed more than one Order). The TotalDue column is the total amount payable by customer for SalesOrder (which is an addition of SubTotal, TaxAmount and Freight charge).

  Sales.SalesOrderHeader.CustomerID = Sales.Customer.CustomerID

- **Sales.SalesOrderDetail:** Table contains SalesOrderID, productID, quantity of product ordered per product, and their unit price. The LineTotal column shows the total amount paid for all the quantity ordered for each Product.

  Sales.SalesOrderDetail.ProductID = Production.Product.ProductID

- **Production.Product:** This Table identifies 504 different products in production so far, by their Name and ProductID. It also shows the date sales began and ended for each of these products as the SellStartDate and SellEndDate.

  Production.Product.ProductID = Sales.SalesOrderDetail.ProductID

- **Sales.Customer:** Unique customer record showing the ID of all customers being shipped to. Table also contains customers respective Store and TerritoryID.

  Sales.Customer.PersonID = Person.Person.BusinessEntityID

- **Sales.SalesTerritory**: The table identifies all the ten territories being shipped to, the country, country region, continent, SalesYTD (which is the total sales revenue generated by each territory from the start of sales to the current date), SalesLastYear (which is the total sales revenue generated last year by the territory).

  Sales.SalesTerritory.TerritoryID = Sales.Customer.TerritoryID

- **Person.Person**: This table stores information about individuals the manufacturing company touches, including employees, customers and vendors. (i.e It gives description of Human beings involved with the company; Employees, customer contacts and vendor contacts).

  **PersonType** defines six distinct person type;

  SC = Store Contact
  IN = Individual (retail) customer
  SP = Sales Person
  EM = Employee (non-sales)
  VC = Vendor Contact
  GC = General Contact

  This categorization helps the company track and manage its relationship with different types of individuals.

The Demographics column stores personal information of persons such as hobbies and income collected from online shoppers.

Person.Person.BusinessEntityID = Sales.Customer.PersonID

- **HumanResources.Employee:** Keeps record of the company's employee information such as JobTitle, OrganizationNode (where the employee is located in the corporate hierarchy), OrganizationLevel (the depth of the employee in the corporate hierarchy), HireDate, NIN, Birthdate

  HumanResources.Employee.BusinessEntityID = Person.Person.BusinessEntityID

## Data Cleaning and Preparation

Data Cleaning was done in powerBI Powerquery. Below are all the transformations carried out for each table.

**Sales_SalesOrderHeader:**

- First row promoted to header.
- Null values for SalesPersonID were replaced with "unknown".
- ShipMethodID contained two distinct values, 1 and 5. These these shipping methods are not clear, they were replaced to "method1" and "method5", and data type was changed to text.

**Production_Product:**

- First row promoted to header.
- Removed "rowguid" and "modifieddate" columns as they are irrelevant to analysis.
- FinishedGoodsFlag column identifies whether a product is sellable or not, but stores its data as True and False (Boolean), visualizing this will add more cognitive load to the report viewer, True and False values were switched to "sellable and "not-sellable".
- MakeFlag column describes in Boolean (true or false) whether the product was manufactured in-house or purchased. To reduce cognitive load, the values were switched to "Purchased" (False) and "Manufactured" (True).
- Columns **ProductLine**, **Class** contained Null values, replaced with "unknown".
- Data in Class column were previously passed incomprehensibly as H, L, M. These were switched to more appropriate expressive "High", "Low", "Medium". Likewise, in ProductLine where types were expressed as R (was replaced to **Road**), M (**Mountain**), T (**Touring**) and S (**Standard**).

**Sales_SalesTerritory**

- First row promoted to header.
- Removed unimportant columns eg, "rowguid" and "modifieddate"

- **CountryRegionCode** contains all the countries being shipped to. But expressed in short form (US, CA, FR, DE, AU, GB), these were switched to more expressive country full name.

**HumanResources_Employee:**

- First row promoted to header.
- Removed "rowguid" and "modifieddate" columns, unimportant to analysis.
- Wrote a measure to convert SickLeaveHours and VacationHours from hours to days.
- Gender was expressed as M and F in **gender** column. These were switched to Male and Female for better expression, likewise in the MaritalStatus Column where S and M were switched to "Single" and "Married".
- SalariedFlag contained Boolean values, "0" to identify workers being paid hourly and "1" to identify workers who earn a monthly income. For better expression on the visual chart, these were replaced with "Hourly" and "salaried". This column likely identifies contractors and differentiates them from the company staffs.

**Person_Person:**

- First row promoted to header.
- Removed unimportant columns ("NameStyle", "Suffix", "ModifiedDate", and "rowguid").
- In **PersonType** column, "EM" was switched to "Employee", GC – **General Contact**, IN - **Individual (retail) customer**, SP - **Sales person**, VC - **Vendor contact**, SC - **Store contact**.

**Sales_Customer:**

- First row promoted to header.
- Removed "modifieddate" and "rowguid" columns as they are not important for analysis.

**Sales_SalesOrderDetail:**

- First row promoted to header.
- Removed "modifieddate" and "rowguid" columns as they are not important for analysis.

## Exploratory Data Analysis/Key Findings

**Sales Overview Page:**

Adventure Works generated $110 million in sales over a four-year period, receiving an estimated 31,470 customer orders. The company serves approximately 20,000 customers across 10 territories, 6 countries, and 3 continents, including 702 customer stores through which orders are placed.

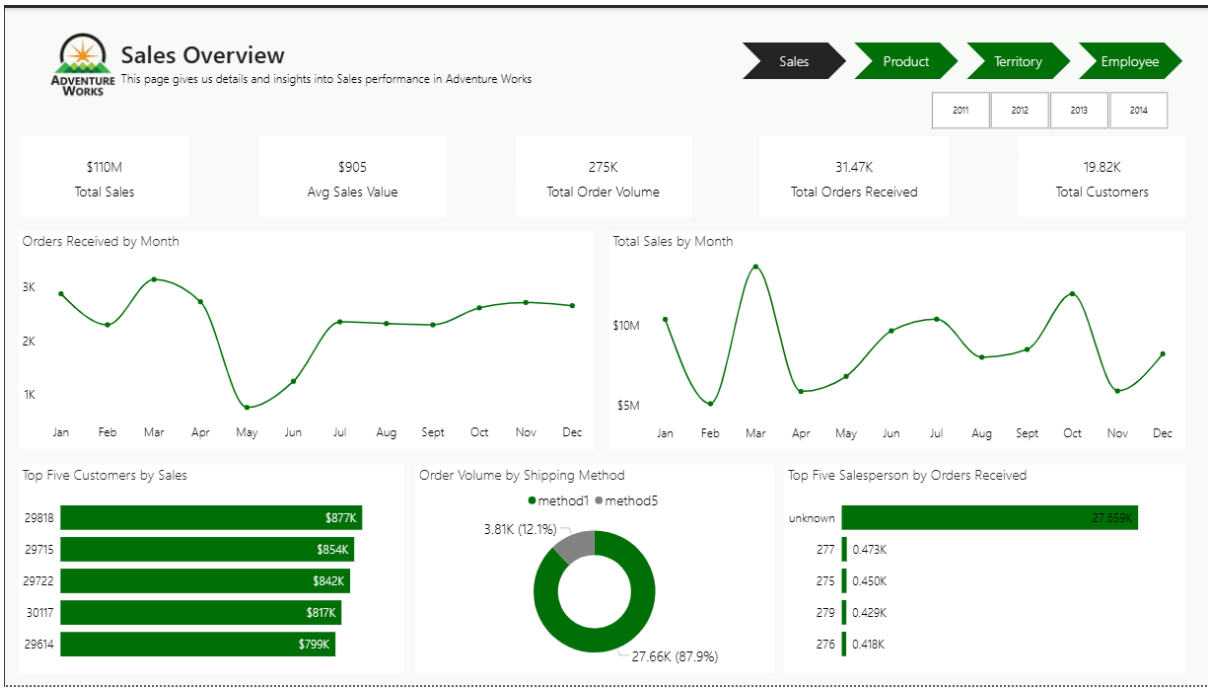
Customers spend an average of $905 per order, typically purchasing 2 items per transaction.

March stands out as the peak month, recording 3,144 orders and driving revenue to $13.6 million, marking it as the highest-earning month. The opposite trend is observed in May, which recorded the lowest activity.

The Top 5 Salespersons by Orders highlights a large discrepancy: the "Unknown" salesperson category—used to replace missing values in the *SalesPersonID* column—received approximately 28,000 orders, accounting for 87.89% of all orders. This dominant share indicates a significant data quality issue.

Shipping analysis shows that Method1 is overwhelmingly preferred, handling 87.9% of orders, compared to Method5, which accounts for only 12.1%.

The Top 10 Customers by Sales show minimal differences in spending levels among them. Together, they contribute only 7.2% of total revenue, suggesting that revenue is well distributed across a broad customer base rather than concentrated among a few.




**Product Overview Page:**

Adventure Works manages 504 products in its inventory. Of these, 47% are manufactured in-house, while 53% are purchased for resale.
Insights further show that 59% of the products are sellable, with the remaining 41% classified as non-sellable.

On average, it takes one day and $259 to manufacture a product, which is then sold at an average price of $439.

**Top Five Products by Orders**

1. Water Bottle – 4,700
2. AWC Logo Cap – 3,400
3. Patch Kit – 3,400
4. Mountain Tire Tube – 3,100
5. Sport Helmet (Blue) – 3,100

Together, these products make up **56% of total orders**, making them high-priority inventory items. They also rank as the top five in overall order quantities.

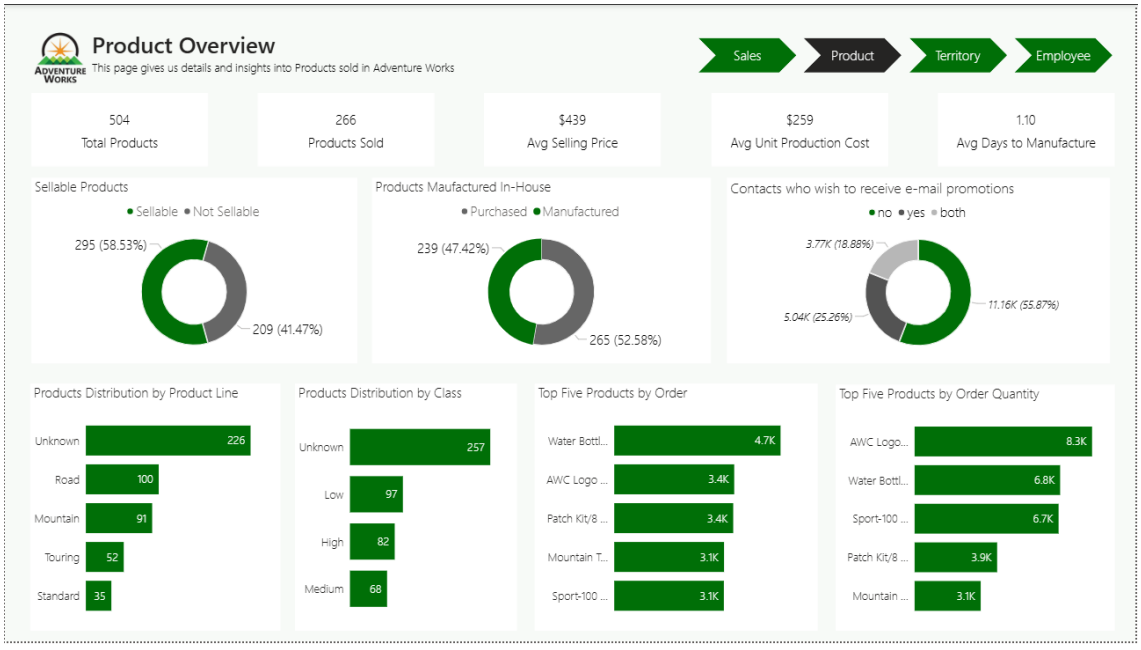
Customer marketing preferences show that only **25%** have opted in for **email promotions**, **56%** **declined**, and **19%** want promotional emails from both Adventure Works and its partners.

Adventure Works organizes its products into **five product lines**, which reflect the terrain or topography for which each product is designed. Their distribution is as follows:

1. Unknown – 224 (44%)
2. Road – 100 (20%)
3. Mountain – 91 (18%)
4. Touring – 52 (10%)
5. Standard – 35 (7%)

Product classification by class shows:

- Low: 97
- High: 82
- Medium: 68
- Unknown: 257




**Territory Overview Page:**

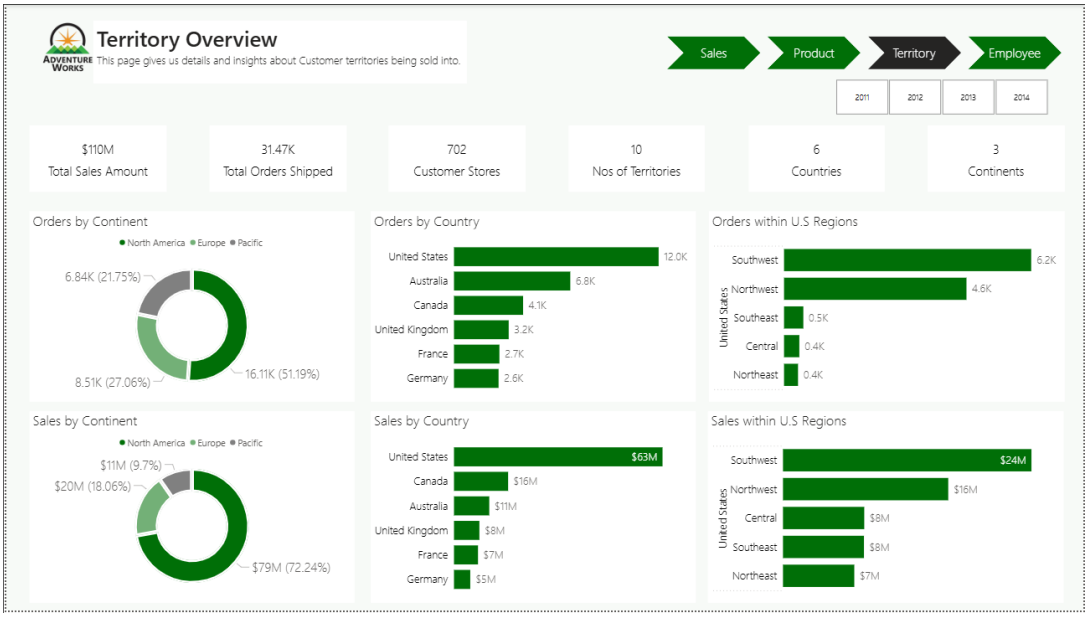
Adventure Works has generated $110 million in sales and shipped 31,470 orders across **10** territories, 6 countries, 3 continents, and 702 customer stores.

North America leads significantly, accounting for $79 million in sales and ~51% of all orders. Europe follows with $20 million in sales and 27% of orders, while the Pacific region ranks lowest

with $11 million in sales and 22% of orders—highlighting North America's critical role in overall performance.

At the country level, the United States is the top performer in both Sales ($63M) and Orders (12,000). Germany records the lowest, with $5M in sales and 12,000 orders. The U.S. dominance is reinforced by its five contributing regions:

- Southwest – $24M
- Northwest – $16M
- Southeast – $8M
- Central – $8M
- Northeast – $7M




**Employee Overview Page:**

Adventure Works employs 290 workers, organized into five organizational levels, with the 5th level representing the CEO.

On average, employees take 45 hours of sick leave and 50 hours of vacation time.

Only 18% of the workforce is salary-based, while the rest are hourly workers, indicating a strategic approach to labour cost control. Hourly workers are likely contractors who do not receive company benefits, further supporting this cost-management strategy.

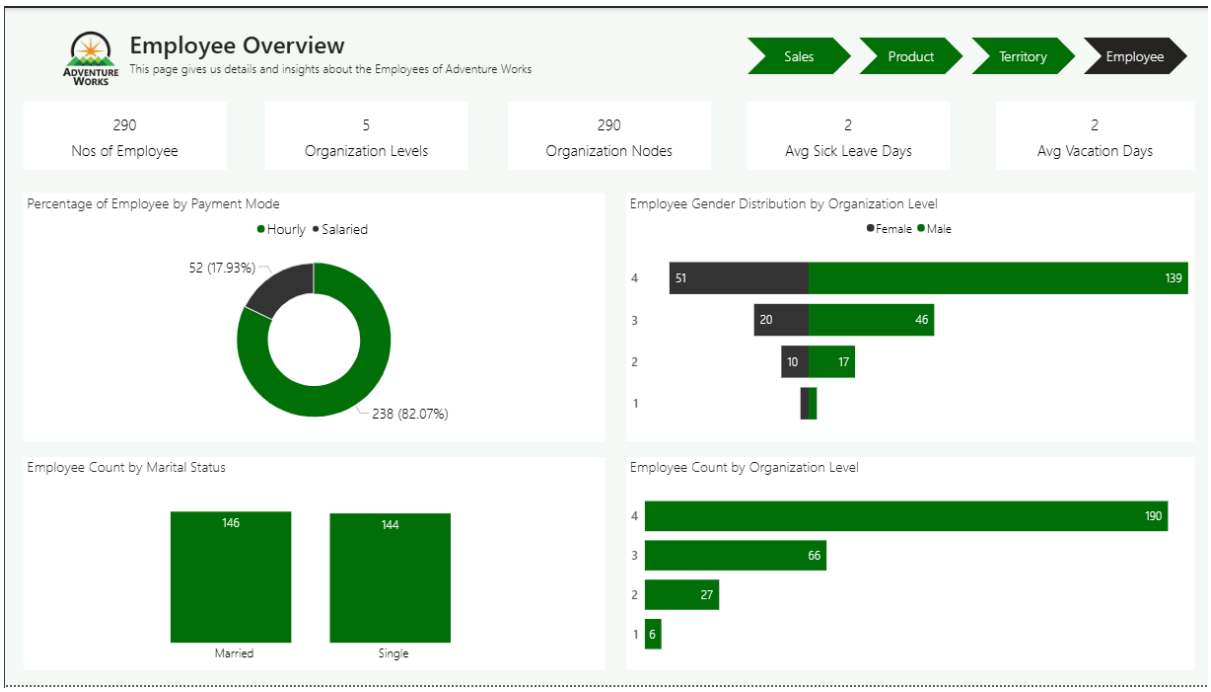
Employee marital status is nearly evenly split:

- Married: 146

- Unmarried: 144

Employee distribution across organizational levels shows a top-heavy structure, with more individuals occupying higher-level positions.

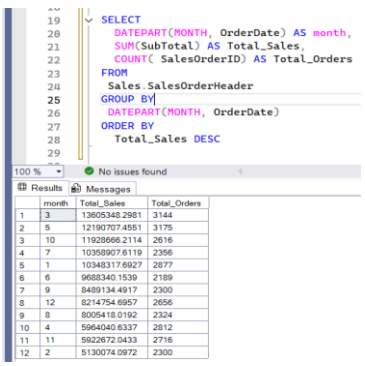
Gender analysis across these levels reveals that female representation falls below 37% at every level, pointing to possible gender imbalance or male dominance in key positions.

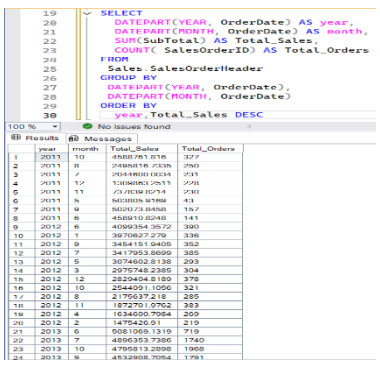


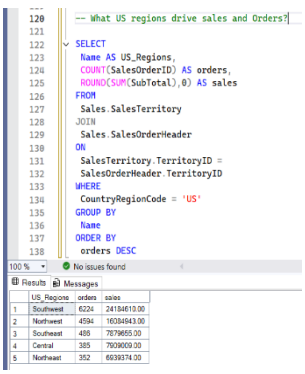

## Summary/Recommendations:

Below are the five high-quality, business-relevant insights

1. **March drives peak revenue and orders.**
   March is the month with highest order overtime, with 3,144 orders and ~$13.6M in total revenue. May is the trough.
   **Note:** This is based on cumulative value for all the period of operation, as each year has varying months for peak and through.
   **Recommendation:** Inventory and marketing should be scaled for peak month demand, while off-peak months should be supported with marketing campaigns and promotions.



2. **Revenue is broadly distributed and not concentrated to few customers.**
   Top 10 customers only contribute ~7.2% of total revenue, making revenue appear widely distributed amongst the business customers.
   **Recommendation:** Growth strategies should focus on scaling wide customer acquisition and retention rather than focusing on a handful of them.

```
25  --What are the top ten customers by revenue?
26  SELECT top 10
27   CustomerID,
28   Round(SUM(LineTotal),0) AS revenue
29  FROM
30   Sales.SalesOrderDetail
31  JOIN
32   Sales.SalesOrderHeader
33  ON
34   SalesOrderDetail.SalesOrderID =
35   SalesOrderHeader.SalesOrderID
36  GROUP BY
37   CustomerID
38  ORDER BY
39   revenue DESC;
```

100 % | 1 | 0

Results | Messages

| | CustomerID | revenue |
|---|---|---|
| 1 | 29818 | 877107.000000 |
| 2 | 29715 | 853849.000000 |
| 3 | 29722 | 841909.000000 |
| 4 | 30117 | 816756.000000 |
| 5 | 29614 | 799278.000000 |
| 6 | 29639 | 787773.000000 |
| 7 | 29701 | 746318.000000 |
| 8 | 29617 | 740986.000000 |
| 9 | 29994 | 730799.000000 |
| 10 | 29646 | 727273.000000 |

3. **Fix Data Quality Problem.**
   It is difficult to identify Salesperson with the highest sales record on account of data quality (missing records). An "unknown" Salesperson masks real sales performance, holds ~28,000 orders (~87.9% of orders).
   **Recommendation:** Sales performance by Salesperson is unreliable. Fixing null values will change top salesperson metric.

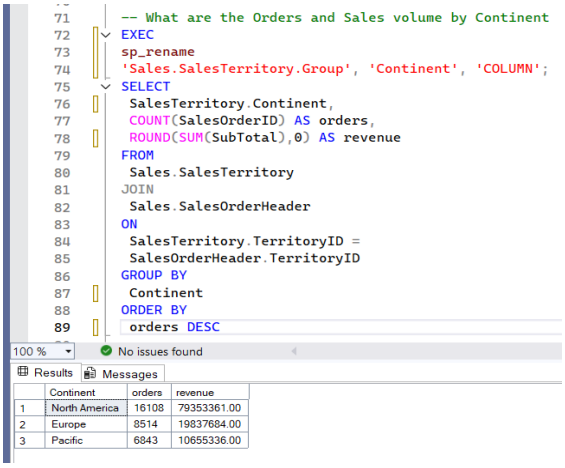
```
41  -- What are the top 5 salesperson by Orders Received
42  SELECT TOP 5
43   COALESCE
44   (CAST(SalesPersonID AS VARCHAR(25)),'unknown')
45      AS SalesPerson,
46   Count (SalesOrderID) AS Orders_Received
47  FROM
48   Sales.SalesOrderHeader
49  GROUP BY
50   SalesPersonID
51  ORDER BY
52   Orders_Received DESC
```

100 % | No issues found

Results | Messages

| | SalesPerson | Orders_Received |
|---|---|---|
| 1 | unknown | 27659 |
| 2 | 277 | 473 |
| 3 | 275 | 450 |
| 4 | 279 | 429 |
| 5 | 276 | 418 |

4. **North America dominates. US regions drive volume.**
With the entire North American continent $79M (and claiming 51% of total orders), while US contributes ~$63M and ~12,000 orders, the United States remains an important destination for the business. Southwest and Northwest are top regions.
**Recommendation:** Invest in North America (logistics, marketing etc), but examine and strategize towards lower-performing regions for possible growth potential.

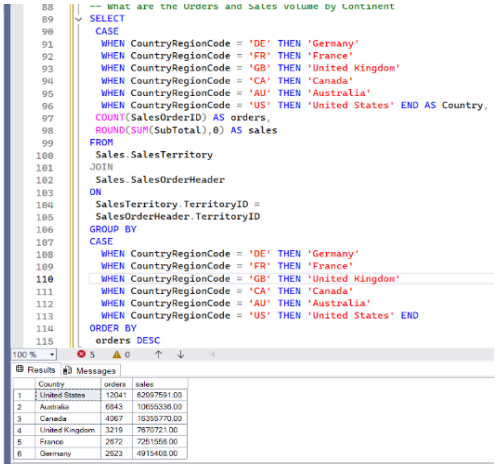


```
-- What are the Orders and Sales volume by Continent
EXEC
sp_rename
'Sales.SalesTerritory.Group', 'Continent', 'COLUMN';
SELECT
 SalesTerritory.Continent,
 COUNT(SalesOrderID) AS orders,
 ROUND(SUM(SubTotal),0) AS revenue
FROM
 Sales.SalesTerritory
JOIN
 Sales.SalesOrderHeader
ON
 SalesTerritory.TerritoryID =
 SalesOrderHeader.TerritoryID
GROUP BY
 Continent
ORDER BY
 orders DESC
```

No issues found

Results | Messages

| | Continent | orders | revenue |
|---|---|---|---|
| 1 | North America | 16108 | 79353361.00 |
| 2 | Europe | 8514 | 19837684.00 |
| 3 | Pacific | 6843 | 10655336.00 |

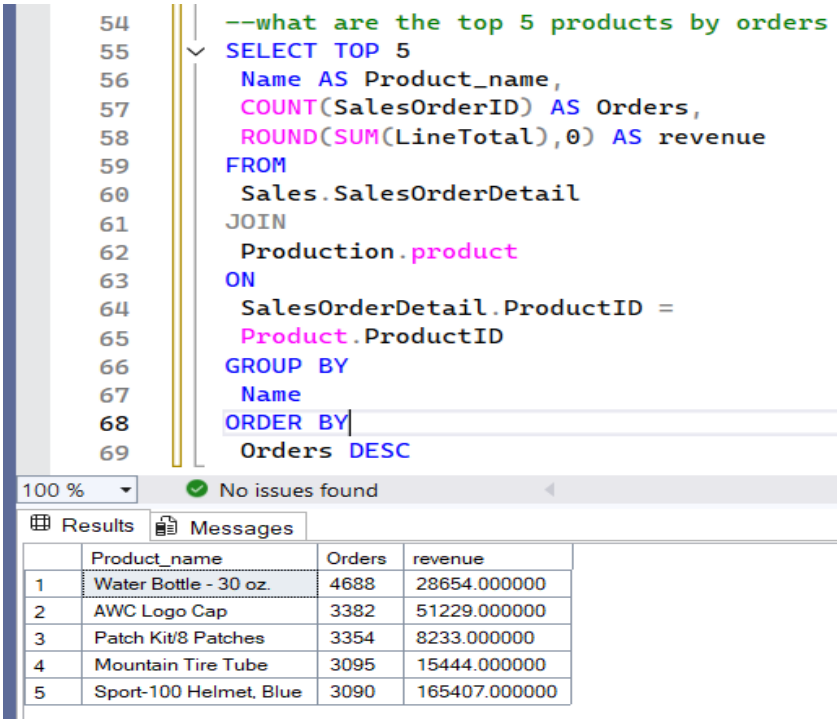


```
-- What are the Orders and Sales volume by Continent
SELECT
 CASE
  WHEN CountryRegionCode = 'DE' THEN 'Germany'
  WHEN CountryRegionCode = 'FR' THEN 'France'
  WHEN CountryRegionCode = 'GB' THEN 'United Kingdom'
  WHEN CountryRegionCode = 'CA' THEN 'Canada'
  WHEN CountryRegionCode = 'AU' THEN 'Australia'
  WHEN CountryRegionCode = 'US' THEN 'United States' END AS Country,
 COUNT(SalesOrderID) AS orders,
 ROUND(SUM(SubTotal),0) AS sales
FROM
 Sales.SalesTerritory
JOIN
 Sales.SalesOrderHeader
ON
 SalesTerritory.TerritoryID =
 SalesOrderHeader.TerritoryID
GROUP BY
 CASE
  WHEN CountryRegionCode = 'DE' THEN 'Germany'
  WHEN CountryRegionCode = 'FR' THEN 'France'
  WHEN CountryRegionCode = 'GB' THEN 'United Kingdom'
  WHEN CountryRegionCode = 'CA' THEN 'Canada'
  WHEN CountryRegionCode = 'AU' THEN 'Australia'
  WHEN CountryRegionCode = 'US' THEN 'United States' END
ORDER BY
 orders DESC
```

Results | Messages

| | Country | orders | sales |
|---|---|---|---|
| 1 | United States | 12041 | 62697591.00 |
| 2 | Australia | 6843 | 10655336.00 |
| 3 | Canada | 4067 | 19359770.00 |
| 4 | United Kingdom | 3219 | 7870721.00 |
| 5 | France | 2672 | 7251556.00 |
| 6 | Germany | 2523 | 4915408.00 |

5. **Top five products drive majority of orders.**
Top 5 products account for ~56% of orders (Water bottle, AWC Logo Cap, Patch Kit, Mountain Tire Tube, Sport Helmet-blue).
**Recommendation:** Prioritize inventory for these products, scale manufacturing, engage in promotions, use discounting strategies to attract more orders.



```
--what are the top 5 products by orders
SELECT TOP 5
 Name AS Product_name,
 COUNT(SalesOrderID) AS Orders,
 ROUND(SUM(LineTotal),0) AS revenue
FROM
 Sales.SalesOrderDetail
JOIN
 Production.product
ON
 SalesOrderDetail.ProductID =
 Product.ProductID
GROUP BY
 Name
ORDER BY
 Orders DESC
```

100 % No issues found

Results | Messages

| | Product_name | Orders | revenue |
|---|---|---|---|
| 1 | Water Bottle - 30 oz. | 4688 | 28654.000000 |
| 2 | AWC Logo Cap | 3382 | 51229.000000 |
| 3 | Patch Kit/8 Patches | 3354 | 8233.000000 |
| 4 | Mountain Tire Tube | 3095 | 15444.000000 |
| 5 | Sport-100 Helmet, Blue | 3090 | 165407.000000 |

Click here to interact with the dashboard report.

# THANK YOU