

# AventureWorks Data Analysis

2019

**Created by:** Stella Malone

**Date:** Apr 12, 2025

---

This report presents a data analysis project based on the AdventureWorks 2019 database, a fictional dataset provided by Microsoft that simulates the operations of a global bicycle manufacturing company. The aim of this project was to explore key business questions related to sales performance, employee behaviour, and store operations. The data analysis workflow began with writing targeted **SQL queries** to extract information from the AdventureWorks database. The query results were then exported as CSV files, which were imported into Python for further analysis and visualisation.

## Objective

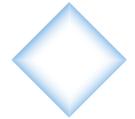


By using structured queries and Python's data analysis libraries -including **Pandas, Matplotlib, and Seaborn** – this project provides insights into the relationships between variables such as annual leave and bonuses, store size and revenue, and country and sales performance. These insights can help business stakeholders identify trends, evaluate performance, and make informed decisions.

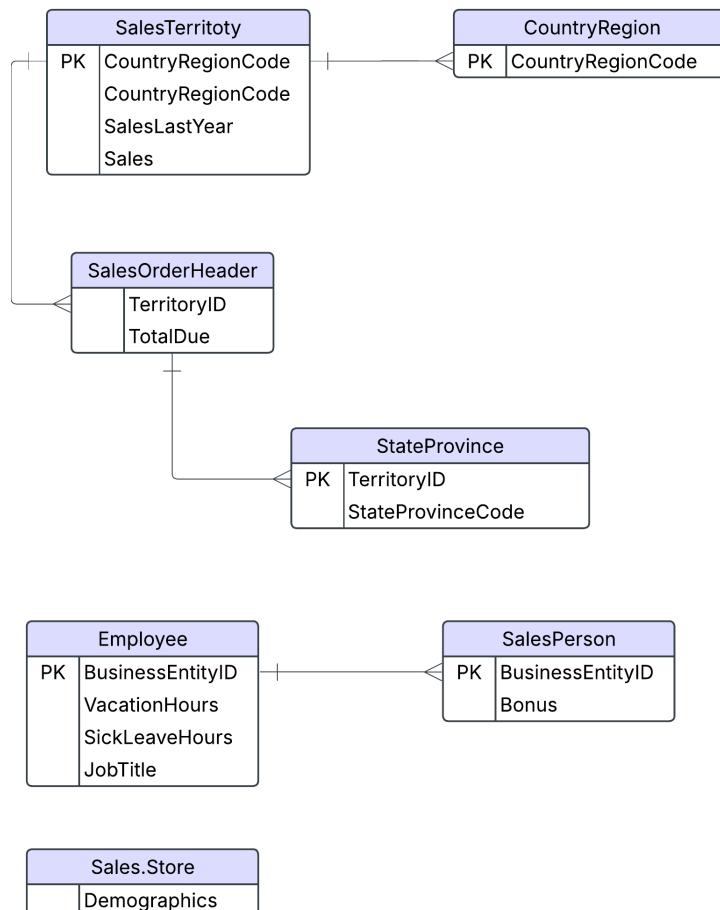
The analysis focused on answering the following key questions:

1. What are the regional sales in the best performing country?
2. What is the relationship between annual leave taken and bonus?
3. What is the relationship between country and revenue?
4. What is the relationship between sick leave and Job Title (PersonType)?
5. What is the relationship between store trading duration and revenue?
6. What is the relationship between the size of the stores, number of employees, and revenue?

# ERD



An entity relationship diagram (ER diagram or ERD) is a visual representation of how items in a database relate to each other – IBM

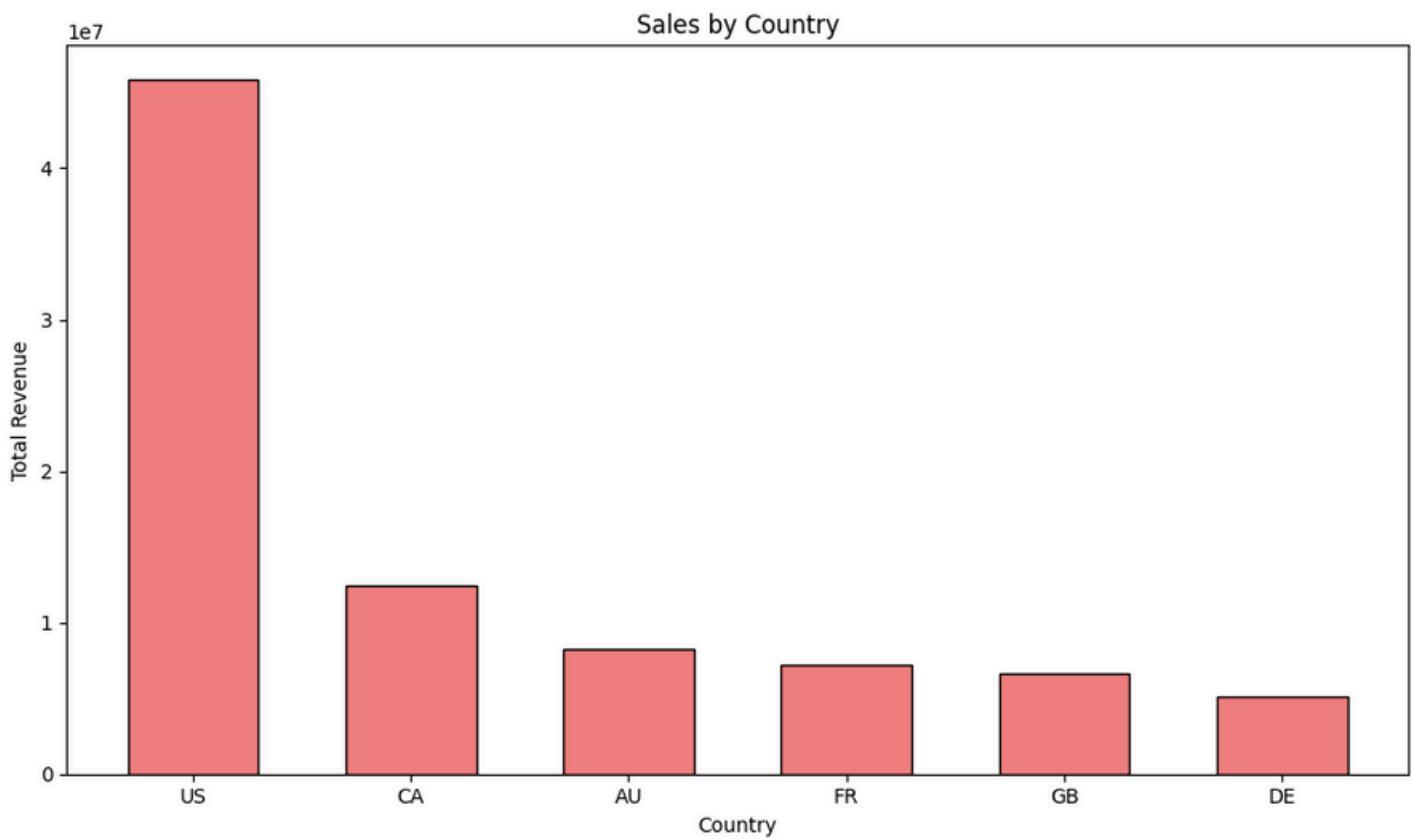


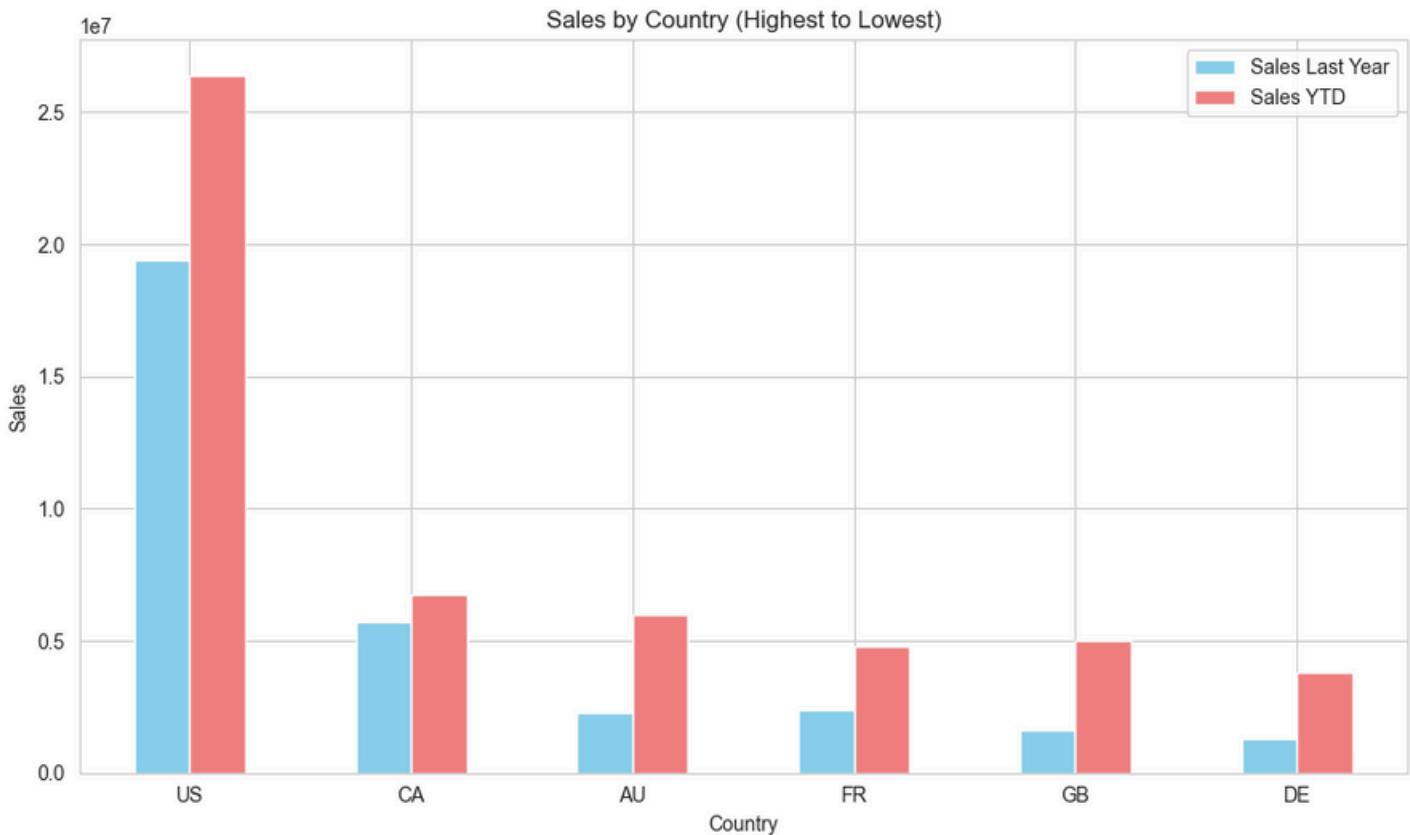
- SalesTerritory is linked to SalesOrderHeader and StateProvince via TerritoryID.
- SalesOrderHeader is indirectly linked to CountryRegion through SalesTerritory using CountryRegionCode.
- SalesPerson is linked to Employee via BusinessEntityID
- Sales.Store stands alone



# Results and analysis

## 1. What are the regional sales in the best performing country?

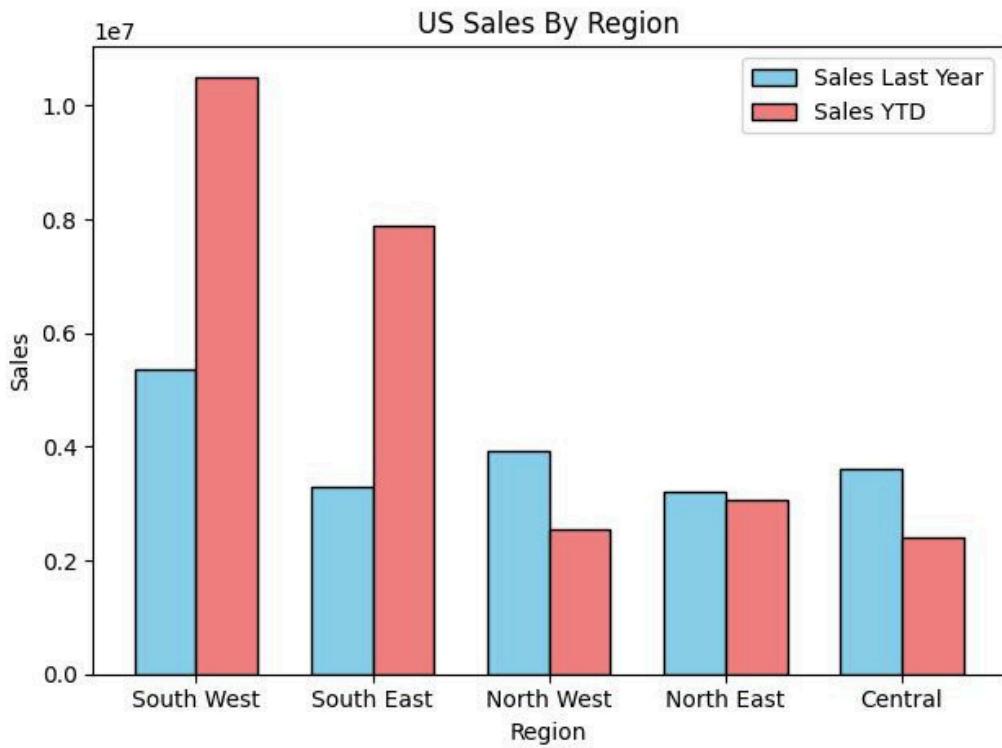




*Analysing by Sales Last Year and Sales Year to Date (YTD)*

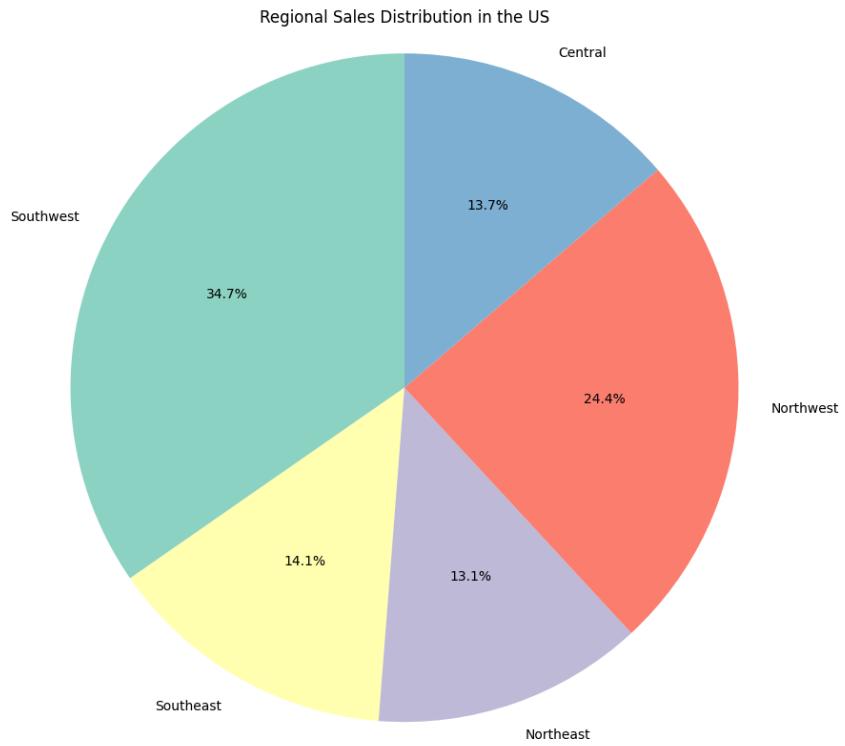
These graphs tell us that USA has the most sales revenue by more than **3 times** the second highest sales (CA).

I then proceeded to graph the sales data to represent each region in the US.

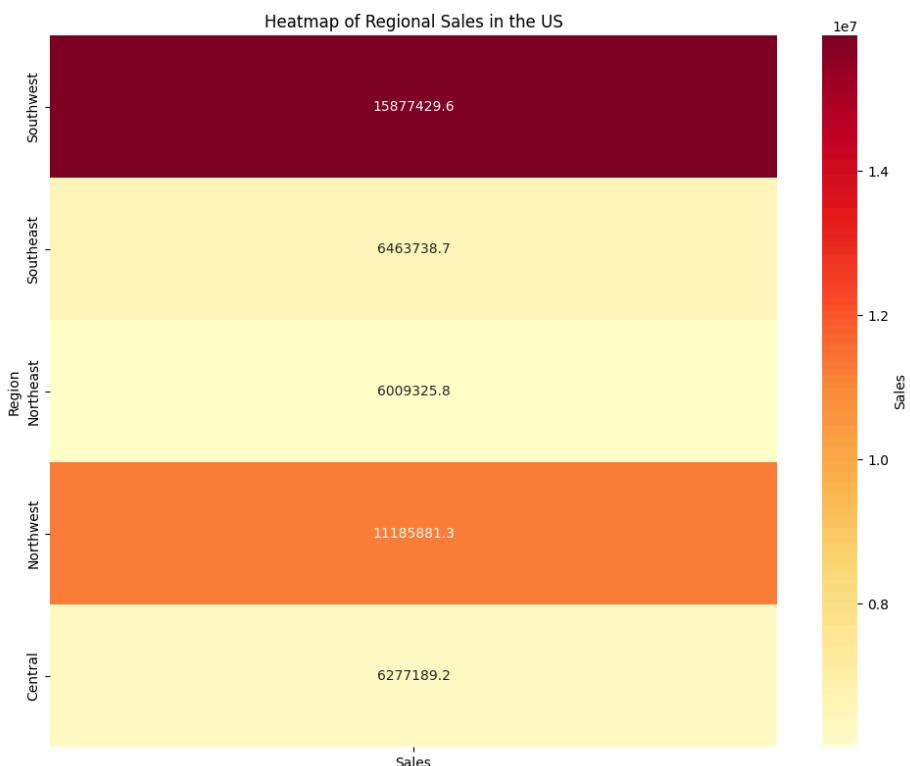


*Analysing by Sales Last Year and Sales Year to Date (YTD)*

- Southwest Region: Leads in sales with significant growth in "Sales YTD" compared to last year.
- Southeast Region: Strong performance with "Sales YTD" surpassing last year's sales.
- Central Region: Slight decline in "Sales YTD," indicating potential challenges.
- Overall Trend: Most regions show positive growth in "Sales YTD," highlighting a strong sales trajectory.



*Pie chart from total sales (sales last year + year to date)*



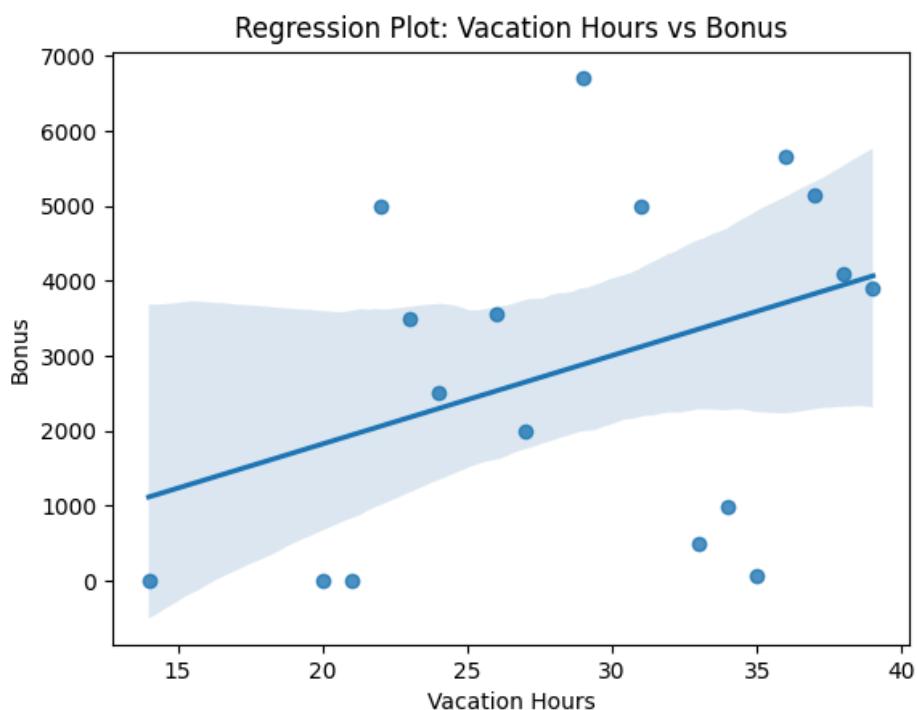
*Heat map of total sales (sales last year + year to date)*

These graphs tell us that:

- Southwest Region has the most sales (34%)

- Northwest has the second most (24%)
  - While Southeast, Northeast and Central are tied at the bottom all with around 13%
- 

## 2. What is the relationship between annual leave taken and bonus?

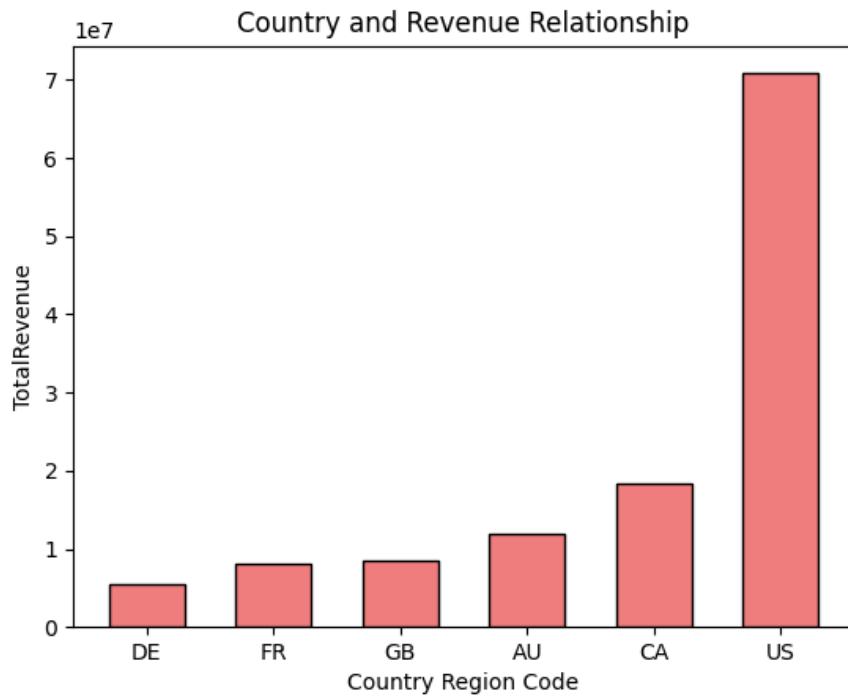


Regression plot of Bonus Revenue by Vacation Hours Taken

The correlation is **0.38** – this indicates a weak positive correlation between annual leave and bonuses. This suggests that while vacation time might play a role, it is very weak and there are probably other factors that have a stronger influence on bonuses.

---

## 3. What is the relationship between country and revenue?

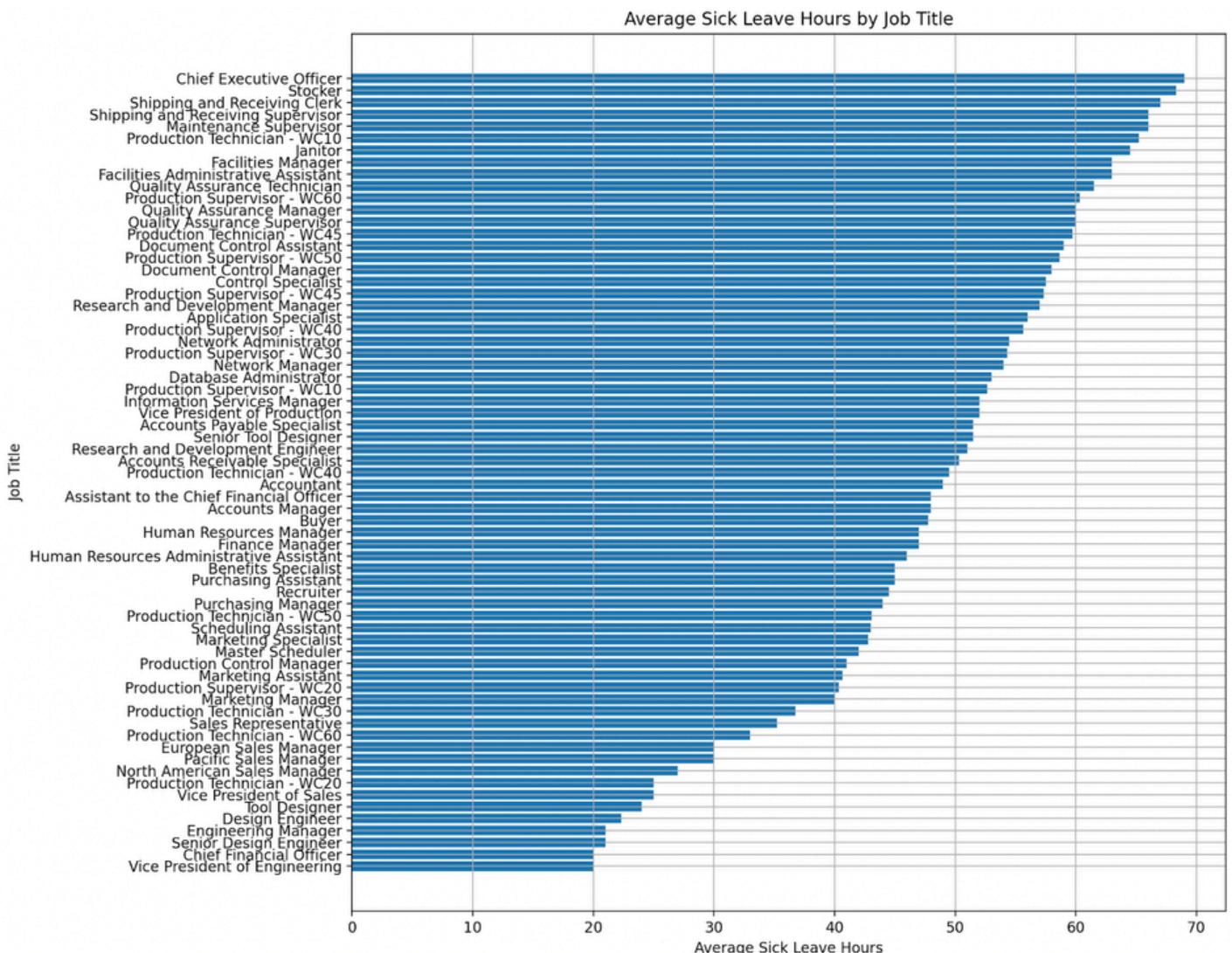


*Bar Chart of TotalRevenue by Country Region Code*

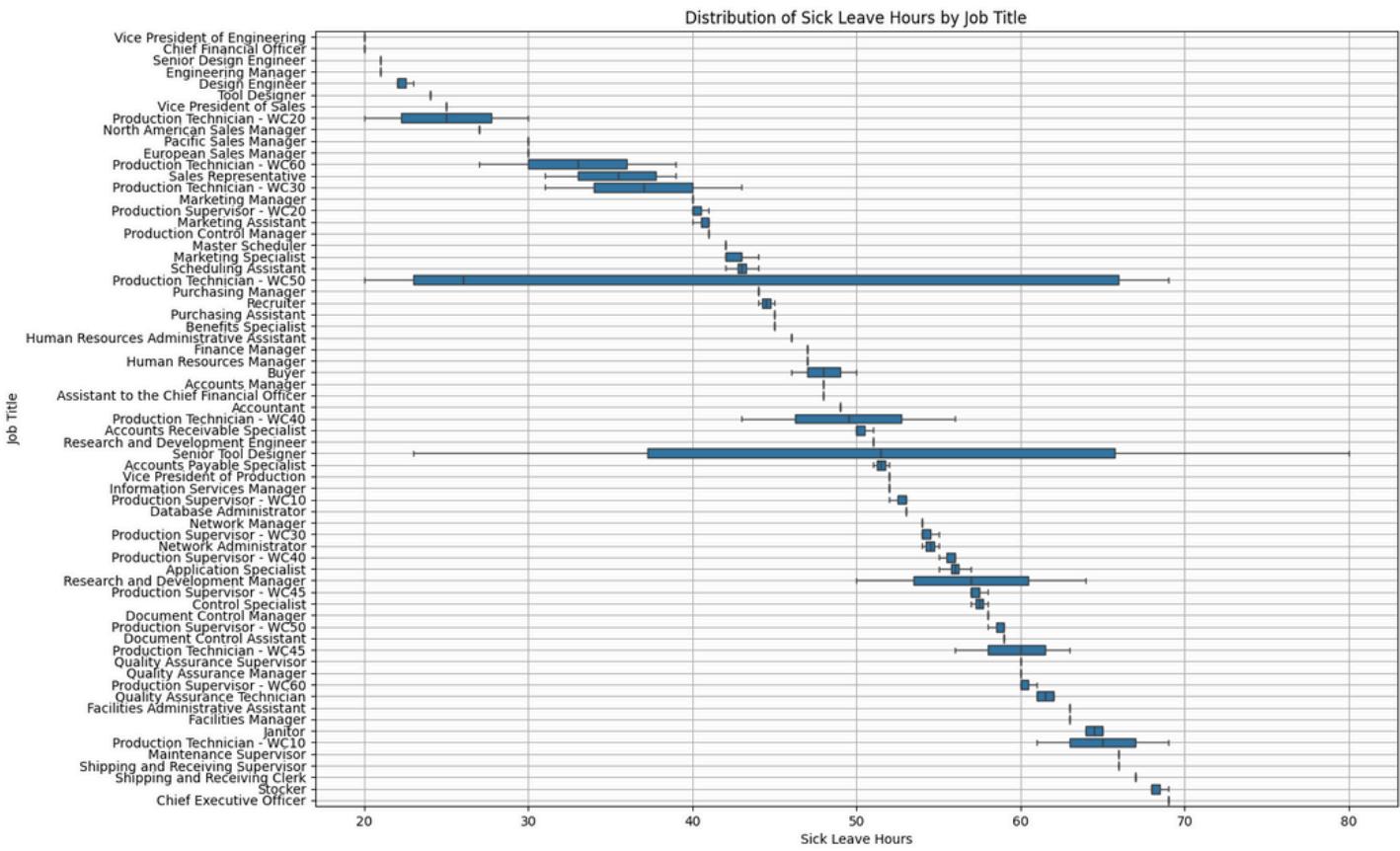
The United States leads significantly, contributing the highest revenue, followed by Canada and Australia. The remaining countries show relatively lower growth in terms of revenue.

---

#### **4. What is the relationship between sick leave and Job Title?**

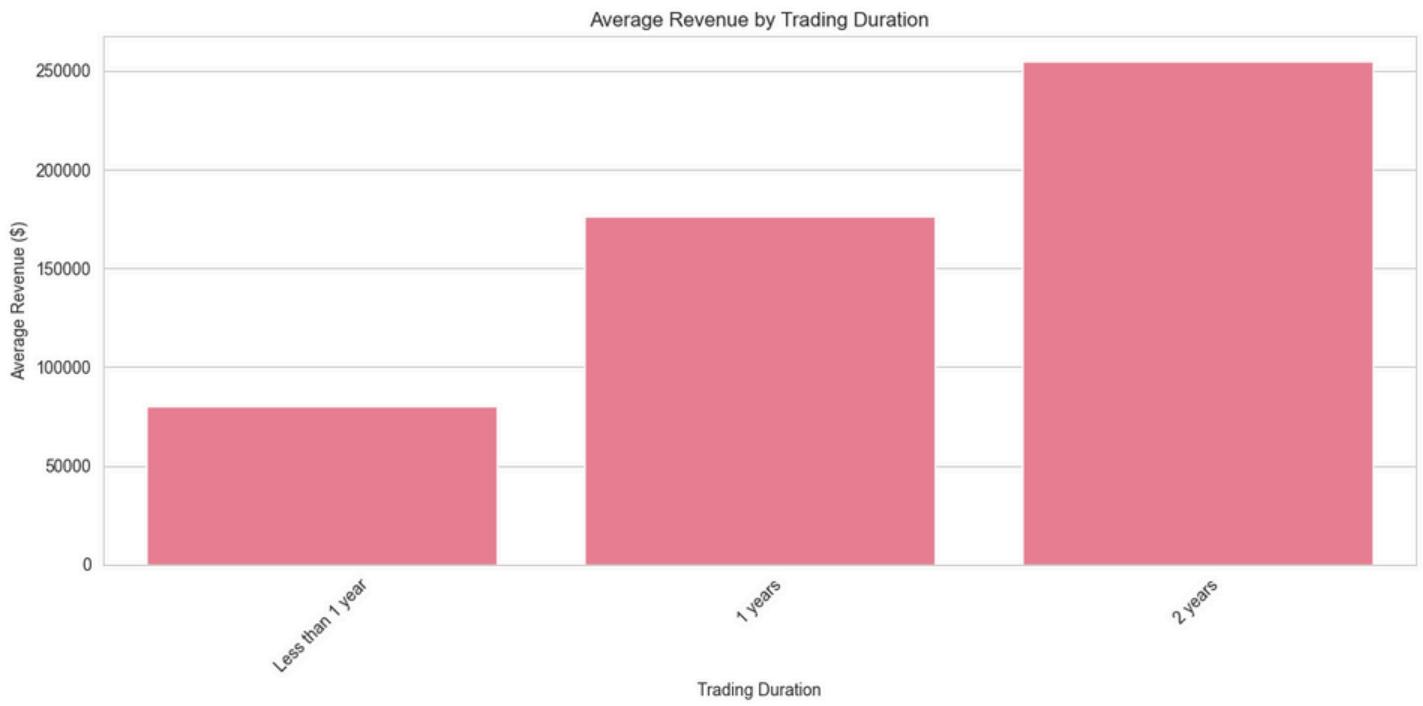


- Average sick leave hours differ significantly by job title
- Employee roles with high averages: Chief Executive Officer, Stocker, Shipping/Receiving Clerk
- Employee roles with low averages : Vice president of Engineer and Chief Financial Officer



- Significant Variation: Some roles consistently take fewer sick leave hours, while others show a wide range, including outliers
- Skewed Distribution: In some job titles, the median is not centred, suggesting a difference between mean and median sick leave hours
- Low Sick Leave Usage: CEO & Vice President of Engineering
- Minimal Absences: Production Technicians & Supervisors (WC10–WC45)
- Higher Absences: Production Technician - WC50 & Accounts Payable Specialist show wider distributions and higher medians

## 5. What is the relationship between store trading duration and revenue?



*Average Store Revenue by Trading Duration, showing increased revenue correlating with longer trading periods*

- There's a clear upward trend – the longer a store trades, the higher its average revenue
- The increase appears to be fairly consistent (roughly \$80–90k increase per year of trading)
- This suggests that stores tend to build their revenue over time

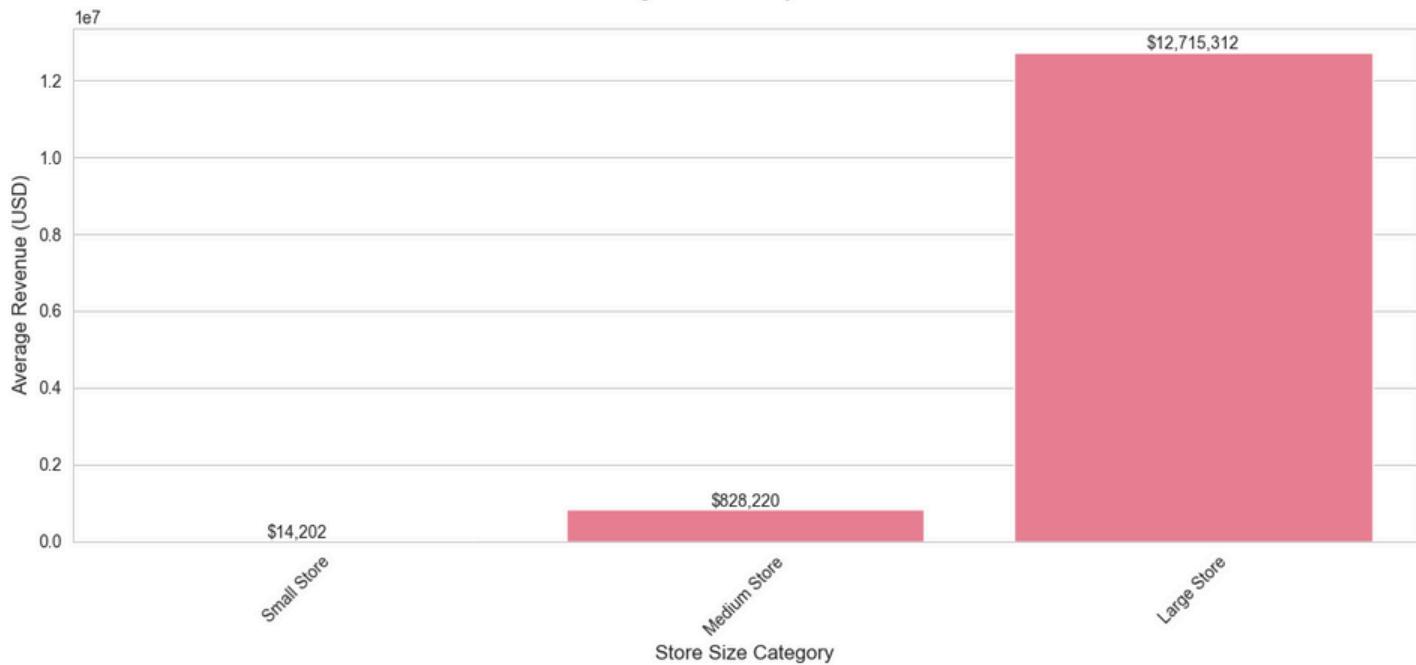


*Distribution of Stores by Trading Duration, indicating a predominance of newer stores in the dataset*

- There's a clear decreasing pattern as trading duration increases
- More than half of the stores are relatively new (less than 1 year)
- This could indicate:
  - Rapid business expansion/growth
  - High turnover rate of stores
  - The business might be in a growth phase
  - Might need different strategies for helping stores survive beyond the first year

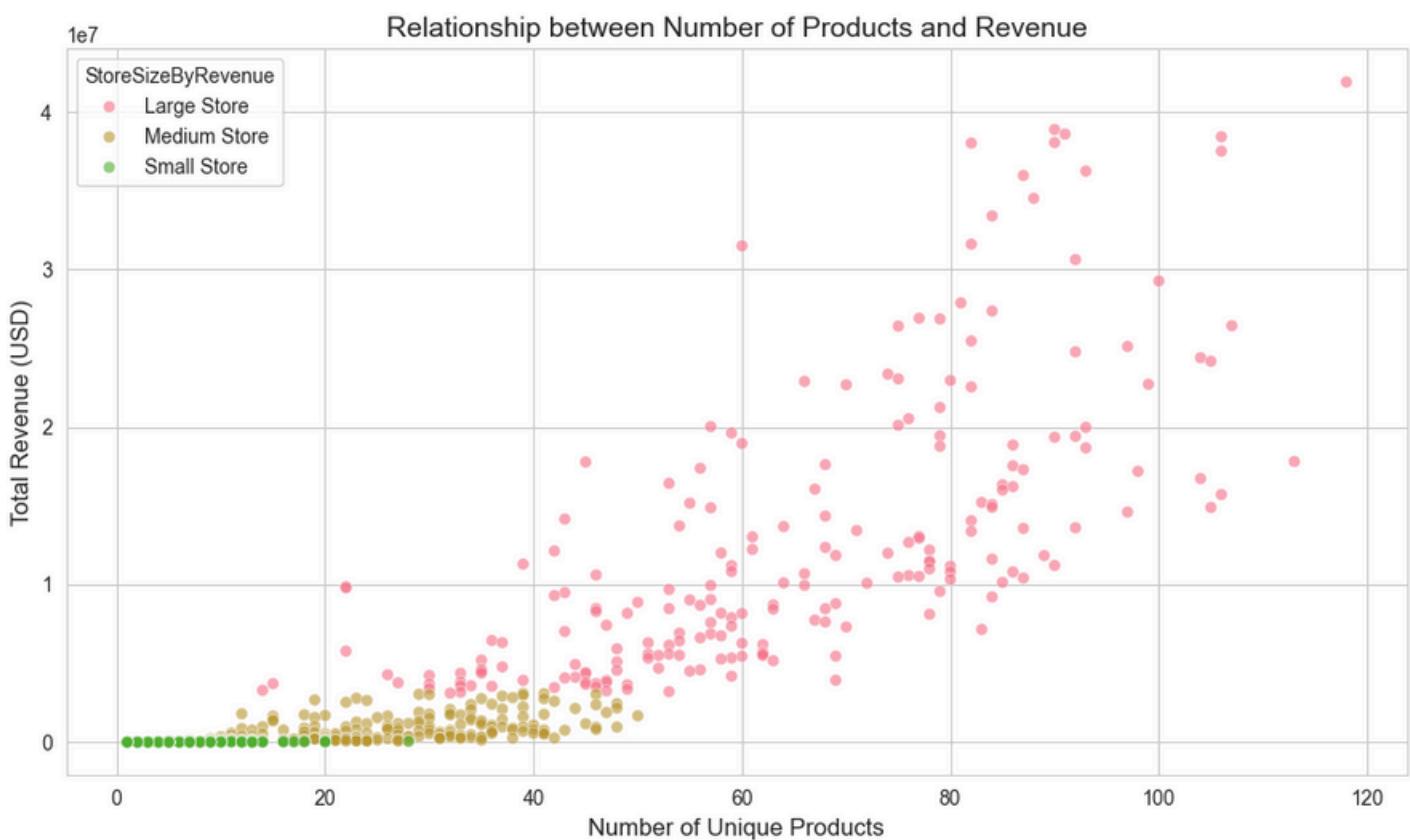
## **6. What is the relationship between the size of the stores, number of employees, and revenue?**

Average Revenue by Store Size



Average Revenue Distribution Across Store Size Categories

- Clear Positive Relationship:
- There's a strong positive relationship between store size and revenue
- Each category shows roughly a 15x increase from the previous category



### *Scatter Plot of Revenue vs Number of Unique Products by Store Size Category*

- Clear positive correlation: More unique products generally correlate with higher revenue
- Exponential rather than linear revenue relationship
- Larger stores have greater capacity to manage product variety
- The relationship suggests that expanding product range could be a pathway to growth for smaller stores

# Conclusion

---



This comprehensive analysis of the AdventureWorks2019 database reveals several key business insights across multiple dimensions.

## Geographic Performance

- The United States dominates sales revenue, exceeding other markets by a factor of three
- Within the US, the Southwest region leads with 34% of sales, followed by Northwest (24%)
- Other regions (Southeast, Northeast, Central) each contribute approximately 13%

## Operational Metrics

- Store Performance Factors:
- Trading duration shows a strong positive correlation with revenue (\$80-90k increase per year)
- Store size significantly impacts revenue, with each category showing roughly 15x growth
- Product diversity correlates positively with revenue, showing an exponential relationship

## Employee-Related Findings

- Weak positive correlation (0.38) between annual leave and bonuses
- Sick leave patterns vary significantly by job title:
- Higher rates: CEO, Stocker, Shipping/Receiving roles
- Lower rates: VP Engineering, CFO positions

## Business Growth Patterns

- Majority of stores are under 1-year old, suggesting:
- Rapid business expansion phase
- Potential concerns about long-term store sustainability
- Need for targeted support strategies for new stores

Larger stores demonstrate:

- Higher revenue generation
- Greater product diversity
- Better capacity for inventory management
- Stronger transaction values

These findings suggest opportunities for strategic growth through store expansion, product diversification, and regional market development, while highlighting areas requiring attention in employee management and store sustainability.

## References

---

- [AdventureWorks dataset](#)
- [ERD diagram video](#)

**Check out my  
other projects!**

[www.reallygreatsite.com](http://www.reallygreatsite.com)

