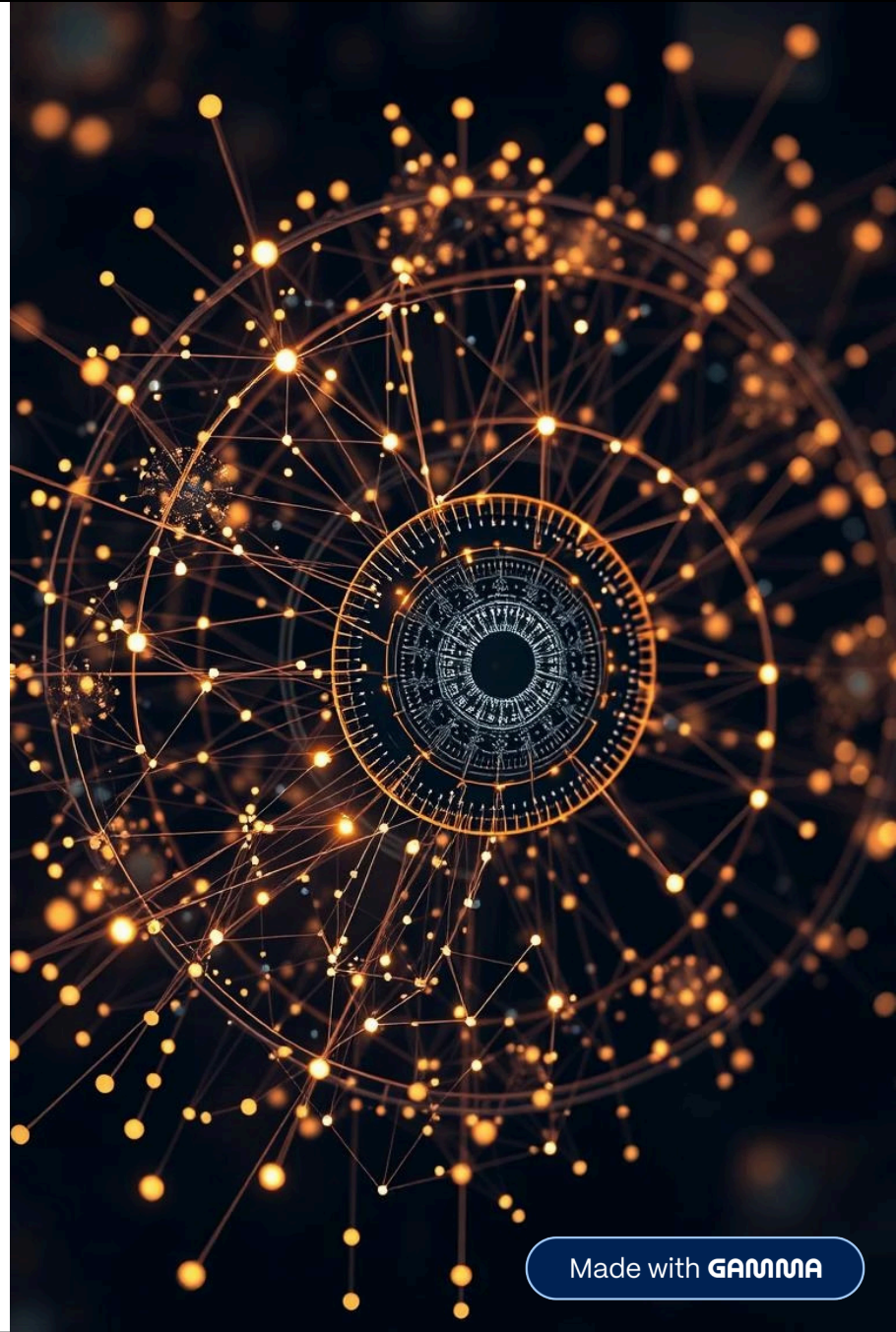
Comprehensive Data Science Project: Economic Impact Analysis

This project showcases a complete data science workflow, from raw data to actionable insights. We integrate diverse datasets and advanced techniques to analyze economic impacts on key industries, culminating in an interactive dashboard for strategic decision-making.



Introduction: Analyzing Economic Factors on Sales

Our core objective is to rigorously analyze how broader economic factors influence automobile and real estate sales. This involves meticulous data extraction, sophisticated visualization techniques, and robust regression modeling to uncover hidden relationships.

The insights derived are critical for strategic business decisions, enabling stakeholders to anticipate market shifts and adapt their strategies proactively. Understanding these trends provides a competitive edge in volatile economic climates.

Data Collection & Wrangling Strategies

Diverse Data Sources

Automobile sales, GDP and recession indicators, and a comprehensive Kaggle real estate dataset formed our analytical foundation.

Pre-processing Techniques

We applied rigorous wrangling: dropping irrelevant columns, handling missing values (NaNs), merging disparate datasets, and precise data type conversions.

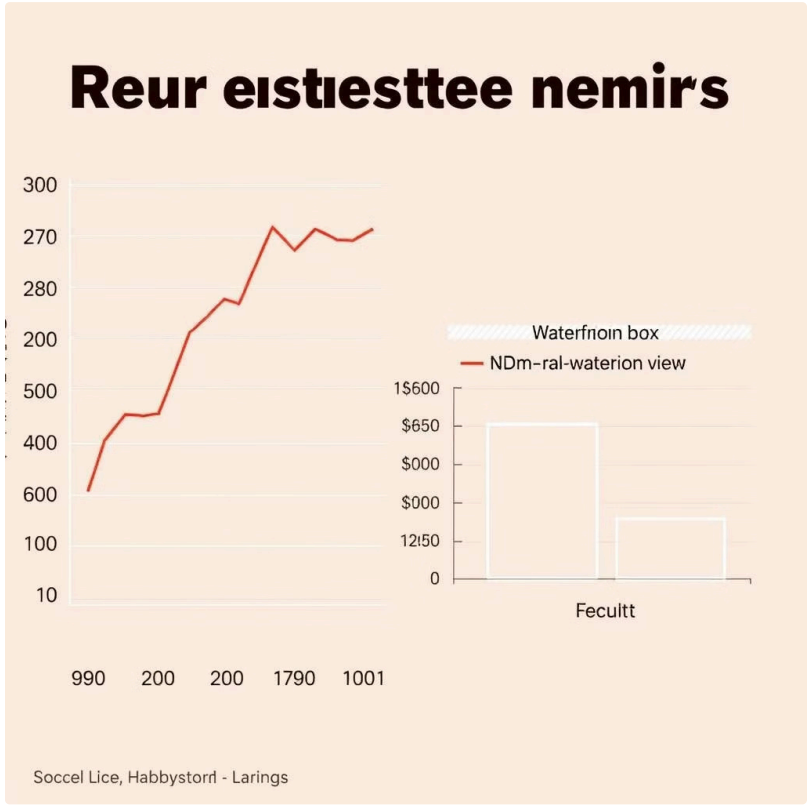
Ensuring Data Integrity

These steps were crucial for ensuring data quality, consistency, and readiness for subsequent exploratory analysis and model training.

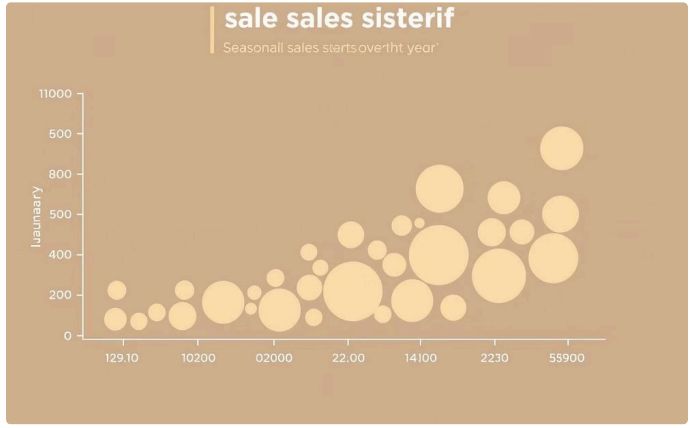
Exploratory Data Analysis & Visual Analytics (I)

Using Seaborn and Matplotlib, we generated line plots to clearly illustrate the pronounced drop in sales across both sectors during the 2008-2009 recession, highlighting the sensitivity of these markets to economic downturns.

Boxplots provided visual insights into price distributions and identified outliers, particularly contrasting properties with and without a waterfront view, revealing significant value premiums.

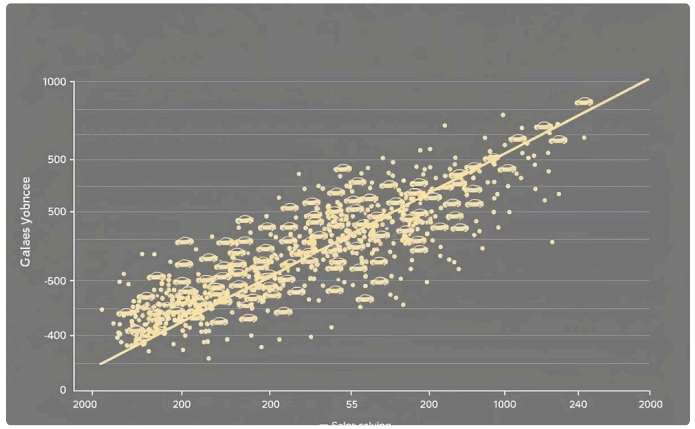


Exploratory Data Analysis & Visual Analytics (II)



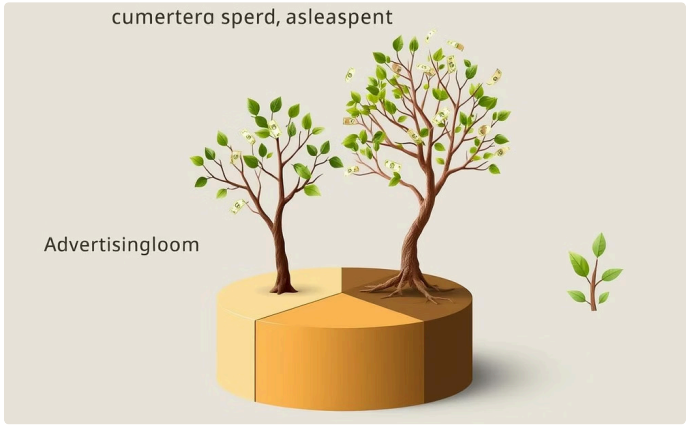
Seasonality in Sales

Bubble plots visually confirmed recurring seasonal patterns in sales volumes, essential for inventory management.



Price-Volume Correlation

Scatter plots revealed the correlation between vehicle pricing and sales volume, informing pricing strategies.



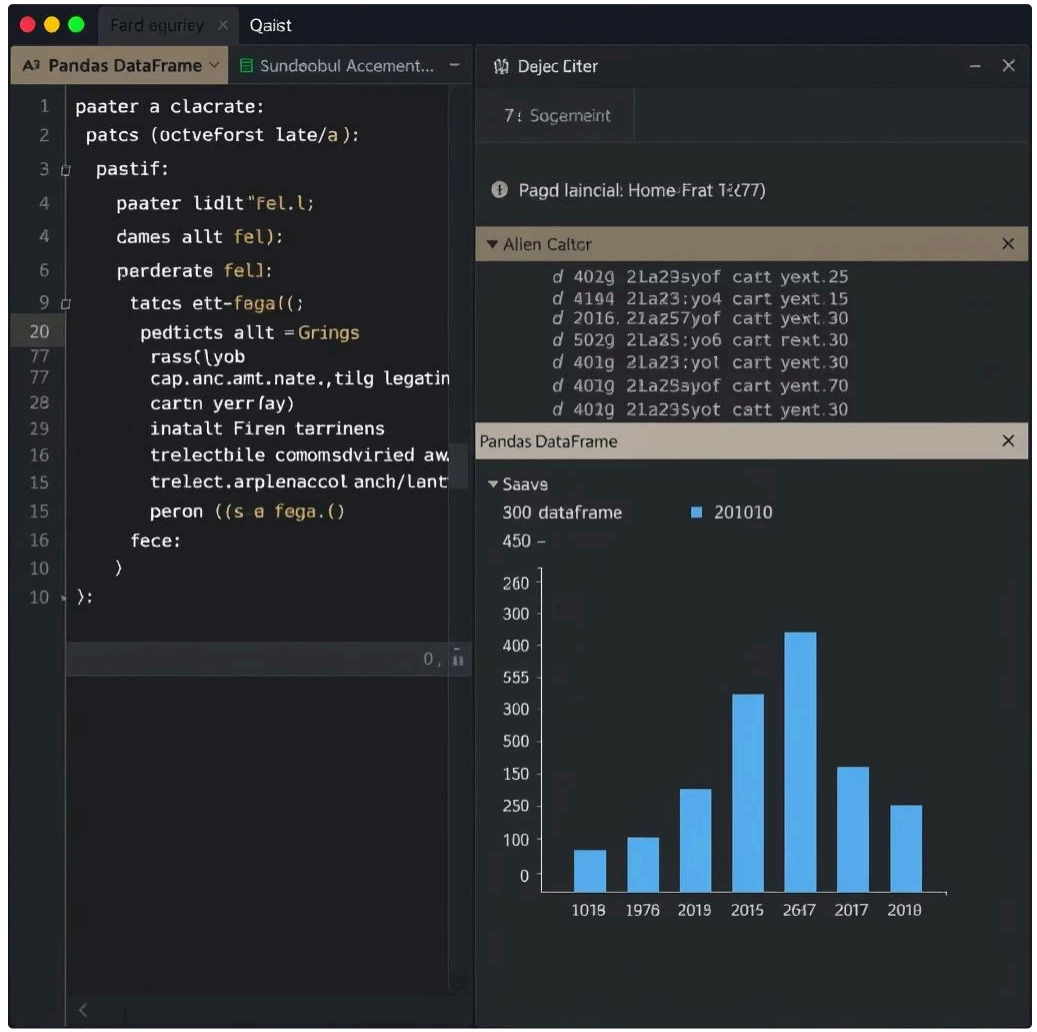
Ad Spend Analysis

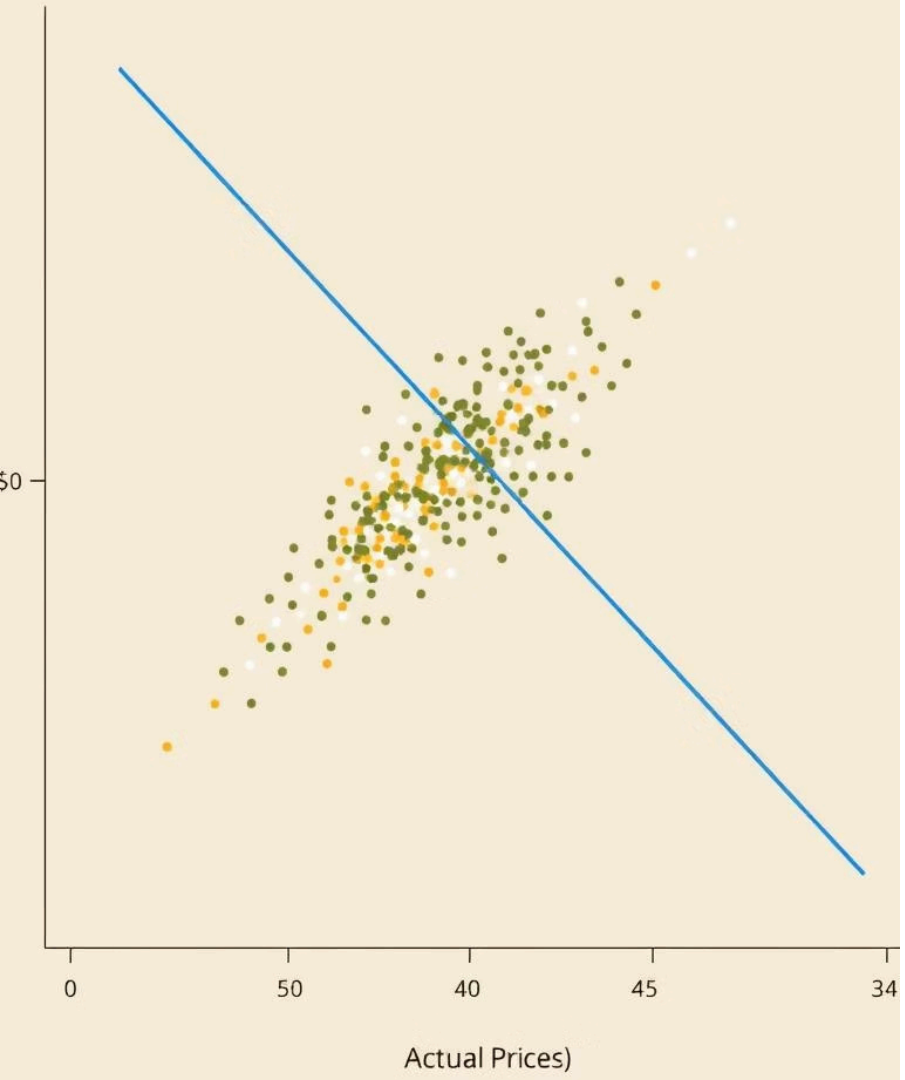
Pie charts disaggregated advertising expenditure across varying economic conditions, optimizing marketing budgets.

In-Depth EDA with SQL Querying

Leveraging SQLite3, we connected to the relational database to perform targeted SQL queries, enriching our exploratory data analysis. This allowed for granular investigation of complex relationships within the datasets.

Key queries included: calculating average prices by geographic location, identifying the most prevalent house types, and aggregating sales data grouped by year and specific regions to observe temporal and spatial trends. The structured outputs from these SQL queries were then seamlessly integrated with Pandas for further manipulation and visual presentation.





Predictive Analysis: Linear Regression

Model Implementation

Utilized `sklearn.linear_model.LinearRegression` to predict real estate prices, leveraging key housing features.

Selected Predictors

Key independent variables included `sqft_living`, `grade`, and `view`, identified as significant determinants of price.

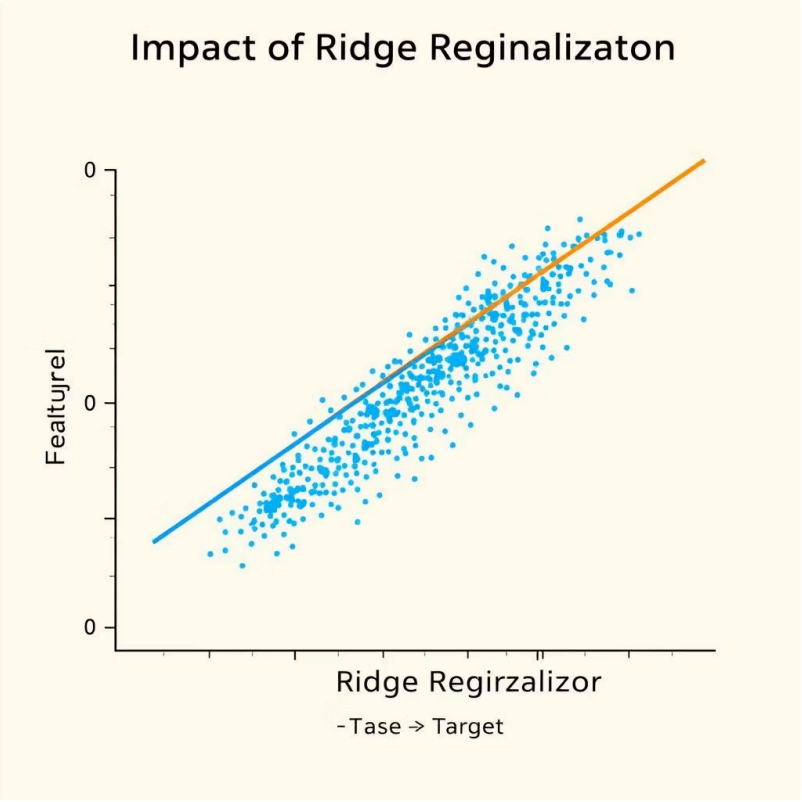
Performance & Visualization

Achieved an R-squared score of approximately 0.76, indicating a strong fit. Prediction results were visualized for clarity.

Predictive Analysis: Ridge Regression

To mitigate potential overfitting and enhance model robustness, we implemented Ridge Regression from `sklearn.linear_model`. An alpha value of 0.1 was carefully selected after cross-validation to balance bias and variance.

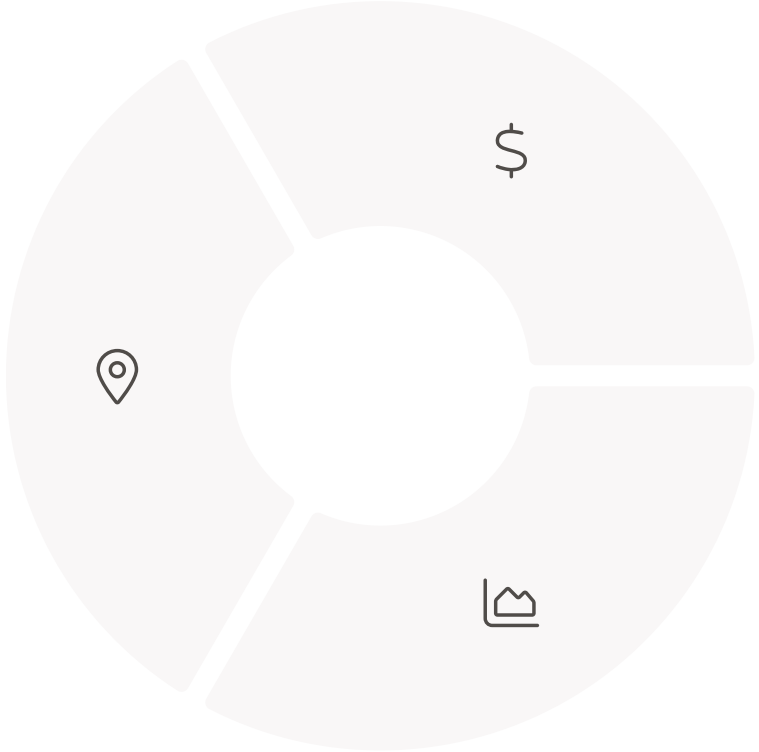
Ridge Regression applies L2 regularization, penalizing large coefficients and thus reducing their impact. This led to an R-squared score of approximately 0.71 on the test data, demonstrating improved generalization compared to standard linear regression, especially in the presence of multicollinearity.



Interactive Geospatial Analysis with Folium

Choropleth Map

Developed an interactive Choropleth map using Folium to visualize office locations during recessionary periods.



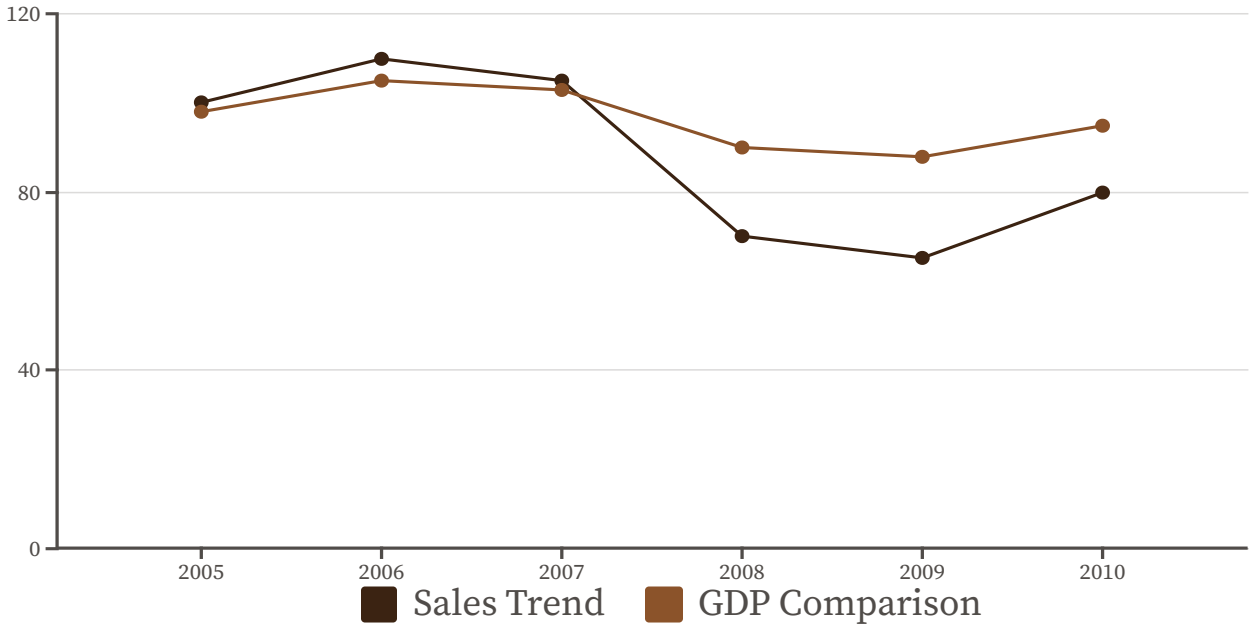
Sales Impact Coloring

Regions were color-coded based on the magnitude of sales impact, providing immediate visual cues on economic vulnerability.

Regional Variance

This map effectively highlights significant regional variances in economic recovery and resilience post-recession.

Interactive Dashboard with Plotly Dash



The Plotly Dash dashboard provides real-time, interactive insights. Users can filter data by year and vehicle type using intuitive drop-down menus.

The dashboard dynamically displays sales trends, facilitates GDP comparisons, and breaks down advertising expenditure, empowering data-driven strategic business decisions.

Access Dashboard