In [194]:
```
import numpy as np
import pandas as pd
from pandas import DataFrame, Series
import sqlite3 as db
```

In [195]:
```
users=pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/master/D
sessions=pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/maste
products=pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/maste
transactions=pd.read_csv('https://raw.githubusercontent.com/ben519/DataWrangling/r
```

In [206]:
```
"""select u.*,t.TransactionID as TransactionID,MIN(t.TransactionDate) as Transacti
```

Out[206]:

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

In [207]:
```
transactions
```

Out[207]:

	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
5	6	2013-12-23	2.0	5	6
6	7	2013-12-30	3.0	4	1
7	8	2014-04-24	NaN	2	3
8	9	2015-04-24	7.0	4	3
9	10	2016-05-08	3.0	4	4

In [205]:
```
pysqldf = lambda x: sqldf(x, globals())
x="""SELECT UserID,MIN(TransactionDate) FROM transactions group by UserID order by
df=pysqldf(x)
df
```

Out[205]:

	UserID	MIN(TransactionDate)
0	7.0	2010-08-21
1	3.0	2011-05-26
2	1.0	2012-08-26
3	2.0	2013-06-06
4	NaN	2014-04-24

In [6]:
```
!pip install pandasql
```

```
Requirement already satisfied: pandasql in c:\users\mgirm\anaconda3\lib\site-p
ackages (0.7.3)
Requirement already satisfied: numpy in c:\users\mgirm\anaconda3\lib\site-pack
ages (from pandasql) (1.15.1)
Requirement already satisfied: pandas in c:\users\mgirm\anaconda3\lib\site-pac
kages (from pandasql) (0.23.4)
Requirement already satisfied: sqlalchemy in c:\users\mgirm\anaconda3\lib\site
-packages (from pandasql) (1.2.11)
Requirement already satisfied: python-dateutil>=2.5.0 in c:\users\mgirm\anacon
da3\lib\site-packages (from pandas->pandasql) (2.7.3)
Requirement already satisfied: pytz>=2011k in c:\users\mgirm\anaconda3\lib\sit
e-packages (from pandas->pandasql) (2018.5)
Requirement already satisfied: six>=1.5 in c:\users\mgirm\anaconda3\lib\site-p
ackages (from python-dateutil>=2.5.0->pandas->pandasql) (1.11.0)
```

In [7]:
```
!python -m pip install --upgrade pip
```

```
Requirement already up-to-date: pip in c:\users\mgirm\anaconda3\lib\site-packa
ges (18.1)
```

In [197]:
```
from pandasql import *
```

15. Join users to transactions, displaying all matching rows AND all non-matching rows (full outer join)

In [156]:
```
pysqldf = lambda q: sqldf(q, globals())
```

In [157]:
```
q="""select * from transactions t left join users u on t.userid=u.userid;"""
```

In [158]:
```
df=pysqldf(q)
df
```

Out[158]:

	TransactionID	TransactionDate	UserID	ProductID	Quantity	UserID	User	Gender	Registere
0	1	2010-08-21	7.0	2	1	NaN	None	None	Nor
1	2	2011-05-26	3.0	4	1	3.0	Caroline	female	2012-10-2
2	3	2011-06-16	3.0	3	1	3.0	Caroline	female	2012-10-2
3	4	2012-08-26	1.0	2	3	1.0	Charles	male	2012-12-2
4	5	2013-06-06	2.0	4	1	2.0	Pedro	male	2010-08-(
5	6	2013-12-23	2.0	5	6	2.0	Pedro	male	2010-08-(
6	7	2013-12-30	3.0	4	1	3.0	Caroline	female	2012-10-2
7	8	2014-04-24	NaN	2	3	NaN	None	None	Nor
8	9	2015-04-24	7.0	4	3	NaN	None	None	Nor
9	10	2016-05-08	3.0	4	4	3.0	Caroline	female	2012-10-2

16. Determine which sessions occurred on the same day each user registered

In [14]:
```
pysqldf = lambda r: sqldf(r, globals())
```

In [15]:
```
r="""select u.*,s.SessionID,s.SessionDate from users u inner join sessions s on u.
```

In [16]:
```
df=pysqldf(r)
df
```

Out[16]:

	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	None	9	2013-07-17

17. Build a dataset with every possible (UserID, ProductID) pair (cross join)

In [18]:
```
pysqldf = lambda s: sqldf(s, globals())
s="""select u.UserId,p.ProductID from users u cross join products p ;"""
df=pysqldf(s)
df
```

Out[18]:

	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
23	5	4
24	5	5

18.Determine how much quntity of each product was purchased by each user

In [147]:
```
pysqldf = lambda t: sqldf(t, globals())
t="""select u.UserId,t.ProductID,sum(Quantity) as Quantity from users u cross join
"""
df=pysqldf(t)
df
```

Out[147]:

	UserID	ProductID	Quantity
0	1	2	3
1	2	4	1
2	2	5	6
3	3	3	1
4	3	4	6

19. For each user, get each possible pair of pair transactions (TransactionID1,TransacationID2)

In [163]:
```
pysqldf = lambda v: sqldf(v, globals())
v="""SELECT X.TransactionID AS TransactionID_X,X.TransactionDate as TransactionDat
"""
df=pysqldf(v)
df
```

Out[163]:

	TransactionID_X	TransactionDate_X	UserId	ProductID_X	Quantity_X	TransactionID_Y	Transac
0	1	2010-08-21	7.0	2	1	1	
1	1	2010-08-21	7.0	2	1	9	
2	9	2015-04-24	7.0	4	3	1	
3	9	2015-04-24	7.0	4	3	9	
4	2	2011-05-26	3.0	4	1	2	
5	2	2011-05-26	3.0	4	1	3	
6	2	2011-05-26	3.0	4	1	7	
7	2	2011-05-26	3.0	4	1	10	
8	3	2011-06-16	3.0	3	1	2	
9	3	2011-06-16	3.0	3	1	3	
10	3	2011-06-16	3.0	3	1	7	
11	3	2011-06-16	3.0	3	1	10	
12	7	2013-12-30	3.0	4	1	2	
13	7	2013-12-30	3.0	4	1	3	
14	7	2013-12-30	3.0	4	1	7	
15	7	2013-12-30	3.0	4	1	10	
16	10	2016-05-08	3.0	4	4	2	
17	10	2016-05-08	3.0	4	4	3	
18	10	2016-05-08	3.0	4	4	7	
19	10	2016-05-08	3.0	4	4	10	
20	5	2013-06-06	2.0	4	1	5	
21	5	2013-06-06	2.0	4	1	6	
22	6	2013-12-23	2.0	5	6	5	
23	6	2013-12-23	2.0	5	6	6	
24	4	2012-08-26	1.0	2	3	4	

20.Join each user to his/her first occuring transaction in the transactions table

In [223]:
```
pysqldf = lambda w: sqldf(w, globals())
w="""select u.*,t.TransactionID as TransactionID,MIN(t.TransactionDate) as Transac
df=pysqldf(w)
df
```

Out[223]:

	UserID	User	Gender	Registered	Cancelled	TransactionID	TransactionDate	ProductID	Qu
0	2	Pedro	male	2010-08-01	2010-08-08	5.0	2013-06-06	4.0	
1	1	Charles	male	2012-12-21	None	4.0	2012-08-26	2.0	
2	3	Caroline	female	2012-10-23	2016-06-07	2.0	2011-05-26	4.0	
3	4	Brielle	female	2013-07-17	None	NaN	None	NaN	
4	5	Benjamin	male	2010-11-25	None	NaN	None	NaN	

In [ ]: