Assignment 13.1

Q. Create a sql db from adult dataset and name it sqladb

Answer:

import numpy as np

import pandas as pd

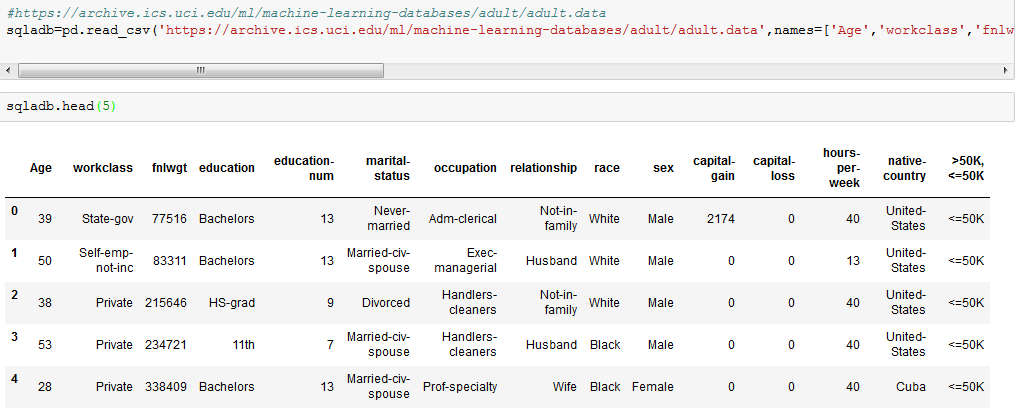
from pandas import DataFrame, Series

import sqlite3 as db

from pandasql import sqldf

pysqldf = lambda q: sqldf(q, globals())

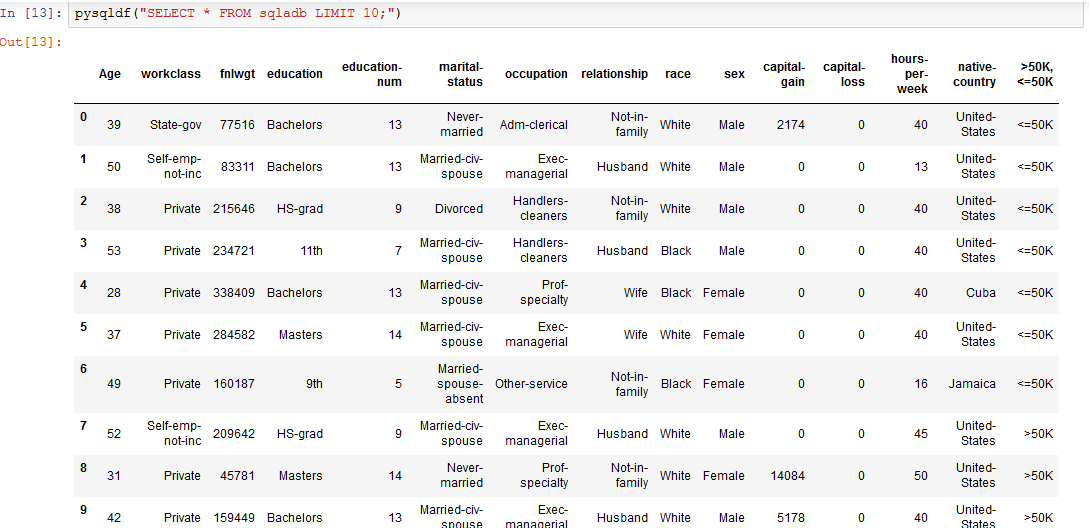
sqladb=pd.read\_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/adult/adult.data',names=['Age','workclass','fnlwgt','education','education-num','marital-status','occupation','relationship','race','sex','capital-gain','capital-loss','hours-per-week','native-country','>50K, <=50K'])



Q1. Select 10 records from the adult sqladb

Answer:

pysqldf("SELECT \* FROM sqladb LIMIT 10;")



Q2. Show me the average hours per week of all men who are working in private sector

Answer:

<<Trim all string(Object) columns for blank spaces >>

def trimAllColumns(df):

"""

Trim whitespace from ends of each value across all series in dataframe

"""

trimStrings = lambda x: x.strip() if type(x) is str else x

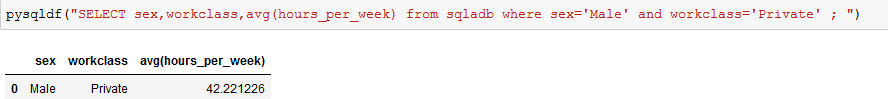
return df.applymap(trimStrings)

sqladb = trimAllColumns(sqladb)

<<Replace column name having “-“ with “\_”>>

sqladb.columns = sqladb.columns.str.replace('-', '\_')

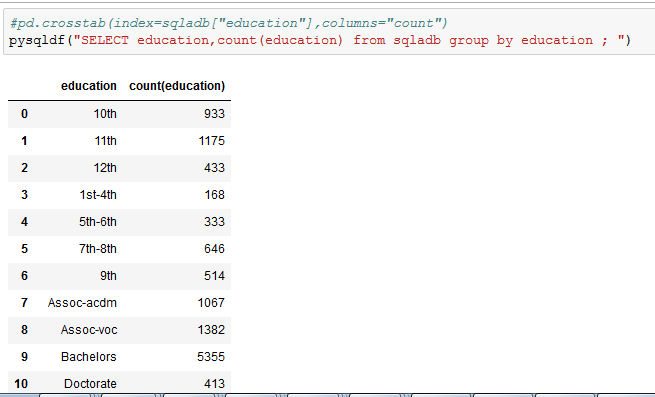
<<Displaying the Average Hours per week for all men in private sector>>



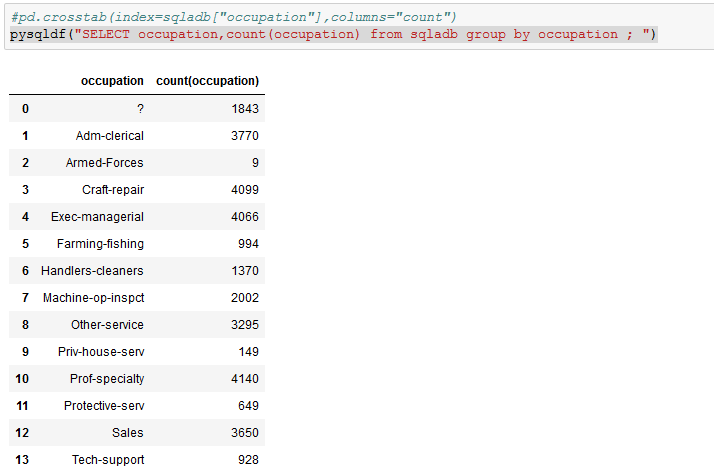
Q3. Show me the frequency table for education, occupation and relationship, separately

Answer:

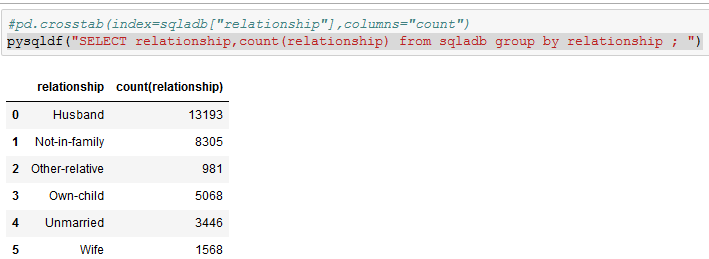
pysqldf("SELECT education,count(education) from sqladb group by education ; ")



pysqldf("SELECT occupation,count(occupation) from sqladb group by occupation ; ")



pysqldf("SELECT relationship,count(relationship) from sqladb group by relationship ; ")



Q4. Are there any people who are married, working in private sector and having a masters

Degree

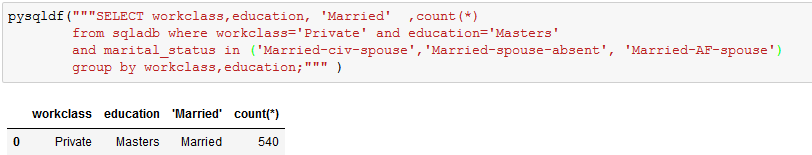
Answer:

pysqldf("""SELECT workclass,education, 'Married' ,count(\*)

from sqladb where workclass='Private' and education='Masters'

and marital\_status in ('Married-civ-spouse','Married-spouse-absent', 'Married-AF-spouse')

group by workclass,education;""" )

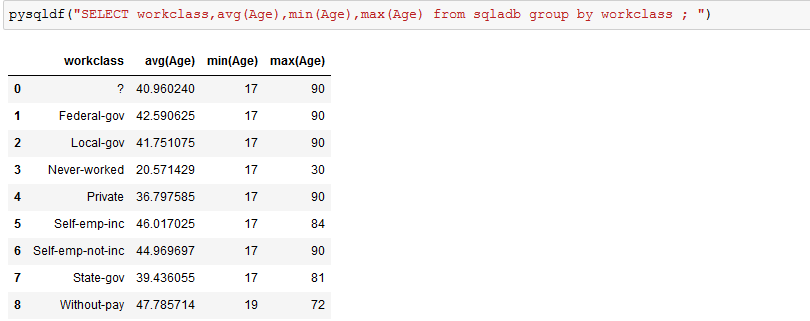


Q5. What is the average, minimum and maximum age group for people working in

different sectors

Answer:

pysqldf("SELECT workclass,avg(Age),min(Age),max(Age) from sqladb group by workclass ; ")



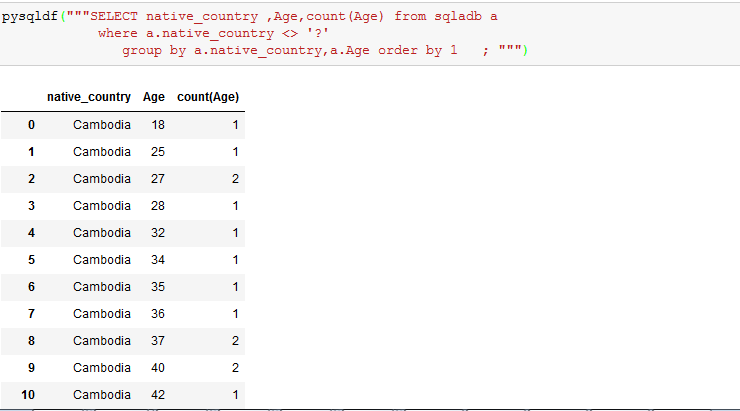
Q6. Calculate age distribution by country

Answer:

pysqldf("""SELECT native\_country ,Age,count(Age) from sqladb a

where a.native\_country <> '?'

group by a.native\_country,a.Age order by 1 ; """)



Q7:Compute a new column as 'Net-Capital-Gain' from the two columns 'capital-gain' and

'capital-loss'

Answer :

pysqldf("""SELECT \*,(capital\_gain-capital\_loss) 'Net-Capital-Gain' from sqladb a

; """)

