COLORADO MOTOR VEHICLE SALES DATA ANALYSIS

A Data Analytics Project Report

Submitted by:- DALWADI MEET

Enrollment /UNID :- UMID12112570217

Internship: - Data Analytics

Submitted to:-

UNIFIED MENTOR

Tools Used: -

Python (Jupyter Notebook), SQL, Power BI, Microsoft Excel

Internship Year

2025 – 2026

Date of Submission

15 / 12 / 2025

1. Introduction:-

Motor vehicle sales play a significant role in measuring economic activity and consumer spending within a region. Analyzing vehicle sales data helps in understanding market demand, economic growth, and changes in consumer behavior over time. The automotive industry also influences employment, infrastructure development, and government revenue, making it an important area for economic analysis.

This project focuses on analyzing motor vehicle sales data across various counties in Colorado. The dataset is structured by year and quarter, allowing the study of both long-term trends and seasonal variations in sales performance. By examining sales values at different time intervals, the project provides insights into how vehicle sales fluctuate throughout the year and across regions.

The analysis aims to identify patterns such as peak sales quarters, high-performing counties, and overall sales growth trends. These insights help businesses understand regional demand, assist policymakers in infrastructure and transportation planning, and support market researchers in evaluating the automotive market landscape.

Through the use of data analysis tools and interactive dashboards, this project transforms raw sales data into meaningful insights. The results enable stakeholders to make data-driven decisions, optimize strategies, and better understand the economic impact of motor vehicle sales in Colorado.

---

2. Project Objectives

The primary objective of this project is to analyze motor vehicle sales data across various counties in Colorado to understand sales trends, seasonal patterns, and regional performance. By examining sales values segmented by year and quarter, the project aims to provide meaningful insights into how vehicle sales change over time and how different counties contribute to overall sales performance.

Another key objective is to identify high-performing and low-performing periods within the automotive market. Through quarterly analysis, the project seeks to determine the best-performing quarters and years, helping to highlight seasonal demand patterns. This analysis supports businesses and policymakers in understanding peak sales periods and planning resources accordingly.

The project also aims to compare vehicle sales across different counties to understand regional market demand. By analyzing county-level sales data and percentage contribution to total sales, the project helps identify counties that play a major role in driving overall revenue. This information is valuable for market research, regional planning, and strategic decision-making.

An additional objective is to develop an interactive and user-friendly dashboard using tools such as Power BI. The dashboard allows users to filter data by year and quarter using slicers and view dynamic KPIs, charts, and visualizations. This enables stakeholders to explore the data easily and gain insights without requiring technical expertise.

Finally, the project aims to transform raw sales data into actionable insights that can support economic analysis, market research, and policy-making decisions. By presenting clear visualizations and performance metrics, the project demonstrates how data analytics can be used to support informed, data-driven decision-making in the automotive industry.

---

3. Dataset Description

The dataset used in this project is titled Colorado Motor Vehicle Sales Data and contains detailed information about motor vehicle sales recorded across various counties in the state of Colorado. The data is structured to capture sales performance over multiple years and quarters, making it suitable for analyzing both long-term trends and seasonal variations in vehicle sales. The dataset is provided in CSV (Comma-Separated Values) format, which allows easy import and processing using data analysis tools such as Python, SQL, and Power BI.

The dataset is organized in a tabular format where each record represents the total motor vehicle sales for a specific county during a particular quarter of a given year. This structure enables granular analysis at both regional and temporal levels. By combining time-based attributes such as year and quarter with geographic information in the form of county names, the dataset supports comparative analysis across regions as well as trend analysis over time.

The dataset consists of four main columns. The Year column represents the calendar year in which the vehicle sales were recorded. This column is essential for identifying annual trends and understanding how vehicle sales evolve over time. The Quarter column indicates the quarter of the year during which the sales occurred. The quarters are divided into four periods: Q1 (January to March), Q2 (April to June), Q3 (July to September), and Q4 (October to December). This quarterly segmentation allows the identification of seasonal sales patterns and peak sales periods.

The County column contains the names of counties within the state of Colorado where the vehicle sales were recorded. This geographic attribute enables county-level analysis and helps identify regions with higher or lower vehicle sales activity. Understanding county-wise performance is important for market research, regional planning, and identifying areas with strong or weak automotive demand. The Sales column represents the total dollar value of

motor vehicle sales for the specified county, year, and quarter. This numeric column is the primary measure used for analysis and visualization throughout the project.

The dataset does not include personally identifiable information, ensuring that the analysis is focused solely on aggregated sales performance rather than individual transactions. The absence of sensitive data makes the dataset suitable for academic and analytical purposes. Additionally, the structured nature of the dataset allows it to be easily transformed, aggregated, and visualized for advanced analysis such as time-series evaluation and percentage contribution analysis.

Overall, the Colorado Motor Vehicle Sales dataset provides a reliable foundation for analyzing automotive sales trends, regional performance, and seasonal behavior. Its clean structure, well-defined columns, and time-based segmentation make it highly suitable for building interactive dashboards and generating insights that support economic analysis, market research, and policy-making decisions related to the automotive industry.

- ## Column Description

The dataset contains four key columns that together provide a structured view of motor vehicle sales across Colorado. Each column plays an important role in enabling time-based, regional, and sales performance analysis.

- Year

The Year column represents the calendar year in which the motor vehicle sales data was recorded. This column is used to analyze long-term sales trends and year-over-year growth or decline in vehicle sales. By grouping and comparing data across years, analysts can identify whether the automotive market is expanding, contracting, or remaining stable over time.

- Quarter

The Quarter column indicates the quarter of the year during which the vehicle sales occurred. The year is divided into four quarters:

Q1: January to March

Q2: April to June

Q3: July to September

Q4: October to December

This column is essential for identifying seasonal patterns in vehicle sales. It allows the analysis of peak and low sales periods within a year and helps understand how factors such as festivals, weather, or economic cycles influence consumer purchasing behavior.

- County

The County column contains the names of counties in the state of Colorado where the vehicle sales were recorded. This geographic dimension enables region-wise analysis of vehicle sales performance. By analyzing county-level data, the project can identify high-performing counties, compare regional demand, and assess how sales activity varies across different parts of the state.

- Sales

The Sales column represents the total dollar value of motor vehicle sales for a specific county, year, and quarter. This is the primary quantitative measure used throughout the analysis. The sales values are aggregated and analyzed to calculate total sales, average sales, percentage contribution, and growth trends. This column forms the foundation for all visualizations, KPIs, and business insights generated in the project.

- Tools and Technologies Used

This project uses a combination of data analysis and visualization tools to effectively analyze and present motor vehicle sales data for the state of Colorado. Each tool was selected based on its strengths in data handling, analysis, and visualization to ensure accurate insights and clear presentation of results.

- Python (Jupyter Notebook)

Python was used for data loading, cleaning, and initial exploration of the dataset. Libraries such as Pandas and NumPy were used to handle missing values, transform data, and perform aggregations efficiently. Python was also useful for validating calculations and preparing the dataset for further analysis. Jupyter Notebook provided an interactive environment that allowed step-by-step analysis and easy documentation of the data preparation process.

- SQL

SQL was used to perform structured data analysis and querying. It enabled efficient filtering, grouping, and aggregation of sales data by year, quarter, and county. SQL was particularly useful for answering analytical questions such as identifying the best-performing quarters, calculating average sales, and comparing county-wise performance. Using SQL helped ensure accuracy and reproducibility of analytical results.

- Power BI

Power BI was used to create interactive dashboards and visualizations. It allowed the creation of KPI cards, line charts, bar charts, and donut charts to present sales trends and percentage contributions clearly. Slicers for Year and Quarter were added to enable dynamic filtering and drill-down analysis. Power BI was chosen for its user-friendly interface and strong capability to transform data into business-ready insights.

- Microsoft Excel (Optional)

Microsoft Excel was used for quick data inspection and basic validation. It helped in understanding the dataset structure and verifying summary calculations before importing the data into other tools.

- Why These Tools Were Chosen

The combination of Python, SQL, and Power BI provides a complete data analytics workflow—from data cleaning and transformation to analysis and visualization. Python ensures flexibility in data processing, SQL enables efficient querying, and Power BI delivers interactive and visually appealing dashboards. Together, these tools help convert raw sales data into actionable insights that support business and policy decision-making.

---

4. Data Cleaning and Preparation

Before performing analysis and building visualizations, the dataset was cleaned and prepared to ensure accuracy and consistency. Proper data preparation is essential to avoid incorrect insights and improve the reliability of results.

First, the dataset was imported from a CSV file into Python and Power BI. The data structure and column types were reviewed to ensure that year, quarter, county, and sales values were correctly recognized. Any missing or null values were identified and handled to maintain data integrity.

Next, the sales column was checked for inconsistencies and standardized for analysis. Sales values were converted into a consistent numerical format, and where required, values were scaled into billions to improve readability in visualizations and KPI cards.

To enable time-based analysis, a date field was created using the year and quarter columns. This allowed the aggregation of sales data over time and supported trend analysis through line charts and time-series visuals. The data was then aggregated by year, quarter, and county to support analytical queries and dashboard creation.

Finally, the cleaned dataset was validated by cross-checking summary statistics and totals across different tools. This ensured that the prepared data was accurate and ready for further analysis and visualization.

- ## Data Cleaning and Preparation (With Process)

This section explains how the raw dataset was transformed into a clean and analysis-ready format. The cleaning process was carried out step by step to ensure data accuracy, consistency, and usability for analysis and dashboard creation.

Step 1: Raw Data Inspection

After importing the CSV file, the dataset was first inspected to understand its structure and identify potential issues.

Raw Data Sample (Before Cleaning):

Year	Quarter	County	Sales
2020	Q1	Denver	1250000000
2020	Q2	Adams	NaN
2021	Q3	Denver	1420000000

Issues Identified:

Missing values in the Sales column

Quarter stored as text (Q1, Q2, etc.)

No single Date column for time-series analysis

Step 2: Handling Missing Values

Missing sales values were checked using null-value detection. Rows with missing or invalid sales values were removed to avoid incorrect calculations.

Action Taken:

Removed rows where sales values were missing

Ensured only valid numeric sales records remained

Result:

Clean dataset with no null values in critical columns

Step 3: Data Type Standardization

The Year column was ensured to be numeric, and the Sales column was converted into a consistent numerical format. This step was necessary for aggregation, averaging, and KPI calculations.

Additionally, sales values were converted into billions (Bn) to improve readability in charts and KPI cards.

Step 4: Creating a Date Column

Since the dataset did not originally contain a date column, a new Date column was created using the Year and Quarter columns.

Quarter-to-Month Mapping:

Q1 → January

Q2 → April

Q3 → July

Q4 → October

Resulting Date Format Example:

Year	Quarter	Date
2020	Q1	2020-01-01
2020	Q2	2020-04-01
2021	Q3	2021-07-01

This enabled proper time-based analysis such as sales trends over time.

Step 5: Data Aggregation

After cleaning, the data was aggregated by:

- Year
- Quarter
- County

This aggregation helped in:

- Identifying best-performing quarters
- Comparing county-wise sales

Creating KPIs and trend charts

Step 6: Final Clean Dataset (After Cleaning)

Clean Data Sample (After Preparation):

Year	Quarter	County	Sales (Bn)	Date
2020	Q1	Denver	1.25	2020-01-01
2021	Q3	Denver	1.42	2021-07-01

The dataset is now consistent, complete, and ready for analysis and visualization.

---

5. Data Analysis and Methodology

After completing the data cleaning and preparation stage, the dataset was analyzed using structured analytical methods to extract meaningful insights from motor vehicle sales data across Colorado. The analysis focused on identifying trends over time, comparing county-level performance, and understanding quarterly sales behavior.

Step 1: Exploratory Data Analysis (EDA)

The first step involved exploring the cleaned dataset to understand the overall distribution and range of sales values. Summary statistics such as total sales, average sales, and minimum and maximum sales were calculated. This helped in gaining an initial understanding of sales magnitude and variability across years and counties.

- Key checks performed during this step included:
- Total sales across all years

	total_sales numeric
1	88205300000

- Average sales per year and quarter

	quarter bigint	sales numeric
1	1	20376846000
2	2	22343294000
3	3	23914934000
4	4	21570226000

- County-wise sales distribution

	county text	total_sales numeric
1	Arapahoe	20142323000
2	El Paso	11926044000
3	Jefferson	9058407000
4	Adams	8902115000
5	Denver	6763613000
6	Larimer	5344367000
7	Weld	5086889000
8	Boulder	4742532000
9	Rest of St...	3582170000
10	Douglas	3236493000

This step ensured that the data was logically consistent and ready for deeper analysis.

Step 2: Time-Based Analysis

To analyze sales trends over time, the data was grouped using the Date, Year, and Quarter fields. A time-series approach was used to study how sales evolved across different years and quarters.

- The following analyses were performed:
  - Year-over-year sales trend analysis

	year bigint	sales numeric
1	2008	8965561000
2	2009	7652500000
3	2010	8556088000
4	2011	12170441000
5	2012	10960876000
6	2013	12000615000
7	2014	13392487000
8	2015	14506732000

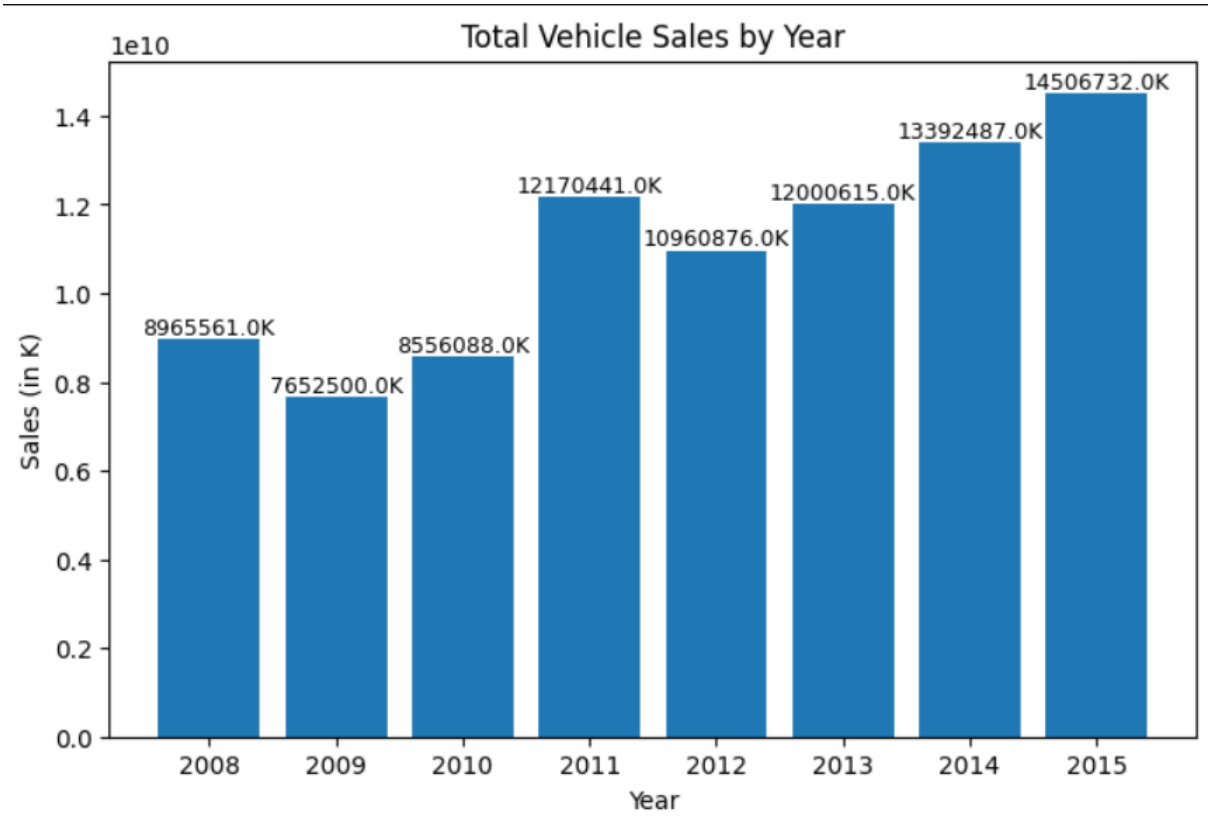
  - Quarterly sales trend analysis

	quarter bigint	sales numeric
1	1	20376846000
2	2	22343294000
3	3	23914934000
4	4	21570226000

- **Identification of peak and low sales periods**

County	Total Sales (USD)	Rank
Arapahoe	3,408,437,000	1
El Paso	1,878,968,000	2
Jefferson	1,523,003,000	3
Adams	1,511,503,000	4
Denver	943,660,000	5
Larimer	931,283,000	6
Boulder	893,050,000	7
Weld	776,080,000	8
Rest of State	642,953,000	9
Douglas	596,706,000	10
Pueblo	392,768,000	11
Mesa	370,497,000	12
Garfield	264,428,000	13
Broomfield	186,445,000	14
La Plata	137,152,000	15
Fremont	49,799,000	16

Line charts were used to visualize these trends, making it easier to identify growth patterns and seasonal fluctuations in motor vehicle sales.



Step 3: County-Level Performance Analysis

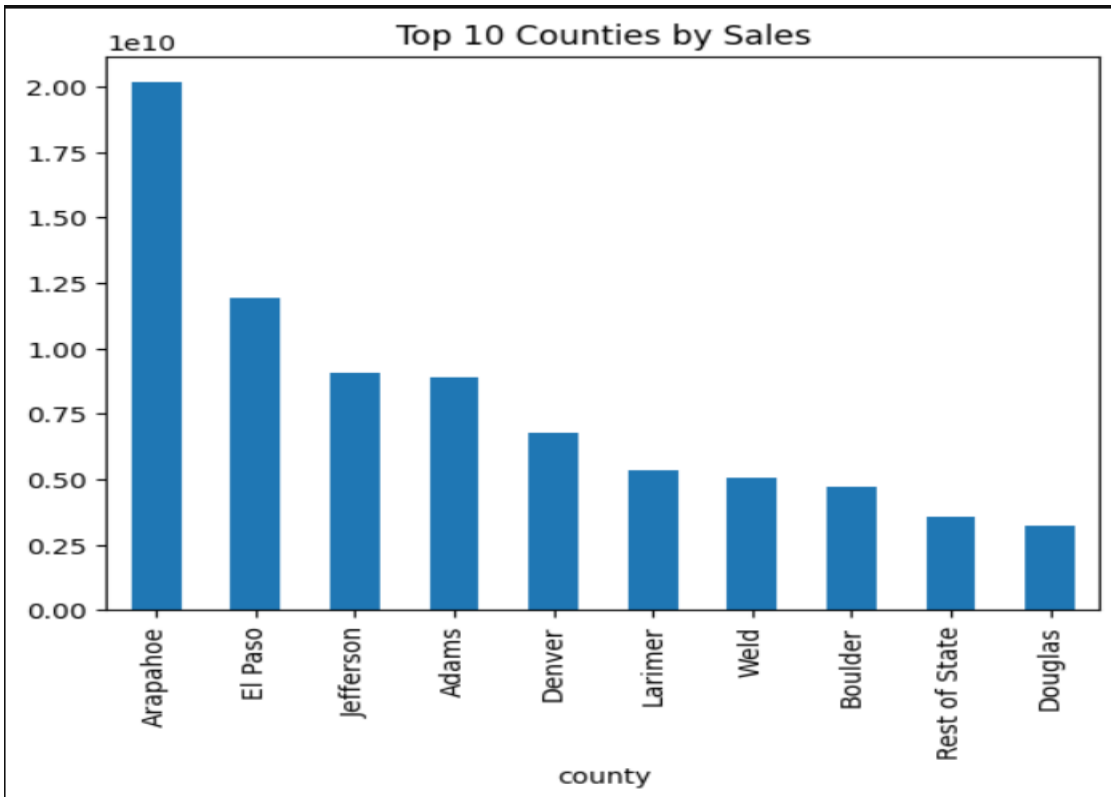
County-wise analysis was conducted to compare sales performance across different regions in Colorado. Sales were aggregated by county to calculate total sales and average sales per county.

This step helped in:

Identifying top-performing counties

	county text	total_sales numeric
1	Arapahoe	20142323000
2	El Paso	11926044000
3	Jefferson	9058407000
4	Adams	8902115000
5	Denver	6763613000
6	Larimer	5344367000
7	Weld	5086889000
8	Boulder	4742532000
9	Rest of St...	3582170000
10	Douglas	3236493000

Bar charts and ranking visuals were used to present county-level comparisons clearly.



Step 4: KPI and Measure Creation

Key Performance Indicators (KPIs) were created to summarize critical business metrics. These included:

Total Sales

Average Sales

Best Performing Quarter

Sales Contribution Percentage

These KPIs dynamically update based on slicer selections for year and quarter, allowing users to interactively explore the data.

Step5: Methodology Validation

To ensure accuracy, the results obtained from Power BI visualizations were cross-verified using Python and SQL aggregations. This validation step ensured that all totals, averages, and percentages were consistent across different tools.

---

6. Dashboard Design and Visualization

The Power BI dashboard was designed to present Colorado motor vehicle sales data in a clear, interactive, and decision-focused manner. The layout follows a top-to-bottom analytical flow, allowing users to quickly understand overall performance and then drill down into detailed insights by year, quarter, and county.

Dashboard Layout and Theme

The dashboard uses a **dark professional theme** to enhance visual clarity and reduce eye strain. High-contrast colors are applied to highlight key metrics and charts, while consistent fonts and spacing ensure a clean and organized appearance. The layout is divided into three main sections: **filters (slicers)**, **key performance indicators (KPIs)**, and **analytical visuals**.

Slicers (Filters)

Interactive slicers are placed on the left side of the dashboard to allow dynamic filtering of the data:

- **Country (County) Slicer** – Enables users to analyze sales performance for a specific county or all counties combined.
- **Quarter Slicer** – Allows selection of a particular quarter (Q1–Q4) to view quarterly sales impact.
- **Year Slicer** – Filters the dashboard to display sales data for selected years.

All visuals and KPIs automatically update based on slicer selections, providing a seamless interactive experience.

Key Performance Indicators (KPIs)

The dashboard includes KPI cards at the top to summarize high-level business metrics:

- **Selected County Name** – Displays the currently selected county.
- **Total Sales** – Shows the total motor vehicle sales value in **billions (Bn)** based on slicer selection.
- **Number of Counties** – Displays the total number of counties included in the analysis.

These KPIs provide an instant overview of sales performance without requiring users to analyze charts.

Charts and Visualizations

1. **Sum of Sales by Quarter (Bar Chart)**
   This chart compares total sales across quarters, helping identify the best and weakest performing quarters. Data labels are displayed in billions to improve readability.
2. **Top Sales by Counties (Horizontal Bar Chart)**
   This visualization ranks counties based on total sales, allowing quick identification of top-performing regions. It is useful for regional performance comparison.
3. **Total Sales by Year (Line Chart)**
   The line chart shows sales trends over time, enabling year-over-year comparison and highlighting growth or decline patterns in vehicle sales.
4. **Percent of Total Sales (Donut Chart)**
   The donut chart illustrates the percentage contribution of each county to total sales. This helps understand which counties contribute the most to overall revenue.
5. **Geographic Map Visualization**
   A map visual is used to represent county-level sales distribution geographically. This adds spatial context and supports location-based analysis.

Interactivity and User Experience

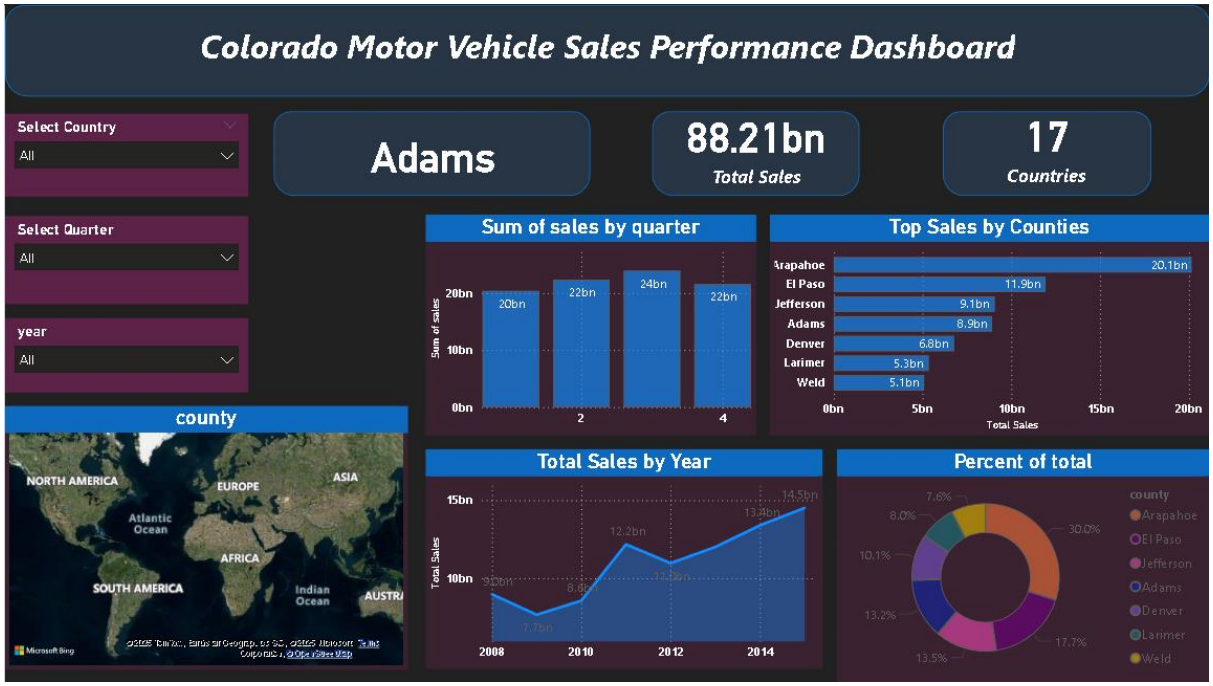
The dashboard is fully interactive, with all visuals responding to slicer selections. Clicking on any chart element highlights related data across other visuals, enabling cross-analysis. This interactivity allows users to explore the data from multiple perspectives and gain deeper insights without switching views.

Purpose of the Dashboard

The dashboard was designed to support:

- Sales trend analysis over time
- Regional performance comparison
- Identification of high-contributing counties
- Quarterly and yearly sales evaluation

Overall, the dashboard transforms raw motor vehicle sales data into actionable insights through intuitive design, effective visualizations, and interactive filtering.



7. Key Insights and Conclusion

Key Insights

The analysis of Colorado Motor Vehicle Sales data reveals several important insights into sales trends, regional performance, and seasonal behavior. One of the major findings is that motor vehicle sales show a **clear upward trend over the years**, indicating steady growth in the automotive market across Colorado. This suggests increasing consumer demand and positive economic activity within the state.

Quarterly analysis highlights noticeable **seasonal variations in sales performance**. Certain quarters consistently outperform others, indicating that vehicle purchases are influenced by seasonal factors such as promotions, economic cycles, or consumer buying behavior. Identifying these high-performing quarters can help businesses plan inventory, marketing campaigns, and resource allocation more effectively.

County-level analysis shows that **sales are not evenly distributed across all counties**. A small number of counties contribute a significant percentage of total sales, making them key drivers of overall revenue. This insight is particularly valuable for market research and regional strategy development, as it highlights areas with strong demand and potential opportunities for expansion.

The percentage contribution analysis further emphasizes the dominance of top-performing counties. The donut chart visualization makes it clear which counties have the highest share of total sales, allowing stakeholders to quickly identify priority regions. The use of interactive slicers enhances this analysis by enabling users to explore changes across different years and quarters.

Conclusion

This project successfully transformed raw motor vehicle sales data into meaningful insights using a structured data analysis approach and interactive visualization techniques. By combining Python, SQL, and Power BI, the project demonstrated how data can be cleaned, analyzed, and presented in a way that supports data-driven decision-making.

The Power BI dashboard serves as an effective analytical tool, providing a comprehensive overview of sales performance while allowing users to drill down into specific years, quarters, and counties. The findings from this analysis can assist businesses, policymakers, and analysts in understanding market behavior and planning future strategies.

Future Scope

The scope of this project can be expanded in several ways. Additional datasets such as population, income levels, or vehicle types could be integrated to provide deeper insights into the factors influencing sales. Forecasting models could also be applied to predict future sales trends based on historical data.

Further enhancements may include more advanced predictive analytics, real-time data integration, and deeper geographic analysis at the city or zip-code level. These improvements would make the analysis more robust and valuable for long-term planning and policy formulation.