BA PROJECT

ONION ANALYTICS

WEBSCRAPPING OF ONION PRICE DATA

Mexico Government

Change the dates directly in the long row. Careful to choose enough rows per page so it fits in only one and it can be easily webscrapped

In [2]:
```
import requests
from bs4 import BeautifulSoup
import csv

#read website with all rows (almost 5000)
response = requests.get("http://www.economia-sniim.gob.mx/NUEVO/Consultas/MercadosNacionales/PreciosDeMercado/Agricolas/ResultadosConsultaFechaFrutasYHortalizas.aspx?fechaInicio=01/01/2017&fechaFinal=28/01/2017&ProductoId=183&OrigenId=-1&Origen=Todos&DestinoId=-1&Destino=Todos&PreciosPorId=2&RegistrosPorPagina=20000")

#parse with beautiful soup
soup = BeautifulSoup(response.content, 'html.parser')

#include in a list (first row to take:22)
#can also be done i a dictionary or dataframe
l = []
count = 0
for tr in soup.findAll('tr'):
    count += 1
    if (count > 21):
        for td in tr.findAll('td'):
            l.append(td.get_text())

#transform to a proper list of lists
dataset = []
for i in range(1,len(l),8):
    k = [l[i],l[i+2],l[i+3],l[i+4],l[i+5],l[i+6]]
    dataset.append(k)

#titles
titles = ['Date','Origin','Destiny','MinPrice','MaxPrice','MeanPrice']

#write to a csv
with open('DataBAOnions.csv','w') as fp:
    datawriter = csv.writer(fp, delimiter = ',')
    datawriter.writerow(titles)
    for i in range(1,len(dataset)):
        datawriter.writerow(dataset[i])

#import pandas
#my_df = pd.DataFrame(columns=['Date','Origin','Destiny','MinPrice','MeanPrice','MaxPrice'])
#for i in range(1,len(l),8):
#    my_df.loc[i]=[l[i],l[i+2],l[i+3],l[i+4],l[i+5],l[i+6]]
# my_df.to_csv('/Users/RamonRodriganez/Escritorio',sep=',')
```