**Jose Dixon**

**12.09.2016**

**Dr. Rahman**

**Assignment 4**

Question 1

Source Code

import csv

from numpy import genfromtxt

from pandas import Series, DataFrame

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

data = pd.read\_csv("weather.csv")

data ##1. Prints out the data as a summary.

data.iloc[364] ##2. Prints out the second to last row.

rangea = pd.read\_csv('weather.csv', usecols=[1])

rtempa = rangea.max(skipna=True, numeric\_only=True)

rtempa

rangeb = pd.read\_csv('weather.csv', usecols=[3])

rtempb = rangeb.min(skipna=True, numeric\_only=True)

rtempb

np.subtract(rtempa, rtempb) #3. The range of two temperatures in a dataset.

data1 = pd.read\_csv('weather.csv', quoting=2, usecols=[1], sep=",") ##4. First histogram of MaxTemperatureF.

data1.hist(bins=2)

plt.xlim([0, 150])

plt.ylim([0, 300])

plt.title("MaxTemperatureF")

plt.xlabel("Range of Time")

plt.ylabel("Temperature Range")

plt.legend()

plt.show()

data2 = pd.read\_csv('weather.csv', quoting=2, usecols=[2], sep=",") ##4. Second histrogram of MeanTemperatureF

data2.hist(bins=2)

plt.xlim([0, 150])

plt.ylim([0, 250])

plt.title("MeanTemperatureF")

plt.xlabel("Range of Time")

plt.ylabel("Temperature Range")

plt.legend()

plt.show()

rangeb = pd.read\_csv('weather.csv', quoting=2, usecols=[20])

rtempc = rangeb.mean(skipna=True, numeric\_only=True)

rtempc ##5.Group the DataFrame, and find the mean temperature of group.

ranged = pd.read\_csv('weather.csv', usecols=[16])

ranged

rangee = pd.read\_csv('weather.csv', usecols=[17])

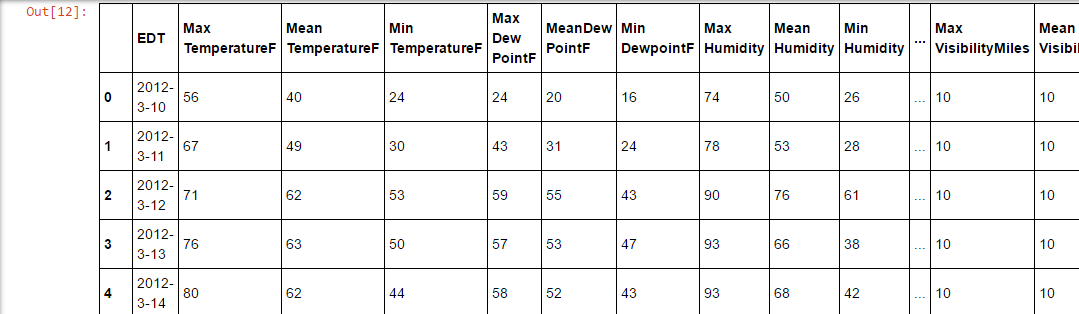
rangee

np.subtract(ranged, rangee)

print "The mean temperature is more variable on days of rain than snow." #6

Output

#1 Display the data out as a summary.



#2. Prints out the second to last row.

EDT 2013-3-9

Max TemperatureF 56

Mean TemperatureF 45

Min TemperatureF 33

Max Dew PointF 32

MeanDew PointF 29

Min DewpointF 23

Max Humidity 75

Mean Humidity 57

Min Humidity 38

Max Sea Level PressureIn 30.32

Mean Sea Level PressureIn 30.16

Min Sea Level PressureIn 29.96

Max VisibilityMiles 10

Mean VisibilityMiles 10

Min VisibilityMiles 10

Max Wind SpeedMPH 16

Mean Wind SpeedMPH 10

Max Gust SpeedMPH 24

PrecipitationIn T

CloudCover 2

Events NaN

WindDirDegrees 141

Name: 364, dtype: object

Max TemperatureF 106

dtype: int64

Min TemperatureF 1

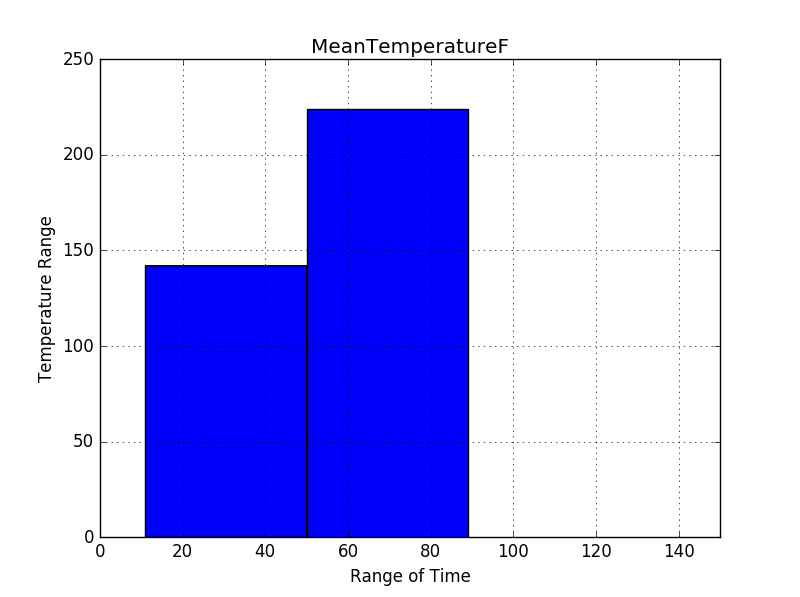
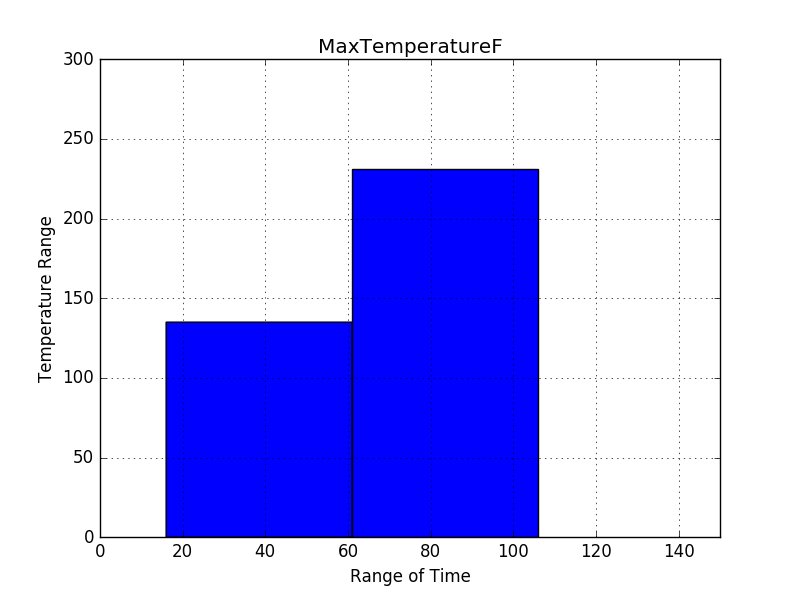
dtype: int64

#3. The range of two temperatures.

Max TemperatureF 105

dtype: int64

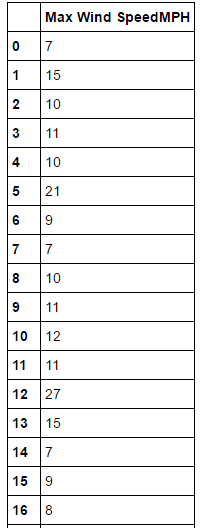
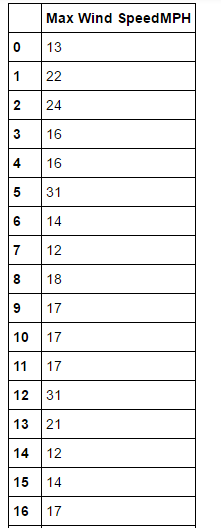
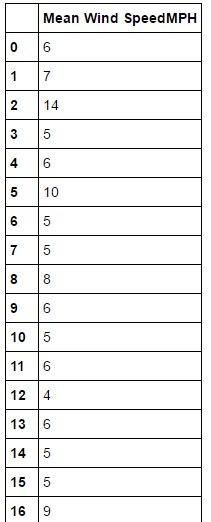
#4. Histogram for the two columns



##5.Group the DataFrame, and find the mean temperature of group.

CloudCover 2.885246

dtype: float64



Question 2

Source Code

import csv

from numpy import genfromtxt

from pandas import Series, DataFrame

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

data = pd.read\_csv("student.csv")

data

rangea = pd.read\_csv('student.csv', usecols=[2, 3, 4]) #1. What does the data tell you. Sum of data.

rtempa = rangea.sum(skipna=True, numeric\_only=True)

rtempa

rtempb = rangea.mean(skipna=True, numeric\_only =True) #1. What does the data tell you. Mean of data.

rtempb

print "The schools with the highest dropout rates are Peace and Education High School and Banner Academy South High School."

rtempc = rangea.max(skipna=True, numeric\_only=True)

rtempc

#2. Schools with highest dropout rates.

print "Network of schools with the higher dropout rates is Alternative Schools. "

rtempg = data.iloc[9]

#3. Network of schools with highest dropout rates.

print "The schools with the highest passing rates are Devry High School, Northside Learning High School, and Northside Learning Prep High School."

rtempd = rangea.min(skipna=True, numeric\_only=True)

rtempd #4. Schools with highest passing rates.

print "The school with the highest passing rate of students is Northside Prep High School."

rtempe = data.iloc[94]

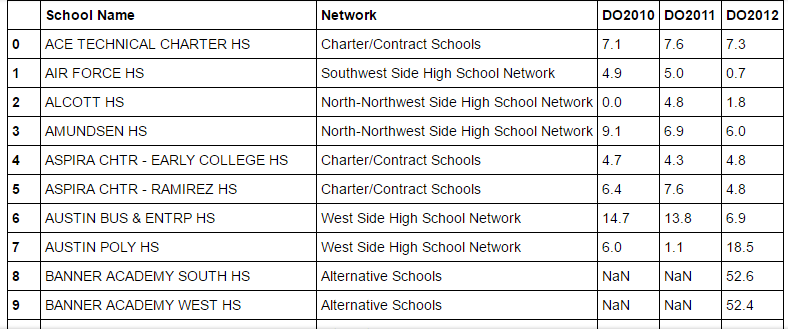
rtempe #5. School with largest passing percentages.

print "The school with highest passing students and percentages is Northside Learning High School."

rtempf = data.iloc[93]

rtempf #6. School with passing percentage and high pass students.

Output



DO2010 1118.1

DO2011 1076.9

DO2012 1293.5

dtype: float64

DO2010 8.102174

DO2011 7.478472

DO2012 8.566225

dtype: float64

DO2010 46.8

DO2011 41.9

DO2012 52.6

dtype: float64

Network of schools with the higher dropout rates is Alternative Schools.

The schools with the highest passing rates are Devry High School, Northside Learning High School, and Northside Learning Prep High School.

DO2010 0.0

DO2011 0.1

DO2012 0.0

dtype: float64

The school with the highest passing rate of students is Northside Learning High School.

Out[26]:

School Name NORTHSIDE PREP HS

Network North-Northwest Side High School Network

DO2010 0.4

DO2011 0.1

DO2012 0.4

Name: 94, dtype: object

The school with highest passing students and percentages is Northside Learning High School.

School Name NORTHSIDE LEARNING HS

Network North-Northwest Side High School Network

DO2010 1

DO2011 1

DO2012 0

Name: 93, dtype: object

Question 3

Source Code

import scipy as sc

import matplotlib.pyplot as plt

import numpy as pd

school= ['amundsen', 'clemente', 'corliss', 'douglass', 'eric solorio academy hs', 'fenger', 'gage park', 'harlan', 'hirsch', 'hubbard', 'juarez', 'kelly']

DO2012=[6, 0.9, 2.3, 0.6, 3, 16.5, 10.6, 10.3, 11.2, 7.4, 5.5, 5.5]

median=[50065, 58987, 40394, 28059, 42809, 40394, 37367, 40394, 40176, 37367, 42575, 42809]

plt.plot(cor\_data['DO2012'], cor\_data['median'], 'bo')

plt.show()

regressionline = sc.stats.linregress(DO2012, median)

m = 0

b = 1

x = np.linspace(START, END, NUMBER\_OF\_POINTS)

plt.plot(x, m\*x + b)

plt.show()

Output

