

```
In [1]: #Name:- Kuldeep Ghorpade
        #Div:- B
        #Roll No. :- 09
        #Experiment Name & No :- Introduction to DataFrames (06)
```

```
In [2]: using Pkg
        Pkg.add("CSV")
```

```

    Updating registry at `C:\Users\kulde\.julia\registries\General.toml`
  Resolving package versions...
  Installed InlineStrings ─────────── v1.3.2
  Installed DataValueInterfaces ───── v1.0.0
  Installed TableTraits ─────────── v1.0.1
  Installed SentinelArrays ───────── v1.3.16
  Installed WeakRefStrings ───────── v1.4.2
  Installed PooledArrays ───────── v1.4.2
  Installed Tables ─────────── v1.10.0
  Installed IteratorInterfaceExtensions ─ v1.0.0
  Installed FilePathsBase ───────── v0.9.20
  Installed CSV ─────────── v0.10.7
    Updating `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
 [336ed68f] + CSV v0.10.7
    Updating `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`
 [336ed68f] + CSV v0.10.7
 [e2d170a0] + DataValueInterfaces v1.0.0
 [48062228] + FilePathsBase v0.9.20
 [842dd82b] + InlineStrings v1.3.2
 [82899510] + IteratorInterfaceExtensions v1.0.0
 [2dfb63ee] + PooledArrays v1.4.2
 [91c51154] + SentinelArrays v1.3.16
 [3783bdb8] + TableTraits v1.0.1
 [bd369af6] + Tables v1.10.0
 [ea10d353] + WeakRefStrings v1.4.2
  Precompiling project...
    ✓ IteratorInterfaceExtensions
    ✓ DataValueInterfaces
    ✓ PooledArrays
    ✓ TableTraits
    ✓ SentinelArrays
    ✓ FilePathsBase
    ✓ Tables
    ✓ InlineStrings
    ✓ WeakRefStrings
    ✓ CSV
  10 dependencies successfully precompiled in 7 seconds. 165 already precompiled.
```

```
In [3]: Pkg.add("DataFrames")
```

```

Resolving package versions...
Installed StringManipulation - v0.3.0
Installed InvertedIndices — v1.1.0
Installed Crayons — v4.1.1
Installed DataFrames — v1.4.3
Installed PrettyTables — v2.2.1
Updating `C:\Users\kulde\.julia\environments\v1.8\Project.toml`
[a93c6f00] + DataFrames v1.4.3
Updating `C:\Users\kulde\.julia\environments\v1.8\Manifest.toml`
[a8cc5b0e] + Crayons v4.1.1
[a93c6f00] + DataFrames v1.4.3
[41ab1584] + InvertedIndices v1.1.0
[08abe8d2] + PrettyTables v2.2.1
[892a3eda] + StringManipulation v0.3.0
Precompiling project...
✓ InvertedIndices
✓ Crayons
✓ StringManipulation
✓ PrettyTables
✓ DataFrames
5 dependencies successfully precompiled in 28 seconds. 175 already precompiled.

```

```

In [4]: using DataFrames
df = DataFrame(A = 1:5, B = ["A", "E", "I", "O", "U"],
C = ["A", "B", "C", "D", "E"])

```

Out[4]: 5×3 DataFrame

Row	A	B	C
	Int64	String	String
1	1	A	A
2	2	E	B
3	3	I	C
4	4	O	D
5	5	U	E

```

In [6]: df2 = DataFrame()
df2.C = 1:5
df2.D = ["A", "E", "I", "O", "U"]
df2

```

Out[6]: 5×2 DataFrame

Row	C	D
	Int64	String
1	1	A
2	2	E
3	3	I
4	4	O
5	5	U

```
In [7]: df3 = DataFrame(E = Int[], F = String[])
push!(df3, (1, "A"))
push!(df3, (2, "E"))
push!(df3, (3, "I"))
push!(df3, (4, "O"))
push!(df3, (5, "U"))
df3
```

Out[7]: 5×2 DataFrame

Row	E	F
	Int64	String
1	1	A
2	2	E
3	3	I
4	4	O
5	5	U

```
In [8]: first(select(df, :B ), 3)
```

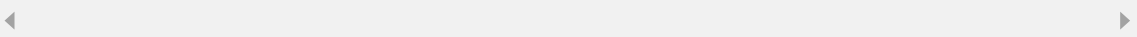
Out[8]: 3×1 DataFrame

Row	B
	String
1	A
2	E
3	I

```
In [9]: first(select(df, Not(:B)), 4)
```

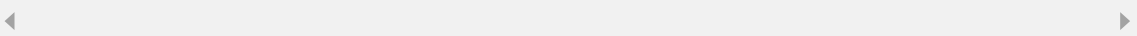
Out[9]: 4×2 DataFrame

Row	A	C
	Int64	String
1	1	A
2	2	B
3	3	C
4	4	D

In [10]: `df[2:4, [:B, :C]]`

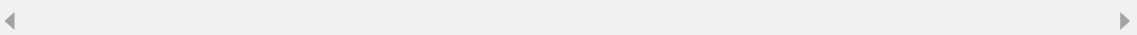
Out[10]: 3×2 DataFrame

Row	B	C
	String	String
1	E	B
2	I	C
3	O	D

In [11]: `replace!(df.A, 4 => 7)`  
`df`

Out[11]: 5×3 DataFrame

Row	A	B	C
	Int64	String	String
1	1	A	A
2	2	E	B
3	3	I	C
4	7	O	D
5	5	U	E

In [12]: `df[:, [:B, :C]] .= ifelse.(df[:, [:B, :C]] .== "E", "None", df[:, [:B, :C]])`  
`df`

Out[12]: 5×3 DataFrame

Row	A	B	C
	Int64	String	String
1	1	A	A
2	2	None	B
3	3	I	C
4	7	O	D
5	5	U	None

In [13]: `df. = ifelse.(df. == "A", "E", df)`

Out[13]: 5×3 DataFrame

Row	A	B	C
	Int64	String	String
1	1	E	E
2	2	None	B
3	3	I	C
4	7	O	D
5	5	U	None

In [14]: `push!(df, [6 "None" "F"])`

Out[14]: 6×3 DataFrame

Row	A	B	C
	Int64	String	String
1	1	E	E
2	2	None	B
3	3	I	C
4	7	O	D
5	5	U	None
6	6	None	F

In [15]: `arr = [2, 3, 5, 7, 11, 13]  
df[!, "E"] = arr  
df`

Out[15]: 6×4 DataFrame

Row	A	B	C	E
	Int64	String	String	Int64
1	1	E	E	2
2	2	None	B	3
3	3	I	C	5
4	7	O	D	7
5	5	U	None	11
6	6	None	F	13

In [16]: `delete!(df, 4)`

Out[16]: 5×4 DataFrame

Row	A	B	C	E
	Int64	String	String	Int64
1	1	E	E	2
2	2	None	B	3
3	3	I	C	5
4	5	U	None	11
5	6	None	F	13

In [17]: `select!(df, Not(:C))`

Out[17]: 5×3 DataFrame

Row	A	B	E
	Int64	String	Int64
1	1	E	2
2	2	None	3
3	3	I	5
4	5	U	11
5	6	None	13

In [18]: `df2 = DataFrame(F = ["A", "E", "I", "O", "U"],  
G = ["G", "E", "E", "K", "S"])  
df4 = DataFrame(H = ["G", "R", "E", "A", "T"])`

Out[18]: 5×1 DataFrame

Row	H
	String
1	G
2	R
3	E
4	A
5	T

In [19]: `hcat(df, df2, df4)`

Out[19]: 5×6 DataFrame

Row	A	B	E	F	G	H
	Int64	String	Int64	String	String	String
1	1	E	2	A	G	G
2	2	None	3	E	E	R
3	3	I	5	I	E	E
4	5	U	11	O	K	A
5	6	None	13	U	S	T

In [20]: `df3 = DataFrame(A = 7, B = "O", C = "G", E = 17)`  
`df5 = DataFrame(A = 8, B = "None", C = "H", E = 19)`  
`vcats(df3, df5)`

Out[20]: 2×4 DataFrame

Row	A	B	C	E
	Int64	String	String	Int64
1	7	O	G	17
2	8	None	H	19

In [24]: `using CSV`  
`using DataFrames`

In [25]: `iris=CSV.read("iris.csv", DataFrame)`

Out[25]: 150×6 DataFrame

125 rows omitted

Row	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
	Int64	Float64	Float64	Float64	Float64	String15
1	1	5.1	3.5	1.4	0.2	Iris-setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa
3	3	4.7	3.2	1.3	0.2	Iris-setosa
4	4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	5.0	3.6	1.4	0.2	Iris-setosa
6	6	5.4	3.9	1.7	0.4	Iris-setosa
7	7	4.6	3.4	1.4	0.3	Iris-setosa
8	8	5.0	3.4	1.5	0.2	Iris-setosa
9	9	4.4	2.9	1.4	0.2	Iris-setosa
10	10	4.9	3.1	1.5	0.1	Iris-setosa
11	11	5.4	3.7	1.5	0.2	Iris-setosa
12	12	4.8	3.4	1.6	0.2	Iris-setosa
13	13	4.8	3.0	1.4	0.1	Iris-setosa
⋮	⋮	⋮	⋮	⋮	⋮	⋮
139	139	6.0	3.0	4.8	1.8	Iris-virginica
140	140	6.9	3.1	5.4	2.1	Iris-virginica
141	141	6.7	3.1	5.6	2.4	Iris-virginica
142	142	6.9	3.1	5.1	2.3	Iris-virginica
143	143	5.8	2.7	5.1	1.9	Iris-virginica
144	144	6.8	3.2	5.9	2.3	Iris-virginica
145	145	6.7	3.3	5.7	2.5	Iris-virginica
146	146	6.7	3.0	5.2	2.3	Iris-virginica
147	147	6.3	2.5	5.0	1.9	Iris-virginica
148	148	6.5	3.0	5.2	2.0	Iris-virginica
149	149	6.2	3.4	5.4	2.3	Iris-virginica
150	150	5.9	3.0	5.1	1.8	Iris-virginica

```
In [26]: iris=CSV.read("iris.csv",DataFrame,normalizenames=true)
```



Out[26]: 150×6 DataFrame

125 rows omitted

Row	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
	Int64	Float64	Float64	Float64	Float64	String15
1	1	5.1	3.5	1.4	0.2	Iris-setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa
3	3	4.7	3.2	1.3	0.2	Iris-setosa
4	4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	5.0	3.6	1.4	0.2	Iris-setosa
6	6	5.4	3.9	1.7	0.4	Iris-setosa
7	7	4.6	3.4	1.4	0.3	Iris-setosa
8	8	5.0	3.4	1.5	0.2	Iris-setosa
9	9	4.4	2.9	1.4	0.2	Iris-setosa
10	10	4.9	3.1	1.5	0.1	Iris-setosa
11	11	5.4	3.7	1.5	0.2	Iris-setosa
12	12	4.8	3.4	1.6	0.2	Iris-setosa
13	13	4.8	3.0	1.4	0.1	Iris-setosa
⋮	⋮	⋮	⋮	⋮	⋮	⋮
139	139	6.0	3.0	4.8	1.8	Iris-virginica
140	140	6.9	3.1	5.4	2.1	Iris-virginica
141	141	6.7	3.1	5.6	2.4	Iris-virginica
142	142	6.9	3.1	5.1	2.3	Iris-virginica
143	143	5.8	2.7	5.1	1.9	Iris-virginica
144	144	6.8	3.2	5.9	2.3	Iris-virginica
145	145	6.7	3.3	5.7	2.5	Iris-virginica
146	146	6.7	3.0	5.2	2.3	Iris-virginica
147	147	6.3	2.5	5.0	1.9	Iris-virginica
148	148	6.5	3.0	5.2	2.0	Iris-virginica
149	149	6.2	3.4	5.4	2.3	Iris-virginica
150	150	5.9	3.0	5.1	1.8	Iris-virginica

In [27]:

typeof(iris)

Out[27]: DataFrame

In [28]:

names(iris)

```
Out[28]: 6-element Vector{String}:
  "Id"
  "SepalLengthCm"
  "SepalWidthCm"
  "PetalLengthCm"
  "PetalWidthCm"
  "Species"
```

```
In [29]: size(iris)
```

Out[29]: (150, 6)

```
In [30]: first(iris,5)
```

Out[30]: 5×6 DataFrame

Row	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
	Int64	Float64	Float64	Float64	Float64	String15
1	1	5.1	3.5	1.4	0.2	Iris-setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa
3	3	4.7	3.2	1.3	0.2	Iris-setosa
4	4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	5.0	3.6	1.4	0.2	Iris-setosa

```
In [31]: last(iris,5)
```

Out[31]: 5×6 DataFrame

Row	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
	Int64	Float64	Float64	Float64	Float64	String15
1	146	6.7	3.0	5.2	2.3	Iris-virginica
2	147	6.3	2.5	5.0	1.9	Iris-virginica
3	148	6.5	3.0	5.2	2.0	Iris-virginica
4	149	6.2	3.4	5.4	2.3	Iris-virginica
5	150	5.9	3.0	5.1	1.8	Iris-virginica

```
In [32]: describe(iris)
```

Out[32]: 6×7 DataFrame

Row	variable	mean	min	median	max	nmissing	eltype
	Symbol	Union...	Any	Union...	Any	Int64	DataType
1	Id	75.5	1	75.5	150	0	Int64
2	SepalLengthCm	5.84333	4.3	5.8	7.9	0	Float64
3	SepalWidthCm	3.054	2.0	3.0	4.4	0	Float64
4	PetalLengthCm	3.75867	1.0	4.35	6.9	0	Float64
5	PetalWidthCm	1.19867	0.1	1.3	2.5	0	Float64
6	Species	Iris-setosa		Iris-virginica		0	String15

In [33]: iris[:,3]

Out[33]: 150-element Vector{Float64}:  
3.5  
3.0  
3.2  
3.1  
3.6  
3.9  
3.4  
3.4  
2.9  
3.1  
3.7  
3.4  
3.0  
:  
3.0  
3.1  
3.1  
3.1  
2.7  
3.2  
3.3  
3.0  
2.5  
3.0  
3.4  
3.0

In [34]: iris[:,[1,2,3]]

Out[34]: 150×3 DataFrame

125 rows omitted

Row	Id	SepalLengthCm	SepalWidthCm
	Int64	Float64	Float64
1	1	5.1	3.5
2	2	4.9	3.0
3	3	4.7	3.2
4	4	4.6	3.1
5	5	5.0	3.6
6	6	5.4	3.9
7	7	4.6	3.4
8	8	5.0	3.4
9	9	4.4	2.9
10	10	4.9	3.1
11	11	5.4	3.7
12	12	4.8	3.4
13	13	4.8	3.0
⋮	⋮	⋮	⋮
139	139	6.0	3.0
140	140	6.9	3.1
141	141	6.7	3.1
142	142	6.9	3.1
143	143	5.8	2.7
144	144	6.8	3.2
145	145	6.7	3.3
146	146	6.7	3.0
147	147	6.3	2.5
148	148	6.5	3.0
149	149	6.2	3.4
150	150	5.9	3.0

```
In [35]: iris[1:5,:]
```

Out[35]: 5×6 DataFrame

Row	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
	Int64	Float64	Float64	Float64	Float64	String15
1	1	5.1	3.5	1.4	0.2	Iris-setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa
3	3	4.7	3.2	1.3	0.2	Iris-setosa
4	4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	5.0	3.6	1.4	0.2	Iris-setosa

In [36]: iris[1:5,1:2]

Out[36]: 5×2 DataFrame

Row	Id	SepalLengthCm
	Int64	Float64
1	1	5.1
2	2	4.9
3	3	4.7
4	4	4.6
5	5	5.0

In [ ]: