
# Data Warehousing & Mining Lab Assignment

# Lab - 1

Sub Code: CSE-326

Date: 12-01-2022

**Vivek Kumar Ahirwar**

**191112419**

**CSE - 3**

Department:

*Computer Science and Engineering*

# Contents

# A. Consider the sample data set (Sample.csv file Review the following attributes of the data set). Convert this into My SQL table:

Feature attributes:
• **Age**: continuous.
• **Work class:** Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Localgov State-Gov, Without-pay, Never-worked.
• **Final Weight**: continuous.
• **Education**: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assocacdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate 5th-6th, Preschool.
• **Education-num**: continuous.
• **Marital-status**: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
• **Occupation**: Tech-support, Craft-repair, Other-service, Sales, Execmanagerial Prof-specialty, Handlers-cleaners, Machine-op-inspct, Admclerical, Farming-fishing, Transport-moving, Priv-house-serv, Protectiveserv, Armed-Forces.
• **Relationship**: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
• **Race**: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
• **Sex**: Female, Male.
• **Capital-gain**: continuous.
• **Capital-loss**: continuous.
• **Hours-per-week**: continuous.
• **Native-country**: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying US (Guam-USVI-etc), India, Japan etc.
• **Income**: >50K, <=50K.

**PHP, Visual C++, C# Java, SQL, R, Python**

## i. Code

```
-- Create Table

CREATE TABLE lab1 (
age varchar(100),
workclass VARCHAR(100),
finalweight varchar(100),
education varchar(100),
educationnum varchar(100),
maritalstatus varchar(100),
occupation varchar(100),
relationship varchar(100),
race varchar(100),
sex varchar(100),
capitalgain varchar(100),
capitalloss varchar(100),
hoursperweek varchar(100),
nativecountry varchar(100),
salary varchar(100)
);


--  Insert data from txt file

load data local infile 'F:/MANIT-Online class/Semester-6/CSE326 DWM Lab/Lab-1 12-01-
2021/sample.txt' into table lab1
fields terminated by ','
lines terminated by '\n'
(age, workclass, finalweight, education, educationnum, maritalstatus, occupation,
relationship, race, sex, capitalgain, capitalloss, hoursperweek, nativecountry, salary);


-- Showing the data
select * from lab1;

-- Count the rows of data
select count(*) from lab1;
```

## ii. Explanation for code

I have used MySQL for importing dataset into MySQL table using 'MySQL Workbench IDE'. Firstly, I have created new Table named '*lab1*' in my database CREATE TABLE command according to given description of data. Then, I have loaded the Sample.txt file using 'load data local' command *(note: we need to first enable loading local file in our server)* and separated field by ',' and lines (entries) are separated by '\n'. Then at last output the data and count of no of rows in the data.

## iii. Output

SQL File 3*

Limit to 1000 rows

```
28
29    -- Showing the data
30 •  select * from lab1;
31
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

| age | workclass | finalweight | education | educationnum | maritalstatus | occupation | relationship | race | sex | capitalgain | capitalloss | hoursperweek | nativecountry | salary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | Private | 122272 | Bachelors | 13 | Never-married | Adm-clerical | Own-child | White | Female | 0 | 0 | 30 | United-States | <=50K |
| 32 | Private | 205019 | Assoc-ac... | 12 | Never-married | Sales | Not-in-family | Black | Male | 0 | 0 | 50 | United-States | <=50K |
| 40 | Private | 121772 | Assoc-voc | 11 | Married-civ-spouse | Craft-repair | Husband | Asia... | Male | 0 | 0 | 40 | ? | >50K |
| 34 | Private | 245487 | 7th-8th | 4 | Married-civ-spouse | Transport-moving | Husband | Amer... | Male | 0 | 0 | 45 | Mexico | <=50K |
| 25 | Self-emp-not-inc | 176756 | HS-grad | 9 | Never-married | Farming-fishing | Own-child | White | Male | 0 | 0 | 35 | United-States | <=50K |
| 32 | Private | 186824 | HS-grad | 9 | Never-married | Machine-op-inspct | Unmarried | White | Male | 0 | 0 | 40 | United-States | <=50K |
| 38 | Private | 28887 | 11th | 7 | Married-civ-spouse | Sales | Husband | White | Male | 0 | 0 | 50 | United-States | <=50K |
| 43 | Self-emp-not-inc | 292175 | Masters | 14 | Divorced | Exec-managerial | Unmarried | White | Female | 0 | 0 | 45 | United-States | >50K |
| 40 | Private | 193524 | Doctorate | 16 | Married-civ-spouse | Prof-specialty | Husband | White | Male | 0 | 0 | 60 | United-States | >50K |
| 54 | Private | 302146 | HS-grad | 9 | Separated | Other-service | Unmarried | Black | Female | 0 | 0 | 20 | United-States | <=50K |
| 35 | Federal-gov | 76845 | 9th | 5 | Married-civ-spouse | Farming-fishing | Husband | Black | Male | 0 | 0 | 40 | United-States | <=50K |
| 43 | Private | 117037 | 11th | 7 | Married-civ-spouse | Transport-moving | Husband | White | Male | 0 | 2042 | 40 | United-States | <=50K |
| 59 | Private | 109015 | HS-grad | 9 | Divorced | Tech-support | Unmarried | White | Female | 0 | 0 | 40 | United-States | <=50K |
| 56 | Local-gov | 216851 | Bachelors | 13 | Married-civ-spouse | Tech-support | Husband | White | Male | 0 | 0 | 40 | United-States | >50K |
| 19 | Private | 168294 | HS-grad | 9 | Never-married | Craft-repair | Own-child | White | Male | 0 | 0 | 40 | United-States | <=50K |
| 54 | ? | 180211 | Some-col... | 10 | Married-civ-spouse | ? | Husband | Asia... | Male | 0 | 0 | 60 | South | >50K |
| 39 | Private | 367260 | HS-grad | 9 | Divorced | Exec-managerial | Not-in-family | White | Male | 0 | 0 | 80 | United-States | <=50K |

lab1 2

Output

Action Output

| # | Time | Action | Message |
|---|---|---|---|
| 1 | 23:15:06 | CREATE TABLE lab1 ( age varchar(100), workclass VARCHAR(100), finalweight varchar(100), education v... | 0 row(s) affected |
| 2 | 23:15:08 | load data local infile 'F:/MANIT-Online class/Semester-6/CSE326 DWM Lab/Lab-1 12-01-2021/sample.txt' i... | 31434 row(s) affected, 13 warning(s): 1261 Row 31434 doesn't contain data for all columns 1261 Row 3143... |
| 3 | 23:15:26 | select * from lab1 LIMIT 0, 1000 | 1000 row(s) returned |