**DATA CLEANING**

The dataset underwent a series of preprocessing steps to enhance its quality and relevance for analysis. Firstly, columns 'notRepairedDamage' and 'nrOfPictures' were removed due to either lacking informative content or having minimal data (with a substantial number of missing values and zero variance, respectively). Subsequently, records containing null values in any column were excluded from the analysis.

Addressing anomalies in the 'yearOfRegistration' column, rows with values exceeding the current year or predating the 1900s were eliminated. Additionally, entries with horsepower ('powerPS') less than 1 or greater than 2400 were discarded, as such extremes are unrealistic for vehicle specifications. Rows below $500 or above $400,000 were also removed to ensure a reasonable price range.

To rectify instances where 'monthOfRegistration' was recorded as 0, these values were replaced with the most frequently occurring month in the dataset. Following these data-cleaning steps, duplicate rows were removed, resulting in a refined dataset containing 280,557 rows and 18 columns suitable for accurate analysis.

**ANALYSIS 1**

1. The processed pandas data frame has been saved as a CSV file named 'autos\_cleaned.csv'.
2. The 'year of registration' feature exhibits a distribution that approximates a normal distribution. With a larger dataset, it is anticipated that the distribution would closely align with a normal distribution.
3. Upon grouping the rows based on the 'vehicleType' column and calculating the average prices for each category, it is observed that SUVs and coupes tend to be more expensive, while kleinwagens are the most budget-friendly compared to other vehicle types.
4. The dataset reveals that limousines are the most abundant vehicle type, whereas coupes and SUVs are relatively less prevalent.
5. The correlation coefficient between price and kilometers is calculated as -0.33. The negative sign suggests a negative correlation, indicating that as kilometers increase, the price tends to decrease (inversely proportional). However, the strength of this correlation is deemed weak, signifying that the impact on price as kilometers increase is relatively low.

**ANALYSIS 2**

Volkswagen cars dominate the listings on the platform, comprising approximately 59,000 listings.

Cars with automatic transmissions are generally priced higher than their manual counterparts. SUVs and coupes stand out as the most expensive vehicle types in both transmission categories.

Most sellers, accounting for approximately 99% of the total, are private individuals.

**ANALYSIS 3**

To optimize memory usage in the dataset, the following strategies have been implemented:

* Downcasting: Numeric columns, both integers (int64) and floats (float64), were downcasted to sub-datatypes with lower memory footprint based on the range of values in each column.
* Columns with zero variance or deemed unnecessary for our analysis were dropped.

When considering different fuel types, automatic cars generally exhibit higher prices than their manual counterparts, except for hybrid fuel types. Hybrid fuel cars are notably more expensive compared to other fuel types, while LPG cars tend to be more affordable.

In terms of horsepower, automatic cars generally have higher power than manual ones. Kleinwagens consistently has the lowest horsepower, and coupes typically boast the highest horsepower, regardless of transmission type.

Across all brands, the majority of cars are priced below $10,000. Among the brands, Porsche's coupe models emerge as the most expensive in this dataset. Jaguar and Porsche feature some of the highest-priced cars overall.