EDS ASSIGNMENT 4

NAME – APOORVA SINGH

ROLL NO. – 301

PRN – 202201070030

DATASET : toyota1.csv

toyota1.csv ×

1 to 25 of 29 entries Filter

Price	Age	KM	FuelType	HP	Weight
13500	23	46986	Diesel	90	1165
13750	23	72937	Diesel	90	1165
13950	24	41711	Diesel	90	1165
14950	26	48000	Diesel	90	1165
13750	30	38500	Diesel	90	1170
12950	32	61000	Diesel	90	1170
16900	27	4522	Diesel	56	1245
18600	30	75889	Diesel	90	1245
21500	27	19700	Petrol	192	1185
12950	23	71138	Diesel	43	1105
20950	25	31461	Petrol	192	1185
19950	22	43610	Petrol	192	1185
19600	25	32189	Petrol	192	1185
21500	31	23000	Petrol	192	1185
22500	32	34131	Petrol	192	1185
22000	28	18739	Petrol	56	1185
22750	30	34000	Petrol	192	1185
17950	24	21716	Petrol	110	1105
16750	24	25563	Petrol	110	1065
16950	30	64359	Petrol	110	1105
15950	30	67660	Petrol	110	1105
16950	29	43905	Petrol	110	1170
15950	28	56349	Petrol	110	1120
16950	28	32220	Petrol	110	1120

Program with queries 1 to 20 :

```
import pandas as pd

# Reading the dataset
df = pd.read_csv("toyota1.csv")

# 1) Calculate the average weight of the engine
mean_av_wt = df['Weight'].mean()
print('1) Average Weight of engine:',mean_av_wt,'\n\n')

# 2) Calculate the median KM can be covered by engine
median_km = df['KM'].median()
print("2) Median KM :",median_km ,'\n\n')

# 3) Find the highest HP of Engine
max_hp = df['HP'].max()
print("3) Maximum HP of Engine  :",max_hp,'\n\n')

# 4) Find the lowest sales sales of engine
min_en_sales = df['Price'].min()
print("4) Minimum Engine Sales :", min_en_sales,'\n\n')

# 5) Count the number of Diesel based engines
diesel_count = df['FuelType'].value_counts()
print("5) Diesel based engines are :\n",diesel_count,'\n\n')

# 6) Count the number of Engine of each fuel type
en_count = df['FuelType'].value_counts()
print("6) Count of engine type:\n",en_count,'\n\n')

# 7) Calculate the total sales of all engines
total_sales = df['Price'].sum()
print("7) Total Sales:",total_sales,'\n\n')

# 8) Group the dataset by Age and calculate the average km for each age
age = df.groupby('Age')['KM'].mean()
print("8) Average km by eacg age:\n",age,'\n\n')

# 9) Calculate the correlation between the Price and KM covered by
engine
na_p_corr = df['Price'].corr(df['KM'])
print("9) Correlation between Price of Engine and KM travelled by
engine:",na_p_corr,'\n')

# 10) Calculate the covariance between the HP and Weight of engine
p_cov = df['HP'].cov(df['Weight'])
print("10) Covariance between HP and Weight of engine :",p_cov,'\n\n')
```

```
# 11) Add an empty column named 'Total Sales'
df['Total_Sales'] = ''
print("11) Upadate DataFrame",df.to_string(),'\n\n')

# 12) Check for missing values in the dataset
isnull = df.isnull().any()
print("12) Missing Values:\n",isnull,'\n\n')

# 13) Drop rows with missing values in the dataset
df_dropped = df.dropna()
print("13) Dataset after dropping rows with missing
values:\n",df_dropped.to_string(),'\n\n')

# 14) Fill missing values witht zero
df_filled = df.fillna(0)
print("14) Dataset after filling missing values with
zero:\n",df_filled.to_string(),'\n\n')

# 15) Check for duplicated rows in the dataset
duplicates = df.duplicated()
print("15) Duplicated Rows:\n",duplicates,'\n\n')

# 16) Drop duplicated rows from the dataset
df_unique = df.drop_duplicates()
print("16) Dataset after dropping duplicated
rows:\n",df_unique.to_string(),'\n\n')

# 17) Convert the 'Age' column to string data type
df['Age'] = df['Age'].astype(str)
print("17) Dataset with 'Age' column converted to int
datatype:\n",df.to_string(),'\n\n')

# 18) Strip leading and trailing whitespace from the 'FuelType' column
df['FuelType'] = df['FuelType'].str.strip()
print("18) Dataset with leading and trailing whitespace removed from
'FuelType' column:\n",df.to_string(),'\n\n')

# 19) Replace 'Petrol' genre with 'CNG' genre in the 'FuelType' column
df['FuelType']  =  df['FuelType'].str.replace('Petrol','CNG')
print("19) Dataset with 'Petrol' genre replaced by 'CNG' genre in
'FuelType' column:\n",df.to_string(),'\n\n')

# 20) Calculate the sum of KM travelled by each engine using groupby
km_sum = df.groupby('FuelType')['KM'].sum()
print("20) Total KM travelled by each engine:\n",km_sum,'\n\n')
```

Output :

```
1) Average Weight of engine: 1153.9655172413793


2) Median KM : 38500.0


3) Maximum HP of Engine  : 192


4) Minimum Engine Sales : 12950


5) Diesel based engines are :
 Petrol    19
Diesel    10
Name: FuelType, dtype: int64


6) Count of engine type:
 Petrol    19
Diesel    10
Name: FuelType, dtype: int64


7) Total Sales: 501895


8) Average km by eacg age:
 Age
22    43610.000000
23    63687.000000
24    29663.333333
25    30700.000000
26    48000.000000
27    19589.000000
28    37862.500000
29    37044.333333
30    56081.600000
31    23000.000000
32    47565.500000
Name: KM, dtype: float64


9) Correlation between Price of Engine and KM travelled by engine: -
0.4887404006306358

10) Covariance between HP and Weight of engine : 398.4975369458126


11) Upadate DataFrame     Price  Age     KM FuelType   HP  Weight
Total_Sales
0   13500   23  46986   Diesel   90    1165
1   13750   23  72937   Diesel   90    1165
2   13950   24  41711   Diesel   90    1165
3   14950   26  48000   Diesel   90    1165
```

```
4   13750   30  38500   Diesel   90     1170
5   12950   32  61000   Diesel   90     1170
6   16900   27   4522   Diesel   56     1245
7   18600   30  75889   Diesel   90     1245
8   21500   27  19700   Petrol  192     1185
9   12950   23  71138   Diesel   43     1105
10  20950   25  31461   Petrol  192     1185
11  19950   22  43610   Petrol  192     1185
12  19600   25  32189   Petrol  192     1185
13  21500   31  23000   Petrol  192     1185
14  22500   32  34131   Petrol  192     1185
15  22000   28  18739   Petrol   56     1185
16  22750   30  34000   Petrol  192     1185
17  17950   24  21716   Petrol  110     1105
18  16750   24  25563   Petrol  110     1065
19  16950   30  64359   Petrol  110     1105
20  15950   30  67660   Petrol  110     1105
21  16950   29  43905   Petrol  110     1170
22  15950   28  56349   Petrol  110     1120
23  16950   28  32220   Petrol  110     1120
24  16250   29  25813   Petrol  110     1120
25  15950   25  28450   Petrol  110     1120
26  17495   27  34545   Diesel  110     1120
27  15750   29  41415   Petrol  110     1120
28  16950   28  44142   Petrol  110     1120
```

```
12) Missing Values:
 Price          False
Age             False
KM              False
FuelType        False
HP              False
Weight          False
Total_Sales     False
dtype: bool
```

```
13) Dataset after dropping rows with missing values:
     Price  Age     KM FuelType   HP  Weight Total_Sales
0   13500   23  46986   Diesel   90     1165
1   13750   23  72937   Diesel   90     1165
2   13950   24  41711   Diesel   90     1165
3   14950   26  48000   Diesel   90     1165
4   13750   30  38500   Diesel   90     1170
5   12950   32  61000   Diesel   90     1170
6   16900   27   4522   Diesel   56     1245
7   18600   30  75889   Diesel   90     1245
8   21500   27  19700   Petrol  192     1185
9   12950   23  71138   Diesel   43     1105
10  20950   25  31461   Petrol  192     1185
11  19950   22  43610   Petrol  192     1185
12  19600   25  32189   Petrol  192     1185
13  21500   31  23000   Petrol  192     1185
14  22500   32  34131   Petrol  192     1185
15  22000   28  18739   Petrol   56     1185
16  22750   30  34000   Petrol  192     1185
17  17950   24  21716   Petrol  110     1105
```

```
18  16750   24  25563   Petrol  110     1065
19  16950   30  64359   Petrol  110     1105
20  15950   30  67660   Petrol  110     1105
21  16950   29  43905   Petrol  110     1170
22  15950   28  56349   Petrol  110     1120
23  16950   28  32220   Petrol  110     1120
24  16250   29  25813   Petrol  110     1120
25  15950   25  28450   Petrol  110     1120
26  17495   27  34545   Diesel  110     1120
27  15750   29  41415   Petrol  110     1120
28  16950   28  44142   Petrol  110     1120
```

```
14) Dataset after filling missing values with zero:
     Price  Age     KM FuelType   HP  Weight Total_Sales
0   13500   23  46986   Diesel   90     1165
1   13750   23  72937   Diesel   90     1165
2   13950   24  41711   Diesel   90     1165
3   14950   26  48000   Diesel   90     1165
4   13750   30  38500   Diesel   90     1170
5   12950   32  61000   Diesel   90     1170
6   16900   27   4522   Diesel   56     1245
7   18600   30  75889   Diesel   90     1245
8   21500   27  19700   Petrol  192     1185
9   12950   23  71138   Diesel   43     1105
10  20950   25  31461   Petrol  192     1185
11  19950   22  43610   Petrol  192     1185
12  19600   25  32189   Petrol  192     1185
13  21500   31  23000   Petrol  192     1185
14  22500   32  34131   Petrol  192     1185
15  22000   28  18739   Petrol   56     1185
16  22750   30  34000   Petrol  192     1185
17  17950   24  21716   Petrol  110     1105
18  16750   24  25563   Petrol  110     1065
19  16950   30  64359   Petrol  110     1105
20  15950   30  67660   Petrol  110     1105
21  16950   29  43905   Petrol  110     1170
22  15950   28  56349   Petrol  110     1120
23  16950   28  32220   Petrol  110     1120
24  16250   29  25813   Petrol  110     1120
25  15950   25  28450   Petrol  110     1120
26  17495   27  34545   Diesel  110     1120
27  15750   29  41415   Petrol  110     1120
28  16950   28  44142   Petrol  110     1120
```

```
15) Duplicated Rows:
 0     False
1      False
2      False
3      False
4      False
5      False
6      False
7      False
8      False
9      False
10     False
```

```
11    False
12    False
13    False
14    False
15    False
16    False
17    False
18    False
19    False
20    False
21    False
22    False
23    False
24    False
25    False
26    False
27    False
28    False
dtype: bool
```

16) Dataset after dropping duplicated rows:

	Price	Age	KM	FuelType	HP	Weight	Total_Sales
0	13500	23	46986	Diesel	90	1165	
1	13750	23	72937	Diesel	90	1165	
2	13950	24	41711	Diesel	90	1165	
3	14950	26	48000	Diesel	90	1165	
4	13750	30	38500	Diesel	90	1170	
5	12950	32	61000	Diesel	90	1170	
6	16900	27	4522	Diesel	56	1245	
7	18600	30	75889	Diesel	90	1245	
8	21500	27	19700	Petrol	192	1185	
9	12950	23	71138	Diesel	43	1105	
10	20950	25	31461	Petrol	192	1185	
11	19950	22	43610	Petrol	192	1185	
12	19600	25	32189	Petrol	192	1185	
13	21500	31	23000	Petrol	192	1185	
14	22500	32	34131	Petrol	192	1185	
15	22000	28	18739	Petrol	56	1185	
16	22750	30	34000	Petrol	192	1185	
17	17950	24	21716	Petrol	110	1105	
18	16750	24	25563	Petrol	110	1065	
19	16950	30	64359	Petrol	110	1105	
20	15950	30	67660	Petrol	110	1105	
21	16950	29	43905	Petrol	110	1170	
22	15950	28	56349	Petrol	110	1120	
23	16950	28	32220	Petrol	110	1120	
24	16250	29	25813	Petrol	110	1120	
25	15950	25	28450	Petrol	110	1120	
26	17495	27	34545	Diesel	110	1120	
27	15750	29	41415	Petrol	110	1120	
28	16950	28	44142	Petrol	110	1120	

17) Dataset with 'Age' column converted to int datatype:

	Price	Age	KM	FuelType	HP	Weight	Total_Sales
0	13500	23	46986	Diesel	90	1165	
1	13750	23	72937	Diesel	90	1165	

2	13950	24	41711	Diesel	90	1165
3	14950	26	48000	Diesel	90	1165
4	13750	30	38500	Diesel	90	1170
5	12950	32	61000	Diesel	90	1170
6	16900	27	4522	Diesel	56	1245
7	18600	30	75889	Diesel	90	1245
8	21500	27	19700	Petrol	192	1185
9	12950	23	71138	Diesel	43	1105
10	20950	25	31461	Petrol	192	1185
11	19950	22	43610	Petrol	192	1185
12	19600	25	32189	Petrol	192	1185
13	21500	31	23000	Petrol	192	1185
14	22500	32	34131	Petrol	192	1185
15	22000	28	18739	Petrol	56	1185
16	22750	30	34000	Petrol	192	1185
17	17950	24	21716	Petrol	110	1105
18	16750	24	25563	Petrol	110	1065
19	16950	30	64359	Petrol	110	1105
20	15950	30	67660	Petrol	110	1105
21	16950	29	43905	Petrol	110	1170
22	15950	28	56349	Petrol	110	1120
23	16950	28	32220	Petrol	110	1120
24	16250	29	25813	Petrol	110	1120
25	15950	25	28450	Petrol	110	1120
26	17495	27	34545	Diesel	110	1120
27	15750	29	41415	Petrol	110	1120
28	16950	28	44142	Petrol	110	1120

18) Dataset with leading and trailing whitespace removed from 'FuelType' column:

	Price	Age	KM	FuelType	HP	Weight	Total_Sales
0	13500	23	46986	Diesel	90	1165	
1	13750	23	72937	Diesel	90	1165	
2	13950	24	41711	Diesel	90	1165	
3	14950	26	48000	Diesel	90	1165	
4	13750	30	38500	Diesel	90	1170	
5	12950	32	61000	Diesel	90	1170	
6	16900	27	4522	Diesel	56	1245	
7	18600	30	75889	Diesel	90	1245	
8	21500	27	19700	Petrol	192	1185	
9	12950	23	71138	Diesel	43	1105	
10	20950	25	31461	Petrol	192	1185	
11	19950	22	43610	Petrol	192	1185	
12	19600	25	32189	Petrol	192	1185	
13	21500	31	23000	Petrol	192	1185	
14	22500	32	34131	Petrol	192	1185	
15	22000	28	18739	Petrol	56	1185	
16	22750	30	34000	Petrol	192	1185	
17	17950	24	21716	Petrol	110	1105	
18	16750	24	25563	Petrol	110	1065	
19	16950	30	64359	Petrol	110	1105	
20	15950	30	67660	Petrol	110	1105	
21	16950	29	43905	Petrol	110	1170	
22	15950	28	56349	Petrol	110	1120	
23	16950	28	32220	Petrol	110	1120	
24	16250	29	25813	Petrol	110	1120	
25	15950	25	28450	Petrol	110	1120	

```
26  17495  27  34545   Diesel  110    1120
27  15750  29  41415   Petrol  110    1120
28  16950  28  44142   Petrol  110    1120
```

19) Dataset with 'Petrol' genre replaced by 'CNG' genre in 'FuelType' column:

```
     Price Age     KM FuelType   HP  Weight Total_Sales
0   13500  23  46986   Diesel   90    1165
1   13750  23  72937   Diesel   90    1165
2   13950  24  41711   Diesel   90    1165
3   14950  26  48000   Diesel   90    1165
4   13750  30  38500   Diesel   90    1170
5   12950  32  61000   Diesel   90    1170
6   16900  27   4522   Diesel   56    1245
7   18600  30  75889   Diesel   90    1245
8   21500  27  19700      CNG  192    1185
9   12950  23  71138   Diesel   43    1105
10  20950  25  31461      CNG  192    1185
11  19950  22  43610      CNG  192    1185
12  19600  25  32189      CNG  192    1185
13  21500  31  23000      CNG  192    1185
14  22500  32  34131      CNG  192    1185
15  22000  28  18739      CNG   56    1185
16  22750  30  34000      CNG  192    1185
17  17950  24  21716      CNG  110    1105
18  16750  24  25563      CNG  110    1065
19  16950  30  64359      CNG  110    1105
20  15950  30  67660      CNG  110    1105
21  16950  29  43905      CNG  110    1170
22  15950  28  56349      CNG  110    1120
23  16950  28  32220      CNG  110    1120
24  16250  29  25813      CNG  110    1120
25  15950  25  28450      CNG  110    1120
26  17495  27  34545   Diesel  110    1120
27  15750  29  41415      CNG  110    1120
28  16950  28  44142      CNG  110    1120
```

20) Total KM travelled by each engine:

```
 FuelType
CNG       688422
Diesel    495228
Name: KM, dtype: int64
```