!pip install pandasql

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import sklearn as sklearn

import scipy

from sklearn.datasets import load\_iris

from pandasql import sqldf

from pandasql import load\_meat, load\_births

import re

births = load\_births()

meat = load\_meat()

iris = load\_iris()

iris\_df = pd.DataFrame(iris.data, columns=iris.feature\_names)

iris\_df['species'] = pd.Categorical.from\_codes(iris.target, iris.target\_names)

iris\_df.columns = [re.sub("[() ]", "", col) for col in iris\_df.columns]

print(sqldf("SELECT \* FROM iris\_df LIMIT 10;", locals()))

print(sqldf("SELECT sepalwidthcm, species FROM iris\_df LIMIT 10;", locals()))

q = """

select

species

, avg(sepalwidthcm)

, min(sepalwidthcm)

, max(sepalwidthcm)

from

iris\_df

group by

species;

"""

print("\*" \* 80)

print("aggregation")

print("-" \* 80)

print(q)

print(sqldf(q, locals()))

def pysqldf(q):

"add this to your script if you get tired of calling locals()"

return sqldf(q, globals())

print("\*" \* 80)

print("calling from a helper function")

print('''def pysqldf(q):)

"add this to your script if you get tired of calling locals()"

return sqldf(q, globals())''')

print("-" \* 80)

print(q)

print(pysqldf(q))

q = """

select

a.\*

from

iris\_df a

inner join

iris\_df b

on a.species = b.species

limit 10;

"""

print("\*" \* 80)

print("joins")

print("-" \* 80)

print(q)

print(pysqldf(q))

q = """

select

\*

from

iris\_df

where

species = 'virginica'

and sepallengthcm > 7.7;

"""

print("\*" \* 80)

print("where clause")

print("-" \* 80)

print(q)

print(pysqldf(q))

iris\_df['id'] = range(len(iris\_df))

q = """

select

\*

from

iris\_df

where

id in (select id from iris\_df where sepalwidthcm\*sepallengthcm > 25);

"""

print("\*" \* 80)

print("subqueries")

print("-" \* 80)

print(q)

print(pysqldf(q))