

# Name of the Candidate: Lalit Kumar Paila

Clootrack Data analysis Assessment

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [2]: auto=pd.read_csv('Data Analyst - Test Data - US.csv')
```

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

Out[3]:

		Review	date	Location
0	I was very impressed with the resort.\n Great ...		2019/08/20	Sebastian
1	The rooms were nice the outside needs work als...		2019/08/20	Los Angeles
2	Great location! I have stayed at this hotel on...		2019/08/20	Georgia
3	The hotel was adequate for my stay. The strips...		2019/08/20	NaN
4	Great location, room was large and spacious. P...		2019/08/19	Palm Harbor

```
In [4]: auto.tail()
```

Out[4]:

		Review	date	Location
6443	·the staff were all so friendly. frank, the ch...		2019/07/12	United States of America
6444	·the internet barely worked or didn't work		2019/07/12	United States of America
6445	·the location was great.\n \n \n \n \n ·the be...		2019/07/10	United States of America
6446	·we were early for check in and the staff was ...		2019/07/10	United States of America
6447	·staff, free coffee, location and activities.\n...		2019/07/10	United States of America

```
In [15]: auto.shape # To Know how many columns and rows are there
```

Out[15]: (6448, 3)

```
In [16]: auto.describe() # To know the details about the data
```

Out[16]:

		Review	date	Location
count		6393	6448	1711
unique		6384	403	1082
top	Amazing\n \n Read more\n Read less	Jun 2019		United States of America
freq		4	145	116

```
In [17]: auto.isnull().sum() # To sum the values
```

```
Out[17]: Review      55
         date        0
         Location    4737
         dtype: int64
```

```
In [18]: auto.drop_duplicates() # To remove duplicates
```

```
Out[18]:
```

		Review	date	Location
0	I was very impressed with the resort.\n Great ...		2019/08/20	Sebastian
1	The rooms were nice the outside needs work als...		2019/08/20	Los Angeles
2	Great location! I have stayed at this hotel on...		2019/08/20	Georgia
3	The hotel was adequate for my stay. The strips...		2019/08/20	NaN
4	Great location, room was large and spacious. P...		2019/08/19	Palm Harbor
...		...	...	...
6443	·the staff were all so friendly. frank, the ch...		2019/07/12	United States of America
6444	·the internet barely worked or didn't work		2019/07/12	United States of America
6445	·the location was great.\n \n \n \n \n ·the be...		2019/07/10	United States of America
6446	·we were early for check in and the staff was ...		2019/07/10	United States of America
6447	·staff, free coffee, location and activities.\n...		2019/07/10	United States of America

6444 rows × 3 columns

```
In [19]: auto.nunique() # to know the unique values
```

```
Out[19]: Review      6384
         date        403
         Location    1082
         dtype: int64
```

```
In [20]: auto.columns # to know column headings
```

```
Out[20]: Index(['Review', 'date', 'Location'], dtype='object')
```

```
In [11]: #Checking for wrong entries like symbols -,?,#,*,etc.  
for col in auto.columns:  
    print('{} : {}'.format(col,auto[col].unique()))
```

Review : ['I was very impressed with the resort.\n Great staff at the main resort pool bar! We had a blast with them.\n Clean, professional staff, great location and very reasonable!\n \n Read more\n Read less'

'The rooms were nice the outside needs work also no free breakfast it would have been nice overall it was ok\n \n Read more\n Read less'

'Great location! I have stayed at this hotel on my last three trips to New York.\n \n Read more\n Read less'

...  
'...the location was great.\n \n \n \n \n \n .the bed was the most uncomfortable bed i have ever been in.'

'...we were early for check in and the staff was very accommodating. we were able to check in several hours before the check in time.'

'...staff, free coffee, location and activities.\n \n \n \n \n \n .the beds were smaller than we thought and not very comfortable.']

date : ['2019/08/20' '2019/08/19' '2019/08/18' '2019/08/17' '2019/08/16'

'2019/08/15' '2019/08/14' '2019/08/13' '2019/08/12' '2019/08/11'

'2019/08/10' '2019/08/09' '2019/08/08' '2019/08/07' '2019/08/06'

'2019/08/05' '2019/08/04' '2019/08/03' '2019/08/02' '2019/08/01'

'2019/07/31' '2019/07/30' '2019/07/29' '2019/07/28' '2019/07/27'

'2019/07/26' '2019/07/25' '2019/07/24' '2019/07/23' '2019/07/22'

'2019/07/21' '2019/07/20' '2019/07/19' '2019/07/18' '2019/07/17'

'2019/07/16' '2019/07/15' '2019/07/14' '2019/07/13' '2019/07/12'

'2019/07/11' '2019/07/10' '2019/07/09' '2019/07/08' '2019/07/07'

'2019/07/06' '2019/07/05' '2019/07/04' '2019/07/03' '2019/07/02'

'2019/07/01' '2019/06/30' '2019/06/29' '2019/06/28' '2019/06/27'

'2019/06/26' '2019/06/25' '2019/06/24' '2019/06/23' '2019/06/22'

'2019/06/21' '2019/06/20' '2019/06/19' '2019/06/18' '2019/06/17'

'2019/06/16' '2019/06/15' '2019/06/14' '2019/06/13' '2019/06/12'

'2019/06/11' '2019/06/10' '2019/06/09' '2019/06/08' '2019/06/07'

'2019/06/06' '2019/06/05' '2019/06/04' '2019/06/03' '2019/06/02'

'2019/06/01' '2019/05/31' '2019/05/30' '2019/05/29' '2019/05/28'

'2019/05/27' '2019/05/26' '2019/05/25' '2019/05/24' '2019/05/23'

'2019/05/22' '2019/05/21' '2019/05/20' '2019/05/19' '2019/05/18'

'2019/05/17' '2019/05/16' '2019/05/15' '2019/05/14' '2019/05/13'

'2019/05/12' '2019/05/11' '2019/05/10' '2019/05/09' '2019/05/08'

'2019/05/07' '2019/05/06' '2019/05/05' '2019/05/04' '2019/05/03'

'2019/05/02' '2019/05/01' '2019/04/30' '2019/04/29' '2019/04/28'

'2019/04/27' '2019/04/26' '2019/04/25' '2019/04/24' '2019/04/23'

'2019/04/22' '2019/04/21' '2019/04/20' '2019/04/19' '2019/04/18'

'2019/04/17' '2019/04/16' '2019/04/15' '2019/04/14' '2019/04/13'

'2019/04/12' '2019/04/11' '2019/04/10' '2019/04/09' '2019/04/08'

'2019/04/07' '2019/04/06' '2019/04/05' '2019/04/04' '2019/04/03'

'2019/04/02' '2019/04/01' '2019/03/31' '2019/03/30' '2019/03/29'

'2019/03/28' '2019/03/27' '2019/03/26' '2019/03/25' '2019/03/24'

'2019/03/23' '2019/03/22' '2019/03/21' '2019/03/20' '2019/03/19'

'2019/03/18' '2019/03/17' '2019/03/16' '2019/03/15' '2019/03/14'

'2019/03/13' '2019/03/12' '2019/03/11' '2019/03/10' '2019/03/09'

'2019/03/08' '2019/03/07' '2019/03/06' '2019/03/05' '2019/03/04'

'2019/3/4' '2019/3/3' '2019/3/2' '2019/3/1' '2019/2/28' '2019/2/27'

'2019/2/26' '2019/2/25' '2019/2/24' '2019/2/23' '2019/2/22' '2019/2/21'

'2019/2/20' '2019/2/19' '2019/2/18' '2019/2/17' '2019/2/16' '2019/2/15'

'2019/2/14' '2019/2/13' '2019/2/12' '2019/2/11' '2019/2/10' '2019/2/9'

'2019/2/8' '2019/2/7' '2019/2/6' '2019/2/5' '2019/2/4' '2019/2/3'

'2019/2/2' '2019/2/1' '2019/1/31' '2019/1/30' '2019/01/30' '2019/01/29'

'2019/01/28' '2019/01/27' '2019/01/26' '2019/01/25' '2019/01/24'

'2019/01/23' '2019/01/22' '2019/1/22' '2019/1/21' '2019/1/20' '2019/1/19'

'2019/1/18' '2019/1/17' '2019/1/16' '2019/1/15' '2019/1/14' '2019/1/13'

'2019/1/12' '2019/1/11' '2019/1/10' '2019/1/9' '2019/1/8' '2019/1/7'

'2019/1/6' '2019/1/5' '2019/1/4' '2019/1/3' '2019/01/03' '2019/01/02'

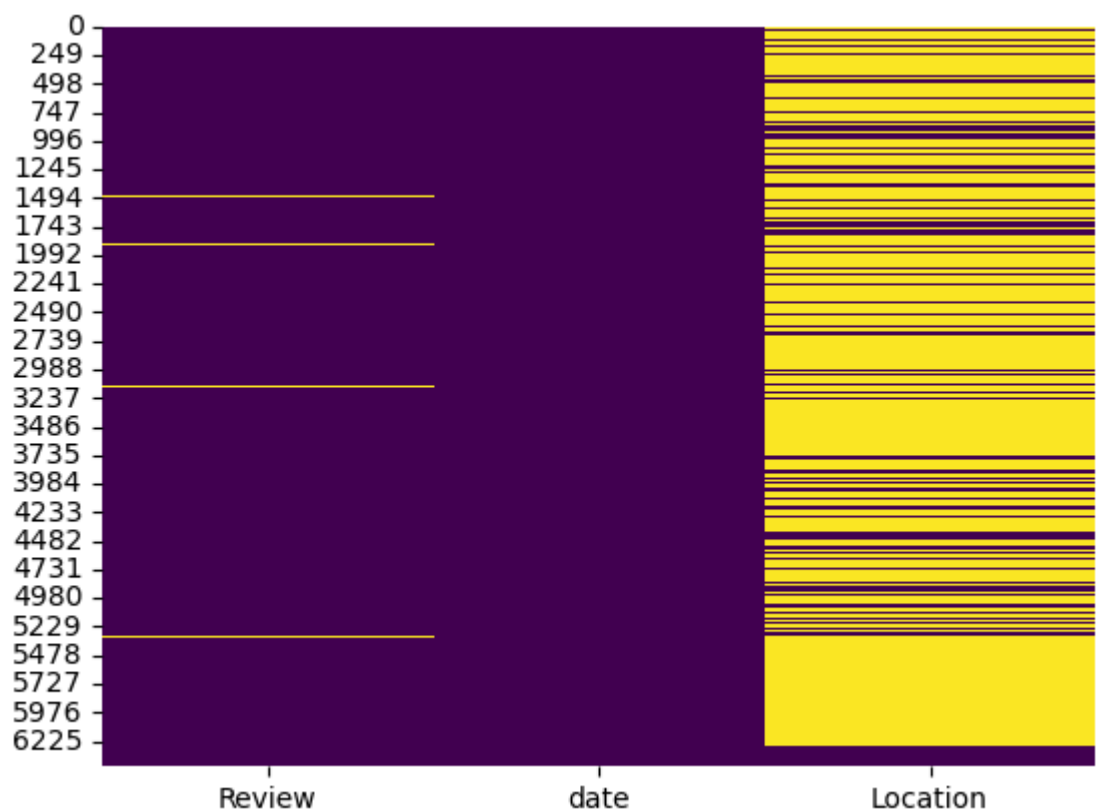
'2019/01/01' '2018/12/31' '2018/12/30' '2018/12/29' '2018/12/28'

'2018/12/27' '2018/12/26' '2018/12/25' '2018/12/24' '2018/12/23'

```
'2018/12/22' '2018/12/21' '2018/12/20' '2018/12/19' '2018/12/18'
'2018/12/17' '2018/12/16' '2018/12/15' '2018/12/14' '2018/12/13'
'2018/12/12' '2018/12/11' '2018/12/10' '2018/12/09' '2018/12/08'
'2018/12/07' '2018/12/06' '2018/12/05' '2018/12/04' '2018/12/03'
'2018/12/02' '2018/12/01' '2018/11/30' '2018/11/29' '2018/11/28'
'2018/11/27' '2018/11/26' '2018/11/25' '2018/11/24' '2018/11/23'
'2018/11/22' '2018/11/21' '2018/11/20' '2018/11/19' '2018/11/18'
'2018/11/17' '2018/11/16' '2018/11/15' '2018/11/14' '2018/11/13'
'2018/11/12' '2018/11/11' '2018/11/10' '2018/11/09' '2018/11/08'
'2018/11/07' '2018/11/06' '2018/11/05' '2018/11/04' '2018/11/03'
'2018/11/02' '2018/11/01' '2018/10/31' '2018/10/30' '2018/10/29'
'2018/10/28' '2018/10/27' '2018/10/26' '2018/10/25' '2018/10/24'
'2018/10/23' '2018/10/22' '2018/10/21' '2018/10/20' '2018/10/19'
'2018/10/18' '2018/10/17' '2018/10/16' '2018/10/15' '2018/10/14'
'2018/10/13' '2018/10/12' '2018/10/11' '2018/10/10' '2018/10/09'
'2018/10/08' '2018/10/07' '2018/10/06' '2018/10/05' '2018/10/04'
'2018/10/03' '2018/10/02' '2018/10/01' '2018/09/30' '2018/09/29'
'2018/09/28' '2018/09/27' '2018/09/26' '2018/09/25' '2018/09/24'
'2018/09/23' '2018/09/22' '2018/09/21' '2018/09/20' '2018/09/19'
'2018/09/18' '2018/09/17' '2018/09/16' '2018/09/15' '2018/09/14'
'2018/09/13' '2018/09/12' '2018/09/11' '2018/09/10' '2018/09/09'
'2018/09/08' '2018/09/07' '2018/09/06' '2018/09/05' '2018/09/04'
'2018/09/03' '2018/09/02' '2018/09/01' '2018/08/31' '2018/08/30'
'2018/08/29' '2018/08/28' '2018/08/27' '2018/08/26' '2018/08/25'
'2018/08/24' '2018/08/23' '2018/08/22' '2018/08/21' '2018/08/20'
'2018/08/19' '2018/08/18' '2018/08/17' '2018/08/16' '2018/08/15'
'2018/08/14' '2018/08/13' '2018/08/12' '2018/08/11' '2018/08/10'
'2018/08/09' '2018/08/08' '2018/08/07' '2018/08/06' '2018/08/05'
'2018/08/04' '2018/08/03' '2018/08/02' '2018/08/01' '2019/02/01'
'2019/03/01' 'Dec 2018' 'Aug 2018' 'Jul 2019' 'Sep 2018' 'Oct 2018'
'Jan 2019' 'Feb 2019' 'Apr 2019' 'Mar 2019' 'May 2019' 'Jun 2019'
'Nov 2018']
Location : ['Sebastian' 'Los Angeles' 'Georgia' ... 'India' 'Malaysia'
'Trinidad and Tobago']
```

```
In [12]: sns.heatmap(auto.isnull(),cbar=False,cmap='viridis')
```

```
Out[12]: <Axes: >
```

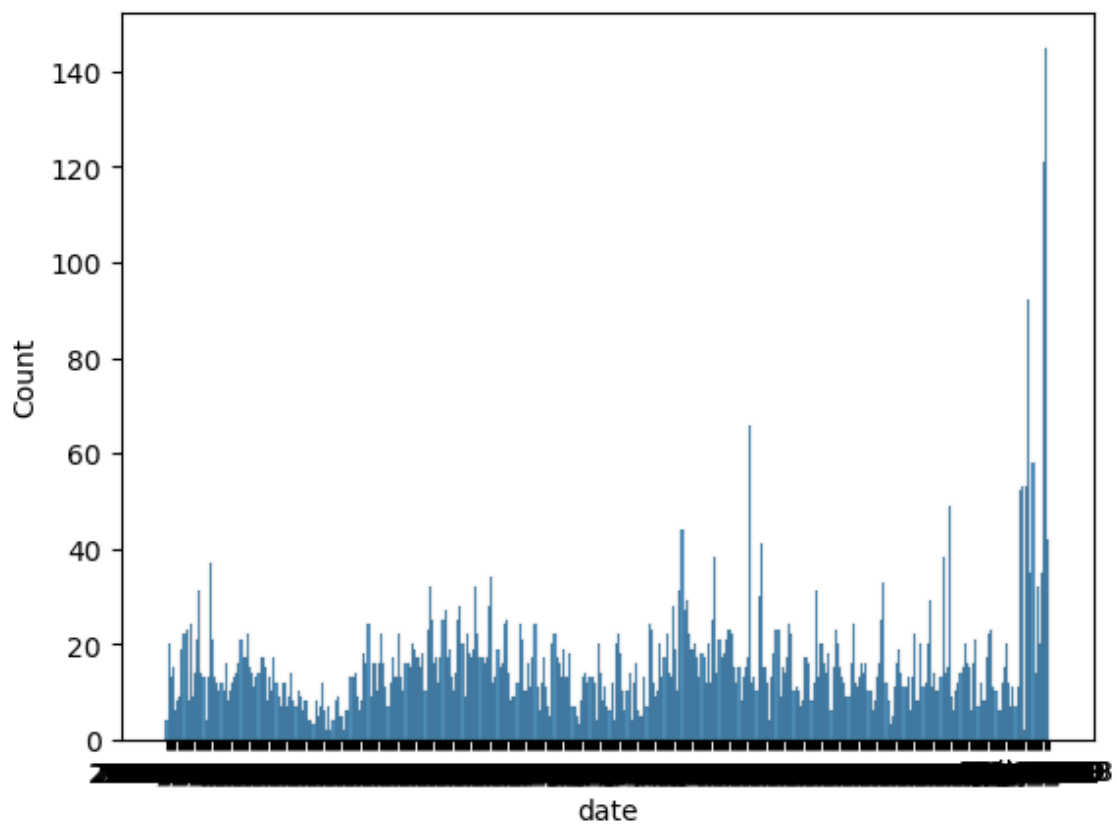


```
In [13]: auto.isnull().sum()
```

```
Out[13]: Review      55  
date          0  
Location    4737  
dtype: int64
```

```
In [14]: sns.histplot(auto.date,bins=10)
```

```
Out[14]: <Axes: xlabel='date', ylabel='Count'>
```



THE END

```
In [ ]:
```