In [1]:
```
import pandas as pd
```

In [2]:
```
raw_data = pd.read_excel("Adops & Data Scientist Sample Data.xlsx",header=0)
```

In [3]:
```
raw_data.head()
```

Out[3]:

	ts	user_id	country_id	site_id
0	2019-02-01 00:01:24	LC36FC	TL6	N0OTG
1	2019-02-01 00:10:19	LC39B6	TL6	N0OTG
2	2019-02-01 00:21:50	LC3500	TL6	N0OTG
3	2019-02-01 00:22:50	LC374F	TL6	N0OTG
4	2019-02-01 00:23:44	LCC1C3	TL6	QGO3G

In [4]:
```
raw_data.shape
```

Out[4]:
```
(3553, 4)
```

Question 1

In [25]:
```
#Consider only the rows with country_id = "BDV"
BDV = raw_data.loc[raw_data['country_id'] == "BDV"]
BDV.shape
```

Out[25]:
```
(844, 4)
```

In [19]:
```
#unique site ids
site_id_BDV = BDV['site_id'].unique()
```

In [23]:
```
for i in site_id_BDV: # for each site id
    data = BDV.loc[BDV['site_id'] == i] # sub dataframe with this site id
    num = data['user_id'].nunique() # number of unique user id for this site
    print("site_id: " + i + ' --> ' + "number of unique user ids: " + str(nu
```

```
site_id: N0OTG --> number of unique user ids: 90
site_id: 5NPAU --> number of unique user ids: 544
site_id: 3POLC --> number of unique user ids: 2
```

Question 2

In [5]:
```
from datetime import datetime
```

In [28]:
```
raw_data['ts'] = pd.to_datetime(raw_data['ts'])
```

In [29]:
```
raw_data
```

Out[29]:

	ts	user_id	country_id	site_id
0	2019-02-01 00:01:24	LC36FC	TL6	N0OTG
1	2019-02-01 00:10:19	LC39B6	TL6	N0OTG
2	2019-02-01 00:21:50	LC3500	TL6	N0OTG
3	2019-02-01 00:22:50	LC374F	TL6	N0OTG
4	2019-02-01 00:23:44	LCC1C3	TL6	QGO3G
...	...	...	...	...
3548	2019-02-07 23:56:57	LC3F13	TL6	QGO3G
3549	2019-02-07 23:58:36	LC3842	HVQ	3POLC
3550	2019-02-07 23:58:56	LC35EB	TL6	QGO3G
3551	2019-02-07 23:59:19	LC3842	HVQ	3POLC
3552	2019-02-07 23:59:37	LC3842	HVQ	3POLC

3553 rows × 4 columns

In [30]:
```
#Between 2019-02-03 00:00:00 and 2019-02-04 23:59:59
mask = (raw_data['ts'] >= "2019-02-03 00:00:00") & (raw_data['ts'] <= "2019-
```

In [31]:
```
data = raw_data.loc[mask].reset_index(drop=True)
```

In [38]:
```
# unique user ids in this time period
user_id = data['user_id'].unique()
```

In [37]:
```
for i in user_id: # for each unique user id
    sub_df = data.loc[data['user_id'] == i].reset_index(drop=True)
    #sub dataframe for this user id

    unique_site = sub_df['site_id'].unique()
    #unique site ids this user id visited

    for j in unique_site: # for each unique site id this unique user id visi
        number_of_visits = len(sub_df[sub_df['site_id'] == j])
        if number_of_visits > 10: # if the user id visit this site id for mo
            print("user_id: " + i + ' , ' + "site_id: " + j + " , " + "numbe
```

```
user_id: LC3C7E , site_id: 3POLC , number of visits: 15
user_id: LC3A59 , site_id: N0OTG , number of visits: 26
user_id: LC06C3 , site_id: N0OTG , number of visits: 25
user_id: LC3C9D , site_id: N0OTG , number of visits: 17
```

Question 3

In [93]:
```
#unique users
user_id = raw_data['user_id'].unique()
```

In [94]:
```
list = []
```

In [95]:
```
for i in user_id:# for each unique users
    sub_df = raw_data.loc[raw_data['user_id'] == i]
    #sub dataframe of this unique users

    site = sub_df['site_id'].loc[sub_df['ts'] == sub_df.ts.max()]
    #site id this unique user's last visit

    list.append(site.values[0]) #append to the list
```

In [96]:
```
from collections import Counter
Counter(list)
# counts the numbers of occurences for each unique site id in the list
```

Out[96]:
```
Counter({'N0OTG': 561,
         'QGO3G': 289,
         '5NPAU': 992,
         'GVOFK': 42,
         '3POLC': 28,
         'RT9Z6': 2,
         'JSUUP': 1,
         'EUZ/Q': 1})
```

Question 4

In [104]:
```
count = 0
for i in user_id:# for each user
    sub_df = raw_data.loc[raw_data['user_id'] == i]
    #sub dataframe of this user

    first_site = sub_df['site_id'].loc[sub_df['ts'] == sub_df.ts.min()].valu
    last_site = sub_df['site_id'].loc[sub_df['ts'] == sub_df.ts.max()].value
    print("user_id: " + i + " , " + "first site: " + first_site + " , " + "l
    if first_site == last_site:
        count += 1
```

```
user_id: LC36FC , first site: N0OTG , last site: N0OTG
user_id: LC39B6 , first site: N0OTG , last site: N0OTG
user_id: LC3500 , first site: N0OTG , last site: N0OTG
user_id: LC374F , first site: N0OTG , last site: N0OTG
user_id: LCC1C3 , first site: QGO3G , last site: QGO3G
user_id: LC3E1D , first site: GVOFK , last site: 5NPAU
user_id: LC3561 , first site: 3POLC , last site: N0OTG
user_id: LC3A01 , first site: N0OTG , last site: N0OTG
user_id: LC3D80 , first site: N0OTG , last site: N0OTG
user_id: LC3B61 , first site: N0OTG , last site: N0OTG
user_id: LCC3C3 , first site: 5NPAU , last site: 5NPAU
user_id: LC39C8 , first site: QGO3G , last site: QGO3G
user_id: LC3C22 , first site: N0OTG , last site: N0OTG
user_id: LC3DA2 , first site: QGO3G , last site: QGO3G
user_id: LC31E1 , first site: N0OTG , last site: N0OTG
user_id: LC39CA , first site: QGO3G , last site: QGO3G
user_id: LC35FB , first site: N0OTG , last site: N0OTG
user_id: LC3EA8 , first site: N0OTG , last site: 5NPAU
user_id: LC3212 , first site: N0OTG , last site: 5NPAU
```

In [105]:
```
print("Number of users whose first/last visits are to the same website: "+ s
```

```
Number of users whose first/last visits are to the same website: 1670
```

In [ ]:
```

```