Analysis of Annual Car Model Sales Data

Introduction

This project analyses annual car model sales data to identify trends and relationships between key numerical variables.

The aim of the analysis is to:

- Examine how sales change over time
- Investigate the relationship between price and demand
- Identify patterns in engine size and fuel efficiency

The dataset contains over 200 real observations.

Dataset Overview

- The dataset contains over 200 rows of real car sales data.
- Each row represents a car model sold in a specific year.
- The dataset includes at least four numerical variables.
- The data was cleaned before analysis to remove errors and inconsistencies.

Variables Used

The key variables analysed in this project are:

- Units_Sold – Number of cars sold annually
- Price – Average selling price
- Engine_Size – Engine size in litres
- Year – Year of sale
- Fuel_Efficiency – Miles per gallon (MPG)

These variables allow for quantitative analysis of trends and relationships.

Mind map

Annual model performance study

Dataset Overview

- 200+ real observations
- Multiple numeric variables
- Cleaned and structured dataset
- No missing values in final version

Key Variables

- Units_Sold
- Price
- Engine_Size
- Fuel_Efficiency
- Year

These variables allow trend and relationship analysis.

Data Preparation

- Removed duplicates
- Standardised column names
- Checked data types
- Verified numeric accuracy

Ensured reliability before analysis.

Sales Trends

- Identified peak sales years
- Observed growth/decline patterns
- Compared performance across time
- Analysed total annual sales

Price vs Demand

- Examined correlation
- Higher price → generally lower sales
- Identified exceptions
- Measured strength of relationship

Website & Visualisation

- Built using HTML
- Styled with Bootstrap
- Charts created using Chart.js
- Deployed on GitHub Pages

Interactive presentation of results.

Data Cleaning Process

Before analysis, the dataset was cleaned:

- Removed duplicate rows
- Standardised column names (no spaces or errors)
- Checked for missing values
- Ensured numeric data types were correctly formatted

This ensures accuracy and reliability in the results.

Descriptive Statistics

Descriptive statistics were calculated to summarise the data:

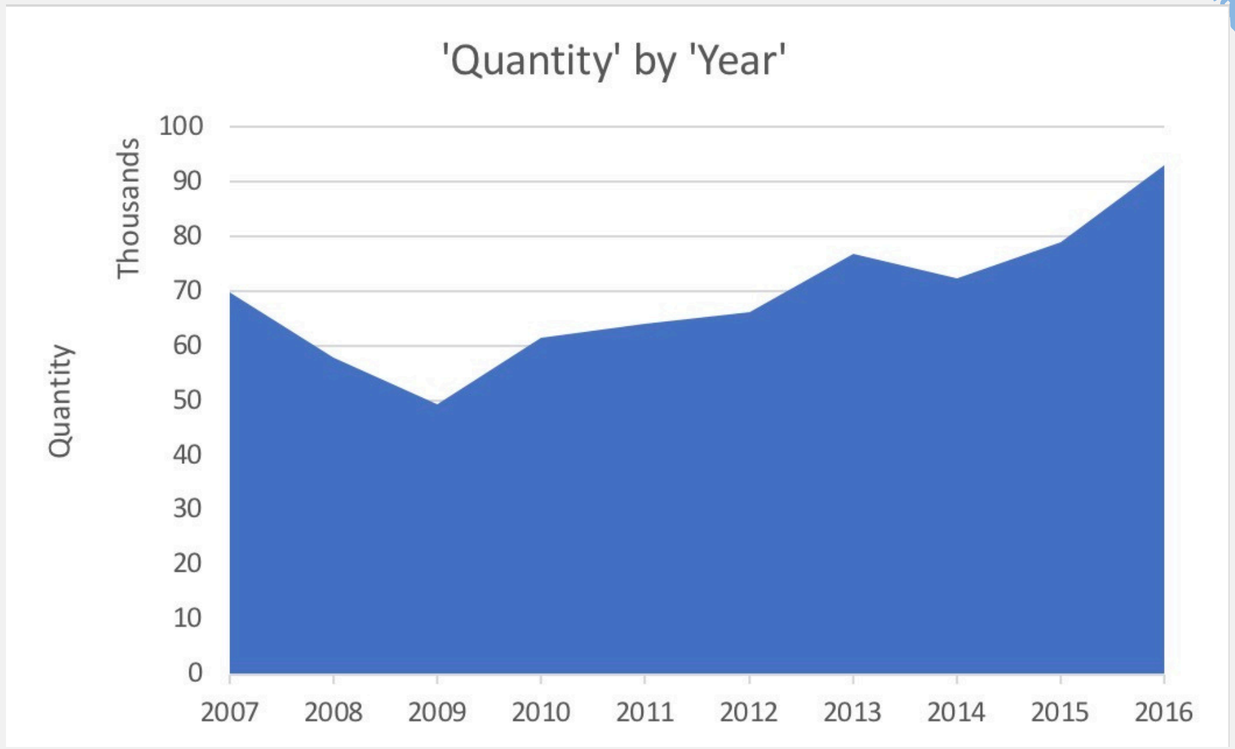
- Mean and median sales
- Highest and lowest selling models
- Average prices across years
- Variation in engine sizes

This provided an overview of overall performance and distribution.

Sales Trends Overtime

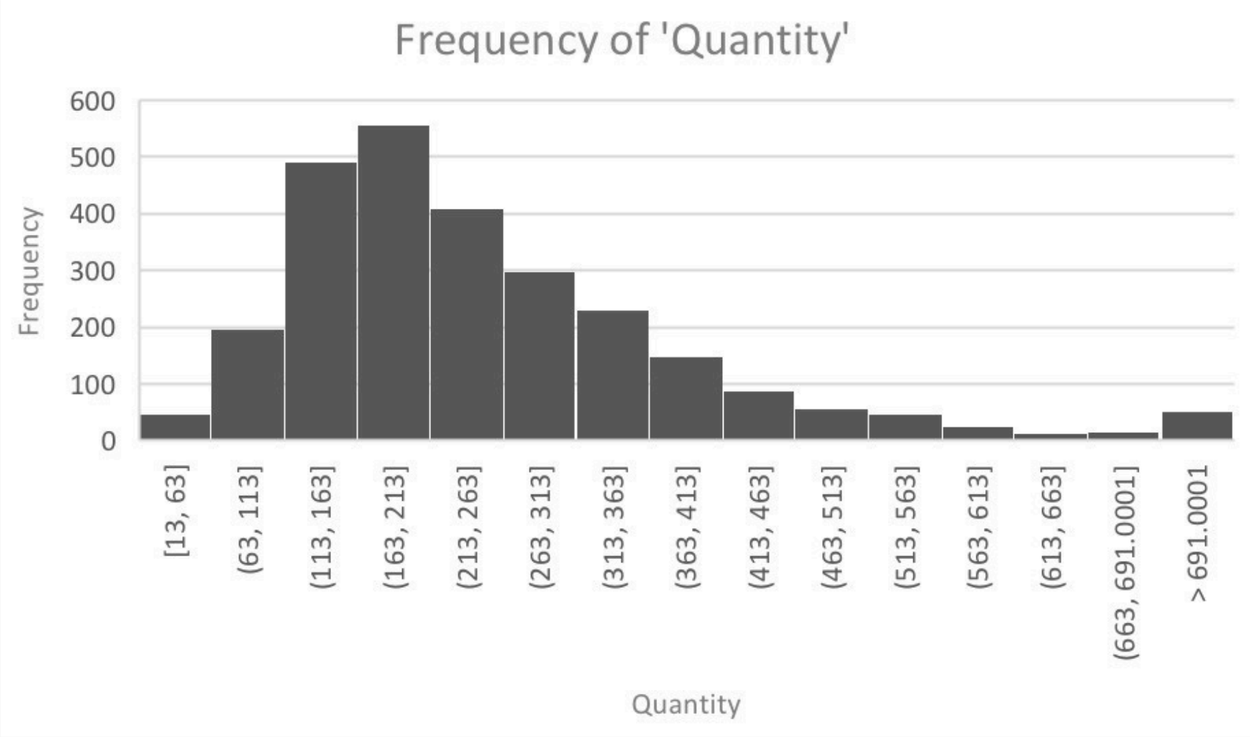
- Sales decreased between 2007 and 2009.
- From 2010 onwards, sales show a steady upward trend.
- There is slight fluctuation around 2013–2014.
- 2016 records the highest sales level in the dataset.

This indicates overall market growth after an early decline.



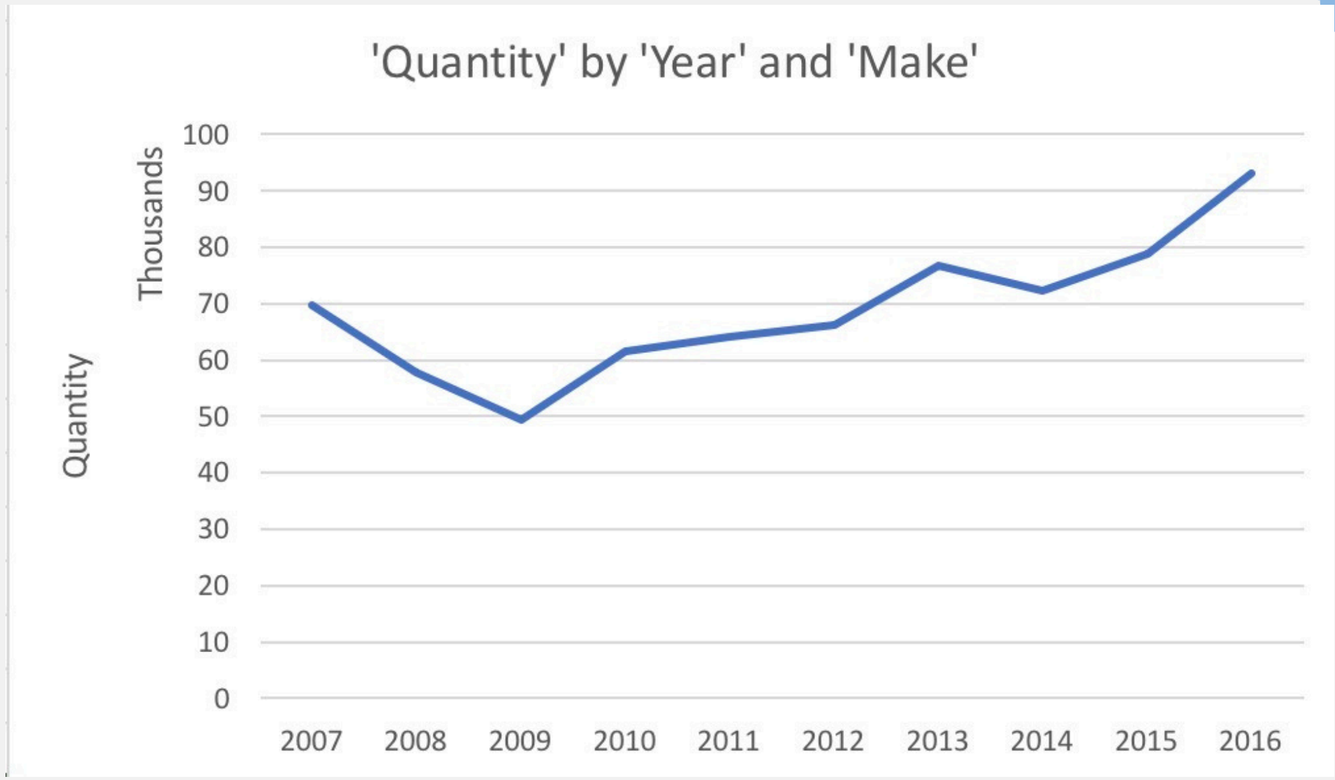
Distribution Of Units Sold

- Most sales values fall within the mid-range categories.
- Fewer models achieve extremely high sales levels.
- The distribution is slightly right-skewed.
- High-sales models are less frequent but significantly impact totals.



Quantity Sold by Year and Make

- Sales declined between 2007 and 2009, reaching a low point in 2009.
- From 2010 onwards, sales steadily increased.
- There was a slight dip around 2014.
- 2016 recorded the highest quantity sold in the period.
- Overall trend shows strong market growth over time.



Ethics and limitations

Ethical Considerations:

- The dataset contains aggregated sales data only.
- No personal or sensitive information was used.
- The analysis avoids misleading visual representations.

Limitations:

- The dataset may not represent all global car markets.
- External factors (e.g., economic conditions) were not included.
- Results are limited to the selected variables.

Thank you very much!