Making a dataframe

a	b	c
1	5	9
2	6	10
3	7	11
4	8	12

```
import pandas as pd
```

- Use 'Dictionary' syntax to create a dataframe

```
df = pd.DataFrame({'a': [1, 2, 3, 4], 'b' : [5, 6, 7, 8], 'c' : [9, 10, 11, 12]})
```

```
df
```

- Use 'List' to create a dataframe

```
a = [[1, 4, 7], [2, 5, 8], [3, 6, 9]]
```

```
df2 = pd.DataFrame(a)
```

```
df2
```

```
df2.columns = ['a', 'b', 'c']
```


- change field names

```
df.columns = ['d', 'e', 'f']
```

```
df
```

- copy dataframes

```
import copy
```

```
df3 = copy.deepcopy(df)
```

```
df3
```

- extracting columns

Extracting columns

- create a series in pandas

```
a = pd.Series([1, 2, 3, 1, 2, 3])
```

```
a
```

- change index

```
a = pd.Series([1, 2, 3, 1, 2, 3], index = ['a', 'b', 'c', 'd', 'e', 'f'])
```

```
a
```

extracting data with conditions

```
df = pd.DataFrame({'a' : [i for i in range(1, 6)], 'b' : [i for i in range(6, 11)], 'c' : [i for i in range(11, 16)]})
```

```
df
```

```
df[['a', 'c']]
```

```
df[df['a'] >= 3]
```

```
df[df['a'] >= 3][['a', 'b']]
```

```
df[(df['a'] >= 3) & (df['b'] < 16)]
```

```
TF = (df['a'] >= 3) & (df['b'] < 16)
```

```
TF
```

```
df[TF]
```

Saved successfully!

```
df.drop(df.index[:4])
```

Working with a big dataset

```
import pandas as pd
titanic_df=pd.read_csv('train.csv')
```

- link googlddrive to colab

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

```
titanic_df= pd.read_csv('/content/drive/MyDrive/Intro ML 2022 Summer/dataset/train.csv')
```

```
type(titanic_df)
```

```
titanic_df
```

```
live = (titanic_df['Survived'] == 1)&(titanic_df['Sex'] == 'female')&(titanic_df['Age']<=10.0)
```

```
live
```

```
titanic_df[live]
```

```
len(titanic_df[live])
```

0s completed at 11:55 AM

Saved successfully!