Colorado Motor Vehicle Sales

Project Presentation

Objective

- Sales Trends Analysis: Identify trends in motor vehicle sales over time monthly, quarterly. Compare sales performance across different quarters to identify seasonal variations.
- County-Level Sales: Analyse sales patterns across different counties in Colorado. Identify high-performing and low-performing counties in terms of sales revenue.
- Forecasting Future Sales: Use time series analysis and machine learning models to predict future vehicle sales. Identify factors influencing sales trends and potential growth opportunities.
- Economic and Market Insights: Offering insights that can be used for economic analysis, market research, and policy-making decisions.

Scope

- Monthly Sales Trends Analysis: Breakdown of sales data by month, with sales figures aggregated on a monthly level for each year. Identify months with the highest and lowest sales, as well as sales trends over time.
- Quarterly Sales Analysis: We will aggregate the data to explore trends by quarter and how that impacts monthly sales performance.
- County-Level Sales Analysis: Deep dive into sales data at the county level, identifying counties with the highest and lowest sales. Compare counties based on their sales over time.
- Predictive Modeling: Build predictive models using machine learning or time-series methods to forecast future sales trends for Colorado.

Scope

- Sales Frequency and Patterns: Analyzing how frequently certain sales numbers appear in the dataset allows us to identify outliers, trends, or typical sales behavior.

- Economic and Market Insights: Investigate the economic health of various counties by analyzing motor vehicle sales trends. Identify any correlations between sales trends and factors like population growth ,economic development, or infrastructure improvements.

Data Collection

1. Colorado Motor Vehicle Sales Dataset

https://www.kaggle.com/datasets/msjahid/colorado-motor-vehicle-sales-data/data

2. Form of Data
   - Data is in csv form having quantitative as well as qualitative data.

3. Rows & Columns

```
df.shape
```

```
(501, 4)
```

Data Preparation

```
import pandas as pd
df = pd.read_csv(r'C:\Users\Hp\Documents\colorado_motor_vehicle_sales1.csv')
df
```

	year	quarter	county	sales
0	2008	1	Adams	231609000
1	2008	1	Arapahoe	550378000
2	2008	1	Boulder/Broomfield	176771000
3	2008	1	Denver	200103000
4	2008	1	Douglas	93259000
...	...	...	...	...
496	2015	4	Larimer	244327000
497	2015	4	Mesa	88202000
498	2015	4	Pueblo	94606000
499	2015	4	Rest of State	157059000
500	2015	4	Weld	192583000

501 rows × 4 columns

- We used python library pandas. We import csv file into(df) dataFrame. Using read_csv method.

Data Preparation

```python
import pandas as pd
df = pd.read_csv(r"C:\Users\Hp\Documents\colorado_motor_vehicle_sales1.csv")

def quarter_to_month(year, quarter):
    if quarter == 1:
        return f"{year}-01"
    elif quarter == 2:
        return f"{year}-04"
    elif quarter == 3:
        return f"{year}-07"
    elif quarter == 4:
        return f"{year}-10"

df['month'] = df.apply(lambda row: quarter_to_month(row['year'], row['quarter']), axis=1)

df.to_csv(r'C:\Users\Hp\Documents\colorado_motor_vehicle_sales2.csv', index=False)
df.head()
```

	year	quarter	county	sales	month
0	2008	1	Adams	231609000	2008-01
1	2008	1	Arapahoe	550378000	2008-01
2	2008	1	Boulder/Broomfield	176771000	2008-01
3	2008	1	Denver	200103000	2008-01
4	2008	1	Douglas	93259000	2008-01

- Created a new month column from the quarter and year columns that is used for time series analysis.

Data Preparation

```
import pandas as pd
df = pd.read_csv(r"C:\Users\Hp\Documents\colorado_motor_vehicle_sales2.csv")

cols = list(df.columns)
cols.insert(0, cols.pop(cols.index('month')))
df = df.reindex(columns=cols)

df.to_csv(r'C:\Users\Hp\Documents\colorado_motor_vehicle_sales2.csv', index=False)
df.head()
```

	month	year	quarter	county	sales
0	2008-01	2008	1	Adams	231609000
1	2008-01	2008	1	Arapahoe	550378000
2	2008-01	2008	1	Boulder/Broomfield	176771000
3	2008-01	2008	1	Denver	200103000
4	2008-01	2008	1	Douglas	93259000

- Month column has been rearranged to the first position.

Data Preparation

```
print(df.dtypes)
```

```
month      object
year        int64
quarter     int64
county     object
sales       int64
dtype: object
```

```
df['month'] = pd.to_datetime(df['month'], format='%Y-%m')
print(df.dtypes)
```

```
month      datetime64[ns]
year                int64
quarter             int64
county             object
sales               int64
dtype: object
```

Data Preparation

```
df.isnull().sum()
```

```
month      0
year       0
quarter    0
county     0
sales      0
dtype: int64
```

Here, We checked data is null or not using isnill() function

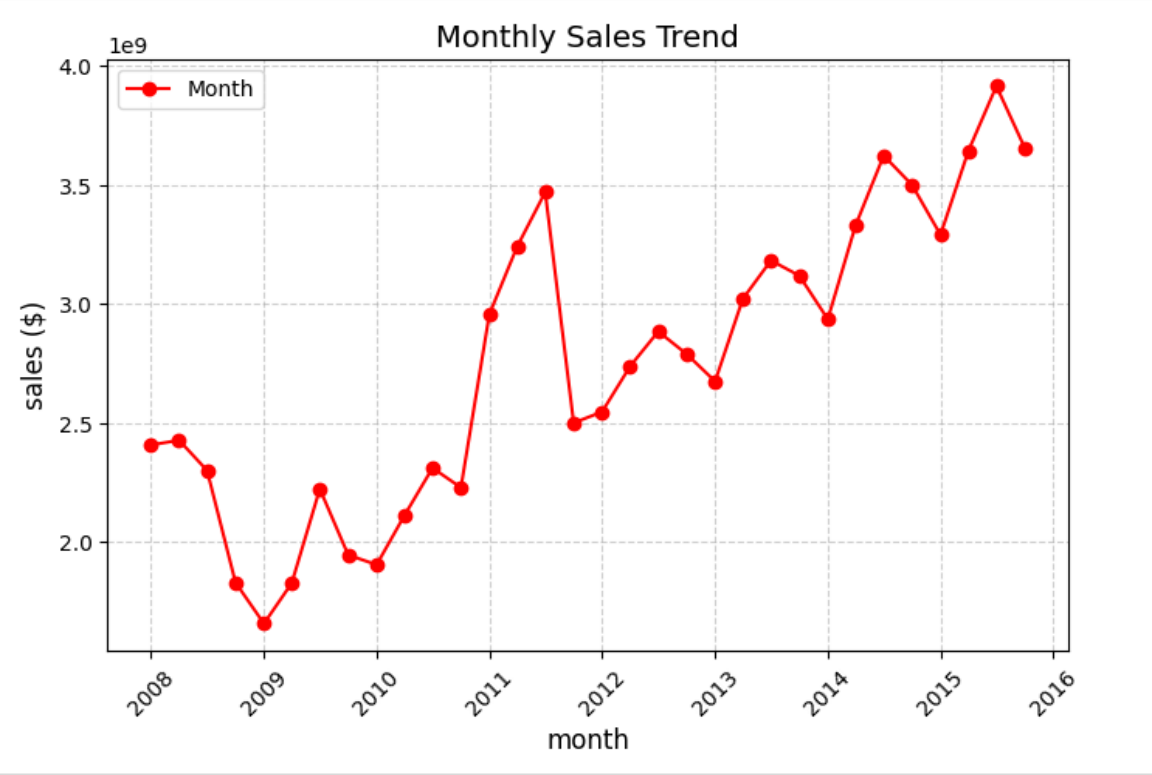
Exploratory Data Analysis

- In the EDA of the Colorado Motor Vehicle Sales data, we aim to uncover underlying patterns, trends, and distributions within the dataset.
- Monthly Sales Trends Analysis: Identify months with the highest and lowest sales, as well as sales trends over time.
- County-Level Sales Analysis: Deep dive into sales data at the county level, identifying county with the highest and lowest sales. Compare counties based on their sales time.

Exploratory Data Analysis

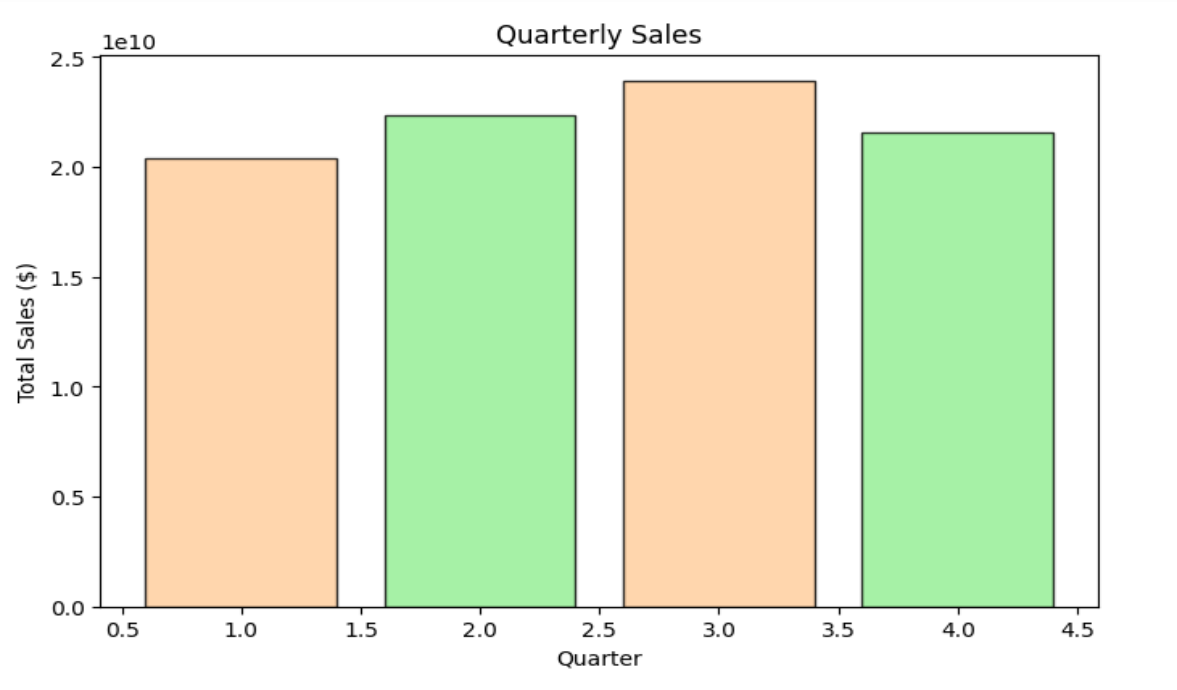
- Quarterly Sales Analysis: We will aggregate the data to explore trends by quarter and how that impacts monthly sales performance.
- Sales Frequency: Analyzing how frequently certain sales numbers appear in the dataset allows us to identify outliers, trends, or typical sales behavior.

Monthly Sales Trend



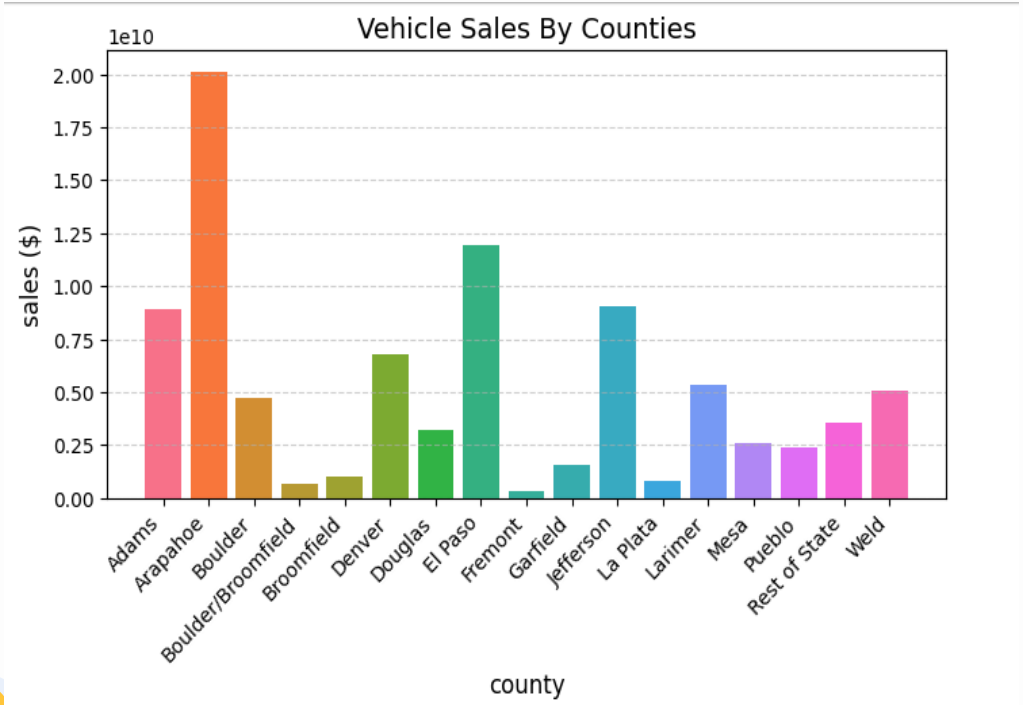
- Monthly Sales Trend : Identify months with the highest and lowest sales, as well as sales trends over time.

Quarterly Sales Trend



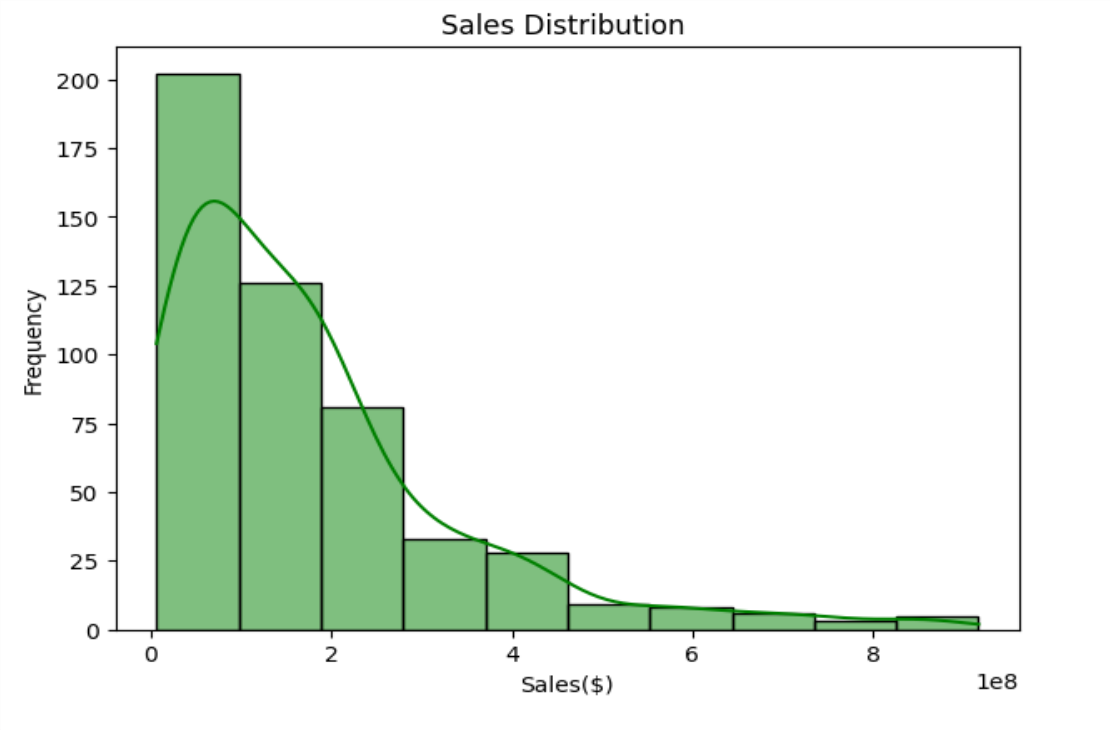
- Quarterly Sales Trend : Compare sales performance across different quarters to identify seasonal variations.

Vehicle Sales Across Counties



- Vehicle Sales Across counties : We analysed sales patterns across different counties in Colorado. Identified high-performing and low-performing counties in terms of sales revenue.

Sales Frequency



- Sales Frequency: This graph illustrates the relationship between sales and frequency.

Statistical Analysis

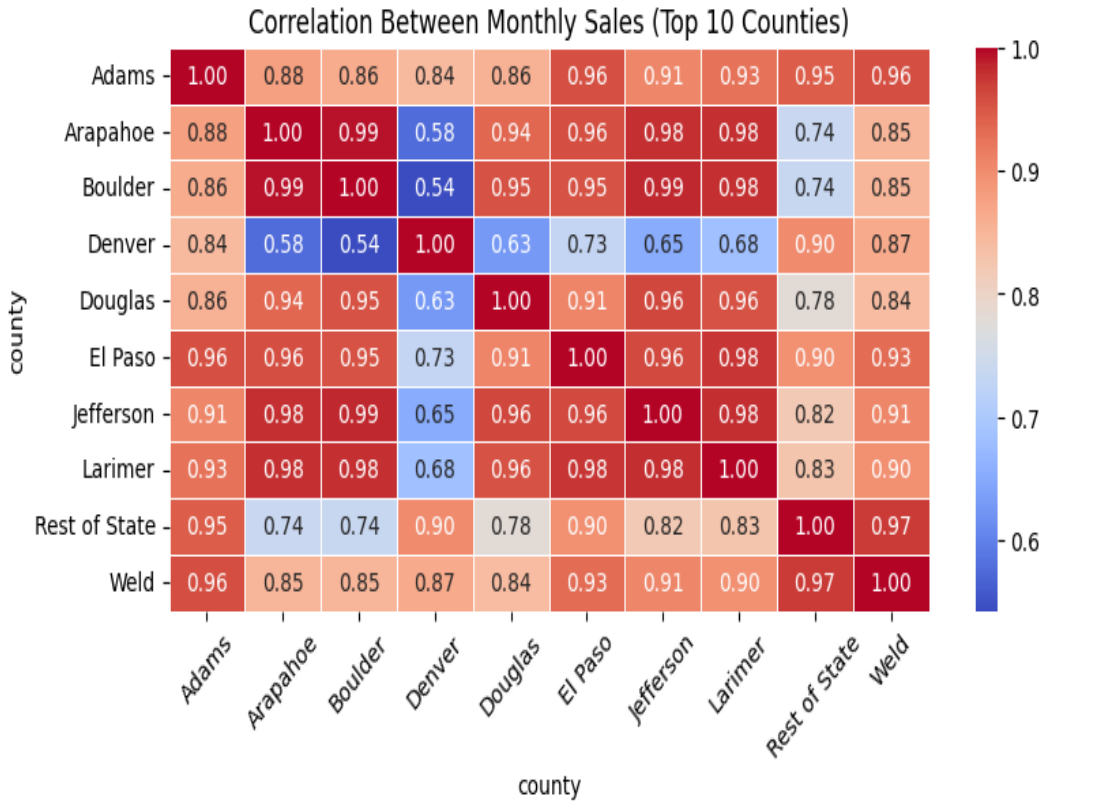
- Perform statistical analysis to understand trends and correlations in the data. Identify correlations between month, sales and county.

```
df.describe()
```

	year	quarter	sales
count	501.000000	501.000000	5.010000e+02
mean	2011.570858	2.502994	1.760585e+08
std	2.266599	1.120041	1.642055e+08
min	2008.000000	1.000000	6.274000e+06
25%	2010.000000	2.000000	6.148200e+07
50%	2012.000000	3.000000	1.385820e+08
75%	2014.000000	4.000000	2.241580e+08
max	2015.000000	4.000000	9.169100e+08

- Here, we define Statical information of quantitative data. Like min, max, mean, count, standard deviation.

Correlation Matrix



- This correlation matrix displays a heatmap of the correlations between monthly vehicle sales among the top 10 counties in Colorado, helping us understand the relationship between sales trends.
- High Correlation
- Moderate Correlation
- Low Correlation

Predictive Modeling

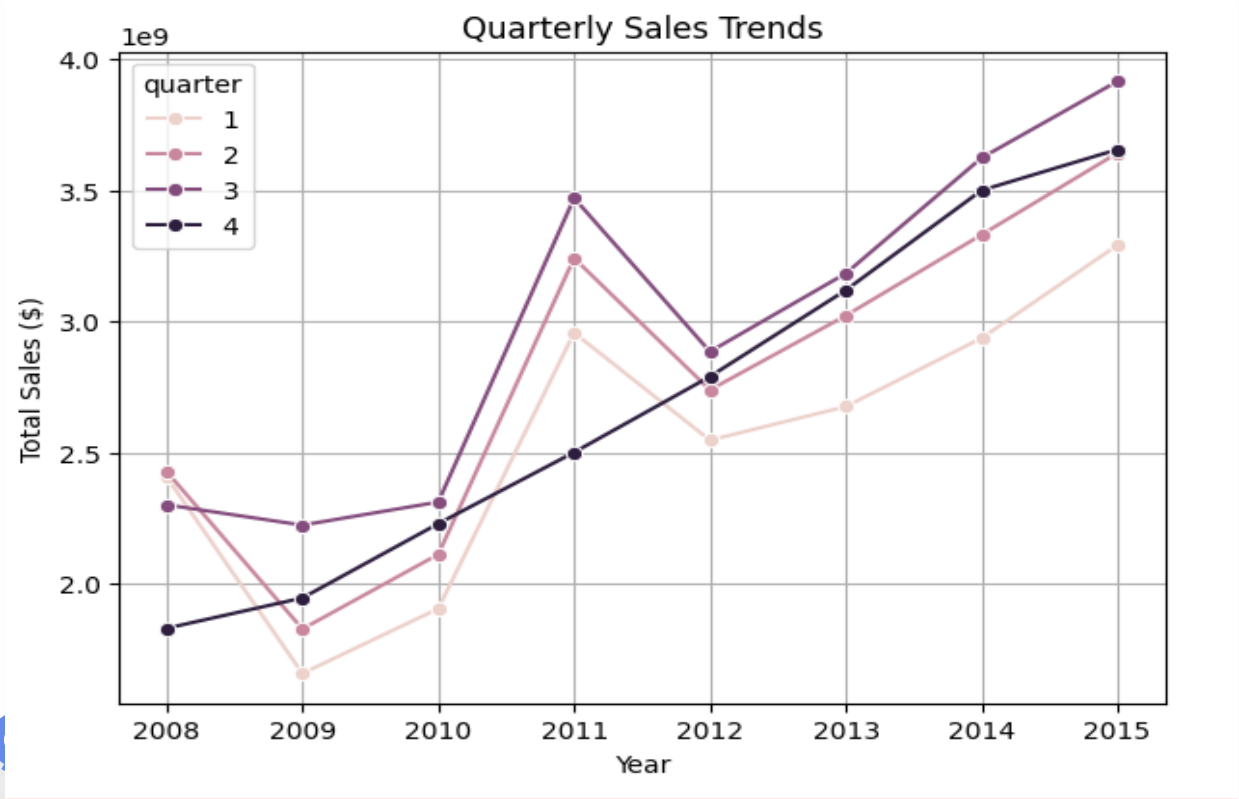
- Build predictive models using machine learning or time - series methods to forecast future sales trends for Colorado.
- Sales Trends Analysis: Group sales by quarter and year and visualize trends over time. Use sns.lineplot() for a smooth representation of quarterly sales trends.
- Counties Sales Distribution: Analyze sales across counties and identify top-performing counties. Plot the counties based on total sales to highlight the ones that contribute most.
- Time Series Forecasting: Fit an ARIMA model and forecast future sales. Revert differencing using cumsum().

Predictive Modeling

- Machine Learning Forecasting (Random Forest): Include county as a feature for predicting sales. Use Random Forest regression to model sales based on county and month features. Plot actual vs. predicted sales to evaluate the model's performance.

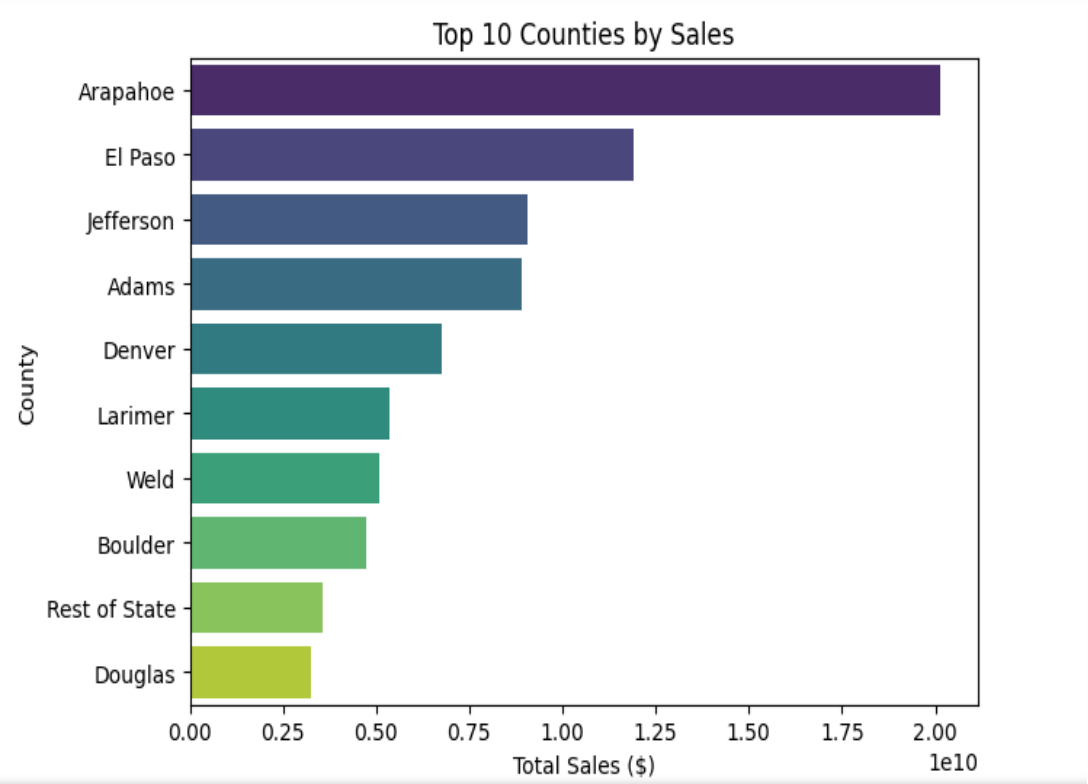
- Identify Influencing Factors (Linear Regression): Use a linear regression model to understand how month and county influence the sales trends. Display the model coefficients to identify which factors have the most impact on vehicle sales.

Quarterly Sales Trends



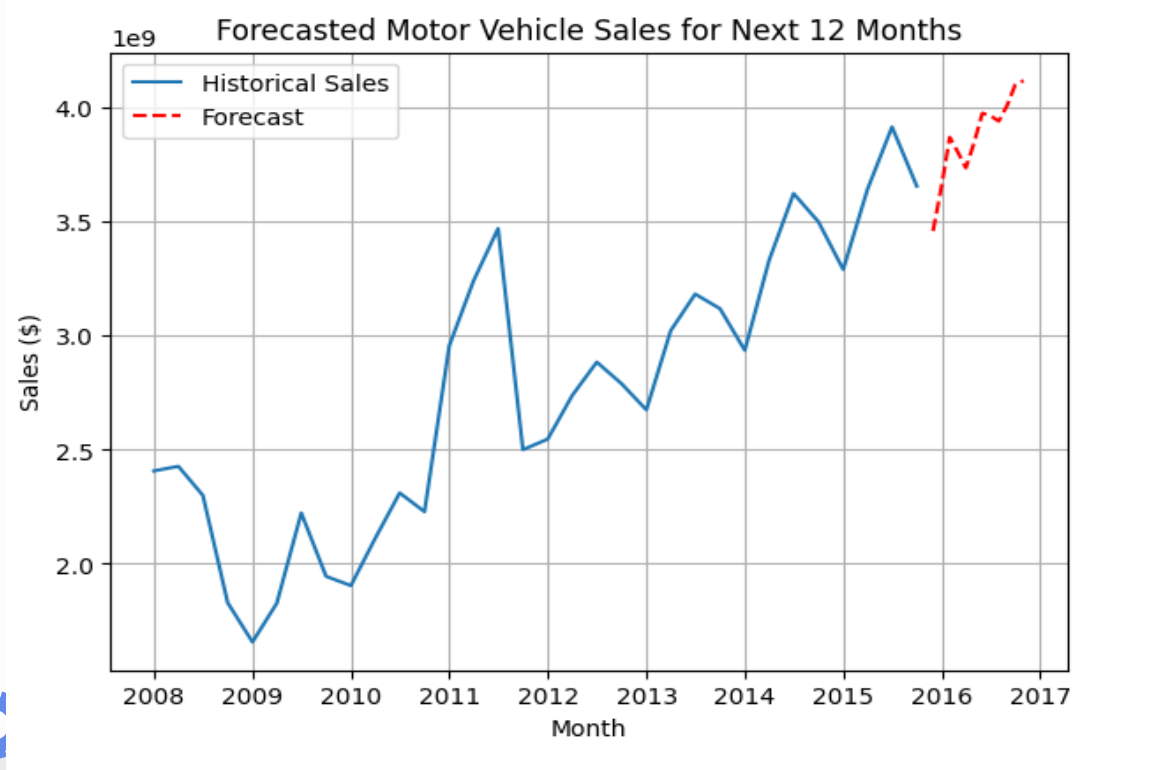
- Group sales by quarter and year. Use sns.lineplot() to visualize quarterly sales trends.

Top 10 Counties by Sales



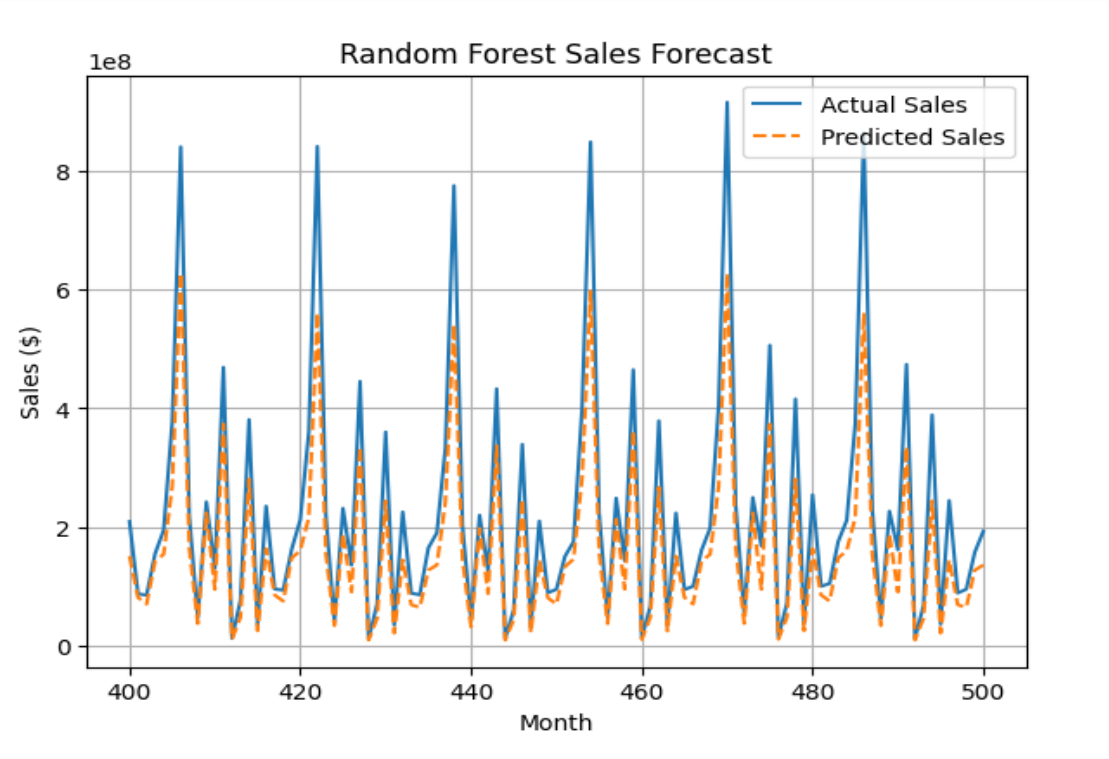
- Here, We analysed sales by county, Identified top-performing counties, and Plotted counties by total sales.

Forecasted Sales for Next 12 Months



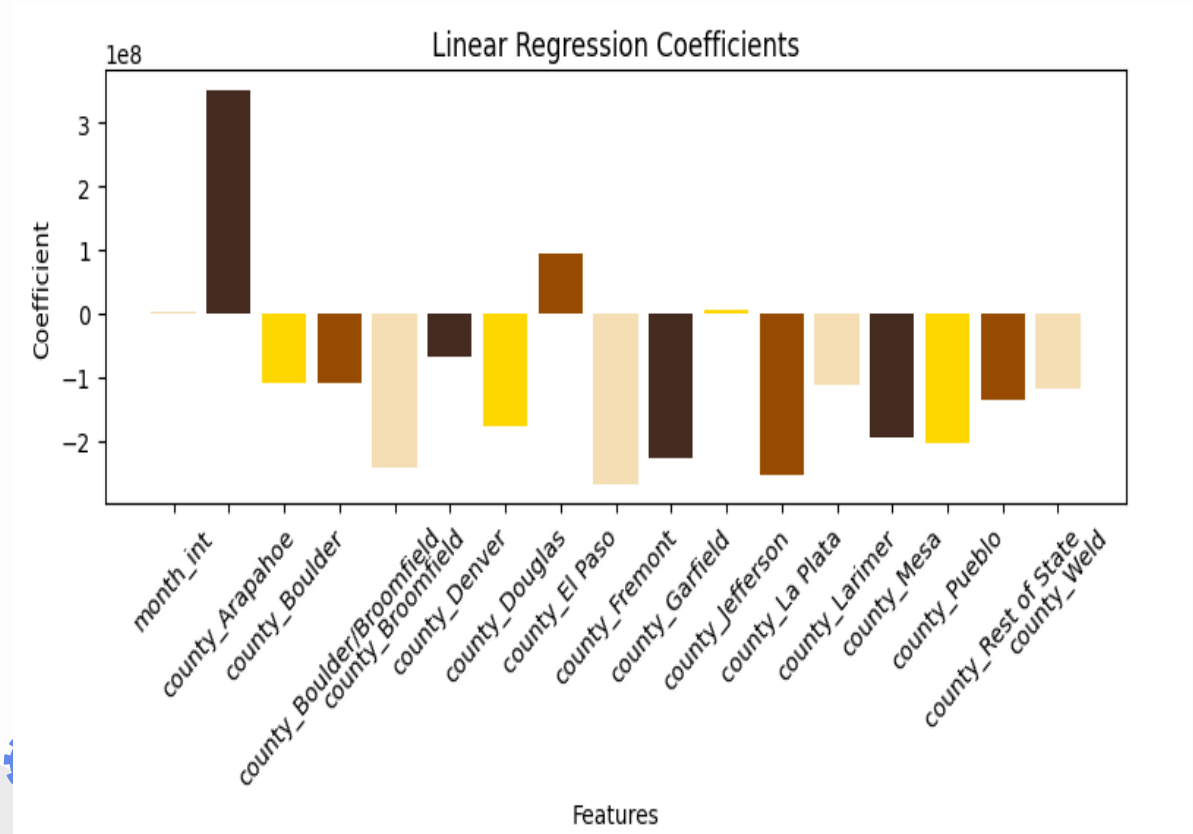
- Fit an ARIMA model and forecast future sales. Revert differencing using cumsum().

Random Forest Sales Forecast



- Include county as a feature. Use Random Forest regression to predict sales. Plot actual vs predicted sales.

Identify Factors Influencing Sales Trends



- Use linear regression to understand the influence of month and county on sales trends. Display model coefficients.

Conclusion

Summary of Achievements

The Colorado Motor Vehicle Sales Analysis successfully uncovered keys trends, seasonality patterns, and county-level sales to provide actionable insights for the industry.
Through EDA, time series forecasting ARIMA, and machine learning models Random Forest, Linear Regression, we identified:

- Seasonal Trends: Sales peak in Q3 (July-September) and Q4 (October-December), aligning with promotional periods and market demand.
- Top-Performing Counties: Certain counties significantly contribute to total sales, highlighting regional market strengths.

Summary of Achievements

- Correlation Analysis: Some counties show strong interdependence in sales patterns, which can be leveraged for regional marketing strategies.
- Sales Forecasting: The ARIMA model and Random Forest regression provide reliable future sales predictions, aiding inventory and sales planning.
- Influencing Factors: County and month play a significant role in determining sales, as confirmed through Linear Regression analysis.
- Data-Driven Decision Making: The findings support better marketing strategies, resource allocation, and business forecasting, leading to improved operational efficiency.

Thank you