Technical Support Internship - Blitz Electronics

Project : Predictive Analysis for Policing

Team-1

In [1]:
```
# Importing libraries

import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder
import seaborn as sns
import matplotlib.pyplot
%matplotlib inline
import matplotlib
matplotlib.rcParams['figsize']
```

In [2]:
```
datasetread.read_csv('data.csv')
```

In [3]:
```
dataset.read_csv('data.csv')
```

In [4]:
```
dataset.head()
```

Out[4]:

	timestamp	id01	id02	id03	id04	id05	id06	latitude	longitude
0	28-02-2016 17:00	1	0	0	0	0	0	37.757000	15.287667
1	28-02-2016 21:75	1	0	0	0	0	0	22.733969	75.876563
2	28-02-2016 10:75	1	0	0	0	0	0	22.734675	75.867148
3	28-02-2016 10:75	0	0	0	0	0	0	22.746021	75.897134
4	28-02-2016 10:75	1	0	0	0	0	0	22.746021	75.897134

In [5]:
```
# Set the date
print (type(dataset[[]]))
```

```
<class 'tuple'>
<class 'numpy.int64'>
<class 'numpy.int64'>
<class 'numpy.int64'>
<class 'numpy.int64'>
<class 'numpy.int64'>
<class 'numpy.int64'>
<class 'numpy.float64'>
<class 'numpy.float64'>
```

In [6]:
```
dataf['timestamp'] = pd.to_datetime(dataf['timestamp'], errors='coerce')
```

In [7]:
```
dataf['timestamp'] = pd.to_datetime(dataf['timestamp'], format = '%Y-%m-%d %H:%M:%S')
```

In [8]:
```
dataf['timestamp']
```

Out[8]:
```
0   2016-02-28 21:00:00
1   2016-02-28 21:75:00
2   2016-02-28 10:75:00
3   2016-02-28 10:75:00
4   2016-02-28 10:75:00
    ...
2285  2016-07-24 23:56:00
2286  2016-07-24 23:56:00
2287  2016-07-24 23:56:00
2288  2016-07-24 23:56:00
2289  2016-07-24 23:56:00
Name: timestamp, Length: 2290, dtype: datetime64[ns]
```

In [9]:
```
# DATE TIME FORMAT FUNCTION
dtime_f = dataf.iloc[:,:]
```

In [10]:
```
dtmpd.DataFrame("year", value_f_dt_year,
                "month", value_f_dt_month,
                "day", value_f_dt_day,
                "hour", value_f_dt_hour,
                "dayofyear", value_f_dt_dayofyear,
                "week", value_f_dt_week,
                "weekofyear", value_f_dt_weekofyear,
                "dayofweek", value_f_dt_dayofweek,
                "weekday", value_f_dt_weekday,
                "quarter", value_f_dt_quarter,
                )
```

In [11]:
```
dataset[dataset.drop['timestamp',axis=1]
```

In [12]:
```
dataf[dtmf.concat([dt,dataset],axis=1)]
```

In [13]:
```
dataf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2290 entries, 0 to 2290
Data columns (total 19 columns):
 #   Column     Non-Non-Count  Dtype
---  ------     --------------  -----
 0   id01       2290 non-null   float64
 1   id02       2290 non-null   float64
 2   id03       2290 non-null   float64
 3   id04       2290 non-null   float64
 4   id05       2290 non-null   float64
 5   id06       2290 non-null   float64
 6   dayofyear  2290 non-null   float64
 7   week       2290 non-null   float64
 8   weekofyear 2290 non-null   float64
 9   weekday    2290 non-null   float64
 10  quarter    2290 non-null   float64
 11  id01T      2290 non-null   int64
 12  id02T      2290 non-null   int64
 13  id03T      2290 non-null   int64
 14  id04T      2290 non-null   int64
 15  id05T      2290 non-null   int64
 16  id06T      2290 non-null   int64
 17  latitude   2290 non-null   float64
 18  longitude  2290 non-null   float64
dtypes: float64(11), int64(8)
memory usage: 234.0 KB
```

In [14]:
```
dataf.dropna(inplace=True)
```

In [15]:
```
dataf.isna().sum()
```

```
year          0
month         0
day           0
hour          0
dayofyear     0
week          0
weekofyear    0
dayofweek     0
weekday       0
quarter       0
id01T         0
id02T         0
id03T         0
id04T         0
id05T         0
id06T         0
latitude      0
longitude     0
dtype: int64
```

In [16]:
```
dataf.head()
```

Out[16]:

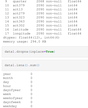
	year	month	day	hour	dayofyear	week	weekofyear	dayofweek	weekday	quarter	id01T	id02T	id03T	id04T	id05T	id06T
0	2016	02	28	21	59	9	9	2	2	1	1	0	0	0	0	0
1	2016	02	28	21	59	9	9	2	2	1	1	0	0	0	0	0
2	2016	02	28	10	59	9	9	2	2	1	1	0	0	0	0	0
3	2016	02	28	10	59	9	9	2	2	1	0	0	0	0	0	0
4	2016	02	28	10	59	9	9	2	2	1	1	0	0	0	0	0

In [17]:
```
sns.boxplot(x='id01', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```

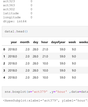
In [18]:
```
sns.boxplot(x='id02', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



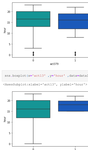
In [19]:
```
sns.boxplot(x='id03', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



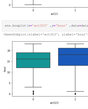
In [20]:
```
sns.boxplot(x='id04', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



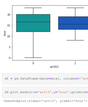
In [21]:
```
sns.boxplot(x='id05', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



In [22]:
```
sns.boxplot(x='id06', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



In [23]:
```
# pd.DataFrame(dataset, columns=['id01', 'hour', 'day'])
```

In [24]:
```
pd.pivot_table(dataset,'id01',index=['hour'],columns=['day'])
```

Out[24]:

day	1	2	3	4	5	6	7	8	9	10	11	12
hour												
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

In [25]:
```
# pd.pivot_table(dataset,'id02',index=['hour'],columns=['day'])
```

In [26]:
```
pd.pivot_table(dataset,'id03',index=['hour'],columns=['day'])
```

Out[26]:

day	1	2	3	4	5	6	7	8	9	10	11	12
hour												
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

In [27]:
```
# pd.pivot_table(dataset,'id04',index=['hour'],columns=['day'])
```

In [28]:
```
pd.pivot_table(dataset,'id05',index=['hour'],columns=['day'])
```

Out[28]:

day	1	2	3	4	5	6	7	8	9	10	11	12
hour												
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

In [29]:
```
# pd.pivot_table(dataset,'id06',index=['hour'],columns=['day'])
```

In [30]:
```
pd.pivot_table(dataset,'id06',index=['hour'],columns=['day'])
```

Out[30]:

day	1	2	3	4	5	6	7	8	9	10	11	12
hour												
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

In [31]:
```
dataset.sample(n=[0,1,2,3,4,5,6,7,8,9,10,11])
```

Out[31]:

	id01	id02	id03	id04	id05	id06	latitude	longitude
1	1	0	0	0	0	0	22.733969	75.876563
2	1	0	0	0	0	0	22.734675	75.867148
3	0	0	0	0	0	0	22.746021	75.897134
4	1	0	0	0	0	0	22.746021	75.897134
5	1	0	0	0	0	0	22.746021	75.897134
6	1	0	0	0	0	0	22.746021	75.897134
7	1	0	0	0	0	0	22.746021	75.897134
8	1	0	0	0	0	0	22.746021	75.897134
9	1	0	0	0	0	0	22.746021	75.897134
10	1	0	0	0	0	0	22.746021	75.897134
11	1	0	0	0	0	0	22.746021	75.897134
12	1	0	0	0	0	0	22.746021	75.897134

In [32]:
```
dataset.sample(n=[0,1,2,3,4,5,6,7,8,9,10,11])
```

Out[32]:

	id01	id02	id03	id04	id05	id06	latitude	longitude
1	1	0	0	0	0	0	22.733969	75.876563
2	1	0	0	0	0	0	22.734675	75.867148
3	0	0	0	0	0	0	22.746021	75.897134
4	1	0	0	0	0	0	22.746021	75.897134
5	1	0	0	0	0	0	22.746021	75.897134
6	1	0	0	0	0	0	22.746021	75.897134
7	1	0	0	0	0	0	22.746021	75.897134
8	1	0	0	0	0	0	22.746021	75.897134
9	1	0	0	0	0	0	22.746021	75.897134
10	1	0	0	0	0	0	22.746021	75.897134
11	1	0	0	0	0	0	22.746021	75.897134
12	1	0	0	0	0	0	22.746021	75.897134

In [33]:
```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2290 entries, 0 to 2290
Data columns (total 19 columns):
 #   Column     Non-Non-Count  Dtype
---  ------     --------------  -----
 0   id01       2290 non-null   float64
 1   id02       2290 non-null   float64
 2   id03       2290 non-null   float64
 3   id04       2290 non-null   float64
 4   id05       2290 non-null   float64
 5   id06       2290 non-null   float64
 6   dayofyear  2290 non-null   float64
 7   week       2290 non-null   float64
 8   weekofyear 2290 non-null   float64
 9   weekday    2290 non-null   float64
 10  quarter    2290 non-null   float64
 11  id01T      2290 non-null   int64
 12  id02T      2290 non-null   int64
 13  id03T      2290 non-null   int64
 14  id04T      2290 non-null   int64
 15  id05T      2290 non-null   int64
 16  id06T      2290 non-null   int64
 17  latitude   2290 non-null   float64
 18  longitude  2290 non-null   float64
dtypes: float64(11), int64(8)
memory usage: 234.0 KB
```

In [34]:
```
dataset.dropna(inplace=True)
```

In [35]:
```
dataset.isna().sum()
```

```
id01       0
id02       0
id03       0
id04       0
id05       0
id06       0
latitude   0
longitude  0
dtype: int64
```

In [36]:
```
dataset.shape
```

```
(2290, 8)
```

In [37]:
```
dataset.hist()
```

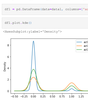
In [38]:
```
dataset.corr()
```

Out[38]:

	id01	id02	id03	id04	id05	id06	latitude	longitude
id01	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
id02	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
id03	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
id04	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
id05	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000
id06	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000
latitude	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
longitude	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000

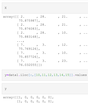
In [39]:
```
sns.boxplot(x='id01', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



In [40]:
```
sns.boxplot(x='id02', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```



In [41]:
```
sns.boxplot(x='id03', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```

In [42]:
```
sns.boxplot(x='id04', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```

In [43]:
```
sns.boxplot(x='id05', y='hour', data=dataset, palette='Paired_r')
```

```
<matplotlib.axes._subplots.AxesSubplot>
```

In [44]: