Question 1

In [1]:
```
import tweepy
consumer_key = "7YSxUteYWORpCV3YvlwVqXvAB"
consumer_secret = "znK2iU94y1kH93SI3NOcDwFY4RyYcbcoIt50seH2PmpGGOzTsE"
access_token = "1411076834012418051-bLma7hywtIq8LnY1wvWZGao4lIKcVf"
access_token_secret = "3fkXkiQZlIIS59mUIgM8xj29orgivWE0TfU1wgVhF1wPe"
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
```

In [2]:
```
import csv
search_tag='#happy'
fname='tweets.csv'
csvFile=open(fname,'a')
csvWriter = csv.writer(csvFile)
try:
    for tweet in tweepy.Cursor(api.search_tweets,search_tag,count=100).items():
        csvWriter.writerow([tweet.id, tweet.lang,tweet.user.screen_name, tweet.text.encode('utf-8')]
except:
    print("Rate exceeded")
csvFile.close()
```

```
Rate exceeded
```

Question 2

In [3]:
```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get("https://archive.ics.uci.edu/ml/datasets.php")
time.sleep(5)
data=driver.find_element(By.XPATH, "/html/body").text
driver.close()
```

In [4]:
```
newdata=data.split()
for i in newdata:
    print(i)
```

```
Center
for
Machine
Learning
and
Intelligent
Systems
About
Citation
Policy
Donate
a
Data
Set
Contact
Repository
Web
View
ALL
```

In [5]:
```
newdata.count("Regression")
```

Out[5]: 116

In [6]:
```
data
```

Out[6]:
```
"         Center for Machine Learning and Intelligent Systems About  Citation Policy  Donate a D
ata Set  Contact\n\n\nRepository Web           \nView ALL Data Sets\n\n×\nCheck out the beta ver
sion of the new UCI Machine Learning Repository we are currently testing! Contact us if you have
any issues, questions, or concerns. Click here to try out the new site.\nBrowse Through:\nDefaul
t Task\nClassification (466)\nRegression (151)\nClustering (121)\nOther (56)\nAttribute Type\nCa
tegorical (38)\nNumerical (422)\nMixed (55)\nData Type\nMultivariate (480)\nUnivariate (30)\nSeq
uential (59)\nTime-Series (126)\nText (69)\nDomain-Theory (23)\nOther (21)\nArea\nLife Sciences
(147)\nPhysical Sciences (57)\nCS / Engineering (234)\nSocial Sciences (41)\nBusiness (45)\nGame
(12)\nOther (81)\n# Attributes\nLess than 10 (166)\n10 to 100 (279)\nGreater than 100 (110)\n# I
nstances\nLess than 100 (38)\n100 to 1000 (210)\nGreater than 1000 (339)\nFormat Type\nMatrix (4
39)\nNon-Matrix (183)\n622 Data Sets\nTable View  List View\nName\nData Types\nDefault Task\nAtt
ribute Types\n# Instances\n# Attributes\nYear\n  Abalone\nMultivariate \nClassification \nCatego
rical, Integer, Real \n4177 \n8 \n1995 \n  Adult\nMultivariate \nClassification \nCategorical, I
nteger \n48842 \n14 \n1996 \n  Annealing\nMultivariate \nClassification \nCategorical, Integer,
Real \n798 \n38 \n    Anonymous Microsoft Web Data\n  Recommender-Systems \nCategorical \n37711
\n294 \n1998 \n  Arrhythmia\nMultivariate \nClassification \nCategorical, Integer, Real \n452 \n
279 \n1998 \n  Artificial Characters\nMultivariate \nClassification \nCategorical, Integer, Real
\n6000 \n7 \n1992 \n  Audiology (Original)\nMultivariate \nClassification \nCategorical \n226 \n
1987 \n  Audiology (Standardized)\nMultivariate \nClassification \nCategorical \n226 \n69 \n1992
```

In [7]:
```
from selenium import webdriver
import chromedriver_autoinstaller
from selenium.webdriver.common.by import By
import time
chromedriver_autoinstaller.install()
driver = webdriver.Chrome()
driver.get("https://archive.ics.uci.edu/ml/datasets.php")
time.sleep(5)
data1=[]
rows=len(driver.find_element(By.XPATH, "/html/body/table[2]/tbody/tr/td[2]/table[2]/tbody/tr").text)
for i in range(1,rows+1):
    data1.append(driver.find_element(By.XPATH, "/html/body/table[2]/tbody/tr/td[2]/table[2]/tbody/tr
# print(data1)
driver.close()
```

In [8]:
```
data2=[]
for i in data1:
    data2.append(i.split("\n"))
fields=data2[0]
data2.sort(key=lambda x:x[len(x)-1],reverse=True)
rows=data2[1:]
```

In [9]:
```
import csv
with open('assignment5', 'w') as f:

    # using csv.writer method from CSV package
    write = csv.writer(f)

    write.writerow(fields)
    write.writerows(rows)
f.close()
```

Question 3

In [10]:
```
# Beautiful Soup is used to pull HTML and XML files from the data.
# It is very easy to work with but not very efficient
# Selenium is used to render web pages and is very quick when it works.
# Its not as easy to work with because of the various drivers it uses.
# It can mimic human behaviour so it can be used to bypass website security
# for bots
# Scrapy is asynchronous and is very hard to work with because of its
# low speed. It has large amount of wait times in it but to bypass it you
# can make a large number of requests in parallel
```

Question 4

In [11]:
```
import webbrowser
webbrowser.open('https://www.indeed.com/?mna=5&aceid=&&aceid=&gclid=Cj0KCQiAn4SeBhCwARIsANeF9DL-X9O-E
```

Out[11]: True