**REPORT**

**prepared by: varshini**

**importing libraries:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

**Reading csv file:**

🡪 data=pd.read\_csv(r"C:\Users\seeth\python 232\Statistics\cleaned data of autos.csv")

🡪df=data.copy()

🡪df

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

**Code:**

df["price"] = df["price"].astype(np.uint32)

df["odometer"] = df["odometer"].astype(np.uint32)

df["power\_ps"] = df["power\_ps"].astype(np.uint16)

df["registration\_month"] = df["registration\_month"].astype(np.uint8)

df["registration\_year"] = df["registration\_year"].astype(np.uint16)

df["postal\_code"] = df["postal\_code"].astype(np.uint32)

df.info()

**Output:**

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 371528 entries, 0 to 371527

Data columns (total 19 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 date\_crawled 371528 non-null object

1 name 371528 non-null object

2 seller 371528 non-null object

3 offer\_type 371528 non-null object

4 price 371528 non-null uint32

5 abtest 371528 non-null object

6 vehicle\_type 371528 non-null object

7 registration\_year 371528 non-null uint16

8 gearbox 371528 non-null object

9 power\_ps 371528 non-null uint16

10 model 371528 non-null object

11 odometer 371528 non-null uint32

12 registration\_month 371528 non-null uint8

13 fuel\_type 371528 non-null object

14 brand 371528 non-null object

15 unrepaired\_damage 371528 non-null object

16 ad\_created 371528 non-null object

17 postal\_code 371528 non-null uint32

18 last\_seen 371528 non-null object

dtypes: object(13), uint16(2), uint32(3), uint8(1)

memory usage: 42.9+ MB

**Explanation:** The memory usage before changing the datatype is 53.9+ MB.The memory is reduced by using astype(), now it is 42.9+ MB.

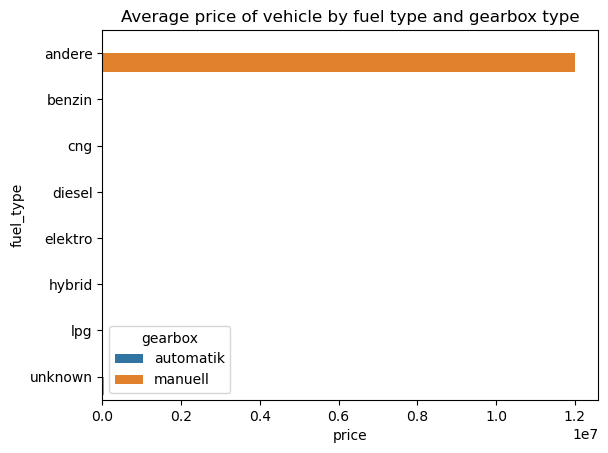
**2.What is the Average price of vehicle by fuel type and gearbox type. Give a plot.**

**Code:**

sns.barplot(data=avg1,x="price",y="fuel\_type",hue="gearbox")

plt.title("Average price of vehicle by fuel type and gearbox type")

**Output:**



**Explanation:** The average price of vehicle by fuel type and gearbox is high for andere.

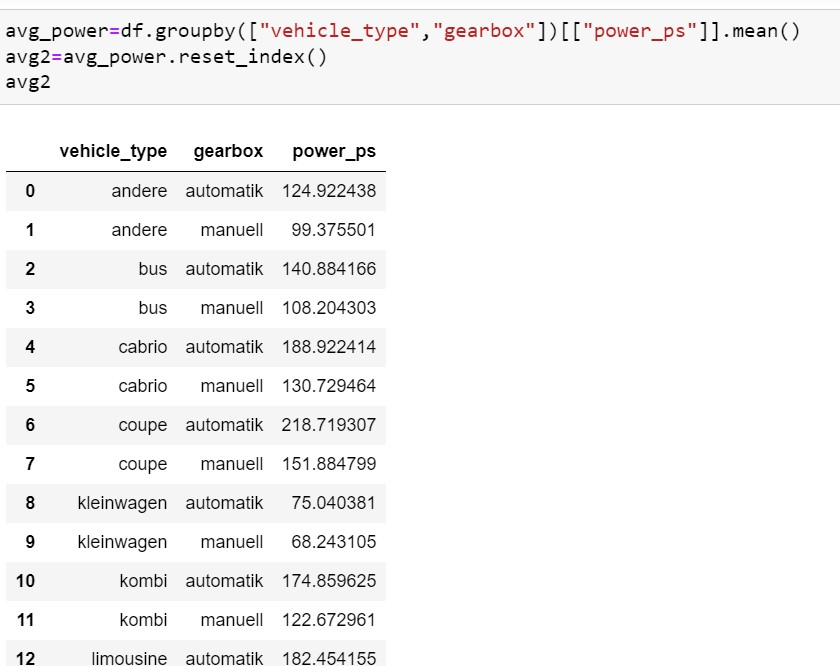
**3. What is the Average power of a vehicle by vehicle type and gearbox type. Give a plot.**

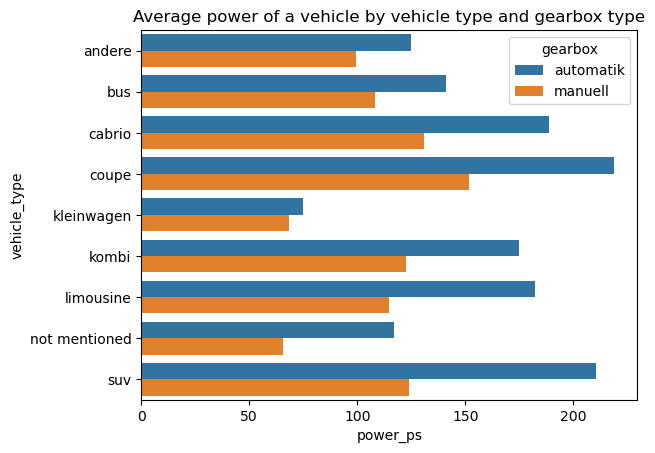
**Code:**

sns.barplot(data=avg2,x="power\_ps",y="vehicle\_type",hue="gearbox")

plt.title("Average power of a vehicle by vehicle type and gearbox type")

**Output:**





**Explanation:** From the above graph the average power consumed vehicle by vehicle\_type and gearbox is high for coupe-automatic and low for kleinwagen-manuell.

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

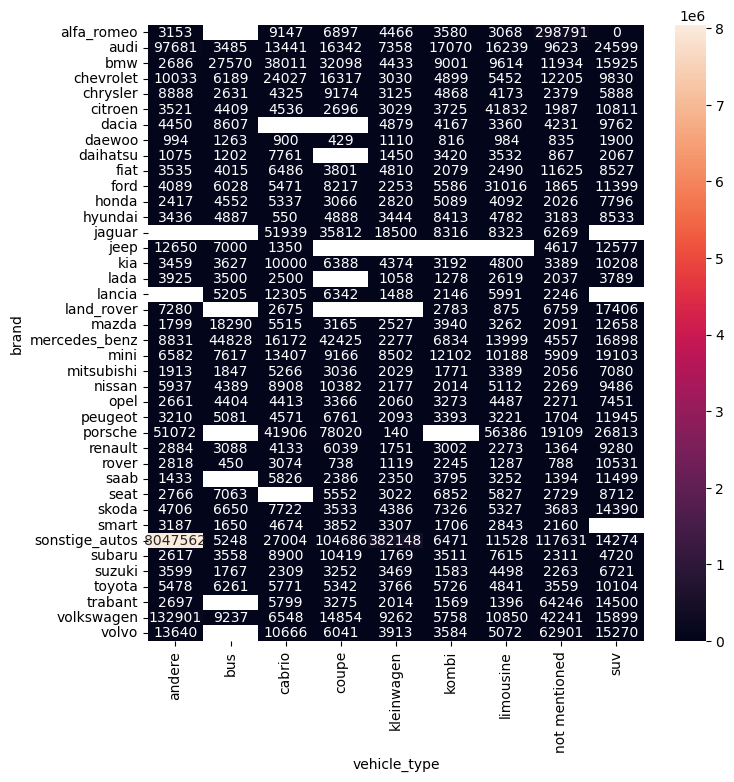
**Code:**

plt.figure(figsize=(8,8))

sns.heatmap(pd.crosstab(df["brand"],df["vehicle\_type"],values=df["price"],aggfunc="mean"),annot=True,fmt=".0f")

**Output:**





**Explanation:** The above graph gives the average price of vehicles by brands as well as

vehicle\_type.