

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

```
#loading the dataset using read_csv function in pandas library
data=pd.read_csv('/content/drive/MyDrive/autos.csv',encoding="ISO-8859-1",low_memory=False)
```

```
#Reading the first five rows in csv file using head function in pandas library
data.head()
```

	dateCrawled	name	seller
offerType \			
0	2016-03-24 11:52:17	Golf_3_1.6	privat
Angebot			
1	2016-03-24 10:58:45	A5_Sportback_2.7_Tdi	privat
Angebot			
2	2016-03-14 12:52:21	Jeep_Grand_Cherokee_"Overland"	privat
Angebot			
3	2016-03-17 16:54:04	GOLF_4_1_4__3TÜRER	privat
Angebot			
4	2016-03-31 17:25:20	Skoda_Fabia_1.4_TDI_PD_Classic	privat
Angebot			

	price	abtest	vehicleType	yearOfRegistration	gearbox	powerPS
model \						
0	480	test	NaN	1993	manuell	0
golf						
1	18300	test	coupe	2011	manuell	190
NaN						
2	9800	test	suv	2004	automatik	163
grand						
3	1500	test	kleinwagen	2001	manuell	75
golf						
4	3600	test	kleinwagen	2008	manuell	69
fabia						

	kilometer	monthOfRegistration	fuelType	brand
notRepairedDamage \				
0	150000	0	benzin	volkswagen
NaN				
1	125000	5	diesel	audi
ja				
2	125000	8	diesel	jeep
NaN				
3	150000	6	benzin	volkswagen
nein				
4	90000	7	diesel	skoda

nein

	dateCreated	nrOfPictures	postalCode	lastSeen
0	2016-03-24 00:00:00	0	70435	2016-04-07 03:16:57
1	2016-03-24 00:00:00	0	66954	2016-04-07 01:46:50
2	2016-03-14 00:00:00	0	90480	2016-04-05 12:47:46
3	2016-03-17 00:00:00	0	91074	2016-03-17 17:40:17
4	2016-03-31 00:00:00	0	60437	2016-04-06 10:17:21

*#Reading the last five rows in csv file using tail function*  
data.tail()

	dateCrawled
name \	
371523	2016-03-14 17:48:27
Suche_t4_vito_ab_6_sitze	
371524	2016-03-05 19:56:21
Smart_smart_leistungssteigerung_100ps	
371525	2016-03-19 18:57:12
Volkswagen_Multivan_T4_TDI_7DC_UY2	
371526	2016-03-20 19:41:08
VW_Golf_Kombi_1_9l_TDI	
371527	2016-03-07 19:39:19
BMW_M135i_vollausgestattet_NP_52.720____Euro	

	seller	offerType	price	abtest	vehicleType
yearOfRegistration \					
371523	privat	Angebot	2200	test	NaN
2005					
371524	privat	Angebot	1199	test	cabrio
2000					
371525	privat	Angebot	9200	test	bus
1996					
371526	privat	Angebot	3400	test	kombi
2002					
371527	privat	Angebot	28990	control	limousine
2013					

	gearbox	powerPS	model	kilometer
monthOfRegistration \				
371523	NaN	0	NaN	20000
1				
371524	automatik	101	fortwo	125000
3				
371525	manuell	102	transporter	150000

```

3
371526    manuell      100      golf    150000
6
371527    manuell      320    m_reihe    50000
8

```

	fuelType	brand	notRepairedDamage	dateCreated
\				
371523	NaN	sonstige_autos	NaN	2016-03-14 00:00:00
371524	benzin	smart	nein	2016-03-05 00:00:00
371525	diesel	volkswagen	nein	2016-03-19 00:00:00
371526	diesel	volkswagen	NaN	2016-03-20 00:00:00
371527	benzin	bmw	nein	2016-03-07 00:00:00

	nrOfPictures	postalCode	lastSeen
371523	0	39576	2016-04-06 00:46:52
371524	0	26135	2016-03-11 18:17:12
371525	0	87439	2016-04-07 07:15:26
371526	0	40764	2016-03-24 12:45:21
371527	0	73326	2016-03-22 03:17:10

*#checking the number of rows and columns in a dataset to understand the size that we are working with*  
data.shape

```
(371528, 20)
```

*#checking the size of the dataset*  
data.size

```
7430560
```

*#list of names of each column in a dataset*  
data.columns

```

Index(['dateCrawled', 'name', 'seller', 'offerType', 'price',
      'abtest',
      'vehicleType', 'yearOfRegistration', 'gearbox', 'powerPS',
      'model',
      'kilometer', 'monthOfRegistration', 'fuelType', 'brand',
      'notRepairedDamage', 'dateCreated', 'nrOfPictures',
      'postalCode',
      'lastSeen'],
      dtype='object')

```

*#checking the every datatype of column in the dataset*  
data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 371528 entries, 0 to 371527
Data columns (total 20 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   dateCrawled                          371528 non-null object
 1   name                                371528 non-null object
 2   seller                              371528 non-null object
 3   offerType                           371528 non-null object
 4   price                               371528 non-null int64
 5   abtest                              371528 non-null object
 6   vehicleType                         333659 non-null object
 7   yearOfRegistration                  371528 non-null int64
 8   gearbox                             351319 non-null object
 9   powerPS                             371528 non-null int64
10   model                               351044 non-null object
11   kilometer                           371528 non-null int64
12   monthOfRegistration                 371528 non-null int64
13   fuelType                           338142 non-null object
14   brand                               371528 non-null object
15   notRepairedDamage                  299468 non-null object
16   dateCreated                         371528 non-null object
17   nrOfPictures                       371528 non-null int64
18   postalCode                         371528 non-null int64
19   lastSeen                           371528 non-null object
dtypes: int64(7), object(13)
memory usage: 56.7+ MB
```

*#knowing the total number of unique values in each column using  
nunique function*  
data.nunique()

dateCrawled	280500
name	233531
seller	2
offerType	2
price	5597
abtest	2
vehicleType	8
yearOfRegistration	155
gearbox	2
powerPS	794
model	251
kilometer	13
monthOfRegistration	13
fuelType	7
brand	40
notRepairedDamage	2

dateCreated	114
nrOfPictures	1
postalCode	8150
lastSeen	182806
dtype: int64	