

Welcome to use the DVM-Car dataset! This dataset is a UK automotive market based research dataset created for business analytics and applications purposes. Specifically, it can be used for topics such as car appearance design, consumer analytics and used for used car sales modelling.

Structure of the dataset

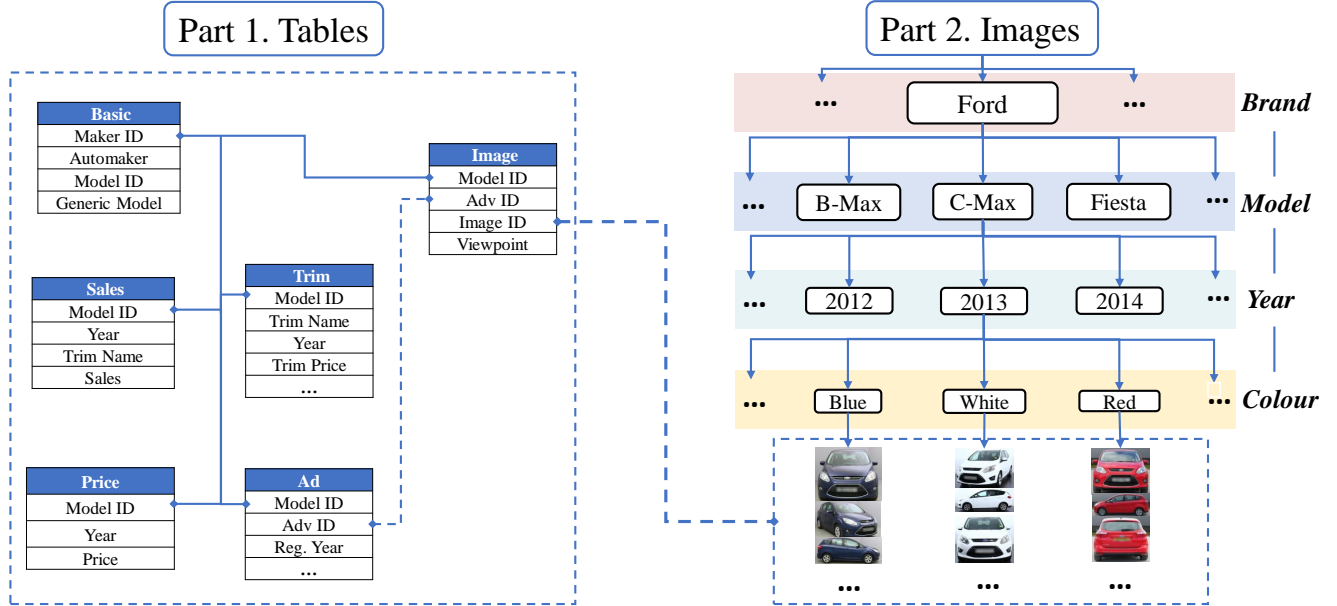


Figure 1: The structure of DVM-Car dataset. This dataset is comprised of two parts: 1.4 million images and six tables.

As the above figure shows, the DVM-Car is comprised of two main parts: the image part (88GB in a single zip file) and the table part (six CSV tables).

Images: Our dataset contains 1,430,096 car images that deliberately collected from automotive classified advertising platforms. It covers 897 car models which have been sold in the UK market over the last 19 years. All provided images are stored under the categorisation “*brand-model-year-colour*”. This structure allows researchers to easily locate images. A table for image indexing (i.e., Image Table) is provided.

Tables: The model/car specification information is provided in the data tables. Following are the details of these tables.

- **Basic Table.** This table is mainly for indexing other tables. It includes 983 generic models from 100 automakers.
- **Sales Table.** It contains over 19 years car sales data of the UK market (based on the government released statics of new car registrations). In sum, it covers the sales of 735 car models from 2000 to 2018.
- **Price Table.** It is designed for users who only need the basic price data of historical models. This table contains the entry-level (i.e., the cheapest trim price) new car prices of 636 models since 1998. For users who are interested in the specific prices of trims, they can use the price values in the trim table.

-
- *Trim Table*. It includes 0.6 million trim-level information such as selling price, fuel type and engine size.
 - *Ad Table*. It shows more than 0.25 million used car advertisements information collected from on-line sources. It consists of variables like the advertisement's creation time and the car's registration year, cumulate mileage and selling price.
 - *Image Table*. It contains images information like colour, viewpoint. Currently, it includes information of 1,430,096 car images, while expansions of this table are applicable in the future.

Using instructions

1. The DVM-Car should be used for research purposes only.
2. Users can use two methods to select the target images from the dataset:
 - I. The provided images have been grouped into eight observation viewpoints by machine. So users can use the *Image Table* to select the images from the target observation angle (e.g., front images). The '-Y' and '-N' in the Image Table indicates manually confirmed results ('Y' means confirmed 'yes', 'N' means confirmed 'no').
 - II. Use the "*brand-model-year-colour*" storing structure to find car images from the target models.
3. For each image, a *PKL* file is provided accordingly, which contains contains the shape of the appeared car in the corresponding image. Users can use the PKL files to eliminate the background from the provided images (e.g., using programming languages like Python to load and process the images accordingly).
4. Users can 'join' different tables via the 'Genmodel ID' or the 'Adv ID' attribute.
5. The unit of price-related attributes in tables is the Great British Pound.
6. The unit of engine size in Trim Table is millilitre.
7. You are welcome to contact us for any usage issues!