pandasql demo-

**!**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

**import** re

data **=** pd.read\_csv("D:/Documents/Data Science/regression/-50Startups.csv")

data.head()

print(sqldf("SELECT \* FROM data LIMIT ;", locals()))

print(sqldf("SELECT State, Administration, Profit FROM data LIMIT 10;", locals()))

q **=** """

select

State

, avg(Profit)

, min(Profit)

, max(Profit)

from

data

group by

State;

"""

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

data a

inner join

data b

on a.State = b.State

limit ;

"""

print("\*" **\*80** )

print("joins")

print("-" **\*80** )

print(q)

print(pysqldf(q))

q **=** """

select

\*

from

data

where

State = 'California'

and Profit > 150000;

"""

print("\*" **\*80** )

print("where clause")

print("-" **\*80** )

print(q)

print(pysqldf(q))

data['id'] **=** range(len(data))

q **=** """

select

\*

from

data

where

id in (select id from data where Profit/Administration > 1.25.);

"""

print("\*" **\*80** )

print("subqueries")

print("-" **\*** 80)

print(q)

print(pysqldf(q))