Data Wrangling Challenge

August 18, 2020

0.0.1 Challenge I

1. Import bike-share excel file, then put the last_updated column to a readable datetime value.

a. Import excel file

[1]:
```
import pandas as pd
BShare_xls=pd.read_excel('bike-share.xlsx')
BShare_xls.iloc[:3]
```

[1]:
```
   Unnamed: 0  bike_id  is_disabled  is_reserved        lat         lon  \
0           0    12749            0            0  33.431500 -111.942720
1           1    12881            0            0  33.521020 -112.201445
2           2     1356            0            0  33.484417 -112.071637

       name              rec_update
0      T723 2020-07-12 00:01:34.386
1      T782 2020-07-12 00:01:34.736
2  GRID 182 2020-07-12 00:01:38.074
```

[2]:
```
BShare_xls=BShare_xls.loc[:,BShare_xls.columns[1:]]
```

2. Using the iot excel dataset, modify the device column by removing the colon that separate each term (for example 1c:bf:ce:15:ec:4d becomes 1cbfce15ec4d).

a. Import excel file

[3]:
```
iot1 = pd.read_excel('iot.xlsx')
iot1=iot1.loc[:,iot1.columns[1:]]
iot1.head(3)
```

[3]:
```
                    ts             device     co  humidity  light    lpg  \
0 2020-07-12 00:01:34.386  b8:27:eb:bf:9d:51  0.005      51.0  False  0.008
1 2020-07-12 00:01:34.736  00:0f:00:70:91:0a  0.003      76.0  False  0.005
2 2020-07-12 00:01:38.074  b8:27:eb:bf:9d:51  0.005      50.9  False  0.008

   motion  smoke  temp
0   False  0.020  22.7
1   False  0.013  19.7
```

```
2   False  0.020  22.6
```

b. Modify device column

[4]:
```
iot1['new_device']=iot1['device'].apply(lambda x: x.split(':'))
iot1['new_device']=iot1['new_device'].apply(lambda x: ''.join(x))
iot1.head(3)
```

[4]:
```
                        ts             device     co  humidity  light    lpg  \
0 2020-07-12 00:01:34.386  b8:27:eb:bf:9d:51  0.005      51.0  False  0.008
1 2020-07-12 00:01:34.736  00:0f:00:70:91:0a  0.003      76.0  False  0.005
2 2020-07-12 00:01:38.074  b8:27:eb:bf:9d:51  0.005      50.9  False  0.008

   motion  smoke  temp    new_device
0   False  0.020  22.7  b827ebbf9d51
1   False  0.013  19.7  000f0070910a
2   False  0.020  22.6  b827ebbf9d51
```

0.0.2 Challenge II

1. Merge the bike-share.xlsx and iot.xlsx depending on time

[5]:
```
BShare_xls['ts']=BShare_xls.rec_update
BS_iot=pd.merge(BShare_xls,iot1,on='ts',how='inner')
BS_iot.head(3)
```

[5]:
```
   bike_id  is_disabled  is_reserved        lat         lon      name  \
0    12749            0            0  33.431500 -111.942720      T723
1    12881            0            0  33.521020 -112.201445      T782
2     1356            0            0  33.484417 -112.071637  GRID 182

               rec_update                      ts             device     co  \
0 2020-07-12 00:01:34.386 2020-07-12 00:01:34.386  b8:27:eb:bf:9d:51  0.005
1 2020-07-12 00:01:34.736 2020-07-12 00:01:34.736  00:0f:00:70:91:0a  0.003
2 2020-07-12 00:01:38.074 2020-07-12 00:01:38.074  b8:27:eb:bf:9d:51  0.005

   humidity  light    lpg  motion  smoke  temp    new_device
0      51.0  False  0.008   False  0.020  22.7  b827ebbf9d51
1      76.0  False  0.005   False  0.013  19.7  000f0070910a
2      50.9  False  0.008   False  0.020  22.6  b827ebbf9d51
```

2. Create a new dataset which is the subset of the one in 2. where the co value is greater than the mean value of the iot.xlsx dataset.

[6]:
```
import numpy as np
BS_iot_m=BS_iot[BS_iot.co>np.mean(iot1.co)]
```