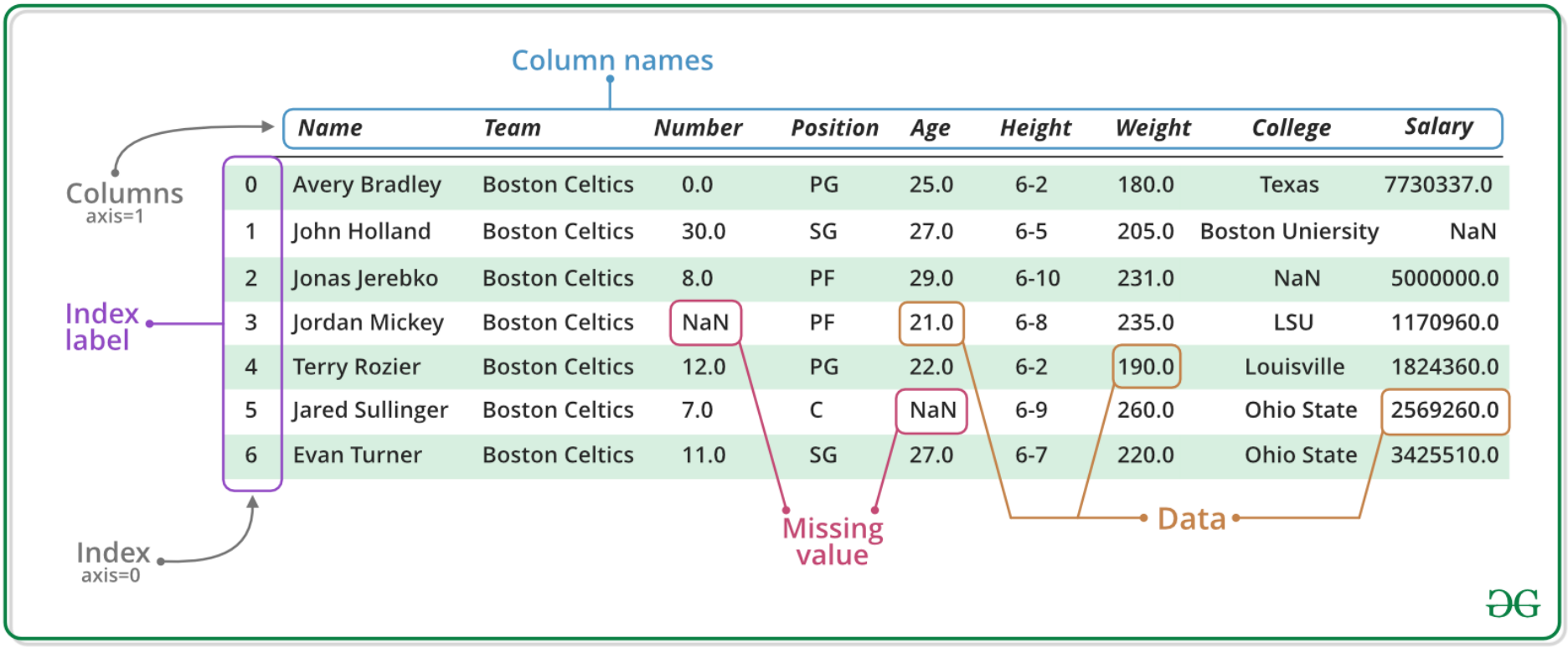
DATA



Manually Create a DataFrame

```
import pandas as pd
weather_data = {
    'day': ['1/1/2017','1/2/2017','1/3/2017','1/4/2017','1/5/2017','1/6/2017'],
    'temperature': [32,35,28,24,32,31],
    'windspeed': [6,7,2,7,4,2],
```

```
    'event': ['Rain', 'Sunny', 'Snow','Snow','Rain', 'Sunny']
}
df = pd.DataFrame(weather_data)

df
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

CSV File from Goole Drive

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

```
Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4n4g3pfee6491h

Enter your authorization code:
··········
Mounted at /content/drive
```

Reading CSV File

```
df = pd.read_csv("/content/drive/My Drive/pandas/weather_data.csv")

df
#'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/c

# df = pd.read_csv('url of your data')

# pd.read_csv('pandas_tutorial_read.csv', delimiter=';')

# pd.read_csv('pandas_tutorial_read.csv', delimiter=';', names = ['my_datetime', 'event
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

Display First Few rows

```
df.head() # df.head(10)
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow

Display last few rows

```
df.tail(3) # df.tail(3)
```

	day	temperature	windspeed	event
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

Random Selection of rows

```
df.sample(3)
```

	day	temperature	windspeed	event
3	1/4/2017	24	7	Snow
2	1/3/2017	28	2	Snow
4	1/5/2017	32	4	Rain

Count number of rows and columns/dataset size

```
#df.shape
rows, columns = df.shape
rows
```

```
6
```

Display the column titles

```
df.columns
```

```
Index(['day', 'temperature', 'windspeed', 'event'], dtype='object')
```

The data type of all coumns

```
df.dtypes
```

```
day            object
temperature     int64
windspeed       int64
event          object
dtype: object
```

Change the default index of a dataframe

```
df.set_index('day')
```

day	temperature	windspeed	event
1/1/2017	32	6	Rain
1/2/2017	35	7	Sunny
1/3/2017	28	2	Snow
1/4/2017	24	7	Snow

Observe the original DataFrame!!!! NOT CHANGED

```
df # NOT CHANGED
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

inplace parameter

```
df.set_index('day', inplace=True)
```

```
df
```

day	temperature	windspeed	event
1/1/2017	32	6	Rain
1/2/2017	35	7	Sunny
1/3/2017	28	2	Snow
1/4/2017	24	7	Snow
1/5/2017	32	4	Rain
1/6/2017	31	2	Sunny

reset the index back

```
df.reset_index(inplace=True)
```

```
df
```

	day	temperature	windspeed	event
0	1/1/2017	32	6	Rain
1	1/2/2017	35	7	Sunny
2	1/3/2017	28	2	Snow
3	1/4/2017	24	7	Snow
4	1/5/2017	32	4	Rain
5	1/6/2017	31	2	Sunny

Select a specific column

```
df['temperature']
```

```
0    32
1    35
2    28
3    24
4    32
5    31
Name: temperature, dtype: int64
```

```
df['temperature']=df['temperature']+1
```

```
df['temperature']
```

```
0    33
1    36
2    29
3    25
4    33
5    32
Name: temperature, dtype: int64
```

Select Multiple columns

```
df[['temperature','event']]
```

day	temperature	event
1/1/2017	32	Rain
1/2/2017	35	Sunny

```
df.event  #df['event']
```

```
0     Rain
1    Sunny
2     Snow
3     Snow
4     Rain
5    Sunny
Name: event, dtype: object
```

Total info about DataFrame

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6 entries, 1/1/2017 to 1/6/2017
Data columns (total 3 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   temperature  6 non-null      int64
 1   windspeed    6 non-null      int64
 2   event        6 non-null      object
dtypes: int64(2), object(1)
memory usage: 352.0+ bytes
```

Change Column name

```
df.columns
```

```
df.columns
```

```
Index(['day', 'temperature', 'windspeed', 'event'], dtype='object')
```

```
df.rename(columns={
        'event': 'Event',
        'temperature': 'TEMPERATURE'
    }, inplace=True)
```

```
df.columns
```

```
Index(['day', 'TEMPERATURE', 'windspeed', 'Event'], dtype='object')
```

```
df.head()
```

	day	TEMPERATURE	windspeed	Event
0	1/1/2017	33	6	Rain
1	1/2/2017	36	7	Sunny
2	1/3/2017	29	2	Snow
3	1/4/2017	25	7	Snow
4	1/5/2017	33	4	Rain

Statistics of DataFrame

```
df.describe()
```

	TEMPERATURE	windspeed
count	6.000000	6.000000
mean	31.333333	4.666667
std	3.829708	2.338090
min	25.000000	2.000000
25%	29.750000	2.500000
50%	32.500000	5.000000
75%	33.000000	6.750000

```
df.describe(include='all')
```

	day	TEMPERATURE	windspeed	Event
count	6	6.000000	6.000000	6
unique	6	NaN	NaN	3
top	1/4/2017	NaN	NaN	Snow
freq	1	NaN	NaN	2
mean	NaN	31.333333	4.666667	NaN
std	NaN	3.829708	2.338090	NaN
min	NaN	25.000000	2.000000	NaN
25%	NaN	29.750000	2.500000	NaN
50%	NaN	32.500000	5.000000	NaN
75%	NaN	33.000000	6.750000	NaN
max	NaN	36.000000	7.000000	NaN

Count of categories

```
df['Event'].value_counts()
```

```
Snow     2
Rain     2
Sunny    2
Name: Event, dtype: int64
```

```
df['windspeed'].value_counts()
```

```
7    2
2    2
6    1
4    1
Name: windspeed, dtype: int64
```

Delete a specific column

```
df.drop(['Event'], axis = 1)
```

	day	TEMPERATURE	windspeed
0	1/1/2017	33	6
1	1/2/2017	36	7
2	1/3/2017	29	2
3	1/4/2017	25	7
4	1/5/2017	33	4
5	1/6/2017	32	2

```
df
```

	day	TEMPERATURE	windspeed	Event
0	1/1/2017	33	6	Rain
1	1/2/2017	36	7	Sunny
2	1/3/2017	29	2	Snow
3	1/4/2017	25	7	Snow
4	1/5/2017	33	4	Rain
5	1/6/2017	32	2	Sunny

```
df.drop(['Event'], axis = 1,inplace=True)
```

```
df
```

	day	TEMPERATURE	windspeed
0	1/1/2017	33	6
1	1/2/2017	36	7
2	1/3/2017	29	2
3	1/4/2017	25	7
4	1/5/2017	33	4
5	1/6/2017	32	2

```
# df.drop(['C', 'D'], axis = 1)
```

```
# df.drop(columns =['C', 'D'])
```

```
# df.drop(df.columns[[0, 4, 2]], axis = 1, inplace = True)
```

TRY these commands for the following dataset https://raw.githubusercontent.com/pplonski/datasets-for-start/master/credit/data.csv

QUIZ