

Datset analysis using pandas

```
In [1]: import numpy as np
import pandas as pd
```

```
In [4]: data=pd.read_csv(r"E:\New folder\fiat500_VehicleSelection_Dataset (1).csv")
data
```

Out[4]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price | Unnamed: 9 | Unnamed: 10 |
|------|-----|--------|--------------|-------------|----------|-----------------|-----------|-------------|----------|------------|-------------|
| 0 | 1.0 | lounge | 51.0 | 882.0 | 25000.0 | 1.0 | 44.907242 | 8.611559868 | 8900 | NaN | NaN |
| 1 | 2.0 | pop | 51.0 | 1186.0 | 32500.0 | 1.0 | 45.666359 | 12.24188995 | 8800 | NaN | NaN |
| 2 | 3.0 | sport | 74.0 | 4658.0 | 142228.0 | 1.0 | 45.503300 | 11.41784 | 4200 | NaN | NaN |
| 3 | 4.0 | lounge | 51.0 | 2739.0 | 160000.0 | 1.0 | 40.633171 | 17.63460922 | 6000 | NaN | NaN |
| 4 | 5.0 | pop | 73.0 | 3074.0 | 106880.0 | 1.0 | 41.903221 | 12.49565029 | 5700 | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1544 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | length | 5 | NaN | NaN |
| 1545 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | concat | lonprice | NaN | NaN |
| 1546 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | Null values | NO | NaN | NaN |
| 1547 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | find | 1 | NaN | NaN |
| 1548 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | search | 1 | NaN | NaN |

1549 rows × 11 columns

```
In [5]: data.head(10)
```

Out[5]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price | Unnamed: 9 | Unnamed: 10 |
|---|------|--------|--------------|-------------|----------|-----------------|-----------|-------------|-------|------------|-------------|
| 0 | 1.0 | lounge | 51.0 | 882.0 | 25000.0 | 1.0 | 44.907242 | 8.611559868 | 8900 | NaN | NaN |
| 1 | 2.0 | pop | 51.0 | 1186.0 | 32500.0 | 1.0 | 45.666359 | 12.24188995 | 8800 | NaN | NaN |
| 2 | 3.0 | sport | 74.0 | 4658.0 | 142228.0 | 1.0 | 45.503300 | 11.41784 | 4200 | NaN | NaN |
| 3 | 4.0 | lounge | 51.0 | 2739.0 | 160000.0 | 1.0 | 40.633171 | 17.63460922 | 6000 | NaN | NaN |
| 4 | 5.0 | pop | 73.0 | 3074.0 | 106880.0 | 1.0 | 41.903221 | 12.49565029 | 5700 | NaN | NaN |
| 5 | 6.0 | pop | 74.0 | 3623.0 | 70225.0 | 1.0 | 45.000702 | 7.68227005 | 7900 | NaN | NaN |
| 6 | 7.0 | lounge | 51.0 | 731.0 | 11600.0 | 1.0 | 44.907242 | 8.611559868 | 10750 | NaN | NaN |
| 7 | 8.0 | lounge | 51.0 | 1521.0 | 49076.0 | 1.0 | 41.903221 | 12.49565029 | 9190 | NaN | NaN |
| 8 | 9.0 | sport | 73.0 | 4049.0 | 76000.0 | 1.0 | 45.548000 | 11.54946995 | 5600 | NaN | NaN |
| 9 | 10.0 | sport | 51.0 | 3653.0 | 89000.0 | 1.0 | 45.438301 | 10.99170017 | 6000 | NaN | NaN |

```
In [6]: data.tail(10)
```

Out[6]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price | Unnamed: 9 | Unnamed: 10 |
|------|-----|-------|--------------|-------------|-----|-----------------|-----|--------------------|-------------|------------|-------------|
| 1539 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | avg | 8576.003901 | NaN | NaN |
| 1540 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | count | 1538 | NaN | NaN |
| 1541 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | countif | 377 | NaN | >10000 |
| 1542 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | sumif | 4017825 | NaN | NaN |
| 1543 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | counta (not empty) | 1538 | NaN | NaN |
| 1544 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | length | 5 | NaN | NaN |
| 1545 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | concat | lonprice | NaN | NaN |
| 1546 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | Null values | NO | NaN | NaN |
| 1547 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | find | 1 | NaN | NaN |
| 1548 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | search | 1 | NaN | NaN |

```
In [8]: data.describe()
```

Out[8]:

| | ID | engine_power | age_in_days | km | previous_owners | lat | Unnamed: 9 |
|-------|-------------|--------------|-------------|---------------|-----------------|-------------|------------|
| count | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 1538.000000 | 0.0 |
| mean | 769.500000 | 51.904421 | 1650.980494 | 53396.011704 | 1.123537 | 43.541361 | NaN |
| std | 444.126671 | 3.988023 | 1289.522278 | 40046.830723 | 0.416423 | 2.133518 | NaN |
| min | 1.000000 | 51.000000 | 366.000000 | 1232.000000 | 1.000000 | 36.855839 | NaN |
| 25% | 385.250000 | 51.000000 | 670.000000 | 20006.250000 | 1.000000 | 41.802990 | NaN |
| 50% | 769.500000 | 51.000000 | 1035.000000 | 39031.000000 | 1.000000 | 44.394096 | NaN |
| 75% | 1153.750000 | 51.000000 | 2616.000000 | 79667.750000 | 1.000000 | 45.467960 | NaN |
| max | 1538.000000 | 77.000000 | 4658.000000 | 235000.000000 | 4.000000 | 46.795612 | NaN |

```
In [10]: data.size
```

Out[10]: 17039

```
In [11]: data.shape
```

Out[11]: (1549, 11)

```
In [12]: data.isna()
```

Out[12]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price | Unnamed: 9 | Unnamed: 10 |
|------|-------|-------|--------------|-------------|-------|-----------------|-------|-------|-------|------------|-------------|
| 0 | False | False | False | False | False | False | False | False | False | True | True |
| 1 | False | False | False | False | False | False | False | False | False | True | True |
| 2 | False | False | False | False | False | False | False | False | False | True | True |
| 3 | False | False | False | False | False | False | False | False | False | True | True |
| 4 | False | False | False | False | False | False | False | False | False | True | True |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1544 | True | True | True | True | True | True | True | False | False | True | True |
| 1545 | True | True | True | True | True | True | True | False | False | True | True |
| 1546 | True | True | True | True | True | True | True | False | False | True | True |
| 1547 | True | True | True | True | True | True | True | False | False | True | True |
| 1548 | True | True | True | True | True | True | True | False | False | True | True |

1549 rows × 11 columns

```
In [13]: data.fillna(value=50)
```

Out[13]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price | Unnamed: 9 | Unnamed: 10 |
|------|------|--------|--------------|-------------|----------|-----------------|-----------|-------------|----------|------------|-------------|
| 0 | 1.0 | lounge | 51.0 | 882.0 | 25000.0 | 1.0 | 44.907242 | 8.611559868 | 8900 | 50.0 | 50 |
| 1 | 2.0 | pop | 51.0 | 1186.0 | 32500.0 | 1.0 | 45.666359 | 12.24188995 | 8800 | 50.0 | 50 |
| 2 | 3.0 | sport | 74.0 | 4658.0 | 142228.0 | 1.0 | 45.503300 | 11.41784 | 4200 | 50.0 | 50 |
| 3 | 4.0 | lounge | 51.0 | 2739.0 | 160000.0 | 1.0 | 40.633171 | 17.63460922 | 6000 | 50.0 | 50 |
| 4 | 5.0 | pop | 73.0 | 3074.0 | 106880.0 | 1.0 | 41.903221 | 12.49565029 | 5700 | 50.0 | 50 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1544 | 50.0 | 50 | 50.0 | 50.0 | 50.0 | 50.0 | 50.000000 | length | 5 | 50.0 | 50 |
| 1545 | 50.0 | 50 | 50.0 | 50.0 | 50.0 | 50.0 | 50.000000 | concat | lonprice | 50.0 | 50 |
| 1546 | 50.0 | 50 | 50.0 | 50.0 | 50.0 | 50.0 | 50.000000 | Null values | NO | 50.0 | 50 |
| 1547 | 50.0 | 50 | 50.0 | 50.0 | 50.0 | 50.0 | 50.000000 | find | 1 | 50.0 | 50 |
| 1548 | 50.0 | 50 | 50.0 | 50.0 | 50.0 | 50.0 | 50.000000 | search | 1 | 50.0 | 50 |

1549 rows × 11 columns

```
In [14]: data[10:20]
```

Out[14]:

| | ID | model | engine_power | age_in_days | km | previous_owners | lat | lon | price | Unnamed: 9 | Unnamed: 10 |
|----|------|--------|--------------|-------------|----------|-----------------|-----------|-------------|-------|------------|-------------|
| 10 | 11.0 | pop | 51.0 | 790.0 | 43286.0 | 1.0 | 40.871429 | 14.43896008 | 8950 | NaN | NaN |
| 11 | 12.0 | lounge | 51.0 | 366.0 | 17500.0 | 1.0 | 45.069679 | 7.704919815 | 10990 | NaN | NaN |
| 12 | 13.0 | lounge | 51.0 | 456.0 | 18450.0 | 1.0 | 45.426571 | 11.78812981 | 9700 | NaN | NaN |
| 13 | 14.0 | pop | 51.0 | 3835.0 | 120000.0 | 1.0 | 40.531590 | 17.43615913 | 4800 | NaN | NaN |
| 14 | 15.0 | lounge | 51.0 | 1035.0 | 40500.0 | 1.0 | 40.911362 | 14.21119976 | 9300 | NaN | NaN |
| 15 | 16.0 | lounge | 51.0 | 1096.0 | 28200.0 | 1.0 | 45.697208 | 9.845970154 | 9500 | NaN | NaN |
| 16 | 17.0 | lounge | 73.0 | 4200.0 | 110000.0 | 1.0 | 41.082352 | 14.25424957 | 5250 | NaN | NaN |
| 17 | 18.0 | pop | 51.0 | 2223.0 | 96848.0 | 1.0 | 43.782372 | 11.25498962 | 7990 | NaN | NaN |
| 18 | 19.0 | lounge | 51.0 | 2861.0 | 31000.0 | 1.0 | 45.069679 | 7.704919815 | 7300 | NaN | NaN |
| 19 | 20.0 | lounge | 51.0 | 425.0 | 20030.0 | 1.0 | 45.354389 | 11.86925983 | 10500 | NaN | NaN |

```
In [15]: data["ID"]
```

Out[15]:

| | |
|------|-----|
| 0 | 1.0 |
| 1 | 2.0 |
| 2 | 3.0 |
| 3 | 4.0 |
| 4 | 5.0 |
| ... | |
| 1544 | NaN |
| 1545 | NaN |
| 1546 | NaN |
| 1547 | NaN |
| 1548 | NaN |

Name: ID, Length: 1549, dtype: float64

```
In [16]: data["km"]
```

Out[16]:

| | |
|------|----------|
| 0 | 25000.0 |
| 1 | 32500.0 |
| 2 | 142228.0 |
| 3 | 160000.0 |
| 4 | 106880.0 |
| ... | |
| 1544 | NaN |
| 1545 | NaN |
| 1546 | NaN |
| 1547 | NaN |
| 1548 | NaN |

Name: km, Length: 1549, dtype: float64

```
In [ ]:
```