mlpc

Sec 3:

#3.0 Introduction

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

dataframe.head(5)

#3.1 creating a Data Frame

import pandas as pd

dataframe = pd.DataFrame()

#add columns

dataframe[“Name“]=[“Jacky Jackson“, “Steven Stevenson“]

dataframe [“Age“]=[38,35]

datframe[“Driver“]= [True,False]

#show dataframe

dataframe

#create row

New\_person= pd.Series([“Molly Mooney“, 40, True], index= [“Name“,“Age“,“Driver“])

#append row

dataframe.append(new\_person, ignore\_index=True)

#3.2 Describing the Data

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

#load data

dataframe=pd.read\_ csv(url)

#show two rows

Dataframe.head(2)

#show dimensions

Dataframe.shape

#show statistics

dataframe.describe()

#3.3 Navigating DataFrames

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

datafame=pd.read\_csv(url)

#select first row

dataframe.iloc[0]

#selec three rows

dataframe.iloc[1:4]

#3.4 Selecting Rows Based on conditionals

Import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#show top two rows where column “Sex“ is “female“

dataframe[dataframe[“Sex“]==“female“].head(2)

#filter rows

dataframe[(dataframe[“Sex“]==“female“)&(dataframe[“Age“] >=65)]

#3.5 Replacing Values

Import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#replace values, show two rows

Dataframe[“Sex“].replace(“female“,“Woman“).head(2)

#Replace “female“ and “male“ with “Woman“ and “Man“

dataframe[“Sex“].replace([“female“, “male“], [“Woman“,“Man“]).head(5)

#Replace values, show two rows

dataframe.replace(1, “One“).head(2)

#replace values, show two rows

dataframe.replace(r “ist“, “First“,regex=True).head(2)

#3.6 Renaming columns

Import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#rename column, show two rows

dataframe.rename(columns = {“PClass“: “Passenger Class “}).head(2)

#rename columns, show two rows

Dataframe.rename(columns={“PClass“: “Passenger Class“, “Sex“: “Gender“}).head(2)

#load library

import collections

#create dictionary

column\_names=collections.defaultdict(str)

#create keys

for name in dataframe. Columns:

column\_names[name]

#show dictionary

column\_names

defautdict(str,

{“Age“: ““),

“Name“: ““),

“ PClass“: ““),

“Sex“: ““),

“SexCode“: ““),

“Survived“: ““}),

#3.7 finding the minmum, maximum, sum, average and count

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#calculate statistics

print(“Maximum: “,dataframe[“Age“].max())

print(“Minimum:“,dataframe[“Age“].min())

print(“Mean:“, dataframe[“Age“].mean())

print(“Sum:“, dataframe[“Age“].sum())

print(“Count:“, dataframe[“Age“]. count())

#3.8 finding unique Values

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#select unique values

dataframe[“Sex“].unique()

array([“female“, “male“], dtype=object)

#show counts

dataframe[“Sex“].value\_counts()

#show counts

dataframe[“PClass“].value\_counts()

#show number of unique values

dataframe[“PClass“].unique()

#3.9 Handling Missing Values

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#select missing values, show two rows

dataframe[dataframe[“Age“].isnull().head(2)]

#attempt to replace vales with NaN

dataframe[“Sex“]= dataframe[“Sex“].replace(“male“,NaN)

import numpy as np

#replace values with NaN

dataframe[“Sex“]=dataframe[“Sex“].replace(“male“, np.man)

#load data, set missing values

dataframe = pd.read\_csv(url, na\_values= [np.man, “NONE“, -999])

#3.10 Deleting a column

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#delete column

dataframe.drop(“Age“, axis=1).head(2)

#drop columns

Dataframe.drop([“Age“, “Sex“], axis=1.head(2))

#3.11 Deleting a Row

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#delete rows, show first two rows of output

Dataframe[dataframe[“Sex“]!= “male“].head(2)

#delete row, show first two rows of output

dataframe[datframe[“Name“]!= “Allison, Miss Helen Loraine“]

#delete row, show first two rows of output

datafame[dataframe.index !=0].head(2)

#3.12 dropping duplicate rows

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#drop duplicates, show first two rows of output

Dataframe.drop\_duplicates().head(2)

#show number of rows

print(“Number Of Rows In The Original DataFrame:“, len(dataframe))

print(“Number Of Rows After Deduping:“, len(dataframe.drop\_duplicates()))

#drop duplicates

Dataframe.drop\_duplicates(subset=[“Sex“],keep=“last“)

#3.13 Grouping Rows by Values

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#group rows by the values oft he column “Sex“, calculate mean of each group

dataframe.groupby(“Sex“).mean()

#Group rows

dataframe..groupby(“Sex“)

<pandas. core.groupby.DataFrameGroupBy object at 0x10efacf28>

#group rows, count rows

dataframe.grouby(“Survived“)[“Name“]. count()

#3.14 Grouping rows by time

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#create data range

time\_index=pd.date\_range(“06/06/2017“, periods=100000, freq=“305“)

dataframe = pd.DataFrame(index=time\_index)

#create column of random values

Dataframe[“Sale\_Amount“] = np.random.randint(1,.10,100000)

#3.15 Looping over a column

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#print first two names uppercased

for name in dataframe[“Name“][0:2]:

print(name.upper())

#3.16

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#create function

def uppercase (x):

return x.upper()

#apply function, show two rows

dataframe[“Name“].apply(uppercase) [0:2]

#3.17 Applying a Function to Groups

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

#group rows, apply function to groups

dataframe.groupby(“Sex“).apply(lambda x: x. count())

#3.18 concatenating DataFrames

import pandas as pd

url= 'https://tinyurl.com/titanic-csv'

dataframe = pd.read\_csv(url)

data\_a= {“id“:[“1“, “2“, “3“],

“first“: [“Alex“, “Amy“, “Allen“]

“last“: [“Anderson“, “Ackerman“, “Ali“]}

Dataframe\_a=pd.DataFrame(data\_a, columns= [“id“, “first“, “last“])

#create Dataframe

Data\_b={ “id“: [“4“, “5“, “6“],

“first“: [“Billy“,“Brian“, “Bran“],

“last“: [“Bonder“, “Black“, “Balwner“]}

dataframe\_b=pd.DataFrame(data\_b, columns = [“id“, “first“, “last“])

#concatenate DataFrames by rows

pd.concat ([dataframe\_a, dataframe\_b], axis=0)

pd.concat([dataframe\_a, dataframe\_b], axis=1)

#3.19 Merging DataFrames

import pandas as pd

employee\_data={ “employee\_id“: [ “1“, “2“, “3“, “4“]

“name“: [“Amy Jones“, “Allen Keys“, “Alice Bees, “Tim Horton“]}

Dataframe\_employees = pd.DataFrame(employee\_data, columns = [“employee\_id“, total\_sales])

#merge dataframes

pd.merge(dataframe\_employees, dataframe\_sales, = on “employee\_id“)

#merge dataframes

Pd.merge(dataframe\_employees, dataframe\_sales, on = “employee\_id“,how=“left“)

#merge dataframes

pd.merge(dataframe\_employees,

dataframe\_sales,

left\_on=“employee\_id“

right\_on=“employee\_id“)