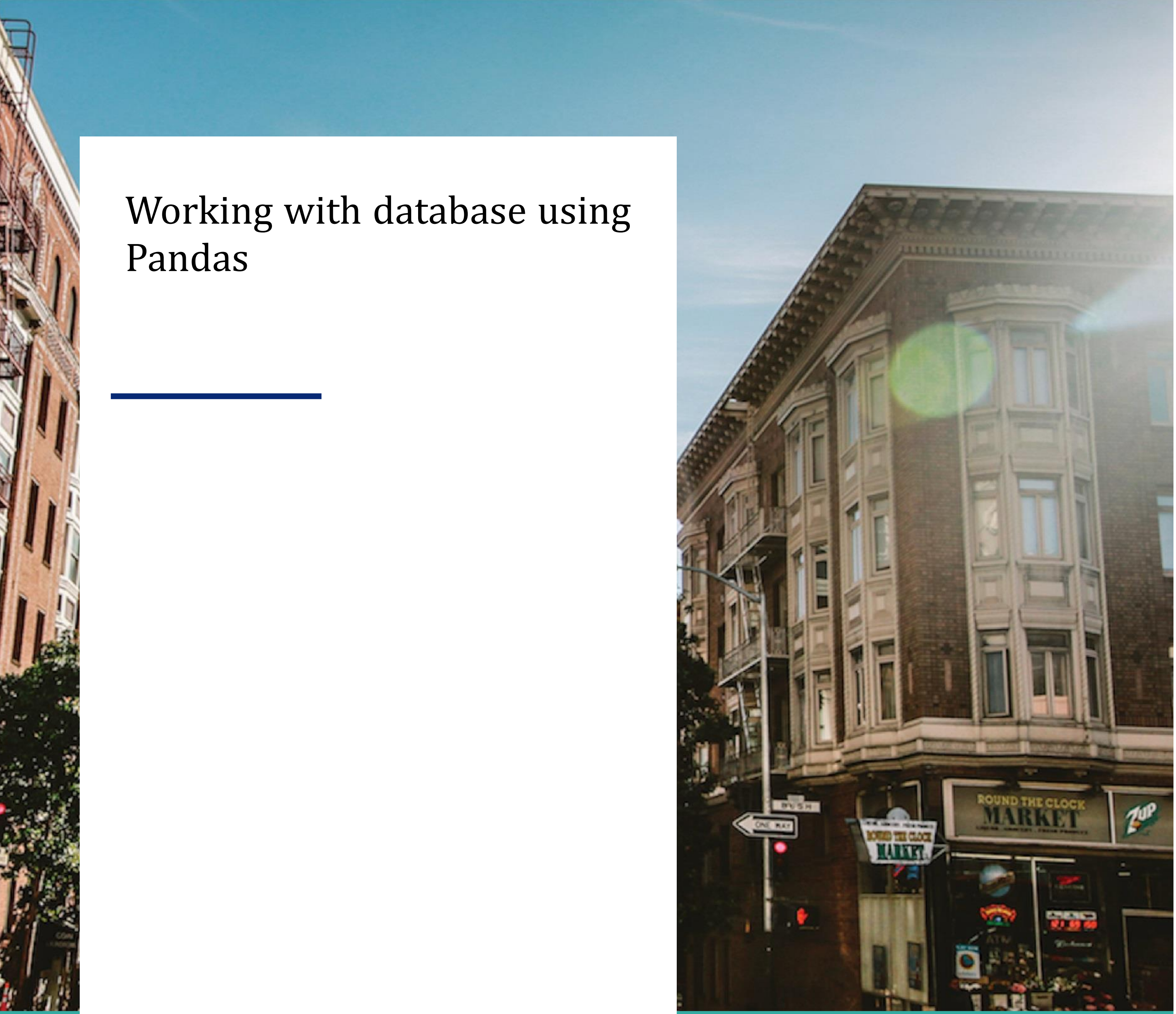


Working with database using Pandas

NOVEMBER 10

ANT

Authored by: Your Name

Logo
Name

https://www.geeksforgeeks.org/working-with-database-using-pandas/

Working with database using Pandas

Last Updated: 07-07-2020

Performing various operations on data saved in SQL might lead to performing very complex queries that are not easy to write. So to make this task easier it is often useful to do the job using pandas which are specially built for data preprocessing and is more simple and user-friendly than SQL.

There might be cases when sometimes the data is stored in SQL and we want to fetch that data from SQL in python and then perform operations using pandas. So let’s see how we can interact with SQL databases using pandas.

This is the database we are going to work with
diabetes_data
Note: Assuming that we the data is stored in sqlite3
Reading the data

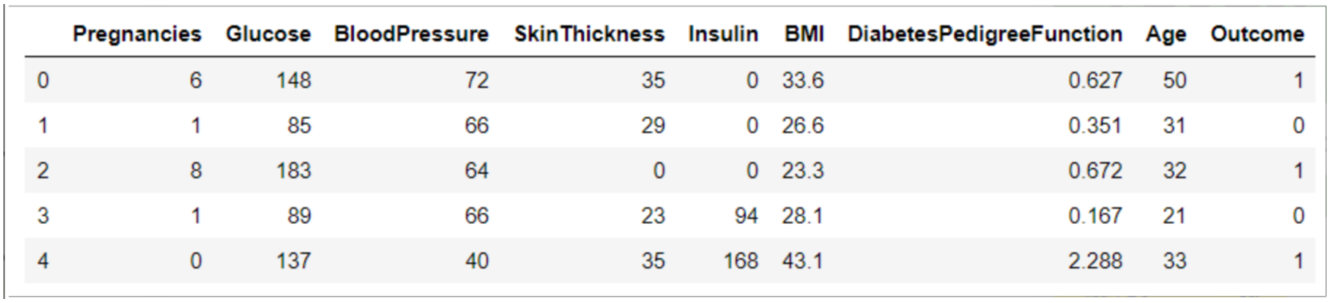
```
# import the libraries
import sqlite3
import pandas as pd

# create a connection
con = sqlite3.connect('Diabetes.db')

# read data from SQL to pandas dataframe.
data = pd.read_sql_query('Select * from Diabetes;', con)

# show top 5 rows
data.head()
```

Output:



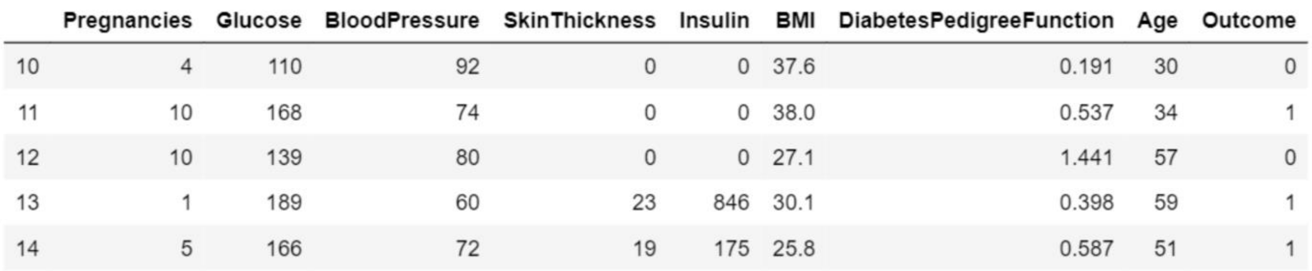
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Basic operation

- **Slicing of rows**
  We can perform slicing operations to get the desired number of rows from within a given range. With the help of slicing, we can perform various operations only on the specific subset of the data

```
# read the data from sql to pandas dataframe.
data = pd.read_sql_query('Select * from Diabetes;', con)

# slicing the number of rows
df1 = data[10:15]
df1
```



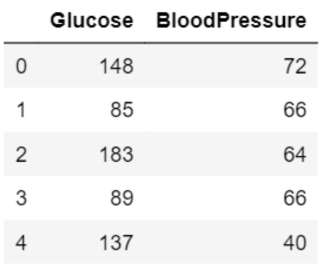
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
10	4	110	92	0	0	37.6	0.191	30	0
11	10	168	74	0	0	38.0	0.537	34	1
12	10	139	80	0	0	27.1	1.441	57	0
13	1	189	60	23	846	30.1	0.398	59	1
14	5	166	72	19	175	25.8	0.587	51	1

Selecting specific columns

To select a particular column or to select number of columns from the dataframe for further processing of data.

```
# read the data from sql to
# pandas dataframe.
data = pd.read_sql_query('Select * from Diabetes;', con)

# selecting specific columns.
df2 = data.loc[:, ['Glucose', 'BloodPressure']].head()
df2
```



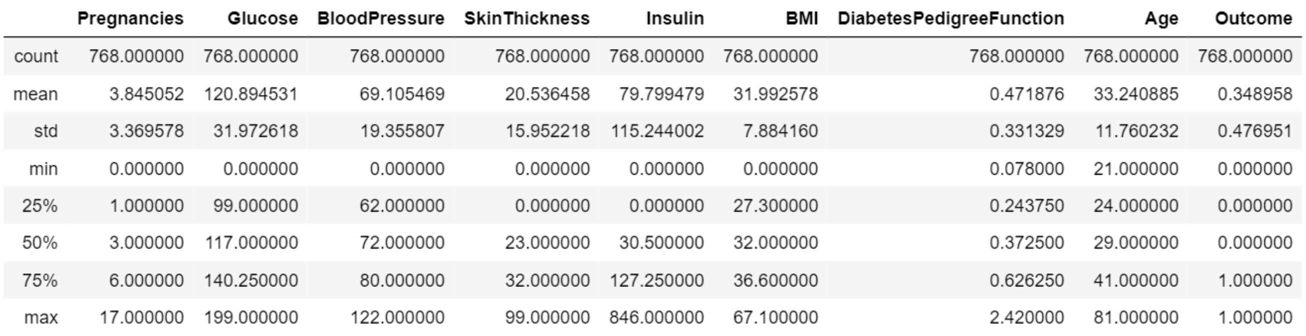
	Glucose	BloodPressure
0	148	72
1	85	66
2	183	64
3	89	66
4	137	40

Summarize the data

In order to get insights from data, we must have a statistical summary of data. To display a statistical summary of the data such as mean, median, mode, std etc. We perform the following operation

```
# read the data from sql
# to pandas dataframe.
data = pd.read_sql_query('Select * from Diabetes;', con)

# summarize the data
data.describe()
```



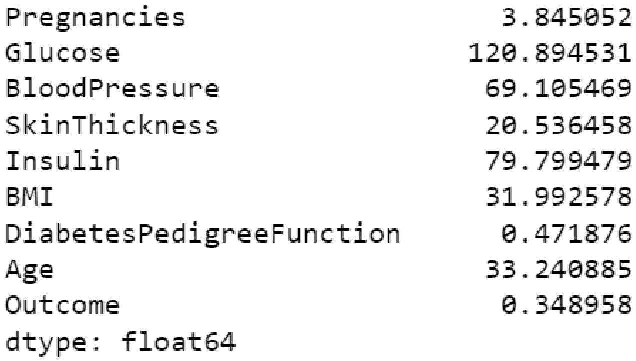
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876	33.240885	0.348958
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750	24.000000	0.000000
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500	29.000000	0.000000
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250	41.000000	1.000000
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000	81.000000	1.000000

Sort data with respect to a column

For sorting the dataframe with respect to a given column values.

```
# read the data from sql
# to pandas dataframe.
data = pd.read_sql_query('Select * from Diabetes;', con)

# count number of rows and columns
data.mean()
```



```
Pregnancies                   3.845052
Glucose                     120.894531
BloodPressure                69.105469
SkinThickness                20.536458
Insulin                      79.799479
BMI                          31.992578
DiabetesPedigreeFunction      0.471876
Age                          33.240885
Outcome                       0.348958
dtype: float64
```