

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

```
In [13]: import pandas as pd

diabetes = pd.read_csv("https://raw.githubusercontent.com/YBI-Foundation/Dataset/main")
print(diabetes.head())
print(diabetes.isnull().sum())
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	\
0	6	148.0	72.0	35.0	NaN	33.6	
1	1	85.0	66.0	29.0	NaN	26.6	
2	8	183.0	64.0	NaN	NaN	23.3	
3	1	89.0	66.0	23.0	94.0	28.1	
4	0	137.0	40.0	35.0	168.0	43.1	

	Diabetes_Pedigree	Age	Class
0	0.627	50	1
1	0.351	31	0
2	0.672	32	1
3	0.167	21	0
4	2.288	33	1
Pregnant	0		
Glucose	5		
Diastolic_BP	35		
Skin_Fold	227		
Serum_Insulin	374		
BMI	11		
Diabetes_Pedigree	0		
Age	0		
Class	0		
dtype:	int64		

```
In [17]: diabetes
```

	Pregnant	Glucose	Diastolic_BP	Skin_Fold	Serum_Insulin	BMI	Diabetes_Pedigre
0	6	148.0	72.0	35.0	NaN	33.6	0.62
1	1	85.0	66.0	29.0	NaN	26.6	0.35
2	8	183.0	64.0	NaN	NaN	23.3	0.67
3	1	89.0	66.0	23.0	94.0	28.1	0.16
4	0	137.0	40.0	35.0	168.0	43.1	2.28
...
763	10	101.0	76.0	48.0	180.0	32.9	0.17
764	2	122.0	70.0	27.0	NaN	36.8	0.34
765	5	121.0	72.0	23.0	112.0	26.2	0.24
766	1	126.0	60.0	NaN	NaN	30.1	0.34
767	1	93.0	70.0	31.0	NaN	30.4	0.31

768 rows × 9 columns

```
In [19]: car_data = pd.read_csv("https://raw.githubusercontent.com/mrdbourke/zero-to-mastery-m")
print(car_data.head())
print(car_data.isna().sum())
```

	Make	Colour	Odometer (KM)	Doors	Price
0	Honda	White	35431.0	4.0	15323.0
1	BMW	Blue	192714.0	5.0	19943.0
2	Honda	White	84714.0	4.0	28343.0
3	Toyota	White	154365.0	4.0	13434.0
4	Nissan	Blue	181577.0	3.0	14043.0
Make			49		
Colour			50		
Odometer (KM)			50		
Doors			50		
Price			50		
			dtype: int64		

In [21]: `car_data`

Out[21]:

	Make	Colour	Odometer (KM)	Doors	Price
0	Honda	White	35431.0	4.0	15323.0
1	BMW	Blue	192714.0	5.0	19943.0
2	Honda	White	84714.0	4.0	28343.0
3	Toyota	White	154365.0	4.0	13434.0
4	Nissan	Blue	181577.0	3.0	14043.0
...
995	Toyota	Black	35820.0	4.0	32042.0
996	NaN	White	155144.0	3.0	5716.0
997	Nissan	Blue	66604.0	4.0	31570.0
998	Honda	White	215883.0	4.0	4001.0
999	Toyota	Blue	248360.0	4.0	12732.0

1000 rows × 5 columns

In [23]: `car_data.head(7)`

Out[23]:

	Make	Colour	Odometer (KM)	Doors	Price
0	Honda	White	35431.0	4.0	15323.0
1	BMW	Blue	192714.0	5.0	19943.0
2	Honda	White	84714.0	4.0	28343.0
3	Toyota	White	154365.0	4.0	13434.0
4	Nissan	Blue	181577.0	3.0	14043.0
5	Honda	Red	42652.0	4.0	23883.0
6	Toyota	Blue	163453.0	4.0	8473.0

In [25]: `car_data.tail(6)`

Out[25]:

	Make	Colour	Odometer (KM)	Doors	Price
994	BMW	Blue	163322.0	3.0	31666.0
995	Toyota	Black	35820.0	4.0	32042.0
996	NaN	White	155144.0	3.0	5716.0
997	Nissan	Blue	66604.0	4.0	31570.0
998	Honda	White	215883.0	4.0	4001.0
999	Toyota	Blue	248360.0	4.0	12732.0

In [27]: car_data.tail(30)

Out[27]:

	Make	Colour	Odometer (KM)	Doors	Price
970	Toyota	Blue	186309.0	4.0	16416.0
971	BMW	Black	178164.0	3.0	24891.0
972	Honda	White	NaN	4.0	17939.0
973	Honda	Green	237627.0	4.0	8430.0
974	NaN	White	155383.0	4.0	14345.0
975	Honda	NaN	22409.0	4.0	10429.0
976	Toyota	Blue	95317.0	4.0	7435.0
977	Toyota	Blue	128016.0	4.0	16835.0
978	BMW	White	85739.0	5.0	48419.0
979	Toyota	Black	17975.0	4.0	17940.0
980	Toyota	Blue	230314.0	4.0	6720.0
981	Toyota	White	129454.0	4.0	6446.0
982	Honda	White	238172.0	4.0	13273.0
983	Toyota	Red	NaN	4.0	14671.0
984	Nissan	Blue	157235.0	4.0	4196.0
985	NaN	Blue	216250.0	4.0	9691.0
986	Honda	White	71934.0	4.0	26882.0
987	Honda	White	215235.0	4.0	3825.0
988	Nissan	Black	248736.0	4.0	8358.0
989	Toyota	Red	41735.0	4.0	13928.0
990	Toyota	White	173408.0	4.0	8082.0
991	Honda	Blue	235985.0	4.0	9184.0
992	Honda	Green	54721.0	4.0	27419.0
993	Nissan	Black	162523.0	4.0	4696.0
994	BMW	Blue	163322.0	3.0	31666.0
995	Toyota	Black	35820.0	4.0	32042.0
996	NaN	White	155144.0	3.0	5716.0
997	Nissan	Blue	66604.0	4.0	31570.0
998	Honda	White	215883.0	4.0	4001.0
999	Toyota	Blue	248360.0	4.0	12732.0

--

```
In [29]: car_data.shape
```

```
Out[29]: (1000, 5)
```

```
In [31]: car_data.size
```

```
Out[31]: 5000
```

```
In [33]: car_data.ndim
```

```
Out[33]: 2
```

```
In [35]: car_data.columns
```

```
Out[35]: Index(['Make', 'Colour', 'Odometer (KM)', 'Doors', 'Price'], dtype='object')
```

```
In [37]: car_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   Make            951 non-null    object
 1   Colour          950 non-null    object
 2   Odometer (KM)   950 non-null    float64
 3   Doors           950 non-null    float64
 4   Price           950 non-null    float64
dtypes: float64(3), object(2)
memory usage: 39.2+ KB
```

```
In [39]: car_data.values
```

```
Out[39]: array([[ 'Honda', 'White', 35431.0, 4.0, 15323.0],
                [ 'BMW', 'Blue', 192714.0, 5.0, 19943.0],
                [ 'Honda', 'White', 84714.0, 4.0, 28343.0],
                ...,
                [ 'Nissan', 'Blue', 66604.0, 4.0, 31570.0],
                [ 'Honda', 'White', 215883.0, 4.0, 4001.0],
                [ 'Toyota', 'Blue', 248360.0, 4.0, 12732.0]], dtype=object)
```

```
In [41]: car_data.describe()
```

```
Out[41]:
```

	Odometer (KM)	Doors	Price
count	950.000000	950.000000	950.000000
mean	131253.237895	4.011579	16042.814737
std	69094.857187	0.382539	8581.695036
min	10148.000000	3.000000	2796.000000
25%	70391.250000	4.000000	9529.250000
50%	131821.000000	4.000000	14297.000000
75%	192668.500000	4.000000	20806.250000
max	249860.000000	5.000000	52458.000000

```
In [43]: car_data.describe(include='all')
```

Out[43]:

	Make	Colour	Odometer (KM)	Doors	Price
count	951	950	950.000000	950.000000	950.000000
unique	4	5	NaN	NaN	NaN
top	Toyota	White	NaN	NaN	NaN
freq	379	390	NaN	NaN	NaN
mean	NaN	NaN	131253.237895	4.011579	16042.814737
std	NaN	NaN	69094.857187	0.382539	8581.695036
min	NaN	NaN	10148.000000	3.000000	2796.000000
25%	NaN	NaN	70391.250000	4.000000	9529.250000
50%	NaN	NaN	131821.000000	4.000000	14297.000000
75%	NaN	NaN	192668.500000	4.000000	20806.250000
max	NaN	NaN	249860.000000	5.000000	52458.000000

In [63]:

car_data[['Make', 'Price']]

Out[63]:

	Make	Price
0	Honda	15323.0
1	BMW	19943.0
2	Honda	28343.0
3	Toyota	13434.0
4	Nissan	14043.0
...
995	Toyota	32042.0
996	NaN	5716.0
997	Nissan	31570.0
998	Honda	4001.0
999	Toyota	12732.0

1000 rows × 2 columns

In [65]:

car_data['Colour']

Out[65]:

0	White
1	Blue
2	White
3	White
4	Blue
...	...
995	Black
996	White
997	Blue
998	White
999	Blue

Name: Colour, Length: 1000, dtype: object

In [67]:

car_data.loc[5]

```
Out[67]: Make Honda
Colour Red
Odometer (KM) 42652.0
Doors 4.0
Price 23883.0
Name: 5, dtype: object
```

```
In [69]: car_data.loc[2:6]
```

Out[69]:

	Make	Colour	Odometer (KM)	Doors	Price
2	Honda	White	84714.0	4.0	28343.0
3	Toyota	White	154365.0	4.0	13434.0
4	Nissan	Blue	181577.0	3.0	14043.0
5	Honda	Red	42652.0	4.0	23883.0
6	Toyota	Blue	163453.0	4.0	8473.0

```
In [71]: car_data.iloc[10]
```

```
Out[71]: Make Nissan
Colour White
Odometer (KM) 167421.0
Doors 4.0
Price 16259.0
Name: 10, dtype: object
```

```
In [73]: car_data.iloc[5:10]
```

Out[73]:

	Make	Colour	Odometer (KM)	Doors	Price
5	Honda	Red	42652.0	4.0	23883.0
6	Toyota	Blue	163453.0	4.0	8473.0
7	Honda	White	NaN	4.0	20306.0
8	NaN	White	130538.0	4.0	9374.0
9	Honda	Blue	51029.0	4.0	26683.0

```
In [75]: car_data[car_data['Odometer (KM)'] > 100000]
```

Out[75]:

	Make	Colour	Odometer (KM)	Doors	Price
1	BMW	Blue	192714.0	5.0	19943.0
3	Toyota	White	154365.0	4.0	13434.0
4	Nissan	Blue	181577.0	3.0	14043.0
6	Toyota	Blue	163453.0	4.0	8473.0
8	NaN	White	130538.0	4.0	9374.0
...
993	Nissan	Black	162523.0	4.0	4696.0
994	BMW	Blue	163322.0	3.0	31666.0
996	NaN	White	155144.0	3.0	5716.0
998	Honda	White	215883.0	4.0	4001.0
999	Toyota	Blue	248360.0	4.0	12732.0

```
In [77]: car_data[car_data['Colour'] == 'Blue']
```

Out[77]:

	Make	Colour	Odometer (KM)	Doors	Price
1	BMW	Blue	192714.0	5.0	19943.0
4	Nissan	Blue	181577.0	3.0	14043.0
6	Toyota	Blue	163453.0	4.0	8473.0
9	Honda	Blue	51029.0	4.0	26683.0
14	Honda	Blue	199833.0	4.0	18946.0
...
985	NaN	Blue	216250.0	4.0	9691.0
991	Honda	Blue	235985.0	4.0	9184.0
994	BMW	Blue	163322.0	3.0	31666.0
997	Nissan	Blue	66604.0	4.0	31570.0
999	Toyota	Blue	248360.0	4.0	12732.0

302 rows × 5 columns

```
In [79]: car_data[car_data['Colour'] == 'Red']
```

Out[79]:

	Make	Colour	Odometer (KM)	Doors	Price
5	Honda	Red	42652.0	4.0	23883.0
16	Toyota	Red	96742.0	4.0	34465.0
50	Toyota	Red	184878.0	4.0	7445.0
57	BMW	Red	162176.0	5.0	19186.0
60	BMW	Red	208124.0	5.0	24668.0
...
939	Nissan	Red	153074.0	4.0	17487.0
941	Toyota	Red	166046.0	4.0	7955.0
945	BMW	Red	156064.0	5.0	17105.0
983	Toyota	Red	NaN	4.0	14671.0
989	Toyota	Red	41735.0	4.0	13928.0

88 rows × 5 columns

```
In [81]: car_data[car_data['Make'].str.startswith('T', na=False)]
```

Out[81]:

	Make	Colour	Odometer (KM)	Doors	Price
3	Toyota	White	154365.0	4.0	13434.0
6	Toyota	Blue	163453.0	4.0	8473.0
15	Toyota	Blue	205592.0	4.0	16290.0
16	Toyota	Red	96742.0	4.0	34465.0
20	Toyota	NaN	124844.0	4.0	24130.0
...
983	Toyota	Red	NaN	4.0	14671.0
989	Toyota	Red	41735.0	4.0	13928.0
990	Toyota	White	173408.0	4.0	8082.0
995	Toyota	Black	35820.0	4.0	32042.0
999	Toyota	Blue	248360.0	4.0	12732.0

379 rows × 5 columns

In [83]:

```
car_data[car_data['Make'].str.startswith('Z', na=False)]
```

Out[83]:

Make	Colour	Odometer (KM)	Doors	Price
------	--------	---------------	-------	-------

In [85]:

```
brands = ['Toyota', 'BMW']
car_data[car_data['Make'].isin(brands)]
```

Out[85]:

	Make	Colour	Odometer (KM)	Doors	Price
1	BMW	Blue	192714.0	5.0	19943.0
3	Toyota	White	154365.0	4.0	13434.0
6	Toyota	Blue	163453.0	4.0	8473.0
15	Toyota	Blue	205592.0	4.0	16290.0
16	Toyota	Red	96742.0	4.0	34465.0
...
989	Toyota	Red	41735.0	4.0	13928.0
990	Toyota	White	173408.0	4.0	8082.0
994	BMW	Blue	163322.0	3.0	31666.0
995	Toyota	Black	35820.0	4.0	32042.0
999	Toyota	Blue	248360.0	4.0	12732.0

476 rows × 5 columns

In [87]:

```
car_data.query('Doors == 5 and Price < 50000')
```


Out[87]:

	Make	Colour	Odometer (KM)	Doors	Price
1	BMW	Blue	192714.0	5.0	19943.0
17	BMW	White	194189.0	5.0	17177.0
55	BMW	White	79937.0	5.0	48686.0
57	BMW	Red	162176.0	5.0	19186.0
60	BMW	Red	208124.0	5.0	24668.0
...
945	BMW	Red	156064.0	5.0	17105.0
953	BMW	White	102773.0	5.0	29551.0
967	BMW	Green	26839.0	5.0	15224.0
969	BMW	Blue	45207.0	5.0	35254.0
978	BMW	White	85739.0	5.0	48419.0

70 rows × 5 columns

```
In [89]: car_data.query('Doors == 4 and Price < 5000')
```

Out[89]:

	Make	Colour	Odometer (KM)	Doors	Price
86	Nissan	Blue	NaN	4.0	3300.0
109	Honda	Blue	219217.0	4.0	4747.0
133	Honda	Green	125802.0	4.0	4754.0
147	Toyota	Blue	103909.0	4.0	4566.0
173	Nissan	Black	240536.0	4.0	3247.0
330	Toyota	NaN	170466.0	4.0	3974.0
345	Toyota	White	127161.0	4.0	4871.0
357	Nissan	White	NaN	4.0	3974.0
361	Toyota	White	164093.0	4.0	3893.0
373	Toyota	Blue	112223.0	4.0	4753.0
392	Toyota	White	227387.0	4.0	3895.0
406	Honda	Blue	210446.0	4.0	3594.0
424	Toyota	White	189194.0	4.0	4946.0
487	Toyota	White	143104.0	4.0	4589.0
497	Nissan	Red	209976.0	4.0	4200.0
523	Honda	Blue	207344.0	4.0	4941.0
536	Honda	White	205597.0	4.0	4239.0
678	Toyota	Green	213861.0	4.0	3751.0
700	Honda	NaN	203447.0	4.0	3509.0
739	Honda	White	187380.0	4.0	4022.0
757	Nissan	Blue	233501.0	4.0	2796.0
759	Nissan	Red	113987.0	4.0	4968.0
768	Honda	Blue	217068.0	4.0	4358.0
781	Honda	Black	158337.0	4.0	4733.0
796	Toyota	Red	190211.0	4.0	4454.0
804	Toyota	White	185146.0	4.0	4566.0
845	Toyota	Green	243177.0	4.0	3125.0
880	Honda	Blue	200219.0	4.0	4606.0
961	Toyota	White	239760.0	4.0	4896.0
984	Nissan	Blue	157235.0	4.0	4196.0
987	Honda	White	215235.0	4.0	3825.0
993	Nissan	Black	162523.0	4.0	4696.0
998	Honda	White	215883.0	4.0	4001.0

```
In [91]: car_data.query('Doors == 5 and Price < 5000')
```

Out[91]:

Make	Colour	Odometer (KM)	Doors	Price
------	--------	---------------	-------	-------

```
In [93]: car_data.isnull().sum()
```

```
Out[93]: Make          49
         Colour        50
         Odometer (KM)  50
         Doors         50
         Price         50
         dtype: int64
```

```
In [95]: car_data.isna().values.any()
```

```
Out[95]: True
```

```
In [97]: car_data.duplicated().sum()
```

```
Out[97]: 1
```

```
In [99]: car_filled = car_data.copy()

         numeric_cols = car_filled.select_dtypes(include=np.number).columns
         car_filled[numeric_cols] = car_filled[numeric_cols].fillna(
             car_filled[numeric_cols].median()
         )
```

```
In [107... categorical_cols = car_filled.select_dtypes(include='object').columns

         for col in categorical_cols:
             car_filled[col].fillna(car_filled[col].mode()[0], inplace=True)
```

```
In [109... car_filled.isna().values.any()
```

```
Out[109... False
```

```
In [111... car_dropped = car_filled.drop(columns=['Colour'])
         car_dropped.head()
```

```
Out[111...
   Make  Odometer (KM)  Doors  Price
0  Honda          35431.0    4.0  15323.0
1   BMW          192714.0    5.0  19943.0
2  Honda          84714.0    4.0  28343.0
3 Toyota         154365.0    4.0  13434.0
4  Nissan         181577.0    3.0  14043.0
```

```
In [113... car_data.head()
```

```
Out[113...
   Make  Colour  Odometer (KM)  Doors  Price
0  Honda   White          35431.0    4.0  15323.0
1   BMW    Blue          192714.0    5.0  19943.0
2  Honda   White          84714.0    4.0  28343.0
3 Toyota   White         154365.0    4.0  13434.0
4  Nissan    Blue         181577.0    3.0  14043.0
```

```
In [115... car_dropped['Price'] = car_dropped['Price'].astype('float')
```

```
In [117... car_dropped.dtypes
```

```
Out[117... Make          object
Odometer (KM)  float64
Doors          float64
Price         float64
dtype: object
```

```
In [119... car_data.dtypes
```

```
Out[119... Make          object
Colour         object
Odometer (KM)  float64
Doors          float64
Price         float64
dtype: object
```

```
In [139... print(car_data)
```

	Make	Colour	Odometer (KM)	Doors	Price
0	Honda	White	35431.0	4.0	15323.0
1	BMW	Blue	192714.0	5.0	19943.0
2	Honda	White	84714.0	4.0	28343.0
3	Toyota	White	154365.0	4.0	13434.0
4	Nissan	Blue	181577.0	3.0	14043.0
...
995	Toyota	Black	35820.0	4.0	32042.0
996	NaN	White	155144.0	3.0	5716.0
997	Nissan	Blue	66604.0	4.0	31570.0
998	Honda	White	215883.0	4.0	4001.0
999	Toyota	Blue	248360.0	4.0	12732.0

[1000 rows x 5 columns]

```
In [143... car_data.index
```

```
Out[143... RangeIndex(start=0, stop=1000, step=1)
```

```
In [145... car_data.values
```

```
Out[145... array([[ 'Honda', 'White', 35431.0, 4.0, 15323.0],
        [ 'BMW', 'Blue', 192714.0, 5.0, 19943.0],
        [ 'Honda', 'White', 84714.0, 4.0, 28343.0],
        ...,
        [ 'Nissan', 'Blue', 66604.0, 4.0, 31570.0],
        [ 'Honda', 'White', 215883.0, 4.0, 4001.0],
        [ 'Toyota', 'Blue', 248360.0, 4.0, 12732.0]], dtype=object)
```

```
In [147... car_data.value_counts()
```

Make	Colour	Odometer (KM)	Doors	Price	
BMW	Black	11049.0	3.0	19500.0	1
Toyota	Black	98409.0	4.0	32863.0	1
	Blue	44815.0	4.0	8460.0	1
		48684.0	4.0	33817.0	1
		51155.0	4.0	15960.0	1
	
Honda	White	92883.0	4.0	14931.0	1
		95481.0	4.0	8687.0	1
		95579.0	4.0	11135.0	1
		100527.0	4.0	22616.0	1
Toyota	White	248815.0	4.0	9785.0	1

Name: count, Length: 773, dtype: int64

```
In [149... car_data.empty
```

```
Out[149... False
```

In [151...

car_data.tail(7)

Out[151...

	Make	Colour	Odometer (KM)	Doors	Price
993	Nissan	Black	162523.0	4.0	4696.0
994	BMW	Blue	163322.0	3.0	31666.0
995	Toyota	Black	35820.0	4.0	32042.0
996	NaN	White	155144.0	3.0	5716.0
997	Nissan	Blue	66604.0	4.0	31570.0
998	Honda	White	215883.0	4.0	4001.0
999	Toyota	Blue	248360.0	4.0	12732.0

In []: