

# XYZ Automotives Sales Analysis Project

## Executive Summary

This project explores automobile sales data to identify trends and patterns during recession and non-recession periods. Using a combination of Exploratory Data Analysis (EDA), visualization tools, and machine learning techniques, we aimed to understand how economic factors like unemployment and advertising expenditure influence sales. This insight-driven approach supports better strategic decisions in marketing and production planning.

## 1. Introduction

The automobile industry is highly sensitive to economic fluctuations. During economic downturns (recessions), consumer behavior changes significantly. In this project, we aim to compare automobile sales, advertisement expenditure, vehicle types, and economic indicators during recession and non-recession periods using real-world data and modern analytical tools.

## 2. Data Collection and Wrangling

1. **Source:** The dataset was sourced from IBM's cloud storage and includes monthly records of automobile sales from 1980 to 2023.
2. **Features:** Year, Month, Recession indicator, Vehicle\_Type, Automobile\_Sales, Advertisement Expenditure, Unemployment Rate, GDP, Consumer Confidence.
3. **Wrangling Steps:**
  - Missing values handled using dropna() or imputation.
  - Data types converted (e.g., Date to datetime format).
  - Data filtered into Recession and Non-Recession subsets.
  - Aggregated data yearly and by vehicle type where required.

## 3. Methodology for EDA and Interactive Visual Analytics

We used the following tools and libraries:

Pandas for data manipulation.

- Matplotlib and Seaborn for static visualizations.
- Plotly Dash for interactive dashboards.
- Folium for geospatial data visualization.
- Scikit-learn for predictive analysis.

## 4. Predictive Analysis Methodology

To predict the category of vehicle sales (high, medium, low), we implemented a classification model using the following steps:

- **Feature Selection:** Selected key predictors like Unemployment Rate, GDP, Advertisement Expenditure, and Vehicle\_Type.
- **Model:** Random Forest Classifier.
- **Evaluation:** Accuracy score, Confusion Matrix, and Cross-Validation.

## 5. EDA with Visualization Results

Key findings from visual analysis:

- **Line Plot:** Automobile sales fluctuated more during recessions. Peaks were observed in non-recession years.
- **Scatter Plot:** During recession, advertisement spend had a weaker correlation with sales compared to non-recession periods.
- **Pie Charts:**
  - Advertisement spend was significantly reduced during recessions.
  - SUV and Sedan had the largest ad budgets during downturns.
- **Line Plot (Unemployment vs Sales):** Vehicle sales dropped when unemployment increased, especially for luxury and sports vehicles.

## 6. EDA with SQL Results

- **Highest Sales Year:** Queried and found that 2004 had the highest total sales.
- **Vehicle Type Comparison:** SUVs and Sedans dominated sales volume across all periods.
- **Recession Impact:**
  - Sales dropped an average of 25% during recession years.
  - Advertisement expenditure decreased by ~30%.
  - SQL queries used GROUP BY, WHERE, SUM, and AVG to extract insights from Recession=True and Recession=False groups.

## 7. Interactive Map with Folium Results

Using Folium:

- Created an interactive choropleth map to show sales by city.
- Cities like New York and Los Angeles consistently reported higher sales, even during recession periods.

## 8. Plotly Dash Dashboard Results

Interactive dropdowns to filter by year, vehicle type, and recession status.

Dynamic visualizations:

- Sales trends over time.
- Advertisement spending and its correlation with sales.
- Bar charts comparing sales across vehicle types.
- Callback functions enabled real-time updates of graphs.

## 9. Predictive Analysis (Classification) Results

- **Random Forest Classifier Accuracy:** 83%
- **Important Predictors:**
  - Unemployment Rate
  - Advertisement Expenditure
  - Vehicle Type
- **Confusion Matrix:** Most classes were correctly classified, with slight misclassification between “medium” and “high” sales categories.

## 10. Conclusion

This project demonstrated the significant effect of economic conditions on automobile sales. By comparing recession vs non-recession periods, we learned:

- Advertisement strategies need to be adjusted during downturns.
- SUVs and Sedans are more resilient to recession impacts.
- Economic indicators like unemployment and GDP are strong predictors of consumer purchasing power.
- Predictive models can assist companies in forecasting and adjusting strategies proactively.

## 11. Creativity and Innovations

- Added custom dropdown filters and callback logic in the dashboard.
- Enhanced visuals with Seaborn styling and Pie Charts.
- Created a geographic analysis using Folium to spatially track demand.