**Data Analysis Project-1**

**Dataset overview:-**

* This dataset contains details of the used cars in Germany which are on sale on eBay.
* The dataset contains 371,528 rows of information, each representing a car listed for sale on eBay online marketplace.
* **dateCrawled** - When this ad was first crawled. All field-values are taken from this date.
* **name** - Name of the car.
* **seller** - Whether the seller is private 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 which year the car was first registered.
* **gearbox** - The transmission type.
* **powerPS**- The power of the car in PS.
* **model** - The car model name.
* **kilometer** - How many kilometers the car has driven.
* **monthOfRegistration**- The month in which which year the car was first registered.
* **fuelType**- What type of fuel the car uses.
* **brand** - The brand of the car.
* **notRepairedDamage**- If the car has a damage which is not yet repaired.
* **dateCreated**- The date on which the eBay listing was created.
* **nrOfPictures**- The number of pictures in the ad.
* **postalCode**- The postal code for the location of the vehicle.
* **lastSeenOnline**- When the crawler saw this ad last online.

**Dataset Link** :-

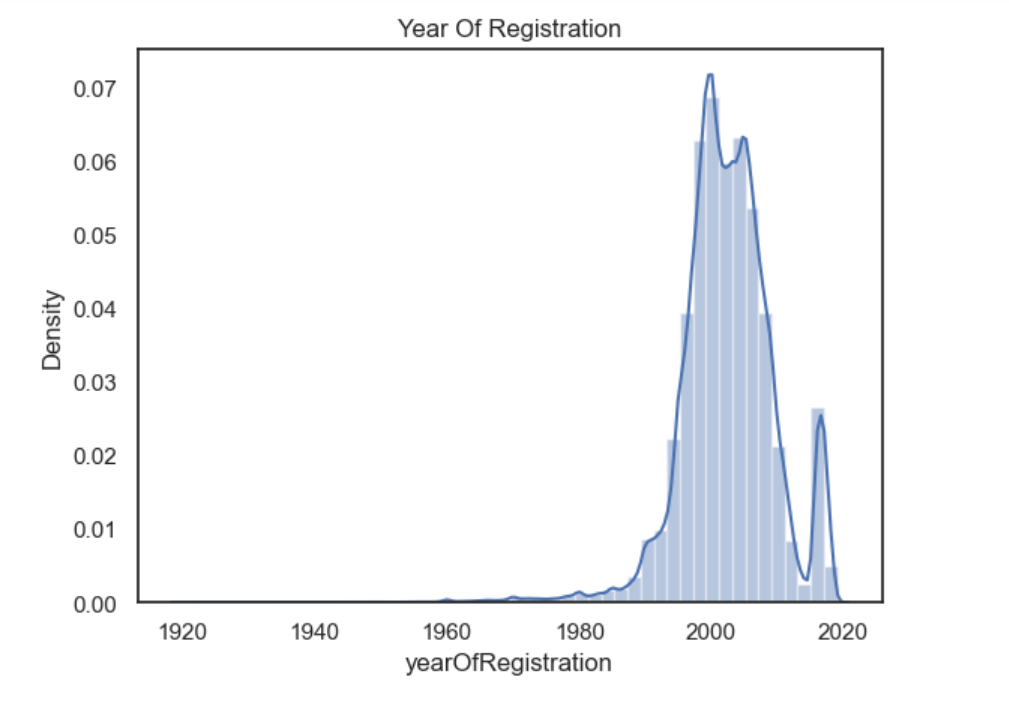
<https://drive.google.com/file/d/1NFpNI-AiSLmta139rYaWrNy8xZ0VrbM9/view?usp=sharing>

**Analysis – 1**

1. **Perform general Data analysis**

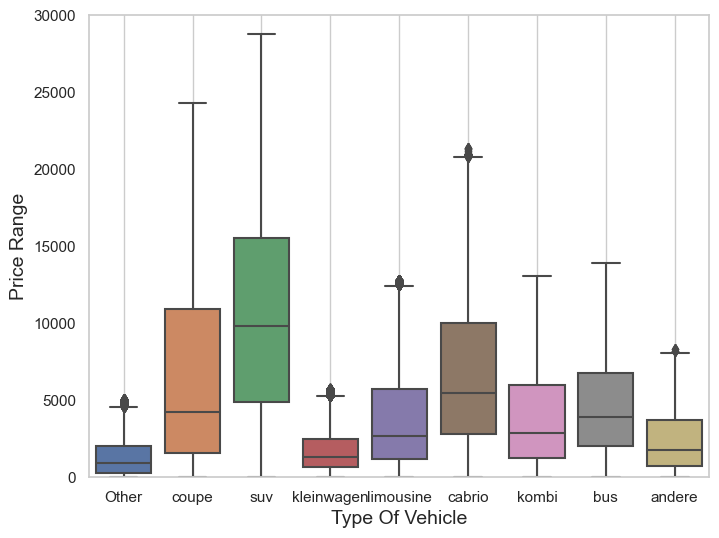
* Importing all the necessary libraries:
* **import pandas as pd**
* **import seaborn as sns**
* **import numpy as np**
* **import matplotlib.pyplot as plt**
* **import re**

2. **Can you tell me the Distribution of Vehicles based on Year of Registration with the help of a plot?**

****

* The number of vehicles registered has increased rapidly since the 1960s.
* From the plot above we can see that the surge of cars is from the years 1990 to 2010.
* The cumulative density line reveals that by 2020, **over** 90% of vehicles had been registered in the previous 30 years.
* Overall, the visualization shows that the number of vehicles registered has increased steadily over period of time and that the fleet of vehicles on the road is relatively young.

# 3. Create a plot based on the Variation of the price range by the vehicle type



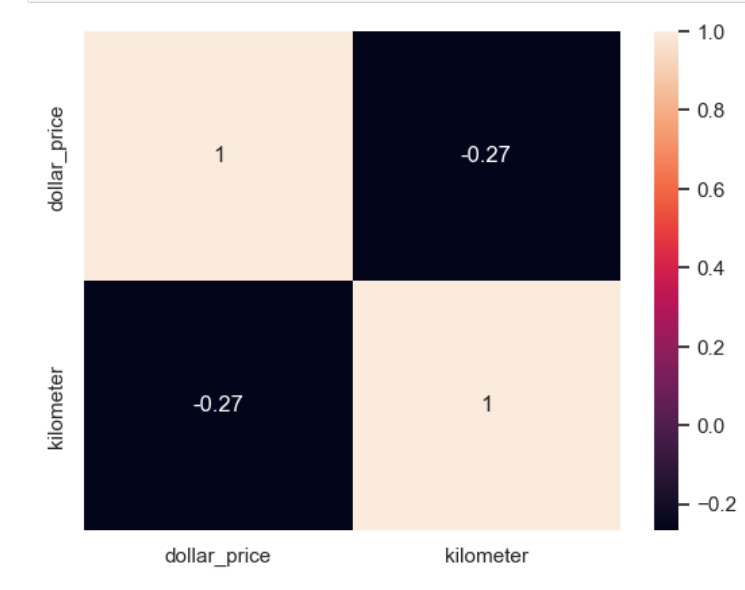
* The above bar graph shows the Average price of vehicles by vehicle types
* From the figure above, we can see, for example, that the median value of an **SUV** is 10,000, with most values between 5000 and 15000, and the maximum price is something close to 30,000.
* Also, we can see that excluding the other type, kleinwagen and andere are the types of vehicles with the lowest price range.

# 4. Find out Total count of vehicles by type available on eBay for sale. As well as create visualization for the client.

# download (2).png

# From the figure above, we can see that the “limousine” is the top type of car for selling, and the “andere” has the least amount of cars for sale.

## 5. is there any relationship between dollar\_price and kilometer? (Explain with appropriate analysis)



* The heatmap shows a slight negative correlation, which means that as the number of kilometers driven on a car increases, the price of the car tend to decrease.
* The correlation coefficient is -0.27, that there is a slight negative relationship between dollar\_price and kilometer.
* The slight negative correlation suggests that mileage is an important factor to consider when pricing a used car.