

REPORT

1) The memory usage of the data is around 6.1 mb. How can we reduce the memory usage of the data set?

Here are several techniques to reduce the memory usage of a dataset:

1. Choose Appropriate Data Types:

- Use more memory-efficient data types (e.g., using int32 instead of int64, float32 instead of float64) for numerical columns to store data efficiently.

2. Downcast Numeric Columns:

- Downcast numeric columns using Pandas to reduce memory usage.
- `df = df.apply(pd.to_numeric, downcast='integer')`

3. Convert Categorical Data:

Convert categorical variables to the Pandas category data type, which can significantly reduce memory usage.

```
df['category_column'] = df['category_column'].astype('category')
```

4. Remove Unnecessary Columns:

Drop columns that are not necessary for analysis to reduce the overall memory footprint.

5. Use File Compression Formats:

Use more memory-efficient file formats such as Parquet, Feather, or HDF5, which can store data more efficiently compared to CSV or Excel files.

6. Load Data in Chunks:

Process data in chunks using Pandas' chunksize parameter while reading large datasets to reduce memory usage.

7. Use External Libraries:

Libraries like Dask can handle larger-than-memory datasets by performing operations on smaller chunks of data.

8. Optimize Text Data:

Optimize text data by removing unnecessary characters or using tokenization, especially for free-text fields.

REPORT

9. Use Sparse Data Structures:

For datasets with lots of zeros, consider using sparse data structures to save memory.

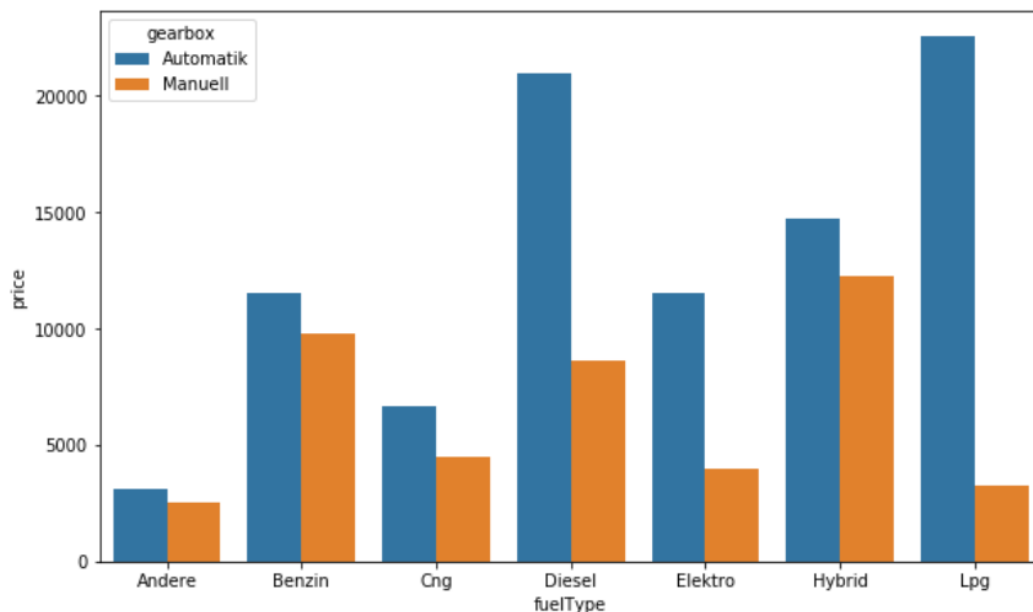
10. Clean Data and Remove Duplicates:

Cleaning data can eliminate unnecessary or duplicate records, reducing overall memory usage.

-- Write the reduced-memory DataFrame to a new file

```
df.to_parquet('reduced_memory_data.parquet')
```

2) What is the Average price of vehicle by fuel type and gearbox type. Give a plot



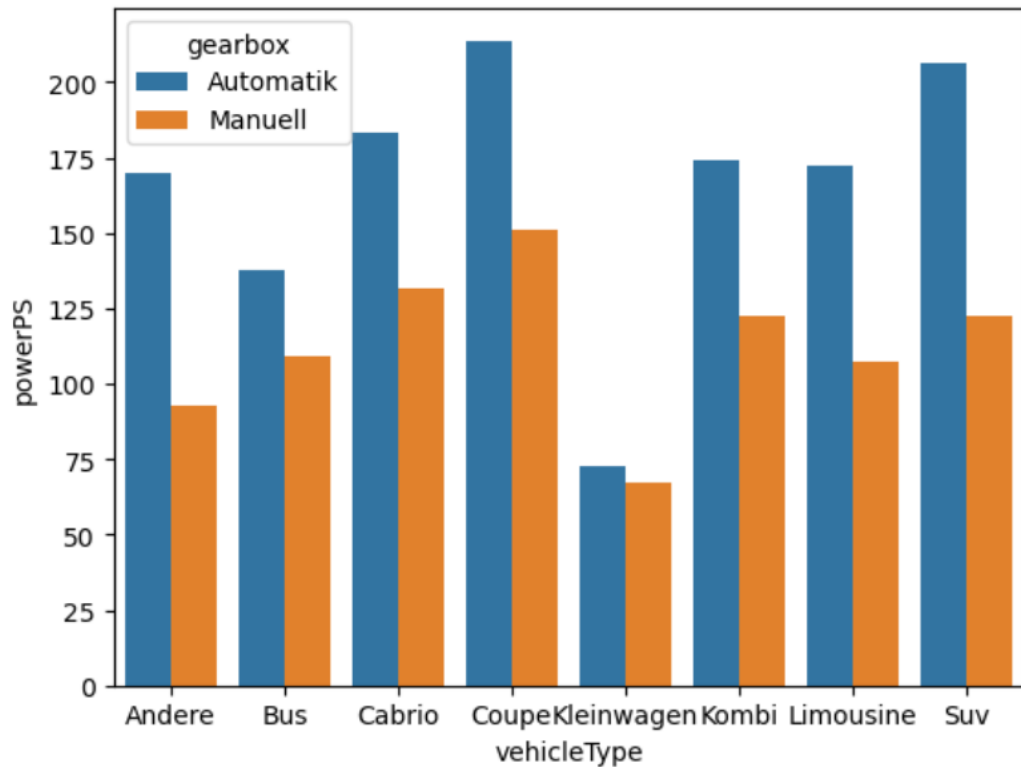
SUMMARY:

- Based on the plot, fuel type Lpg and gearbox Automatik have the highest price. And fuel type Andere and gearbox Automatik have the lowest price.
- Based on the plot, fuel type Hybrid and gearbox Manuell have the highest price. And fuel type Andere and gearbox Manuell have the lowest price.
- Based on the plot, gearbox and fuel type of Andere have a low price.
- So people are liking to buy the combination of Diesel and Automatic and the combination of Diesel and Manual gear box.

REPORT

3) What is the Average power of a vehicle by vehicle type and gearbox type. Give a plot

```
Out[7]: <Axes: xlabel='vehicleType', ylabel='powerPS'>
```



SUMMARY:

- Based on the plot ,vehicle type Suv and gearbox Automatik have high power And vehicle type kleinwagen and gearbox Automatik have low power
- Based on the plot ,vehicle type coupe and gearbox Manuell have high power And vehicle type kleinwagen and gearbox Manuell have low power
- The combination of ,gearbox and vehicleType kleinwagen have low power

4) What is the Average price of a vehicle by brand as well as vehicle type. Use heatmap to explain this

REPORT



SUMMARY:

- The generated heatmap displays the average prices of vehicles categorized by both brands and types.
- Each cell in the heatmap represents the average price of a specific brand and vehicle type combination. Darker color indicate higher average prices and lighter color indicate lower average prices.
- This plot in Andere vehicle type category , Audi appears to have the highest average price among the considered brands.
- This visualization helps to identify which brands generally higher or lower prices within each vehicle type category. It offers a clear comparison of average prices across different brands and types