

# Abdul Basit\_Assignment\_Week 11 & 12 Exercises

File: Abdul Basit\_Assignment\_Week 11 & 12 Exercises.ipynb Name: Abdul Basit Date: 05/28/2020

Course: DSC 540 Data Preparation Activity 11 page 320

```
In [6]: from __future__ import print_function, division
from bs4 import BeautifulSoup
```

```
!pip install plotly
!pip install cufflinks

import matplotlib.pyplot as plt
import statsmodels.formula.api as smf
import pandas as pd
import plotly.offline as py
import plotly.graph_objs as go
import string
import os
import numpy as np
import sys
import math
import scipy.stats
import density
import random
import hypothesis
import scatter
import requests
import re, string
import random
import sklearn
import twitter
```

```
!apt-get update
!apt-get install -y default-jdk
!pip install tabula-py xlrd lxml
!pip install tweepy
!pip install Tkinter
!pip install Textblob
!pip install -r requirements.txt
!pip install virtualenv
```

```
py.init_notebook_mode(connected=True)
```

```
Requirement already satisfied: plotly in c:\users\basiab1\appdata\local\co
ntinuum\anaconda3\lib\site-packages (4.6.0)
Requirement already satisfied: retrying>=1.3.3 in c:\users\basiab1\appdata
\local\continuum\anaconda3\lib\site-packages (from plotly) (1.3.3)
Requirement already satisfied: six in c:\users\basiab1\appdata\local\conti
```

```

num\anaconda3\lib\site-packages (from plotly) (1.11.0)
Requirement already satisfied: cufflinks in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (0.17.3)
Requirement already satisfied: colorlover>=0.2.1 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (0.3.0)
Requirement already satisfied: ipython>=5.3.0 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (7.1.1)
Requirement already satisfied: numpy>=1.9.2 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (1.15.4)
Requirement already satisfied: ipywidgets>=7.0.0 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (7.4.2)
Requirement already satisfied: setuptools>=34.4.1 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (40.5.0)
Requirement already satisfied: plotly>=4.1.1 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (4.6.0)
Requirement already satisfied: six>=1.9.0 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (1.11.0)
Requirement already satisfied: pandas>=0.19.2 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from cufflinks) (1.0.3)
Requirement already satisfied: pygments in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (2.2.0)
Requirement already satisfied: backcall in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.1.0)
Requirement already satisfied: colorama; sys_platform == "win32" in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.4.0)
Requirement already satisfied: traitlets>=4.2 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (4.3.2)
Requirement already satisfied: prompt-toolkit<2.1.0,>=2.0.0 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (2.0.7)
Requirement already satisfied: decorator in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (4.3.0)
Requirement already satisfied: pickleshare in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.7.5)
Requirement already satisfied: jedi>=0.10 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipython>=5.3.0->cufflinks) (0.13.1)
Requirement already satisfied: widgetsnbextension~=3.4.0 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (3.4.2)
Requirement already satisfied: nbformat>=4.2.0 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (4.4.0)
Requirement already satisfied: ipykernel>=4.5.1 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from ipywidgets>=7.0.0->cufflinks) (5.1.0)
Requirement already satisfied: retrying>=1.3.3 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from plotly>=4.1.1->cufflinks) (1.3.3)
Requirement already satisfied: python-dateutil>=2.6.1 in c:\users\basiabl\appdata\local\continuum\anaconda3\lib\site-packages (from pandas>=0.19.2->

```

```

cufflinks) (2.7.5)
Requirement already satisfied: pytz>=2017.2 in c:\users\basiab1\appdata\lo
cal\continuum\anaconda3\lib\site-packages (from pandas>=0.19.2->cufflinks)
(2018.7)
Requirement already satisfied: ipython-genutils in c:\users\basiab1\appdat
a\local\continuum\anaconda3\lib\site-packages (from traitlets>=4.2->ipytho
n>=5.3.0->cufflinks) (0.2.0)
Requirement already satisfied: wcwidth in c:\users\basiab1\appdata\local\c
ontinuum\anaconda3\lib\site-packages (from prompt-toolkit<2.1.0,>=2.0.0->i
python>=5.3.0->cufflinks) (0.1.7)
Requirement already satisfied: parso>=0.3.0 in c:\users\basiab1\appdata\lo
cal\continuum\anaconda3\lib\site-packages (from jedi>=0.10->ipython>=5.3.0
->cufflinks) (0.3.1)
Requirement already satisfied: notebook>=4.4.1 in c:\users\basiab1\appdata
\local\continuum\anaconda3\lib\site-packages (from widgetsnbextension~=3.4
.0->ipywidgets>=7.0.0->cufflinks) (5.7.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\users\basiab1
\appdata\local\continuum\anaconda3\lib\site-packages (from nbformat>=4.2.0
->ipywidgets>=7.0.0->cufflinks) (2.6.0)
Requirement already satisfied: jupyter-core in c:\users\basiab1\appdata\lo
cal\continuum\anaconda3\lib\site-packages (from nbformat>=4.2.0->ipywidget
s>=7.0.0->cufflinks) (4.4.0)
Requirement already satisfied: tornado>=4.2 in c:\users\basiab1\appdata\lo
cal\continuum\anaconda3\lib\site-packages (from ipykernel>=4.5.1->ipywidge
ts>=7.0.0->cufflinks) (5.1.1)
Requirement already satisfied: jupyter-client in c:\users\basiab1\appdata\
local\continuum\anaconda3\lib\site-packages (from ipykernel>=4.5.1->ipywid
gets>=7.0.0->cufflinks) (5.2.3)
Requirement already satisfied: nbconvert in c:\users\basiab1\appdata\local
\continuum\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbext
ension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (5.3.1)
Requirement already satisfied: terminado>=0.8.1 in c:\users\basiab1\appdat
a\local\continuum\anaconda3\lib\site-packages (from notebook>=4.4.1->widge
tsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (0.8.1)
Requirement already satisfied: jinja2 in c:\users\basiab1\appdata\local\co
ntinuum\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbextens
ion~=3.4.0->ipywidgets>=7.0.0->cufflinks) (2.10)
Requirement already satisfied: pyzmq>=17 in c:\users\basiab1\appdata\local
\continuum\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbext
ension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (17.1.2)
Requirement already satisfied: Send2Trash in c:\users\basiab1\appdata\loca
l\continuum\anaconda3\lib\site-packages (from notebook>=4.4.1->widgetsnbex
tension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (1.5.0)
Requirement already satisfied: prometheus-client in c:\users\basiab1\appda
ta\local\continuum\anaconda3\lib\site-packages (from notebook>=4.4.1->widg
etsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (0.4.2)
Requirement already satisfied: mistune>=0.7.4 in c:\users\basiab1\appdata\
local\continuum\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4
.1->widgetsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (0.8.4)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\basiab1\ap
pdata\local\continuum\anaconda3\lib\site-packages (from nbconvert->noteboo
k>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (1.4.2)
Requirement already satisfied: testpath in c:\users\basiab1\appdata\local\
continuum\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->wi
dgetsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (0.4.2)
Requirement already satisfied: entrypoints>=0.2.2 in c:\users\basiab1\appd
ata\local\continuum\anaconda3\lib\site-packages (from nbconvert->notebook>

```

```
=4.4.1->widgetsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (0.2.3)
Requirement already satisfied: bleach in c:\users\basiab1\appdata\local\co
ntinuum\anaconda3\lib\site-packages (from nbconvert->notebook>=4.4.1->widg
etsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (3.0.2)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\basiab1\appdat
a\local\continuum\anaconda3\lib\site-packages (from jinja2->notebook>=4.4.
1->widgetsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (1.1.0)
Requirement already satisfied: webencodings in c:\users\basiab1\appdata\lo
cal\continuum\anaconda3\lib\site-packages (from bleach->nbconvert->noteboo
k>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets>=7.0.0->cufflinks) (0.5.1)
```

```
'apt-get' is not recognized as an internal or external command,
operable program or batch file.
'apt-get' is not recognized as an internal or external command,
operable program or batch file.
```

```
Requirement already satisfied: tabula-py in c:\users\basiab1\appdata\local
\continuum\anaconda3\lib\site-packages (2.1.0)
Requirement already satisfied: xlrd in c:\users\basiab1\appdata\local\cont
inuum\anaconda3\lib\site-packages (1.1.0)
Requirement already satisfied: lxml in c:\users\basiab1\appdata\local\cont
inuum\anaconda3\lib\site-packages (4.2.5)
Requirement already satisfied: pandas>=0.25.3 in c:\users\basiab1\appdata\
local\continuum\anaconda3\lib\site-packages (from tabula-py) (1.0.3)
Requirement already satisfied: distro in c:\users\basiab1\appdata\local\co
ntinuum\anaconda3\lib\site-packages (from tabula-py) (1.5.0)
Requirement already satisfied: numpy in c:\users\basiab1\appdata\local\con
tinuum\anaconda3\lib\site-packages (from tabula-py) (1.15.4)
Requirement already satisfied: pytz>=2017.2 in c:\users\basiab1\appdata\lo
cal\continuum\anaconda3\lib\site-packages (from pandas>=0.25.3->tabula-py)
(2018.7)
Requirement already satisfied: python-dateutil>=2.6.1 in c:\users\basiab1\
appdata\local\continuum\anaconda3\lib\site-packages (from pandas>=0.25.3->
tabula-py) (2.7.5)
Requirement already satisfied: six>=1.5 in c:\users\basiab1\appdata\local\
continuum\anaconda3\lib\site-packages (from python-dateutil>=2.6.1->pandas
>=0.25.3->tabula-py) (1.11.0)
Requirement already satisfied: tweepy in c:\users\basiab1\appdata\local\co
ntinuum\anaconda3\lib\site-packages (3.8.0)
Requirement already satisfied: requests>=2.11.1 in c:\users\basiab1\appdat
a\local\continuum\anaconda3\lib\site-packages (from tweepy) (2.20.1)
Requirement already satisfied: requests-oauthlib>=0.7.0 in c:\users\basiab
1\appdata\local\continuum\anaconda3\lib\site-packages (from tweepy) (1.3.0
)
Requirement already satisfied: six>=1.10.0 in c:\users\basiab1\appdata\loc
al\continuum\anaconda3\lib\site-packages (from tweepy) (1.11.0)
Requirement already satisfied: PySocks>=1.5.7 in c:\users\basiab1\appdata\
local\continuum\anaconda3\lib\site-packages (from tweepy) (1.6.8)
Requirement already satisfied: idna<2.8,>=2.5 in c:\users\basiab1\appdata\
local\continuum\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy
) (2.7)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in c:\users\basiab1\appdata\local\continuum\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy) (1.23)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\basiab1\appdata\local\continuum\anaconda3\lib\site-packages (from requests>=2.11.1->tweepy) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\basiab1\appd
```

```
ata\local\continuum\anaconda3\lib\site-packages (from requests>=2.11.1->tw
eepy) (2019.6.16)
Requirement already satisfied: oauthlib>=3.0.0 in c:\users\basiabl\appdata
\local\continuum\anaconda3\lib\site-packages (from requests-oauthlib>=0.7.
0->tweepy) (3.1.0)
Collecting Tkinter
```

Could not find a version that satisfies the requirement Tkinter (from ve  
rsions: )  
No matching distribution found for Tkinter

```
Requirement already satisfied: Textblob in c:\users\basiabl\appdata\local\
continuum\anaconda3\lib\site-packages (0.15.3)
Requirement already satisfied: nltk>=3.1 in c:\users\basiabl\appdata\local
\continuum\anaconda3\lib\site-packages (from Textblob) (3.3)
Requirement already satisfied: six in c:\users\basiabl\appdata\local\conti
nuum\anaconda3\lib\site-packages (from nltk>=3.1->Textblob) (1.11.0)
Collecting matplotlib==1.4.3 (from -r requirements.txt (line 1))
```

```
Using cached https://files.pythonhosted.org/packages/bb/ac/485df0ecb15aa
6fec1991945dc0cabfeb724a64f6729e34bab3c6a766813/matplotlib-1.4.3.tar.gz
```

```
Complete output from command python setup.py egg_info:
=====
=====
Edit setup.cfg to change the build options
```

```
BUILDING MATPLOTLIB
    matplotlib: yes [1.4.3]
    python: yes [3.7.1 (default, Oct 28 2018, 08:39:03) [M
SC
    v.1912 64 bit (AMD64)]]
    platform: yes [win32]
```

```
REQUIRED DEPENDENCIES AND EXTENSIONS
    numpy: yes [version 1.15.4]
    six: yes [using six version 1.11.0]
    dateutil: yes [using dateutil version 2.7.5]
    pytz: yes [using pytz version 2018.7]
    tornado: yes [using tornado version 5.1.1]
    pyparsing: yes [using pyparsing version 2.3.0]
    pycxx: yes [Official versions of PyCXX are not compat
```

```
ible
    with matplotlib on Python 3.x, since they lack
    support for the buffer object. Using local co
```

```
py]
    libagg: yes [pkg-config information for 'libagg' could
not
    be found. Using local copy.]
```

```
freetype: no [The C/C++ header for freetype (ft2build.h
)
    could not be found. You may need to install t
he
    development package.]
```

```
png: no [The C/C++ header for png (png.h) could no
t be
    found. You may need to install the developmen
t
    package.]
```

```
qhull: yes [pkg-config information for 'qhull' could
```

```
not be
found. Using local copy.]

OPTIONAL SUBPACKAGES
    sample_data: yes [installing]
    toolkits: yes [installing]
    tests: yes [using nose version 1.3.7 / using unittest
.mock]
    toolkits_tests: yes [using nose version 1.3.7 / using unittest
.mock]

OPTIONAL BACKEND EXTENSIONS
    macosx: no [Mac OS-X only]
    qt5agg: yes [installing, Qt: 5.9.6, PyQt: 5.9.6]
    qt4agg: no [PyQt4 not found]
    pyside: no [PySide not found]
    gtk3agg: no [Requires pygobject to be installed.]
    gtk3cairo: no [Requires cairocffi or pycairo to be insta
lled.]
    gtkagg: no [Requires pygtk]
    tkagg: no [The C/C++ header for Tk (tk.h) could not
be
found. You may need to install the developmen
t
package.]
    wxagg: no [requires wxPython]
    gtk: no [Requires pygtk]
    agg: yes [installing]
    cairo: no [cairocffi or pycairo not found]
    windowing: yes [installing, installing]

OPTIONAL LATEX DEPENDENCIES
    dvipng: no
    ghostscript: no
    latex: no
    pdftops: no
```

```
=====
* The following required packages can not be b
uilt:
* freetype, png
-----
```

```
Command "python setup.py egg_info" failed with error code 1 in C:\Users\basiab1\AppData\Local\Temp\pip-install-r72ktpou\matplotlib\
Requirement already satisfied: virtualenv in c:\users\basiab1\appdata\local\continuum\anaconda3\lib\site-packages (20.0.18)
Requirement already satisfied: appdirs<2,>=1.4.3 in c:\users\basiab1\appdata\local\continuum\anaconda3\lib\site-packages (from virtualenv) (1.4.3)
Requirement already satisfied: importlib-metadata<2,>=0.12; python_version < "3.8" in c:\users\basiab1\appdata\local\continuum\anaconda3\lib\site-packages (from virtualenv) (1.6.0)
Requirement already satisfied: six<2,>=1.9.0 in c:\users\basiab1\appdata\local\continuum\anaconda3\lib\site-packages (from virtualenv) (1.11.0)
Requirement already satisfied: filelock<4,>=3.0.0 in c:\users\basiab1\appd
```

```
ata\local\continuum\anaconda3\lib\site-packages (from virtualenv) (3.0.10)
Requirement already satisfied: distlib<1,>=0.3.0 in c:\users\basiab1\appda
ta\local\continuum\anaconda3\lib\site-packages (from virtualenv) (0.3.0)
Requirement already satisfied: zipp>=0.5 in c:\users\basiab1\appdata\local
\continuum\anaconda3\lib\site-packages (from importlib-metadata<2,>=0.12;
python_version < "3.8"->virtualenv) (0.5.2)
```

## TASK 1 : Connect to the supplied petsDB, and write a function to check if the connection is done

```
In [7]: import sqlite3
```

```
In [8]: conn = sqlite3.connect("petsdb")
```

```
In [9]: # a tiny function to make sure the connection is successful
def is_opened(conn):
    try:
        conn.execute("SELECT * FROM persons LIMIT 1")
        return True
    except sqlite3.ProgrammingError as e:
        print("Connection closed {}".format(e))
        return False
```

```
In [10]: print(is_opened(conn))
```

True

```
In [11]: print(is_opened(conn))
```

True

```
In [12]: conn.close()
```

```
In [13]: print(is_opened(conn))
```

Connection closed Cannot operate on a closed database.

False

## TASK 2: What are the different age groups in the persons database

```
In [14]: conn = sqlite3.connect("petsdb")
```

```
In [15]: c = conn.cursor()
```

```
In [16]: for ppl, age in c.execute("SELECT count(*), age FROM persons GROUP BY age"
):
    print("We have {} people aged {}".format(ppl, age))
```

We have 2 people aged 5

We have 1 people aged 6  
We have 1 people aged 7  
We have 3 people aged 8  
We have 1 people aged 9  
We have 2 people aged 11  
We have 3 people aged 12  
We have 1 people aged 13  
We have 4 people aged 14  
We have 2 people aged 16  
We have 2 people aged 17  
We have 3 people aged 18  
We have 1 people aged 19  
We have 3 people aged 22  
We have 2 people aged 23  
We have 3 people aged 24  
We have 2 people aged 25  
We have 1 people aged 27  
We have 1 people aged 30  
We have 3 people aged 31  
We have 1 people aged 32  
We have 1 people aged 33  
We have 2 people aged 34  
We have 3 people aged 35  
We have 3 people aged 36  
We have 1 people aged 37  
We have 2 people aged 39  
We have 1 people aged 40  
We have 1 people aged 42  
We have 2 people aged 44  
We have 2 people aged 48  
We have 1 people aged 49  
We have 1 people aged 50  
We have 2 people aged 51  
We have 2 people aged 52  
We have 2 people aged 53  
We have 2 people aged 54  
We have 1 people aged 58  
We have 1 people aged 59  
We have 1 people aged 60  
We have 1 people aged 61  
We have 2 people aged 62  
We have 1 people aged 63  
We have 2 people aged 65  
We have 2 people aged 66  
We have 1 people aged 67  
We have 3 people aged 68  
We have 1 people aged 69  
We have 1 people aged 70  
We have 4 people aged 71  
We have 1 people aged 72  
We have 5 people aged 73  
We have 3 people aged 74

**Task 3: Which age group has maximum number of people?**



```
In [17]: for ppl, age in c.execute("SELECT count(*), age FROM persons GROUP BY age
ORDER BY count(*) DESC"):
    print("Highest number of people {} came from {} age group".format(ppl,
age))
    break
```

Highest number of people 5 came from 73 age group

## Task 4: How many people do not have a full name (Last name is blank/null)

```
In [18]: res = c.execute("SELECT count(*) FROM persons WHERE last_name IS null")
for row in res:
    print(row)
```

(60,)

## Task 5: How many people has more than one pet?

```
In [19]: res = c.execute("SELECT count(*) FROM (SELECT count(owner_id) FROM pets GR
OUP BY owner_id HAVING count(owner_id) >1)")
for row in res:
    print("{} People has more than one pets".format(row[0]))
```

43 People has more than one pets

## Task 6: How many pets have received treatments?

```
In [20]: res = c.execute("SELECT count(*) FROM pets WHERE treatment_done=1")
for row in res:
    print(row)
```

(36,)

## Task 7: How many pets have received treatment that we know the type of?

```
In [21]: res = c.execute("SELECT count(*) FROM pets WHERE treatment_done=1 AND pet_
type IS NOT null")
for row in res:
    print(row)
```

(16,)

## Task 8: How many pets are there from the city called "east port"

```
In [22]: res = c.execute("SELECT count(*) FROM pets JOIN persons ON pets.owner_id =
           persons.id WHERE persons.city='east port'")
           for row in res:
               print(row)

(49,)
```

## Task 9: How many pets are there from the city called "east port" and who received a treatment?

```
In [23]: res = c.execute("SELECT count(*) FROM pets JOIN persons ON pets.owner_id =
           persons.id WHERE persons.city='east port' AND pets.treatment_done=1")
           for row in res:
               print(row)

(11,)
```