NHRDF - This is the website of National Horticultural Research & Development Foundation and maintains a database on Market Arrivals and Price, Area and Production and Export Data for three commodities - Garlic, Onion and Potatoes. We are in luck! It also has data from 1996 onwards and has only got one form to fill to get the data in a tabular form. Further it also has production and export data. Excellent. Lets use this. Here is the best link to got to get all that is available - http://nhrdf.org/en-us/DatabaseReports (http://nhrdf.org/en-us/DatabaseReports)

1 Scrapping the Data

So let us fill the form to get data and test our scraping process.(http://nhrdf.org/en-us/MonthWiseMarketArrivals (http://nhrdf.org/en-us/MonthWiseMarketArrivals)).

1. Crop Name: Onion
2. Month: All
3. Market: Chennai
4. Year: All

2 Importing library and loading html file

In [10]:

```
import pandas as pd
```

In [11]:

```
table_1 = pd.read_html('MonthWiseMarketArrivals.html',
                       attrs = {'id' : 'dnn_ctr974_MonthWiseMarketArrivals_GridView1'})
```

In [12]:

```
table_1[0].head(10)
```

Out[12]:

	Market	Month Name	Year	Arrival (q)	Price Minimum (Rs/q)	Price Maximum (Rs/q)	Modal Price (Rs/q)
0	CHENNAI	January	2004	103400	798	1019	910
1	CHENNAI	January	2005	120500	430	638	533
2	CHENNAI	January	2006	111900	428	621	524
3	CHENNAI	January	2007	84800	900	1370	1129
4	CHENNAI	January	2008	127400	588	1000	797
5	CHENNAI	January	2009	111320	1428	2028	1762
6	CHENNAI	January	2010	110000	1639	2259	1980
7	CHENNAI	January	2011	102000	3583	4583	4083
8	CHENNAI	January	2012	126000	771	1013	892
9	CHENNAI	January	2013	116700	1786	2132	1964

3 Creating Data Frame

In [13]:

```
df = table_1[0]
```

In [14]:

```
df.shape
```

Out[14]:

```
(196, 7)
```

In [15]:

```
pd.unique(df['Year'])
```

Out[15]:

```
array(['2004', '2005', '2006', '2007', '2008', '2009', '2010', '2011',
       '2012', '2013', '2014', '2015', '2016', '2017', '2018', '2019',
       '2020', 'Total'], dtype=object)
```

In [16]:

```
df.columns = ['market', 'month', 'year', 'quantity', 'priceMin', 'priceMax', 'priceMod']
```

In [17]:

```
df.head(10)
```

Out[17]:

	market	month	year	quantity	priceMin	priceMax	priceMod
0	CHENNAI	January	2004	103400	798	1019	910
1	CHENNAI	January	2005	120500	430	638	533
2	CHENNAI	January	2006	111900	428	621	524
3	CHENNAI	January	2007	84800	900	1370	1129
4	CHENNAI	January	2008	127400	588	1000	797
5	CHENNAI	January	2009	111320	1428	2028	1762
6	CHENNAI	January	2010	110000	1639	2259	1980
7	CHENNAI	January	2011	102000	3583	4583	4083
8	CHENNAI	January	2012	126000	771	1013	892
9	CHENNAI	January	2013	116700	1786	2132	1964

4 Saving Dataframe to csv file

In [18]:

```
df.to_csv('MonthWiseMarketArrivals_Chennai.csv', index = False)
```