Introduction[¶](https://www.kaggle.com/code/suugaku/dataquest-exploring-ebay-car-sales-data/notebook#Introduction)

In this guided project, we work with a dataset of used cars from *eBay Kleinanzeigen*, a classifieds section of the German eBay website. While the full dataset consisting of over 370,000 listings was cleaned an uploaded to [Kaggle](https://www.kaggle.com/orgesleka/used-cars-database/data), this guided project will use a subset of 50,000 observations that has also been dirtied by the DataQuest team in order to more closely mimic what could be expected from the originally scraped data set.

The set contains the following columns.

* dateCrawled: When the ad was first crawled. All other field values for the corresponding row were scraped on this date.
* name: Name of the car listing.
* seller: Whether the seller is a private owner or a dealer.
* offerType: The type of listing.
* price: The price on the ad to sell the car.
* abtest: Whether the listing is included in an A/B test.
* vehicleType: The vehicle type.
* yearOfRegistration: The year in which the car was first registered.
* gearbox: The transmission type.
* powerPS: The power of the car in [PS](https://www.carwow.co.uk/guides/glossary/what-is-horsepower).
* model: The car model name.
* odometer: The odometer reading on the car, in kilometers.
* monthOfRegistration: The month in which the car was first registered.
* fuelType: What type of fuel the car uses.
* brand: The brand of the car.
* notRepairedDamage: Whether or not the car has damage which is not yet repaired.
* dateCreated: The date on which the eBay listing was created.
* nrOfPictures: The number of pictures in the listing.
* postalCode: The postal code for the location of the vehicle.
* lastSeenOnline: When the crawler last saw this listing online.

The goal of this project is to clean the data and then use pandas to perform some basic initial analysis of the listings. To start, we first import the NumPy and pandas libraries, and then will attempt read the CSV file which contains the data and load it into a pandas DataFrame.