

+ <> + ⌂



RAM

Disk



✓  
0s

▶ import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt

✓  
0s

▶ import os  
for dirname, \_, filenames in os.walk(  
 for filename in filenames:  
 print(os.path.join(dirname,

✓  
1s

[8] df=pd.read\_csv('/archive.zip')

✓  
1s

▶ df.head()



Car Make	Car Model	Year	Engin	Si	(
----------	-----------	------	-------	----	---

0	Porsche	911	2022		
---	---------	-----	------	--	--

1	Lamborghini	Huracan	2021	5	
---	-------------	---------	------	---	--

2	Ferrari	488 GTB	2022	3	
---	---------	---------	------	---	--

3	Audi	R8	2022	5	
---	------	----	------	---	--

4	McLaren	720S	2021		
---	---------	------	------	--	--

+ <> + ⌂



RAM



Disk



✓  
0s



df.info()

```
→ <class 'pandas.core.frame.DataFrame'
RangeIndex: 1007 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  -- 
 0   Car Make         1007 non-null
 1   Car Model        1007 non-null
 2   Year             1007 non-null
 3   Engine Size (L) 997 non-null
 4   Horsepower       1007 non-null
 5   Torque (lb-ft)  1007 non-null
 6   0-60 MPH Time (seconds) 1007 non-null
 7   Price (in USD)  1007 non-null
dtypes: int64(1), object(7)
memory usage: 63.1+ KB
```

✓  
0s



df.nunique()



0

Car Make

38

Car Model

176

Year

9

Engine Size (L)

45

Horsepower

124

+ <> + ⌂



RAM  
Disk



✓  
0s



0



Car Make

38

Car Model

176

Year

9

Engine Size (L)

45

Horsepower

124

Torque (lb-ft)

93

0-60 MPH Time (seconds)

43

Price (in USD)

367

**dtype:** int64

✓  
0s



df['Engine Size (L)'].describe()



Engine Size (L)

count

997

unique

45

top

4

freq

219

+ <> + ⌂



RAM

Disk



0s



→ dtype: object

✓  
0s



null\_data=df[df['Engine Size (L)'].isnull()]  
null\_data



		Car Make	Car Model	Year	Engin Siz (L)
168	Rimac	C_Two	2022	Nan	
171	Tesla	Model S Plaid	2021	Nan	
222	Porsche	Taycan Turbo S	2021	Nan	
247	Tesla	Model S Plaid	2022	Nan	
387	Rimac	C_Two	2022	Nan	
389	Tesla	Roadster	2022	Nan	
686	Rimac	C_Two	2022	Nan	
697	Lotus	Evija	2022	Nan	
752	Porsche	Taycan	2022	Nan	
916	Tesla	Roadster	2022	Nan	

+ <> + ⌂



RAM



Disk



✓  
0s



```
df['Engine Size (L)'].fillna('Electr...')
```

✓  
0s



```
df.info()
```



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1007 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  -- 
 0   Car Make         1007 non-null
 1   Car Model        1007 non-null
 2   Year             1007 non-null
 3   Engine Size (L) 1007 non-null
 4   Horsepower       1007 non-null
 5   Torque (lb-ft)  1007 non-null
 6   0-60 MPH Time (seconds) 1007 non-null
 7   Price (in USD)  1007 non-null
dtypes: int64(1), object(7)
memory usage: 63.1+ KB
```

✓  
0s



```
df[df['Torque (lb-ft)'].isnull()]==1
```



Car Make	Car Model	Year
----------	-----------	------

642

Tesla

Model S  
Plaid

2021

El

+ <> + ⌂



RAM



Disk



✓  
0s



```
df[df['Torque (lb-ft)'].isnull() == 1]
```



En

Car Make	Car Model	Year
-------------	-----------	------

642	Tesla	Model S Plaid	2021	El
-----	-------	------------------	------	----

878	Maserati	GranTurismo	2021	El
-----	----------	-------------	------	----

916	Tesla	Roadster	2022	El
-----	-------	----------	------	----

✓  
0s



```
df['Torque (lb-ft)'].fillna(0, inplace=True)
```

✓  
0s



```
df.info()
```



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

RangeIndex: 1007 entries, 0 to 1006

Data columns (total 8 columns):

#	Column	Non-null count
---	--------	----------------

0	Car Make	1007
---	----------	------

1	Car Model	1007
---	-----------	------

2	Year	1007
---	------	------

3	Engine Size (L)	1007
---	-----------------	------

4	Horsepower	1007
---	------------	------

+ <> + TT



RAM



Disk



✓  
0s



df.info()

```
→ <class 'pandas.core.frame.DataFrame'
RangeIndex: 1007 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 --- 
 0   Car Make        1007 non-null
 1   Car Model       1007 non-null
 2   Year            1007 non-null
 3   Engine Size (L) 1007 non-null
 4   Horsepower      1007 non-null
 5   Torque (lb-ft)  1007 non-null
 6   0-60 MPH Time (seconds) 1007 non-null
 7   Price (in USD)  1007 non-null
dtypes: int64(1), object(7)
memory usage: 63.1+ KB
```

✓  
0s

```
[19] unique_year=df["Year"].sort_values()
unique_year
```



```
array([1965, 2014, 2015, 2017,
       2019, 2020, 2021, 2022, 2023])
```

✓  
0s



```
df['Car Model'].groupby(df['Year'])
plt.title('Year wise released car model')
plt.xlabel('Year', fontdict={'fontsize': 12})
```



```
Text(0.5, 0, 'Year')
```

```
WARNING:matplotlib.font_manager: FontWarning: findfont: Could not find family 'serif' in
```

```
WARNING:matplotlib.font_manager: findfont: Could not find family 'serif' in
```

+ <> + T



RAM



Disk



✓  
0s



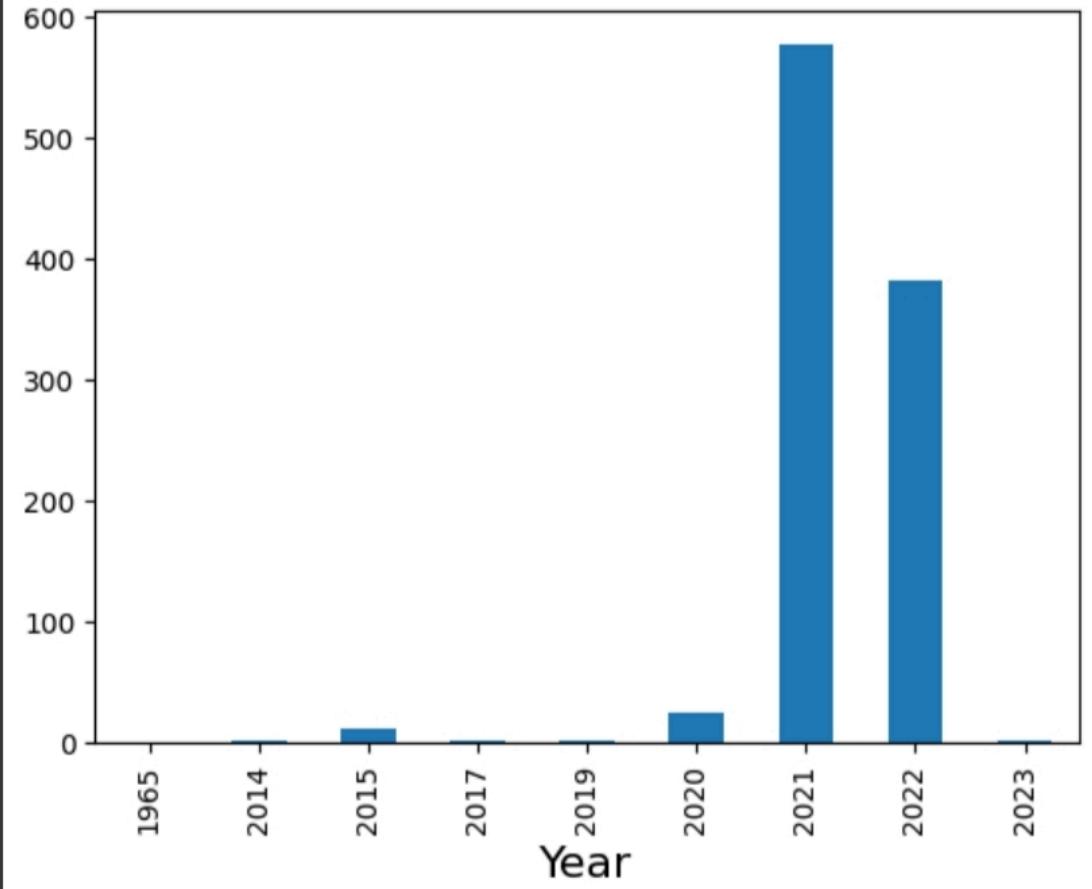
```
df['Car Model'].groupby(df['Year'])  
plt.title('Year wise released car models')  
plt.xlabel('Year', fontdict={'fontsize': 12})
```



Text(0.5, 0, 'Year')

WARNING:matplotlib.font\_manager:FontWarning: Font-family not found in fontset: serif

Year wise released car models



✓  
0s



df.describe()

+ <> + ⌂



RAM



Disk



✓  
0s



df.describe()



Year



count 1007.000000

mean 2021.201589

std 2.019802

min 1965.000000

25% 2021.000000

50% 2021.000000

75% 2022.000000

max 2023.000000

✓  
0s



df.info()



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

RangeIndex: 1007 entries, 0 to 1006

Data columns (total 8 columns):

#	Column	Non-Null Count
---	---	---
0	Car Make	1007
1	Car Model	1007
2	Year	1007
3	Engine Size (L)	1007
4	Horsepower	1007

+ <> + TT



RAM



Disk



✓  
0s



df.info()

```
→ <class 'pandas.core.frame.DataFrame'
RangeIndex: 1007 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  -- 
 0   Car Make        1007 non-null
 1   Car Model       1007 non-null
 2   Year            1007 non-null
 3   Engine Size (L) 1007 non-null
 4   Horsepower      1007 non-null
 5   Torque (lb-ft)  1007 non-null
 6   0-60 MPH Time (seconds) 1007 non-null
 7   Price (in USD)  1007 non-null
dtypes: int64(1), object(7)
memory usage: 63.1+ KB
```

✓  
0s



```
for column in df.columns:
    print(column.upper())
    print(df[column].value_counts())
    print("-"*50)
```

```
→ 2.9          62
  2.8          58
  3.6          51
  3.1          50
  2.5          48
  4             42
  3             37
  4.4          37
```

+ <> + ⌂



RAM



Disk



✓  
0s

```
▶ for column in df.columns:  
    print(column.upper())  
    print(df[column].value_counts())  
    print("-"*50)
```

2.9	62
2.8	58
3.6	51
3.1	50
2.5	48
4	42
3	37
4.4	37
3.2	35
4.2	32
4.9	26
4.1	25
3.3	23
2.7	22
3.7	21
3.9	21
3.4	21
4.7	20
1.9	19
4.8	15
2.4	14
2.6	13
5	11
5.4	10
4.5	8
4.3	8
5.3	7
1.85	7

+ <> + ⌂



RAM



Disk



✓  
0s



PRICE (IN USD)

Price (in USD)



500,000 34

3,000,000 24

625,000 22

58,900 17

114,000 16

..

64,100 1

81,550 1

62,250 1

100,550 1

27,205 1

Name: count, Length: 367, dtype:

✓  
0s



df[df["0-60 MPH Time (seconds)"] ==



Car Make	Car Model	Year	Engine Size (L)
----------	-----------	------	-----------------

364 Tesla Roadster 2023 Electric

✓  
0s



df.loc[df["0-60 MPH Time (seconds)"] ==

✓  
0s



df[df["0-60 MPH Time (seconds)"] ==

+ < > + ⌂



RAM



Disk



✓  
0s



df[df["0-60 MPH Time (seconds)"] ==



Eng

S

Car  
Make

Car  
Model

Year

99

Tesla

Roadster

2022

Elec

168

Rimac

C\_Two

2022

Elec  
Mo

247

Tesla

Model S  
Plaid

2022

Elec  
Mo

280

Pininfarina

Battista

2022

Elec

320

Tesla

Model S  
Plaid

2021

Elec

335

Tesla

Roadster

2022

354

Tesla

Roadster

2022

Elec

364

Tesla

Roadster

2023

Elec

389

Tesla

Roadster

2022

Elec  
Mo

401

Tesla

Model S  
Plaid

2021

Elec  
mo

509

Rimac

C\_Two

2021

Elec

+ <> + ⌂



RAM



Disk



✓	0s	▶	511	Tesla	Model S Plaid	2021	Elec
➡			526	Rimac	C_Two	2022	Elec
			640	Rimac	Nevera	2021	Elec
			642	Tesla	Model S Plaid	2021	Elec
			821	Tesla	Roadster	2022	Elec
			885	Tesla	Roadster	2022	
			916	Tesla	Roadster	2022	Elec Mo
			965	Tesla	Model S Plaid	2022	Elec
			988	Pininfarina	Battista	2021	Elec

✓  
0s



df.columns

```
➡ Index(['Car Make', 'Car Model',
       'Year', 'Engine Size (L)',
       'Horsepower',
       'Torque (lb-ft)', '0-60
       MPH Time (seconds)', 'Price (in
       USD)'],
      dtype='object')
```

+ <> + ⌂



RAM



Disk



✓  
0s

▶ bad\_cols = ['Horsepower', 'Torque (

```
df[bad_cols] = df[bad_cols].astype(
```

✓  
0s

▶ def remove\_unwanted(characters, cols):  
 for character in characters:  
 charCol = []  
 for col in cols:  
 for i in df[col]:  
 if character in i:  
 charCol.append(i)  
  
 charCol = list(set(charCol))  
  
 for i in charCol:  
 df[i] = df[i].str.replace(  
  
remove\_unwanted(["+", ",", "-", "="], ba

✓  
0s

▶ df



EI

Car	Make	Car Model	Year
-----	------	-----------	------

0	Porsche	911	2022
---	---------	-----	------

1	Lamborghini	Huracan	2021
---	-------------	---------	------

2	Ferrari	488 GTB	2022
---	---------	---------	------

+ <> + ⌂



RAM Disk



✓  
0s



df



EI

Car Make Car Model Year

0 Porsche 911 2022

1 Lamborghini Huracan 2021

2 Ferrari 488 GTB 2022

3 Audi R8 2022

4 McLaren 720S 2021

... ... ... ...

1002 Koenigsegg Jesko 2022

1003 Lotus Evija 2021 E

1004 McLaren Senna 2021

1005 Pagani Huayra 2021

1006 Rimac Nevera 2021 E

1007 rows × 8 columns

+ &lt;&gt; + ⌂



RAM

Disk



0s



```
df["Horsepower"] = df["Horsepower"]
df["Torque (lb-ft)"] = df["Torque (lb-ft)"]
df["0-60 MPH Time (seconds)"] = df["0-60 MPH Time (seconds)"]
df["Price (in USD)"] = df["Price (USD)"]
df["Year"] = df["Year"]
```



E

	Car	Make	Car Model	Year
--	-----	------	-----------	------

0	Porsche	911	2022
---	---------	-----	------

1	Lamborghini	Huracan	2021
---	-------------	---------	------

2	Ferrari	488 GTB	2022
---	---------	---------	------

3	Audi	R8	2022
---	------	----	------

4	McLaren	720S	2021
---	---------	------	------

...	...	...	...
-----	-----	-----	-----

1002	Koenigsegg	Jesko	2022
------	------------	-------	------

1003	Lotus	Evija	2021
------	-------	-------	------

1004	McLaren	Senna	2021
------	---------	-------	------

1005	Pagani	Huayra	2021
------	--------	--------	------

1006	Rimac	Nevera	2021
------	-------	--------	------

E

E

+ <> + ⌂



RAM Disk



✓  
0s



df.describe()



Year Horsepower

<b>count</b>	1007.000000	1007.000000	1
<b>mean</b>	2021.201589	657.984111	
<b>std</b>	2.019802	593.017842	
<b>min</b>	1965.000000	181.000000	
<b>25%</b>	2021.000000	454.000000	
<b>50%</b>	2021.000000	591.000000	
<b>75%</b>	2022.000000	708.500000	
<b>max</b>	2023.000000	10000.000000	10

✓  
0s



df.info()



<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1007 entries, 0 to 1006  
Data columns (total 8 columns):  
 # Column Non-Null Count  
 --- --  
 0 Car Make 1007 non-null

+ <> + ⌂



RAM

Disk



✓  
0s



df.info()

```
→ <class 'pandas.core.frame.DataFrame'
RangeIndex: 1007 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  --   --   --
 0   Car Make        1007 non-null
 1   Car Model       1007 non-null
 2   Year            1007 non-null
 3   Engine Size (L) 1007 non-null
 4   Horsepower      1007 non-null
 5   Torque (lb-ft)  1007 non-null
 6   0-60 MPH Time (seconds) 1007 non-null
 7   Price (in USD)  1007 non-null
dtypes: float64(2), int64(3), object(3)
memory usage: 63.1+ KB
```

✓  
0s



df.dropna(inplace=True)

✓  
0s



df.info()

```
→ <class 'pandas.core.frame.DataFrame'
Index: 1006 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  --   --   --
 0   Car Make        1006 non-null
 1   Car Model       1006 non-null
 2   Year            1006 non-null
```

+ <> + ⌂



RAM   
Disk 



✓  
0s



df.dropna(inplace=True)

✓  
0s



df.info()



```
<class 'pandas.core.frame.DataFrame'
Index: 1006 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  -- 
 0   Car Make        1006 non-null
 1   Car Model       1006 non-null
 2   Year            1006 non-null
 3   Engine Size (L) 1006 non-null
 4   Horsepower      1006 non-null
 5   Torque (lb-ft)  1006 non-null
 6   0-60 MPH Time (seconds) 1006 non-null
 7   Price (in USD)  1006 non-null
dtypes: float64(2), int64(3), object(3)
memory usage: 70.7+ KB
```

✓  
0s



df.dropna(inplace=True)

✓  
0s



df.info()



```
<class 'pandas.core.frame.DataFrame'
Index: 1006 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  -- 
 0   Car Make        1006 non-null
 1   Car Model       1006 non-null
 2   Year            1006 non-null
 3   Engine Size (L) 1006 non-null
 4   Horsepower      1006 non-null
 5   Torque (lb-ft)  1006 non-null
 6   0-60 MPH Time (seconds) 1006 non-null
 7   Price (in USD)  1006 non-null
dtypes: float64(2), int64(3), object(3)
memory usage: 70.7+ KB
```

+ <> + ⌂



RAM



Disk



✓  
0s



df.info()

```
→ <class 'pandas.core.frame.DataFrame'
Index: 1006 entries, 0 to 1006
Data columns (total 8 columns):
 #   Column           Non-Null Count
 ---  -- 
 0   Car Make        1006 non-null
 1   Car Model       1006 non-null
 2   Year            1006 non-null
 3   Engine Size (L) 1006 non-null
 4   Horsepower      1006 non-null
 5   Torque (lb-ft)  1006 non-null
 6   0-60 MPH Time (seconds) 1006 non-null
 7   Price (in USD)  1006 non-null
dtypes: float64(2), int64(3), object(3)
memory usage: 70.7+ KB
```