

In [1]: # Import the Packages.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

import sqlite3 as db
from pandasql import sqldf

pysqldf = lambda q: sqldf(q, globals())
```

In [2]: # Read the file from a Git Url

```
df = pd.read_csv('https://raw.githubusercontent.com/jackiekazil/data-wrangling/master/data/chp3/data-text.csv')
df.head(3)
```

Out[2]:

	Indicator	PUBLISH STATES	Year	WHO region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
0	Life expectancy at birth (years)	Published	1990	Europe	High-income	Andorra	Both sexes	77	77.0	NaN	NaN	NaN
1	Life expectancy at birth (years)	Published	2000	Europe	High-income	Andorra	Both sexes	80	80.0	NaN	NaN	NaN
2	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Andorra	Female	28	28.0	NaN	NaN	NaN

```
In [3]: # Read the file from a Git Url  
df2 = pd.read_csv('https://raw.githubusercontent.com/kjam/data-wrangling-pycon/master/data/berlin_weather_oldest.csv')  
df2.head(3)
```

Out[3]:

	STATION	STATION_NAME	DATE	PRCP	SNWD	SNOW	TMAX	TMIN	WDFG	PGTM	...	WT09	WT07	W
0	GHCND:GME00111445	BERLIN TEMPELHOF GM	19310101	46	-9999	-9999	-9999	-11	-9999	-9999	...	-9999	-9999	-9
1	GHCND:GME00111445	BERLIN TEMPELHOF GM	19310102	107	-9999	-9999	50	11	-9999	-9999	...	-9999	-9999	-9
2	GHCND:GME00111445	BERLIN TEMPELHOF GM	19310103	-9999	-9999	-9999	28	11	-9999	-9999	...	-9999	-9999	-9

3 rows × 21 columns



```
In [4]: # 1. Get the Metadata for first Data frame 1  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4656 entries, 0 to 4655  
Data columns (total 12 columns):  
 Indicator           4656 non-null object  
 PUBLISH STATES     4656 non-null object  
 Year                4656 non-null int64  
 WHO region          4656 non-null object  
 World Bank income group 4656 non-null object  
 Country              4656 non-null object  
 Sex                  4656 non-null object  
 Display Value        4656 non-null int64  
 Numeric              4656 non-null float64  
 Low                  0 non-null float64  
 High                 0 non-null float64  
 Comments             0 non-null float64  
 dtypes: float64(4), int64(2), object(6)  
 memory usage: 436.6+ KB
```

```
In [5]: # 1. Get the Metadata for first Data frame 1  
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 117208 entries, 0 to 117207  
Data columns (total 21 columns):  
STATION           117208 non-null object  
STATION_NAME      117208 non-null object  
DATE              117208 non-null int64  
PRCP              117208 non-null int64  
SNWD              117208 non-null int64  
SNOW              117208 non-null int64  
TMAX              117208 non-null int64  
TMIN              117208 non-null int64  
WDFG              117208 non-null int64  
PGTM              117208 non-null int64  
WSFG              117208 non-null int64  
WT09              117208 non-null int64  
WT07              117208 non-null int64  
WT01              117208 non-null int64  
WT06              117208 non-null int64  
WT05              117208 non-null int64  
WT04              117208 non-null int64  
WT16              117208 non-null int64  
WT08              117208 non-null int64  
WT18              117208 non-null int64  
WT03              117208 non-null int64  
dtypes: int64(19), object(2)  
memory usage: 18.8+ MB
```

```
In [6]: # 2. Get the row names from the above files.  
dfRowNames = np.where(df["Indicator"].isnull() != True)  
dfRowNames
```

```
Out[6]: (array([    0,     1,     2, ..., 4653, 4654, 4655], dtype=int64),)
```

```
In [7]: # 2. Get the row names from the above files.  
dfRowNames = np.where(df2["STATION"].isnull() != True)  
dfRowNames
```

```
Out[7]: (array([      0,       1,       2, ..., 117205, 117206, 117207], dtype=int64),)
```

In [8]: # 3 & 4. Change the column name from any of the above file.  
df.rename(columns={'Indicator':'Indicator\_id'}, inplace = True)  
df.head(1)

Out[8]:

	Indicator_id	PUBLISH STATES	Year	WHO region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
0	Life expectancy at birth (years)	Published	1990	Europe	High-income	Andorra	Both sexes	77	77.0	NaN	NaN	NaN

In [9]: # 5.Change the column name from any of the above file and store the changes made permanently.

```
df.rename(columns = {'PUBLISH STATES':'Publication Status' , 'WHO region' : 'WHO Region' }, inplace = True )
df.head(1)
```

Out[9]:

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
0	Life expectancy at birth (years)	Published	1990	Europe	High-income	Andorra	Both sexes	77	77.0	NaN	NaN	NaN

In [10]: #6. Arrange values of a particular column in ascending order  
df.sort\_values(['Year'], ascending=[True])

Out[10]:

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
0	Life expectancy at birth (years)	Published	1990	Europe	High-income	Andorra	Both sexes	77	77.0	NaN	NaN	NaN
1270	Life expectancy at birth (years)	Published	1990	Europe	High-income	Germany	Male	72	72.0	NaN	NaN	NaN
3193	Life expectancy at birth (years)	Published	1990	Europe	Lower-middle-income	Republic of Moldova	Male	65	65.0	NaN	NaN	NaN
3194	Life expectancy at birth (years)	Published	1990	Europe	Lower-middle-income	Republic of Moldova	Both sexes	68	68.0	NaN	NaN	NaN
3197	Life expectancy at age 60 (years)	Published	1990	Europe	Lower-middle-income	Republic of Moldova	Male	15	15.0	NaN	NaN	NaN
1264	Life expectancy at birth (years)	Published	1990	Europe	High-income	Cyprus	Both sexes	76	76.0	NaN	NaN	NaN
3199	Life expectancy at age 60 (years)	Published	1990	Europe	Lower-middle-income	Republic of Moldova	Both sexes	17	17.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
1262	Life expectancy at age 60 (years)	Published	1990	Western Pacific	High-income	Cook Islands	Male	17	17.0	NaN	NaN	NaN
1259	Life expectancy at birth (years)	Published	1990	Western Pacific	High-income	Cook Islands	Male	67	67.0	NaN	NaN	NaN
3203	Life expectancy at age 60 (years)	Published	1990	South-East Asia	Lower-middle-income	Maldives	Female	12	12.0	NaN	NaN	NaN
1273	Life expectancy at age 60 (years)	Published	1990	Europe	High-income	Denmark	Both sexes	20	20.0	NaN	NaN	NaN
3204	Life expectancy at birth (years)	Published	1990	Western Pacific	Lower-middle-income	Marshall Islands	Female	65	65.0	NaN	NaN	NaN
1253	Life expectancy at birth (years)	Published	1990	Western Pacific	High-income	Brunei Darussalam	Both sexes	73	73.0	NaN	NaN	NaN
1247	Life expectancy at age 60 (years)	Published	1990	Americas	High-income	Bahamas	Male	17	17.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
3219	Life expectancy at age 60 (years)	Published	1990	Western Pacific	Lower-middle-income	Vanuatu	Both sexes	16	16.0	NaN	NaN	NaN
3226	Life expectancy at birth (years)	Published	1990	Europe	Upper-middle-income	Bulgaria	Both sexes	71	71.0	NaN	NaN	NaN
1240	Life expectancy at age 60 (years)	Published	1990	Europe	High-income	Belgium	Female	23	23.0	NaN	NaN	NaN
1239	Life expectancy at birth (years)	Published	1990	Europe	High-income	Belgium	Both sexes	76	76.0	NaN	NaN	NaN
1238	Life expectancy at birth (years)	Published	1990	Europe	High-income	Belgium	Female	79	79.0	NaN	NaN	NaN
1237	Life expectancy at birth (years)	Published	1990	Europe	High-income	Austria	Both sexes	76	76.0	NaN	NaN	NaN
1236	Life expectancy at birth (years)	Published	1990	Europe	High-income	Austria	Male	72	72.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
3207	Life expectancy at birth (years)	Published	1990	Western Pacific	Lower-middle-income	Mongolia	Female	64	64.0	NaN	NaN	NaN
3231	Life expectancy at birth (years)	Published	1990	Europe	Upper-middle-income	Belarus	Female	76	76.0	NaN	NaN	NaN
3188	Life expectancy at age 60 (years)	Published	1990	Africa	Lower-middle-income	Lesotho	Female	17	17.0	NaN	NaN	NaN
1277	Life expectancy at birth (years)	Published	1990	Europe	High-income	Estonia	Both sexes	70	70.0	NaN	NaN	NaN
1302	Life expectancy at birth (years)	Published	1990	Europe	High-income	Hungary	Male	65	65.0	NaN	NaN	NaN
3158	Life expectancy at birth (years)	Published	1990	Europe	Lower-middle-income	Georgia	Male	67	67.0	NaN	NaN	NaN
1300	Life expectancy at birth (years)	Published	1990	Europe	High-income	Croatia	Male	69	69.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
3159	Life expectancy at age 60 (years)	Published	1990	Europe	Lower-middle-income	Georgia	Both sexes	19	19.0	NaN	NaN	NaN
3160	Life expectancy at age 60 (years)	Published	1990	Americas	Lower-middle-income	Guatemala	Female	19	19.0	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
3175	Life expectancy at age 60 (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Iran (Islamic Republic of)	Male	19	19.0	NaN	NaN	NaN
3174	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Iran (Islamic Republic of)	Female	76	76.0	NaN	NaN	NaN
1285	Life expectancy at birth (years)	Published	2012	Europe	High-income	France	Both sexes	82	82.0	NaN	NaN	NaN
1286	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	France	Both sexes	25	25.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
3171	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Iran (Islamic Republic of)	Male	72	72.0	NaN	NaN	NaN
1288	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	United Kingdom of Great Britain and Northern I...	Female	25	25.0	NaN	NaN	NaN
1290	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	United Kingdom of Great Britain and Northern I...	Both sexes	24	24.0	NaN	NaN	NaN
1292	Life expectancy at birth (years)	Published	2012	Africa	High-income	Equatorial Guinea	Female	57	57.0	NaN	NaN	NaN
3166	Life expectancy at age 60 (years)	Published	2012	Americas	Lower-middle-income	Honduras	Male	21	21.0	NaN	NaN	NaN
3165	Life expectancy at birth (years)	Published	2012	Americas	Lower-middle-income	Honduras	Both sexes	74	74.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
3163	Life expectancy at age 60 (years)	Published	2012	Americas	Lower-middle-income	Guyana	Male	13	13.0	NaN	NaN	NaN
3162	Life expectancy at birth (years)	Published	2012	Americas	Lower-middle-income	Guyana	Female	67	67.0	NaN	NaN	NaN
1301	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Croatia	Both sexes	21	21.0	NaN	NaN	NaN
3137	Life expectancy at birth (years)	Published	2012	Africa	Lower-middle-income	Cameroon	Male	55	55.0	NaN	NaN	NaN
1303	Life expectancy at birth (years)	Published	2012	Europe	High-income	Hungary	Both sexes	75	75.0	NaN	NaN	NaN
3155	Life expectancy at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Micronesia (Federated States of)	Male	68	68.0	NaN	NaN	NaN
3154	Life expectancy at age 60 (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Egypt	Male	16	16.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
1304	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Hungary	Both sexes	20	20.0	NaN	NaN	NaN
1306	Life expectancy at birth (years)	Published	2012	Europe	High-income	Ireland	Female	83	83.0	NaN	NaN	NaN
3150	Life expectancy at age 60 (years)	Published	2012	Americas	Lower-middle-income	Ecuador	Male	21	21.0	NaN	NaN	NaN
3148	Life expectancy at birth (years)	Published	2012	Americas	Lower-middle-income	Ecuador	Female	78	78.0	NaN	NaN	NaN
3147	Life expectancy at age 60 (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Djibouti	Female	17	17.0	NaN	NaN	NaN
3146	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Djibouti	Both sexes	61	61.0	NaN	NaN	NaN
3145	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	Lower-middle-income	Djibouti	Female	63	63.0	NaN	NaN	NaN

	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Country	Sex	Display Value	Numeric	Low	High	Comments
1309	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Ireland	Female	25	25.0	NaN	NaN	NaN
1316	Life expectancy at birth (years)	Published	2012	Europe	High-income	Italy	Both sexes	83	83.0	NaN	NaN	NaN
3141	Life expectancy at birth (years)	Published	2012	Africa	Lower-middle-income	Cabo Verde	Both sexes	74	74.0	NaN	NaN	NaN
3139	Life expectancy at age 60 (years)	Published	2012	Africa	Lower-middle-income	Cameroon	Female	17	17.0	NaN	NaN	NaN
3156	Life expectancy at age 60 (years)	Published	2012	Western Pacific	Lower-middle-income	Micronesia (Federated States of)	Male	16	16.0	NaN	NaN	NaN
4655	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Low-income	Zimbabwe	Female	51	51.0	NaN	NaN	NaN

4656 rows × 12 columns

In [11]: #7. Arrange multiple column values in ascending order

```
dfcopy = df[['Indicator_id', 'Country','Year','WHO Region', 'Publication Status']].sort_values(['Indicator_id', 'Country', 'Year' ], ascending=[False, True, True])
dfcopy
```

Out[11]:

	Indicator_id	Country	Year	WHO Region	Publication Status
554	Life expectancy at birth (years)	Afghanistan	1990	Eastern Mediterranean	Published
965	Life expectancy at birth (years)	Afghanistan	1990	Eastern Mediterranean	Published
1792	Life expectancy at birth (years)	Afghanistan	1990	Eastern Mediterranean	Published
146	Life expectancy at birth (years)	Afghanistan	2000	Eastern Mediterranean	Published
1393	Life expectancy at birth (years)	Afghanistan	2000	Eastern Mediterranean	Published
2957	Life expectancy at birth (years)	Afghanistan	2000	Eastern Mediterranean	Published
966	Life expectancy at birth (years)	Afghanistan	2012	Eastern Mediterranean	Published
1394	Life expectancy at birth (years)	Afghanistan	2012	Eastern Mediterranean	Published
2958	Life expectancy at birth (years)	Afghanistan	2012	Eastern Mediterranean	Published
299	Life expectancy at birth (years)	Albania	1990	Europe	Published
689	Life expectancy at birth (years)	Albania	1990	Europe	Published
3113	Life expectancy at birth (years)	Albania	1990	Europe	Published
1087	Life expectancy at birth (years)	Albania	2000	Europe	Published
1520	Life expectancy at birth (years)	Albania	2000	Europe	Published
1929	Life expectancy at birth (years)	Albania	2000	Europe	Published
300	Life expectancy at birth (years)	Albania	2012	Europe	Published

	Indicator_id	Country	Year	WHO Region	Publication Status
688	Life expectancy at birth (years)	Albania	2012	Europe	Published
3112	Life expectancy at birth (years)	Albania	2012	Europe	Published
2145	Life expectancy at birth (years)	Algeria	1990	Africa	Published
2510	Life expectancy at birth (years)	Algeria	1990	Africa	Published
4358	Life expectancy at birth (years)	Algeria	1990	Africa	Published
2146	Life expectancy at birth (years)	Algeria	2000	Africa	Published
3968	Life expectancy at birth (years)	Algeria	2000	Africa	Published
3969	Life expectancy at birth (years)	Algeria	2000	Africa	Published
3583	Life expectancy at birth (years)	Algeria	2012	Africa	Published
3584	Life expectancy at birth (years)	Algeria	2012	Africa	Published
4357	Life expectancy at birth (years)	Algeria	2012	Africa	Published
0	Life expectancy at birth (years)	Andorra	1990	Europe	Published
818	Life expectancy at birth (years)	Andorra	1990	Europe	Published
2799	Life expectancy at birth (years)	Andorra	1990	Europe	Published
...	...	...	...	...	...
2417	Healthy life expectancy (HALE) at birth (years)	Venezuela (Bolivarian Republic of)	2000	Americas	Published
2418	Healthy life expectancy (HALE) at birth (years)	Venezuela (Bolivarian Republic of)	2000	Americas	Published
2790	Healthy life expectancy (HALE) at birth (years)	Venezuela (Bolivarian Republic of)	2000	Americas	Published
2419	Healthy life expectancy (HALE) at birth (years)	Venezuela (Bolivarian Republic of)	2012	Americas	Published
4255	Healthy life expectancy (HALE) at birth (years)	Venezuela (Bolivarian Republic of)	2012	Americas	Published

	<b>Indicator_id</b>	<b>Country</b>	<b>Year</b>	<b>WHO Region</b>	<b>Publication Status</b>
4648	Healthy life expectancy (HALE) at birth (years)	Venezuela (Bolivarian Republic of)	2012	Americas	Published
2791	Healthy life expectancy (HALE) at birth (years)	Viet Nam	2000	Western Pacific	Published
3539	Healthy life expectancy (HALE) at birth (years)	Viet Nam	2000	Western Pacific	Published
4256	Healthy life expectancy (HALE) at birth (years)	Viet Nam	2000	Western Pacific	Published
2420	Healthy life expectancy (HALE) at birth (years)	Viet Nam	2012	Western Pacific	Published
2792	Healthy life expectancy (HALE) at birth (years)	Viet Nam	2012	Western Pacific	Published
3880	Healthy life expectancy (HALE) at birth (years)	Viet Nam	2012	Western Pacific	Published
2793	Healthy life expectancy (HALE) at birth (years)	Yemen	2000	Eastern Mediterranean	Published
2794	Healthy life expectancy (HALE) at birth (years)	Yemen	2000	Eastern Mediterranean	Published
3882	Healthy life expectancy (HALE) at birth (years)	Yemen	2000	Eastern Mediterranean	Published
2424	Healthy life expectancy (HALE) at birth (years)	Yemen	2012	Eastern Mediterranean	Published
3883	Healthy life expectancy (HALE) at birth (years)	Yemen	2012	Eastern Mediterranean	Published
4652	Healthy life expectancy (HALE) at birth (years)	Yemen	2012	Eastern Mediterranean	Published
3544	Healthy life expectancy (HALE) at birth (years)	Zambia	2000	Africa	Published

	Indicator_id	Country	Year	WHO Region	Publication Status
3885	Healthy life expectancy (HALE) at birth (years)	Zambia	2000	Africa	Published
4654	Healthy life expectancy (HALE) at birth (years)	Zambia	2000	Africa	Published
2796	Healthy life expectancy (HALE) at birth (years)	Zambia	2012	Africa	Published
3545	Healthy life expectancy (HALE) at birth (years)	Zambia	2012	Africa	Published
4260	Healthy life expectancy (HALE) at birth (years)	Zambia	2012	Africa	Published
2426	Healthy life expectancy (HALE) at birth (years)	Zimbabwe	2000	Africa	Published
2797	Healthy life expectancy (HALE) at birth (years)	Zimbabwe	2000	Africa	Published
3886	Healthy life expectancy (HALE) at birth (years)	Zimbabwe	2000	Africa	Published
3546	Healthy life expectancy (HALE) at birth (years)	Zimbabwe	2012	Africa	Published
4261	Healthy life expectancy (HALE) at birth (years)	Zimbabwe	2012	Africa	Published
4655	Healthy life expectancy (HALE) at birth (years)	Zimbabwe	2012	Africa	Published

4656 rows × 5 columns

In [12]: # 8. Make country as the first column of the dataframe

```
df = df[['Country', 'Indicator_id', 'Publication Status', 'Year', 'WHO Region', 'World Bank income group', 'Sex', 'Display Value', 'Numeric','Low','High','Comments']]
df.head(5)
```

Out[12]:

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
0	Andorra	Life expectancy at birth (years)	Published	1990	Europe	High-income	Both sexes	77	77.0	NaN	NaN	NaN
1	Andorra	Life expectancy at birth (years)	Published	2000	Europe	High-income	Both sexes	80	80.0	NaN	NaN	NaN
2	Andorra	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Female	28	28.0	NaN	NaN	NaN
3	Andorra	Life expectancy at age 60 (years)	Published	2000	Europe	High-income	Both sexes	23	23.0	NaN	NaN	NaN
4	United Arab Emirates	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	High-income	Female	78	78.0	NaN	NaN	NaN

In [13]: #9. Get the column array using a variable

```
dfcolumns = df.loc[df['WHO Region'].isnull() != True,'WHO Region'].values
dfcolumns
```

Out[13]: array(['Europe', 'Europe', 'Europe', ..., 'Africa', 'Africa', 'Africa'],
 dtype=object)

```
In [14]: #10. Get the subset rows 11, 24, 37  
df.loc[[11, 24, 37]]
```

Out[14]:

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
11	Austria	Life expectancy at birth (years)	Published	2012	Europe	High-income	Female	83	83.0	NaN	NaN	NaN
24	Brunei Darussalam	Life expectancy at age 60 (years)	Published	2012	Western Pacific	High-income	Female	21	21.0	NaN	NaN	NaN
37	Cyprus	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Female	26	26.0	NaN	NaN	NaN

```
In [15]: #11. Get the subset rows excluding 5, 12, 23, and 56  
df.drop(df.index[[5,12,23,56]])
```

Out[15]:

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
0	Andorra	Life expectancy at birth (years)	Published	1990	Europe	High-income	Both sexes	77	77.0	NaN	NaN	NaN
1	Andorra	Life expectancy at birth (years)	Published	2000	Europe	High-income	Both sexes	80	80.0	NaN	NaN	NaN
2	Andorra	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Female	28	28.0	NaN	NaN	NaN
3	Andorra	Life expectancy at age 60 (years)	Published	2000	Europe	High-income	Both sexes	23	23.0	NaN	NaN	NaN
4	United Arab Emirates	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	High-income	Female	78	78.0	NaN	NaN	NaN
6	Antigua and Barbuda	Life expectancy at age 60 (years)	Published	1990	Americas	High-income	Male	17	17.0	NaN	NaN	NaN
7	Antigua and Barbuda	Life expectancy at age 60 (years)	Published	2012	Americas	High-income	Both sexes	22	22.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
8	Australia	Life expectancy at birth (years)	Published	2012	Western Pacific	High-income	Male	81	81.0	NaN	NaN	NaN
9	Australia	Life expectancy at birth (years)	Published	2000	Western Pacific	High-income	Both sexes	80	80.0	NaN	NaN	NaN
10	Australia	Life expectancy at birth (years)	Published	2012	Western Pacific	High-income	Both sexes	83	83.0	NaN	NaN	NaN
11	Austria	Life expectancy at birth (years)	Published	2012	Europe	High-income	Female	83	83.0	NaN	NaN	NaN
13	Belgium	Life expectancy at birth (years)	Published	2012	Europe	High-income	Female	83	83.0	NaN	NaN	NaN
14	Bahrain	Life expectancy at birth (years)	Published	2000	Eastern Mediterranean	High-income	Male	73	73.0	NaN	NaN	NaN
15	Bahrain	Life expectancy at birth (years)	Published	1990	Eastern Mediterranean	High-income	Female	74	74.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
16	Bahrain	Life expectancy at age 60 (years)	Published	1990	Eastern Mediterranean	High-income	Male	17	17.0	NaN	NaN	NaN
17	Bahamas	Life expectancy at birth (years)	Published	2012	Americas	High-income	Male	72	72.0	NaN	NaN	NaN
18	Bahamas	Life expectancy at age 60 (years)	Published	2000	Americas	High-income	Both sexes	21	21.0	NaN	NaN	NaN
19	Barbados	Life expectancy at birth (years)	Published	1990	Americas	High-income	Male	71	71.0	NaN	NaN	NaN
20	Barbados	Life expectancy at age 60 (years)	Published	2012	Americas	High-income	Female	25	25.0	NaN	NaN	NaN
21	Barbados	Life expectancy at age 60 (years)	Published	2012	Americas	High-income	Both sexes	23	23.0	NaN	NaN	NaN
22	Brunei Darussalam	Life expectancy at age 60 (years)	Published	1990	Western Pacific	High-income	Female	20	20.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
24	Brunei Darussalam	Life expectancy at age 60 (years)	Published	2012	Western Pacific	High-income	Female	21	21.0	NaN	NaN	NaN
25	Canada	Life expectancy at birth (years)	Published	2000	Americas	High-income	Female	82	82.0	NaN	NaN	NaN
26	Canada	Life expectancy at age 60 (years)	Published	2000	Americas	High-income	Male	21	21.0	NaN	NaN	NaN
27	Canada	Life expectancy at age 60 (years)	Published	1990	Americas	High-income	Female	24	24.0	NaN	NaN	NaN
28	Switzerland	Life expectancy at birth (years)	Published	1990	Europe	High-income	Male	74	74.0	NaN	NaN	NaN
29	Switzerland	Life expectancy at birth (years)	Published	2012	Europe	High-income	Both sexes	83	83.0	NaN	NaN	NaN
30	Switzerland	Life expectancy at age 60 (years)	Published	2000	Europe	High-income	Both sexes	23	23.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
31	Cook Islands	Life expectancy at birth (years)	Published	2012	Western Pacific	High-income	Both sexes	76	76.0	NaN	NaN	NaN
32	Cook Islands	Life expectancy at age 60 (years)	Published	2012	Western Pacific	High-income	Female	22	22.0	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
4626	Serbia	Healthy life expectancy (HALE) at birth (years)	Published	2012	Europe	Upper-middle-income	Female	67	67.0	NaN	NaN	NaN
4627	Suriname	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Both sexes	66	66.0	NaN	NaN	NaN
4628	Sweden	Healthy life expectancy (HALE) at birth (years)	Published	2012	Europe	High-income	Both sexes	72	72.0	NaN	NaN	NaN
4629	Swaziland	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Lower-middle-income	Female	47	47.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4630	Seychelles	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Upper-middle-income	Male	61	61.0	NaN	NaN	NaN
4631	Syrian Arab Republic	Healthy life expectancy (HALE) at birth (years)	Published	2000	Eastern Mediterranean	Lower-middle-income	Female	64	64.0	NaN	NaN	NaN
4632	Chad	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Low-income	Female	44	44.0	NaN	NaN	NaN
4633	Thailand	Healthy life expectancy (HALE) at birth (years)	Published	2000	South-East Asia	Lower-middle-income	Male	59	59.0	NaN	NaN	NaN
4634	Thailand	Healthy life expectancy (HALE) at birth (years)	Published	2000	South-East Asia	Lower-middle-income	Female	65	65.0	NaN	NaN	NaN
4635	Tajikistan	Healthy life expectancy (HALE) at birth (years)	Published	2000	Europe	Low-income	Both sexes	56	56.0	NaN	NaN	NaN
4636	Tajikistan	Healthy life expectancy (HALE) at birth (years)	Published	2012	Europe	Low-income	Female	60	60.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4637	Tonga	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Female	61	61.0	NaN	NaN	NaN
4638	Trinidad and Tobago	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	High-income	Female	64	64.0	NaN	NaN	NaN
4639	Trinidad and Tobago	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	High-income	Both sexes	61	61.0	NaN	NaN	NaN
4640	Tunisia	Healthy life expectancy (HALE) at birth (years)	Published	2000	Eastern Mediterranean	Lower-middle-income	Male	63	63.0	NaN	NaN	NaN
4641	Tuvalu	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Upper-middle-income	Male	57	57.0	NaN	NaN	NaN
4642	Uganda	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Low-income	Female	40	40.0	NaN	NaN	NaN
4643	Ukraine	Healthy life expectancy (HALE) at birth (years)	Published	2000	Europe	Lower-middle-income	Both sexes	60	60.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4644	Uruguay	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Male	65	65.0	NaN	NaN	NaN
4645	Uruguay	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Female	70	70.0	NaN	NaN	NaN
4646	Uruguay	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Both sexes	68	68.0	NaN	NaN	NaN
4647	Saint Vincent and the Grenadines	Healthy life expectancy (HALE) at birth (years)	Published	2000	Americas	Upper-middle-income	Both sexes	61	61.0	NaN	NaN	NaN
4648	Venezuela (Bolivarian Republic of)	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Both sexes	66	66.0	NaN	NaN	NaN
4649	Vanuatu	Healthy life expectancy (HALE) at birth (years)	Published	2000	Western Pacific	Lower-middle-income	Male	59	59.0	NaN	NaN	NaN
4650	Samoa	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Male	62	62.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4651	Samoa	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Female	66	66.0	NaN	NaN	NaN
4652	Yemen	Healthy life expectancy (HALE) at birth (years)	Published	2012	Eastern Mediterranean	Low-income	Both sexes	54	54.0	NaN	NaN	NaN
4653	South Africa	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Upper-middle-income	Male	49	49.0	NaN	NaN	NaN
4654	Zambia	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Low-income	Both sexes	36	36.0	NaN	NaN	NaN
4655	Zimbabwe	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Low-income	Female	51	51.0	NaN	NaN	NaN

4652 rows × 12 columns

In [16]: df

Out[16]:

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
0	Andorra	Life expectancy at birth (years)	Published	1990	Europe	High-income	Both sexes	77	77.0	NaN	NaN	NaN
1	Andorra	Life expectancy at birth (years)	Published	2000	Europe	High-income	Both sexes	80	80.0	NaN	NaN	NaN
2	Andorra	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Female	28	28.0	NaN	NaN	NaN
3	Andorra	Life expectancy at age 60 (years)	Published	2000	Europe	High-income	Both sexes	23	23.0	NaN	NaN	NaN
4	United Arab Emirates	Life expectancy at birth (years)	Published	2012	Eastern Mediterranean	High-income	Female	78	78.0	NaN	NaN	NaN
5	Antigua and Barbuda	Life expectancy at birth (years)	Published	2000	Americas	High-income	Male	72	72.0	NaN	NaN	NaN
6	Antigua and Barbuda	Life expectancy at age 60 (years)	Published	1990	Americas	High-income	Male	17	17.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
7	Antigua and Barbuda	Life expectancy at age 60 (years)	Published	2012	Americas	High-income	Both sexes	22	22.0	NaN	NaN	NaN
8	Australia	Life expectancy at birth (years)	Published	2012	Western Pacific	High-income	Male	81	81.0	NaN	NaN	NaN
9	Australia	Life expectancy at birth (years)	Published	2000	Western Pacific	High-income	Both sexes	80	80.0	NaN	NaN	NaN
10	Australia	Life expectancy at birth (years)	Published	2012	Western Pacific	High-income	Both sexes	83	83.0	NaN	NaN	NaN
11	Austria	Life expectancy at birth (years)	Published	2012	Europe	High-income	Female	83	83.0	NaN	NaN	NaN
12	Austria	Life expectancy at age 60 (years)	Published	2012	Europe	High-income	Female	25	25.0	NaN	NaN	NaN
13	Belgium	Life expectancy at birth (years)	Published	2012	Europe	High-income	Female	83	83.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
14	Bahrain	Life expectancy at birth (years)	Published	2000	Eastern Mediterranean	High-income	Male	73	73.0	NaN	NaN	NaN
15	Bahrain	Life expectancy at birth (years)	Published	1990	Eastern Mediterranean	High-income	Female	74	74.0	NaN	NaN	NaN
16	Bahrain	Life expectancy at age 60 (years)	Published	1990	Eastern Mediterranean	High-income	Male	17	17.0	NaN	NaN	NaN
17	Bahamas	Life expectancy at birth (years)	Published	2012	Americas	High-income	Male	72	72.0	NaN	NaN	NaN
18	Bahamas	Life expectancy at age 60 (years)	Published	2000	Americas	High-income	Both sexes	21	21.0	NaN	NaN	NaN
19	Barbados	Life expectancy at birth (years)	Published	1990	Americas	High-income	Male	71	71.0	NaN	NaN	NaN
20	Barbados	Life expectancy at age 60 (years)	Published	2012	Americas	High-income	Female	25	25.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
21	Barbados	Life expectancy at age 60 (years)	Published	2012	Americas	High-income	Both sexes	23	23.0	NaN	NaN	NaN
22	Brunei Darussalam	Life expectancy at age 60 (years)	Published	1990	Western Pacific	High-income	Female	20	20.0	NaN	NaN	NaN
23	Brunei Darussalam	Life expectancy at age 60 (years)	Published	2000	Western Pacific	High-income	Female	22	22.0	NaN	NaN	NaN
24	Brunei Darussalam	Life expectancy at age 60 (years)	Published	2012	Western Pacific	High-income	Female	21	21.0	NaN	NaN	NaN
25	Canada	Life expectancy at birth (years)	Published	2000	Americas	High-income	Female	82	82.0	NaN	NaN	NaN
26	Canada	Life expectancy at age 60 (years)	Published	2000	Americas	High-income	Male	21	21.0	NaN	NaN	NaN
27	Canada	Life expectancy at age 60 (years)	Published	1990	Americas	High-income	Female	24	24.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
28	Switzerland	Life expectancy at birth (years)	Published	1990	Europe	High-income	Male	74	74.0	NaN	NaN	NaN
29	Switzerland	Life expectancy at birth (years)	Published	2012	Europe	High-income	Both sexes	83	83.0	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
4626	Serbia	Healthy life expectancy (HALE) at birth (years)	Published	2012	Europe	Upper-middle-income	Female	67	67.0	NaN	NaN	NaN
4627	Suriname	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Both sexes	66	66.0	NaN	NaN	NaN
4628	Sweden	Healthy life expectancy (HALE) at birth (years)	Published	2012	Europe	High-income	Both sexes	72	72.0	NaN	NaN	NaN
4629	Swaziland	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Lower-middle-income	Female	47	47.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4630	Seychelles	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Upper-middle-income	Male	61	61.0	NaN	NaN	NaN
4631	Syrian Arab Republic	Healthy life expectancy (HALE) at birth (years)	Published	2000	Eastern Mediterranean	Lower-middle-income	Female	64	64.0	NaN	NaN	NaN
4632	Chad	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Low-income	Female	44	44.0	NaN	NaN	NaN
4633	Thailand	Healthy life expectancy (HALE) at birth (years)	Published	2000	South-East Asia	Lower-middle-income	Male	59	59.0	NaN	NaN	NaN
4634	Thailand	Healthy life expectancy (HALE) at birth (years)	Published	2000	South-East Asia	Lower-middle-income	Female	65	65.0	NaN	NaN	NaN
4635	Tajikistan	Healthy life expectancy (HALE) at birth (years)	Published	2000	Europe	Low-income	Both sexes	56	56.0	NaN	NaN	NaN
4636	Tajikistan	Healthy life expectancy (HALE) at birth (years)	Published	2012	Europe	Low-income	Female	60	60.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4637	Tonga	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Female	61	61.0	NaN	NaN	NaN
4638	Trinidad and Tobago	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	High-income	Female	64	64.0	NaN	NaN	NaN
4639	Trinidad and Tobago	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	High-income	Both sexes	61	61.0	NaN	NaN	NaN
4640	Tunisia	Healthy life expectancy (HALE) at birth (years)	Published	2000	Eastern Mediterranean	Lower-middle-income	Male	63	63.0	NaN	NaN	NaN
4641	Tuvalu	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Upper-middle-income	Male	57	57.0	NaN	NaN	NaN
4642	Uganda	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Low-income	Female	40	40.0	NaN	NaN	NaN
4643	Ukraine	Healthy life expectancy (HALE) at birth (years)	Published	2000	Europe	Lower-middle-income	Both sexes	60	60.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4644	Uruguay	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Male	65	65.0	NaN	NaN	NaN
4645	Uruguay	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Female	70	70.0	NaN	NaN	NaN
4646	Uruguay	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Both sexes	68	68.0	NaN	NaN	NaN
4647	Saint Vincent and the Grenadines	Healthy life expectancy (HALE) at birth (years)	Published	2000	Americas	Upper-middle-income	Both sexes	61	61.0	NaN	NaN	NaN
4648	Venezuela (Bolivarian Republic of)	Healthy life expectancy (HALE) at birth (years)	Published	2012	Americas	Upper-middle-income	Both sexes	66	66.0	NaN	NaN	NaN
4649	Vanuatu	Healthy life expectancy (HALE) at birth (years)	Published	2000	Western Pacific	Lower-middle-income	Male	59	59.0	NaN	NaN	NaN
4650	Samoa	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Male	62	62.0	NaN	NaN	NaN

	Country	Indicator_id	Publication Status	Year	WHO Region	World Bank income group	Sex	Display Value	Numeric	Low	High	Comments
4651	Samoa	Healthy life expectancy (HALE) at birth (years)	Published	2012	Western Pacific	Lower-middle-income	Female	66	66.0	NaN	NaN	NaN
4652	Yemen	Healthy life expectancy (HALE) at birth (years)	Published	2012	Eastern Mediterranean	Low-income	Both sexes	54	54.0	NaN	NaN	NaN
4653	South Africa	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Upper-middle-income	Male	49	49.0	NaN	NaN	NaN
4654	Zambia	Healthy life expectancy (HALE) at birth (years)	Published	2000	Africa	Low-income	Both sexes	36	36.0	NaN	NaN	NaN
4655	Zimbabwe	Healthy life expectancy (HALE) at birth (years)	Published	2012	Africa	Low-income	Female	51	51.0	NaN	NaN	NaN

4656 rows × 12 columns

```
In [17]: # Load datasets from CSV
users = pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/master/Data/users.csv')
```

```
In [18]: sessions = pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/master/Data/sessions.csv')
```

```
In [19]: products = pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/master/Data/products.csv')
```

```
In [20]: transactions = pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/master/Data/transactions.csv')
```

```
In [21]: users.head()
```

Out[21]:

	<b>UserID</b>	<b>User</b>	<b>Gender</b>	<b>Registered</b>	<b>Cancelled</b>
<b>0</b>	1	Charles	male	2012-12-21	NaN
<b>1</b>	2	Pedro	male	2010-08-01	2010-08-08
<b>2</b>	3	Caroline	female	2012-10-23	2016-06-07
<b>3</b>	4	Brielle	female	2013-07-17	NaN
<b>4</b>	5	Benjamin	male	2010-11-25	NaN

```
In [22]: sessions.head()
```

Out[22]:

	<b>SessionID</b>	<b>SessionDate</b>	<b>UserID</b>
<b>0</b>	1	2010-01-05	2
<b>1</b>	2	2010-08-01	2
<b>2</b>	3	2010-11-25	2
<b>3</b>	4	2011-09-21	5
<b>4</b>	5	2011-10-19	4

In [23]: `products.head()`

Out[23]:

	ProductID	Product	Price
0	1	A	14.16
1	2	B	33.04
2	3	C	10.65
3	4	D	10.02
4	5	E	29.66

In [24]: `transactions.head()`

Out[24]:

	TransactionID	TransactionDate	UserID	ProductID	Quantity
0	1	2010-08-21	7.0	2	1
1	2	2011-05-26	3.0	4	1
2	3	2011-06-16	3.0	3	1
3	4	2012-08-26	1.0	2	3
4	5	2013-06-06	2.0	4	1

In [25]: #12. Join users to transactions, keeping all rows from transactions and only matching rows from users(Leftjoin)

```
q = """ SELECT t.*, u.user, u.gender, u.Registered, u.Cancelled FROM transactions t LEFT JOIN users u on t.UserID = u.UserID ; """
```

In [26]:

```
result = pysqldf(q)
result
```

Out[26]:

	TransactionID	TransactionDate	UserID	ProductID	Quantity	User	Gender	Registered	Cancelled
0	1	2010-08-21	7.0	2	1	None	None	None	None
1	2	2011-05-26	3.0	4	1	Caroline	female	2012-10-23	2016-06-07
2	3	2011-06-16	3.0	3	1	Caroline	female	2012-10-23	2016-06-07
3	4	2012-08-26	1.0	2	3	Charles	male	2012-12-21	None
4	5	2013-06-06	2.0	4	1	Pedro	male	2010-08-01	2010-08-08
5	6	2013-12-23	2.0	5	6	Pedro	male	2010-08-01	2010-08-08
6	7	2013-12-30	3.0	4	1	Caroline	female	2012-10-23	2016-06-07
7	8	2014-04-24	NaN	2	3	None	None	None	None
8	9	2015-04-24	7.0	4	3	None	None	None	None
9	10	2016-05-08	3.0	4	4	Caroline	female	2012-10-23	2016-06-07

In [27]: #13. Which transactions have a UserID not in users?

```
q = """ SELECT t.* FROM transactions t LEFT JOIN users u on t.UserID = u.UserID WHERE u.userID is null; """
```

In [28]:

```
result = pysqldf(q)
result
```

Out[28]:

	TransactionID	TransactionDate	UserID	ProductID	Quantity
0	1	2010-08-21	7.0	2	1
1	8	2014-04-24	NaN	2	3
2	9	2015-04-24	7.0	4	3

```
In [29]: #14 Join users to transactions, keeping only rows from transactions and users that match via UserID (inner join)
pd.merge(transactions, users, on='UserID', how='inner')
```

Out[29]:

	TransactionID	TransactionDate	UserID	ProductID	Quantity	User	Gender	Registered	Cancelled
0	2	2011-05-26	3.0	4	1	Caroline	female	2012-10-23	2016-06-07
1	3	2011-06-16	3.0	3	1	Caroline	female	2012-10-23	2016-06-07
2	7	2013-12-30	3.0	4	1	Caroline	female	2012-10-23	2016-06-07
3	10	2016-05-08	3.0	4	4	Caroline	female	2012-10-23	2016-06-07
4	4	2012-08-26	1.0	2	3	Charles	male	2012-12-21	NaN
5	5	2013-06-06	2.0	4	1	Pedro	male	2010-08-01	2010-08-08
6	6	2013-12-23	2.0	5	6	Pedro	male	2010-08-01	2010-08-08

In [30]: #15. Join users to transactions, displaying all matching rows AND all non-matching rows (full outer join) - FULL OUTER JOIN IS NOT SUPPORTED  
pd.merge(transactions, users, on='UserID', how='outer')

Out[30]:

	TransactionID	TransactionDate	UserID	ProductID	Quantity	User	Gender	Registered	Cancelled
0	1.0	2010-08-21	7.0	2.0	1.0	NaN	NaN	NaN	NaN
1	9.0	2015-04-24	7.0	4.0	3.0	NaN	NaN	NaN	NaN
2	2.0	2011-05-26	3.0	4.0	1.0	Caroline	female	2012-10-23	2016-06-07
3	3.0	2011-06-16	3.0	3.0	1.0	Caroline	female	2012-10-23	2016-06-07
4	7.0	2013-12-30	3.0	4.0	1.0	Caroline	female	2012-10-23	2016-06-07
5	10.0	2016-05-08	3.0	4.0	4.0	Caroline	female	2012-10-23	2016-06-07
6	4.0	2012-08-26	1.0	2.0	3.0	Charles	male	2012-12-21	NaN
7	5.0	2013-06-06	2.0	4.0	1.0	Pedro	male	2010-08-01	2010-08-08
8	6.0	2013-12-23	2.0	5.0	6.0	Pedro	male	2010-08-01	2010-08-08
9	8.0	2014-04-24	NaN	2.0	3.0	NaN	NaN	NaN	NaN
10	NaN	NaN	4.0	NaN	NaN	Brielle	female	2013-07-17	NaN
11	NaN	NaN	5.0	NaN	NaN	Benjamin	male	2010-11-25	NaN

In [31]: #16. Determine which sessions occurred on the same day each user registered  
pd.merge(users, sessions, left\_on=['Registered', 'UserID'], right\_on = ['SessionDate', 'UserID'] , how ='inner' )

Out[31]:

	UserID	User	Gender	Registered	Cancelled	SessionID	SessionDate
0	2	Pedro	male	2010-08-01	2010-08-08	2	2010-08-01
1	4	Brielle	female	2013-07-17	NaN	9	2013-07-17

In [32]: #17. Build a dataset with every possible (UserID, ProductID) pair (cross join)  
df1 = pd.DataFrame({'key': np.repeat(1, users.shape[0]), 'UserID': users.UserID})  
df2 = pd.DataFrame({'key': np.repeat(1, products.shape[0]), 'ProductID': products.ProductID})  
pd.merge(df1, df2, on='key')[['UserID', 'ProductID']]

Out[32]:

	UserID	ProductID
0	1	1
1	1	2
2	1	3
3	1	4
4	1	5
5	2	1
6	2	2
7	2	3
8	2	4
9	2	5
10	3	1
11	3	2
12	3	3
13	3	4
14	3	5
15	4	1
16	4	2
17	4	3
18	4	4
19	4	5
20	5	1
21	5	2
22	5	3

	<b>UserID</b>	<b>ProductID</b>
<b>23</b>	5	4
<b>24</b>	5	5

```
In [33]: #18. Determine how much quantity of each product was purchased by each user
df1 = pd.DataFrame({'key': np.repeat(1, users.shape[0]), 'UserID': users.UserID})
df2 = pd.DataFrame({'key': np.repeat(1, products.shape[0]), 'ProductID': products.ProductID})

user_products = pd.merge(df1, df2, on='key')[['UserID', 'ProductID']]

pd.merge(user_products, transactions, how='left', on=['UserID', 'ProductID']).groupby(['UserID', 'ProductID']).apply(
    lambda x: pd.Series(dict(Quantity = x.Quantity.sum() )) ).reset_index().fillna(0)
```

Out[33]:

	UserID	ProductID	Quantity
0	1	1	0.0
1	1	2	3.0
2	1	3	0.0
3	1	4	0.0
4	1	5	0.0
5	2	1	0.0
6	2	2	0.0
7	2	3	0.0
8	2	4	1.0
9	2	5	6.0
10	3	1	0.0
11	3	2	0.0
12	3	3	1.0
13	3	4	6.0
14	3	5	0.0
15	4	1	0.0
16	4	2	0.0
17	4	3	0.0
18	4	4	0.0
19	4	5	0.0
20	5	1	0.0
21	5	2	0.0
22	5	3	0.0

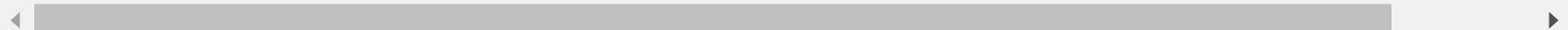
	<b>UserID</b>	<b>ProductID</b>	<b>Quantity</b>
<b>23</b>	5	4	0.0
<b>24</b>	5	5	0.0

In [34]: #19. For each user, get each possible pair of pair transactions (TransactionID1, TransactionID2)  
pd.merge(transactions, transactions, on='UserID')

Out[34]:

	TransactionID_x	TransactionDate_x	UserID	ProductID_x	Quantity_x	TransactionID_y	TransactionDate_y	ProductID_y
0	1	2010-08-21	7.0	2	1	1	2010-08-21	2
1	1	2010-08-21	7.0	2	1	9	2015-04-24	4
2	9	2015-04-24	7.0	4	3	1	2010-08-21	2
3	9	2015-04-24	7.0	4	3	9	2015-04-24	4
4	2	2011-05-26	3.0	4	1	2	2011-05-26	4
5	2	2011-05-26	3.0	4	1	3	2011-06-16	3
6	2	2011-05-26	3.0	4	1	7	2013-12-30	4
7	2	2011-05-26	3.0	4	1	10	2016-05-08	4
8	3	2011-06-16	3.0	3	1	2	2011-05-26	4
9	3	2011-06-16	3.0	3	1	3	2011-06-16	3
10	3	2011-06-16	3.0	3	1	7	2013-12-30	4
11	3	2011-06-16	3.0	3	1	10	2016-05-08	4
12	7	2013-12-30	3.0	4	1	2	2011-05-26	4
13	7	2013-12-30	3.0	4	1	3	2011-06-16	3
14	7	2013-12-30	3.0	4	1	7	2013-12-30	4
15	7	2013-12-30	3.0	4	1	10	2016-05-08	4
16	10	2016-05-08	3.0	4	4	2	2011-05-26	4
17	10	2016-05-08	3.0	4	4	3	2011-06-16	3
18	10	2016-05-08	3.0	4	4	7	2013-12-30	4
19	10	2016-05-08	3.0	4	4	10	2016-05-08	4
20	4	2012-08-26	1.0	2	3	4	2012-08-26	2
21	5	2013-06-06	2.0	4	1	5	2013-06-06	4
22	5	2013-06-06	2.0	4	1	6	2013-12-23	5

	TransactionID_x	TransactionDate_x	UserID	ProductID_x	Quantity_x	TransactionID_y	TransactionDate_y	ProductID_y
23	6	2013-12-23	2.0	5	6	5	2013-06-06	4
24	6	2013-12-23	2.0	5	6	6	2013-12-23	5
25	8	2014-04-24	NaN	2	3	8	2014-04-24	2



```
In [35]: #20. Join each user to his/her first occurring transaction in the transactions table\
first_occuring_transactions = transactions.groupby('UserID').first().reset_index()

pd.merge(users,first_occuring_transactions, how='left', on='UserID')
```

Out[35]:

	UserID	User	Gender	Registered	Cancelled	TransactionID	TransactionDate	ProductID	Quantity
0	1	Charles	male	2012-12-21	NaN	4.0	2012-08-26	2.0	3.0
1	2	Pedro	male	2010-08-01	2010-08-08	5.0	2013-06-06	4.0	1.0
2	3	Caroline	female	2012-10-23	2016-06-07	2.0	2011-05-26	4.0	1.0
3	4	Brielle	female	2013-07-17	NaN	NaN	NaN	NaN	NaN
4	5	Benjamin	male	2010-11-25	NaN	NaN	NaN	NaN	NaN

```
In [36]: #21. Test to see if we can drop columns
```

```
#Display the columns
data = pd.merge(users,first_occuring_transactions, how='left', on='UserID')
my_columns = list(data.columns)
my_columns
```

```
Out[36]: ['UserID',
 'User',
 'Gender',
 'Registered',
 'Cancelled',
 'TransactionID',
 'TransactionDate',
 'ProductID',
 'Quantity']
```

```
In [37]: #Display the columns having na.  
list(data.dropna(thresh=int(data.shape[0] * .9), axis=1).columns) #set threshold to drop NAs
```

```
Out[37]: ['UserID', 'User', 'Gender', 'Registered']
```

```
In [38]: # Display the columns if it has nan data  
missing_info = list(data.columns[data.isnull().any()])  
missing_info
```

```
Out[38]: ['Cancelled', 'TransactionID', 'TransactionDate', 'ProductID', 'Quantity']
```

```
In [39]: # Count of missing data  
for col in missing_info:  
    num_missing = data[data[col].isnull() == True].shape[0]  
    print('number missing for column {}: {}'.format(col, num_missing))
```

```
number missing for column Cancelled: 3  
number missing for column TransactionID: 2  
number missing for column TransactionDate: 2  
number missing for column ProductID: 2  
number missing for column Quantity: 2
```

```
In [40]: #Percentage missing.  
for col in missing_info:  
    percent_missing = data[data[col].isnull() == True].shape[0] / data.shape[0]  
    print('percent missing for column {}: {}'.format(  
        col, percent_missing))
```

```
percent missing for column Cancelled: 0.6  
percent missing for column TransactionID: 0.4  
percent missing for column TransactionDate: 0.4  
percent missing for column ProductID: 0.4  
percent missing for column Quantity: 0.4
```