**PANDAS**

df**.**shape

df.columns

pd**.**set\_option('display.max\_columns', 85)

pd**.**set\_option('display.max\_rows', 85)

**# Selecting a column**

df['Hobbyist']

**# Selecting multiple columns**

df[['last', 'email']]

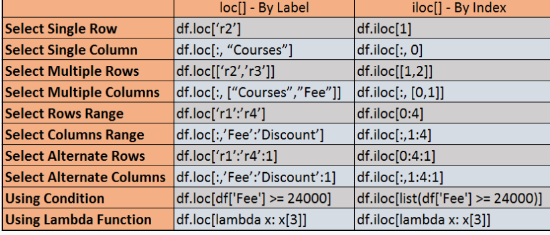
**# Selecting row and column using loc and iloc**

# The main difference between pandas loc[] vs iloc[] is loc gets DataFrame rows & columns by labels/names and iloc[] gets by integer Index/position. For loc[], if the label is not present it gives a key error. For iloc[], if the position is not present it gives an index error.

# Selecting columns from DataFrame results in a new DataFrame containing only specified selected columns from the original DataFrame.

df**.**loc[0:2, 'Hobbyist':'Employment']

df**.**iloc[[0, 1], 2]



# Select Multiple Rows by Label

print(df.loc[['r2','r3']])

# Select Multiple Rows by Index

print(df.iloc[[1,2]])

**# Converting a dictionary to a dataframe**

import pandas as pd

technologies = {

'Courses':["Spark","PySpark","Hadoop","Python","pandas"],

'Fee' :[20000,25000,26000,22000,24000],

'Duration':['30day','40days','35days','40days','60days'],

'Discount':[1000,2300,1200,2500,2000]

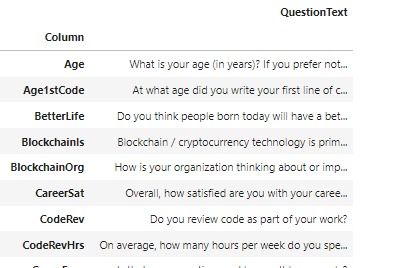
}

index\_labels=['r1','r2','r3','r4','r5']

df = pd.DataFrame(technologies,index=index\_labels)

**# Sorting data frame columns alphabetically and saving it**

schema\_df**.**sort\_index(inplace**=True**)



**# Setting email as measure to choose which row**

df**.**set\_index('email', inplace**=True)**

**# Getting values of the column which will be used as a measure to choose which row**

df**.**index

**# Removing the index we choose earlier as a measure of selecting row**

df**.**reset\_index(inplace**=True**)