Datathon

Team name : Tokyo Banana

Soma Suzuki - h2so4cos0@gmail.com & George Reed - reedgeo@hotmail.com

It was found that the data contained a large number of cancelled transactions. For this reason, it was decided that the cancelled transaction data should be analysed in order to determine the motivation for cancellation. Included at the conclusion of this report are various suggestions as to how users might be discouraged from cancelling their bets.

In [5]:

```python
#import module
import csv
import pandas as pd
import matplotlib.pyplot as plt
import math
from matplotlib import style
import datetime
import seaborn as sns
style.use('ggplot')
%matplotlib inline
```

In [15]:

```python
#Using Excel, We refined the data to include only cancelled.  unnecessary data
has also been removed for the purpose of analysis.
df_cancelled = pd.read_csv('output.csv')
```

In [14]:

```python
#Here, we check the data.
df_cancelled
```

Out[14]:

	"BET_ID"	BET_TRANS_ID	ACCOUNT_ID	COUNTRY_OF_RESIDENCE_N
0	4574469576	5698291880	1016144	India
1	4755525134	5899236919	1016903	United Kingdom
2	4709776106	5846031545	1016268	India
3	4753227895	5896451797	1017070	Seychelles

4	4755437878	5899050541	1016903	United Kingdom
5	4755518907	5899245753	1016903	United Kingdom
6	4755534182	5899245751	1016903	United Kingdom
7	4755525769	5899245752	1016903	United Kingdom
8	4755445845	5899056208	1016903	United Kingdom
9	4755449679	5899068356	1016903	United Kingdom
10	4755819194	5899491444	1016924	UAE
11	4753235544	5896468210	1017543	Dominica
12	4755895791	5899578656	1017543	Dominica
13	4753816087	5897127421	1001376	Seychelles
14	4753816156	5897127422	1001376	Seychelles
15	4753226587	5896449980	1017857	Malta
16	4754244195	5897628092	1016903	United Kingdom
17	4755477622	5899245754	1016903	United Kingdom
18	4755332670	5898931160	1016903	United Kingdom
19	4756051018	5899763631	1017543	Dominica
20	4754254575	5897639336	1016903	United Kingdom
21	4753147463	5896457229	1016924	UAE
22	4740657470	5881776121	1016924	UAE
23	4754550302	5899554826	1017543	Dominica

24	4753401705	5897129448	1006651	Malta
25	4755947759	5899633066	1017543	Dominica
26	4754207209	5897587047	1016903	United Kingdom
27	4756059861	5899978311	1016903	United Kingdom
28	4733768644	5873816880	1017328	Seychelles
29	4574469577	5698291881	1016144	India
...	...	...	...	...
570075	4859142304	6020930122	1003421	UAE
570076	4858596933	6020242132	1003421	UAE
570077	4858701689	6020178994	1007186	Malta
570078	4858706700	6020233330	1007814	Malta
570079	4858784878	6020283231	1008924	UAE
570080	4858641287	6020097196	1017328	Seychelles
570081	4857443573	6018691657	1021906	Malta
570082	4858448490	6021198687	1017709	India
570083	4858421964	6019956710	1017692	Seychelles
570084	4857268400	6018482851	1002140	India
570085	4857918170	6019316592	1018033	United Kingdom
570086	4857442419	6018688709	1018033	United Kingdom
570087	4857763778	6019074409	1018033	United Kingdom

570088	4857442643	6018693827	1018033	United Kingdom
570089	4857953136	6019308252	1005229	Virgin Islands, British
570090	4858355706	6019765225	1002437	India
570091	4857443831	6018690193	1006367	Virgin Islands, British
570092	4857443233	6018700605	1015374	Seychelles
570093	4857769194	6019513412	1006775	Virgin Islands, British
570094	4857246536	6018462571	1006045	Virgin Islands, British
570095	4857587834	6018864519	1006045	Virgin Islands, British
570096	4857629450	6018913627	1006045	Virgin Islands, British
570097	4858285942	6019683992	1006045	Virgin Islands, British
570098	4857778140	6019380072	1005326	Malta
570099	4857539488	6018808852	1017933	Malta
570100	4857783170	6019624079	1000604	Malta
570101	4857788506	6019103727	1007152	Seychelles
570102	4858103604	6019567383	1017649	UAE
570103	4858474165	6019936167	1005999	Seychelles
570104	4857443264	6018688109	1003629	Virgin Islands, British

570105 rows × 19 columns

In [7]:

```
#Formatting data.
date1 = pd.to_datetime(df_cancelled['OFF_DT'])
date2 = pd.to_datetime(df_cancelled['CANCELLED_DATE'])
date1.dt
date2.dt
#Caluculation. Subtracting 'OFF_DT' from 'CANCELLED_DATE' to determine when us
ers cancelled.
x = ((date1.dt.month * 24 *30) + (date1.dt.day * 24) + (date1.dt.hour) + (date
1.dt.minute /60) + (date1.dt.second / 3600)) -((date2.dt.month * 24 *30) + (da
te2.dt.day * 24) + (date2.dt.hour) + (date2.dt.minute /60) + (date2.dt.second
/ 3600))
```
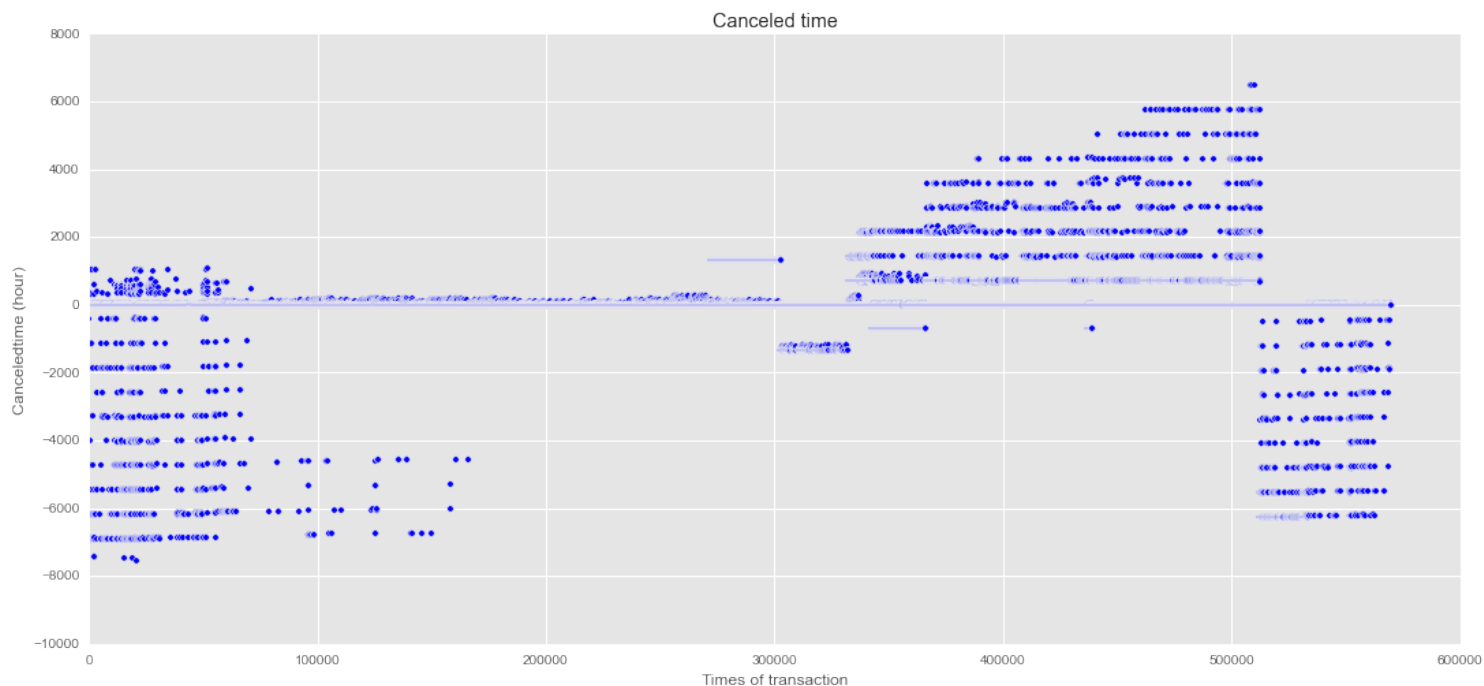
Out[7]:

```
<pandas.tseries.common.DatetimeProperties object at 0x1182878d0>
```

In [9]:

```
#Plot when users cancelled transaction.
y = x.index.tolist()
z = x.tolist()
plt.figure(figsize=(18,8))

plt.title("Canceled time")
plt.xlabel("Times of transaction")
plt.ylabel("Canceledtime (hour)")
plt.xlim(0, 600000)
plt.scatter(y,z)
plt.show()
```
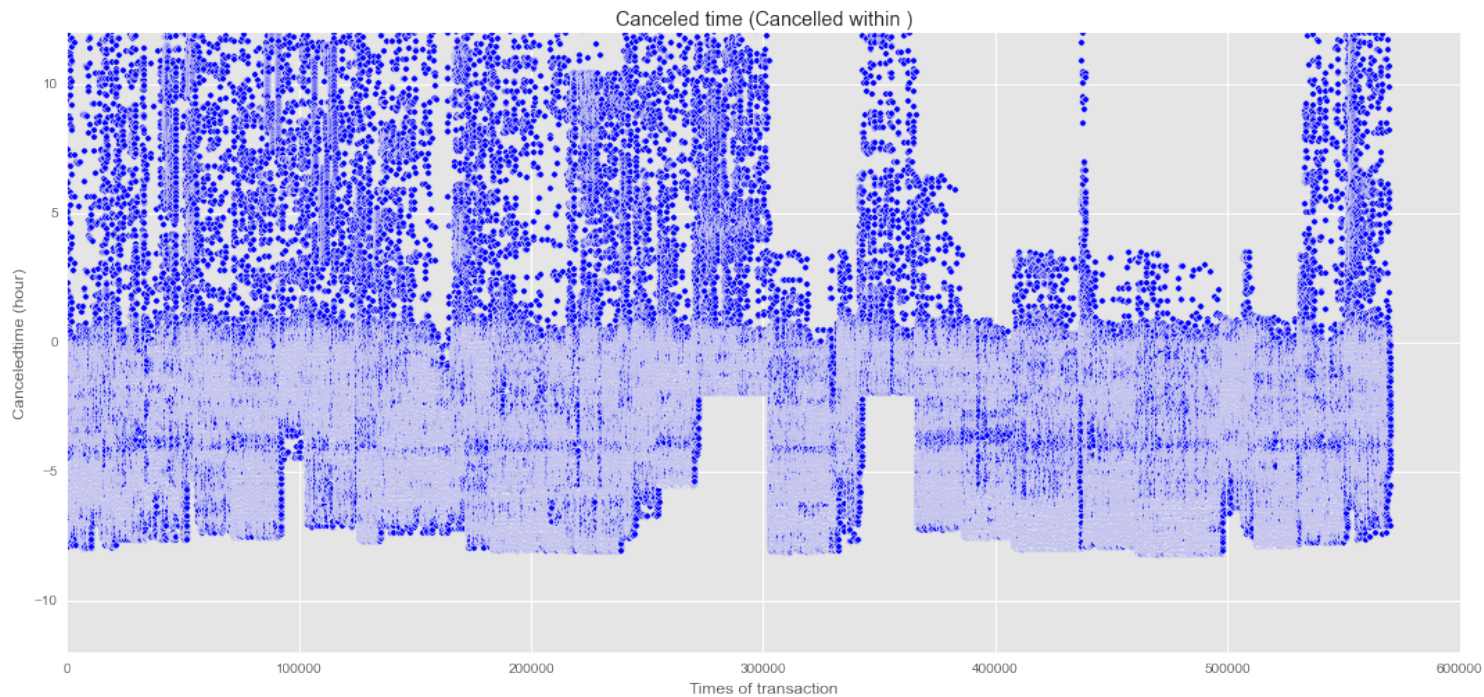

The issue of users cancelling their bet prior to the start of the event cannot be helped as this is likely due to reasons beyond Betfair's control. We have formed our prediction based upon the assumption that those who cancelled two or three days post off date were unable to be matched. Therefore, as one-day cricket matches last for six hours or more; it was decided to plot this data where transactions were cancelled only 12 (6 X2) hours prior to a game, and 12 hours post game.

In [10]:

```python
#Plot when users cancelled transaction. (Within 12 hours.)
y = x.index.tolist()
z = x.tolist()
plt.figure(figsize=(18,8))

plt.title("Canceled time (Cancelled within )")
plt.xlabel("Times of transaction")
plt.ylabel("Canceledtime (hour)")
plt.xlim(0, 600000)
plt.ylim(-12, 12)
plt.scatter(y,z)
plt.show()
```


Canceled time (Cancelled within)

Conclusion

As can be seen the last plot. A large number of users cancelled their transaction just prior to the match start time. As users cannot predict the play by play outcome of the game, their cancellations can generally be considered baseless. This may however, be negated by the introduction of a chance based system, thereby, encouraging users to not cancel. This could be implemented via regular promotions, allowing users a (one in fifty, one in one hundred etc.) chance to have a losing bet negated. This may encourage users to keep their bets live throughout the game, rather than cancelling before match start. This promotion could be made live in the twelve hours prior to the match start time, as this is when the vast majority of users cancel.