

Used Car Data Analysis Project

by Sam Buwalda | Portfolio Project, 2025

Dataset Description

This project uses a large, publicly available dataset containing 426,880 used car listings scraped from Craigslist across the United States. The dataset was sourced from Kaggle, and is commonly used in the data analytics community due to its size, realism, and relevance. It provides a reliable starting point for showcasing analytical skills within a practical business context.

- **Link to raw data source:** [Kaggle - Used Car Dataset](#)
- **Link to data files used:**
 - **Raw dataset:** <https://drive.google.com/file/d/1TOBFqXA3n7xa0W-MI7C9dF7JRIE1OCTy/view?usp=sharing>
 - **Cleaned dataset:** <https://drive.google.com/file/d/1cRSfon5fKlvNKiB6BffyfoM-im7p92me/view?usp=sharing>
- **Size:** 426,880 rows and 26 columns (before cleaning)

The dataset includes a wide range of real-world variables, making it well-suited for analyzing used car pricing, brand value retention, and vehicle traits. Key features used in this project include:

- **price** – Listed sale price
- **year** – Year of manufacture
- **manufacturer** – Car brand
- **fuel** – Type of fuel (e.g., gas, diesel, electric)
- **transmission** – Transmission type (automatic or manual)
- **odometer** – Mileage reading
- **type** – Vehicle type (e.g., sedan, SUV, truck)
- **state** – U.S. state where the car was listed

The dataset's volume, complexity, and diverse set of features make it ideal for applying core data analytics techniques—ranging from descriptive statistics to predictive modeling—in a realistic and business-oriented setting.