DATA SCi

[PANDAS CHEATSHEET]

import/export

IMPORT DATA
from google.colab import files
uploaded = files.upload()

LOAD DATA INTO A DF
df = pd.read_csv ("path/file-name-as-it-was-uploaded.csv")

CREATE DATA
data = {'Name':['areej_abuali', 'rvtheverett', 'lilyray.nyc', 'AlexisKSanders'],
'Power':['TrailBlazer', 'PythonWiz', 'DJFire', 'Pioneer'],
'Qualifications':['GeNius', 'GoOgle', 'SpIN', 'GoOgle'] }

NAME DATA INTO DF
df = pd.DataFrame(data)

EXPORT DF
df.to_csv('export-name.csv')

modify

RENAME COLUMN HEADERS
df = df.rename(columns={'Current':'New', 'Current1':'New1'})

DROP COLUMNS/ROWS
df.drop('column', axis=1, inplace=True)

DROP COLUMNS/ROWS IF THEY = 'X'
df.drop(df.loc[df['column']>20].index, inplace=True)

CREATE DF W/ONLY CERTAIN COLUMNS
df_new = df[['column1', 'column2', 'column3']].copy()

CREATE DF W/ALL INSTANCES OF 'X'
df_new = df[(df['column'].str.contains("keyword|keyword2",
regex=True)==True)]

ADD NEW COLUMN CATEGORIZING 'X'
df["New Column"]=df.column.str.contains("X|brandX")

JOIN 2 DF WHERE VALUES MATCH
df_all = df.merge(df_gsc, how='inner', on=['Keywords'])

FILL NA VALUES
df=df.fillna({"column":"0"})

RESET INDEX
df.reset.index

CLEAR OUT ODD CHARACTERS
spec_chars = ["!","","#","%"]
for char in spec_chars:
df['column'] = df['column'].str.replace(char, ' ')

explore

LOOK AT THE HEAD
df.head()

LOOK AT THE TAIL
df.tail()

DESCRIBE THE DATA
df.describe(include="all")

LOAD FILTERABLE TABLE
%load_ext google.colab.data_table

DISABLE FILTERABLE TABLE
%unload_ext google.colab.data_table

SUM OF A COLUMN
df.column.sum()

MEAN OF A COLUMN
df['column'].mean()

STANDARD DEVIATION
df['column'].std()

PIVOT TABLE
df.pivot_table(df, index=['Name', 'Qualifications'],
columns=['Power'], aggfunc=len)

FIND ALL UNIQUE VALUES
df.column.unique()

visualize

HISTOGRAM
sns.distplot(df['Column'])

COMPARISON HISTOGRAM
df.hist(by='Category', column='Column')

CORRELATION MATRIX
corr = df.corr()
corr