

✓ vehicle_images_dataset

The unit of observation is a vehicle image, therefore each row pertains to a vehicle image. Each subsection in quotations, e.g. "image_path", below pertains to a variable describing the observations, with the heading being the variable name. The reported numbers in each variable subsection in the format n(m) represent n total observations for the variable and m number of missing values for the variable.

```
! git clone https://github.com/jillianhaig/Project3_DS4002 # so we can access data loaded from shared github
```

> _

[Show code](#)

✓ "image_path"

This variable takes on string values and represents the path to access the image content when mounted to google colab.

```
print(df["image_path"].count(), "(" , "0",")") # we already know all file paths are valid and there are no NAs because of our code working pre
```

```
↵ 4356 ( 0 )
```

```
df['image_path'].describe()
```

```
↵
```

	image_path
count	4356
unique	4356
top	/content/vehicleimages/sedan/PHOTO_744.jpg
freq	1

dtype: object

✓ "label"

This variable takes on values representing the vehicle type of the image being represented.

```
print(df["label"].count(), "(" , sum(df["label"].isna()),")")
df["label"].unique()
```

```
↵ 4356 ( 0 )
array(['sedan', 'hatchback', 'other', 'pickup', 'suv'], dtype=object)
```

```
df["label"].describe()
```

```
↵
```

	label
count	4356
unique	5
top	pickup
freq	1240

dtype: object

> Class Distributions

[Show code](#)

 <ipython-input-12-f75f063f0d90>:9: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend`

```
sns.barplot(x=label_counts.index, y=label_counts.values, palette='coolwarm')
```

