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

In [2]:
```
df = pd.DataFrame({'From_To': ['LoNDon_paris', 'MAdrid_miLAN', 'londON_StockhOlm'
'Budapest_PaRis', 'Brussels_londOn'],
'FlightNumber': [10045, np.nan, 10065, np.nan, 10085],
'RecentDelays': [[23, 47], [], [24, 43, 87], [13], [67, 32]],
'Airline': ['KLM(!)', '<Air France> (12)', '(British Airways. )',
'12. Air France', '"Swiss Air"']})
```

In [3]:
```
df
```

Out[3]:

	From_To	FlightNumber	RecentDelays	Airline
0	LoNDon_paris	10045.0	[23, 47]	KLM(!)
1	MAdrid_miLAN	NaN	[]	<Air France> (12)
2	londON_StockhOlm	10065.0	[24, 43, 87]	(British Airways.)
3	Budapest_PaRis	NaN	[13]	12. Air France
4	Brussels_londOn	10085.0	[67, 32]	"Swiss Air"

In [4]:
```
# 1. Some values in the the FlightNumber column are missing. These numbers are me

# Replacing NAN values with the preceding values + 10 in FlightNumber Column

for i in range(df.FlightNumber.count() + 1):
    print(df.FlightNumber.loc[i,])
    if pd.isnull(df.FlightNumber.loc[i,]):
        df.loc[i,'FlightNumber'] = df.FlightNumber.loc[i-1,] + 10
        print(df.FlightNumber.loc[i,])
```

```
10045.0
nan
10055.0
10065.0
nan
10075.0
```

In [5]:
```
df
```

Out[5]:

	From_To	FlightNumber	RecentDelays	Airline
0	LoNDon_paris	10045.0	[23, 47]	KLM(!)
1	MAdrid_miLAN	10055.0	[]	<Air France> (12)
2	londON_StockhOlm	10065.0	[24, 43, 87]	(British Airways.)
3	Budapest_PaRis	10075.0	[13]	12. Air France
4	Brussels_londOn	10085.0	[67, 32]	"Swiss Air"

In [6]:
```
# 2. The From_To column would be better as two separate columns! Split each strin

df_tmp = df.copy()
df_tmp[['From','To']] = df_tmp.From_To.str.split("_",expand=True)
```

In [7]:
```
df
```

Out[7]:

	From_To	FlightNumber	RecentDelays	Airline
0	LoNDon_paris	10045.0	[23, 47]	KLM(!)
1	MAdrid_miLAN	10055.0	[]	<Air France> (12)
2	londON_StockhOlm	10065.0	[24, 43, 87]	(British Airways.)
3	Budapest_PaRis	10075.0	[13]	12. Air France
4	Brussels_londOn	10085.0	[67, 32]	"Swiss Air"

In [8]:
```
df_tmp
```

Out[8]:

	From_To	FlightNumber	RecentDelays	Airline	From	To
0	LoNDon_paris	10045.0	[23, 47]	KLM(!)	LoNDon	paris
1	MAdrid_miLAN	10055.0	[]	<Air France> (12)	MAdrid	miLAN
2	londON_StockhOlm	10065.0	[24, 43, 87]	(British Airways.)	londON	StockhOlm
3	Budapest_PaRis	10075.0	[13]	12. Air France	Budapest	PaRis
4	Brussels_londOn	10085.0	[67, 32]	"Swiss Air"	Brussels	londOn

In [9]:
```
# 3. 3. Notice how the capitalisation of the city names is all mixed up in this t

print(df_tmp)
df_tmp.From = df_tmp.From.str.capitalize()
df_tmp.To = df_tmp.To.str.capitalize()
df_tmp.From_To = df_tmp.From_To.str.capitalize()
print('-'*80)
print(df_tmp)
```

```
            From_To  FlightNumber  RecentDelays              Airline  \
0      LoNDon_paris       10045.0      [23, 47]               KLM(!)
1      MAdrid_miLAN       10055.0            []   <Air France> (12)
2  londON_StockhOlm       10065.0  [24, 43, 87]  (British Airways. )
3    Budapest_PaRis       10075.0          [13]       12. Air France
4   Brussels_londOn       10085.0      [67, 32]          "Swiss Air"

       From         To
0    LoNDon      paris
1    MAdrid      miLAN
2    londON  StockhOlm
3  Budapest      PaRis
4  Brussels     londOn
--------------------------------------------------------------------------------
-
            From_To  FlightNumber  RecentDelays              Airline  \
0      London_paris       10045.0      [23, 47]               KLM(!)
1      Madrid_milan       10055.0            []   <Air France> (12)
2  London_stockholm       10065.0  [24, 43, 87]  (British Airways. )
3    Budapest_paris       10075.0          [13]       12. Air France
4   Brussels_london       10085.0      [67, 32]          "Swiss Air"

       From         To
0    London      Paris
1    Madrid      Milan
2    London  Stockholm
3  Budapest      Paris
4  Brussels     London
```

In [10]:
```
# 4.Delete the From_To column from df and attach the temporary DataFrame from the

print(df)
df.drop('From_To',axis=1,inplace=True)
print('-'*80)
print(df)
df['From_To'] = df_tmp['From_To']
print('-'*80)
print(df)
```

```
            From_To  FlightNumber  RecentDelays              Airline
0      LoNDon_paris       10045.0      [23, 47]               KLM(!)
1      MAdrid_miLAN       10055.0            []    <Air France> (12)
2  londON_StockhOlm       10065.0  [24, 43, 87]  (British Airways. )
3    Budapest_PaRis       10075.0          [13]       12. Air France
4   Brussels_londOn       10085.0      [67, 32]          "Swiss Air"
--------------------------------------------------------------------------------
-
   FlightNumber  RecentDelays              Airline
0       10045.0      [23, 47]               KLM(!)
1       10055.0            []    <Air France> (12)
2       10065.0  [24, 43, 87]  (British Airways. )
3       10075.0          [13]       12. Air France
4       10085.0      [67, 32]          "Swiss Air"
--------------------------------------------------------------------------------
-
   FlightNumber  RecentDelays              Airline          From_To
0       10045.0      [23, 47]               KLM(!)      London_paris
1       10055.0            []    <Air France> (12)      Madrid_milan
2       10065.0  [24, 43, 87]  (British Airways. )  London_stockholm
3       10075.0          [13]       12. Air France    Budapest_paris
4       10085.0      [67, 32]          "Swiss Air"   Brussels_london
```

In [11]:
```
# 5. In the RecentDelays column, the values have been entered into the DataFrame

rows = []
_ = df.apply(lambda row: [rows.append([row['Airline'], row['FlightNumber'],nn,row
                          for nn in row.RecentDelays], axis=1)
```

In [12]:
```
rows
```

Out[12]:
```
[['KLM(!)', 10045.0, 23, 'London_paris'],
 ['KLM(!)', 10045.0, 47, 'London_paris'],
 ['(British Airways. )', 10065.0, 24, 'London_stockholm'],
 ['(British Airways. )', 10065.0, 43, 'London_stockholm'],
 ['(British Airways. )', 10065.0, 87, 'London_stockholm'],
 ['12. Air France', 10075.0, 13, 'Budapest_paris'],
 ['"Swiss Air"', 10085.0, 67, 'Brussels_london'],
 ['"Swiss Air"', 10085.0, 32, 'Brussels_london']]
```

In [13]:
```
df_new = pd.DataFrame(rows, columns=df.columns)
```

In [14]:
```
print("Original DataFrame :  \n")
print(df)
print('*'*80)
print("New DataFrame :  \n")
print(df_new)
```

```
Original DataFrame :

   FlightNumber  RecentDelays              Airline          From_To
0       10045.0      [23, 47]               KLM(!)      London_paris
1       10055.0            []    <Air France> (12)      Madrid_milan
2       10065.0  [24, 43, 87]  (British Airways. )  London_stockholm
3       10075.0          [13]       12. Air France    Budapest_paris
4       10085.0      [67, 32]          "Swiss Air"   Brussels_london
********************************************************************************
*
New DataFrame :

          FlightNumber  RecentDelays  Airline          From_To
0              KLM(!)       10045.0       23      London_paris
1              KLM(!)       10045.0       47      London_paris
2  (British Airways. )       10065.0       24  London_stockholm
3  (British Airways. )       10065.0       43  London_stockholm
4  (British Airways. )       10065.0       87  London_stockholm
5       12. Air France       10075.0       13    Budapest_paris
6          "Swiss Air"       10085.0       67   Brussels_london
7          "Swiss Air"       10085.0       32   Brussels_london
```

In [15]:
```
# Expand the Series of lists into a DataFrame named delays, rename the columns de


df3 = pd.DataFrame(df['RecentDelays'].values.tolist())
```

In [16]:
```
df3
```

Out[16]:

	0	1	2
0	23.0	47.0	NaN
1	NaN	NaN	NaN
2	24.0	43.0	87.0
3	13.0	NaN	NaN
4	67.0	32.0	NaN

In [17]:
```
length_cols = df3.shape[1]
```

In [18]:
```
length_cols
```

Out[18]:
```
3
```

In [19]:
```
df3.columns[0]
```

Out[19]:
```
0
```

In [20]:
```
col_list = []
col_dict ={}
for i in range(length_cols):
    Key = df3.columns[i]
    #print(key,i)
    Value = "Delay" + str(i+1)
    col_dict[Key] = Value
```

In [21]:
```
col_dict
```

Out[21]:
```
{0: 'Delay1', 1: 'Delay2', 2: 'Delay3'}
```

In [22]:
```
df3.rename(columns=col_dict,inplace=True)
```

In [23]:
```
df3
```

Out[23]:

	Delay1	Delay2	Delay3
0	23.0	47.0	NaN
1	NaN	NaN	NaN
2	24.0	43.0	87.0
3	13.0	NaN	NaN
4	67.0	32.0	NaN

In [24]:
```
df[["Delay1","Delay2","Delay3"]] = df3[["Delay1","Delay2","Delay3"]]
```

In [25]:
```
print(df)
print('*'*80)
df.drop('RecentDelays', axis=1, inplace=True)
print(df)
```

```
   FlightNumber  RecentDelays              Airline          From_To  Delay1  \
0       10045.0      [23, 47]               KLM(!)      London_paris    23.0
1       10055.0            []    <Air France> (12)      Madrid_milan     NaN
2       10065.0  [24, 43, 87]  (British Airways. )  London_stockholm    24.0
3       10075.0          [13]       12. Air France    Budapest_paris    13.0
4       10085.0      [67, 32]          "Swiss Air"   Brussels_london    67.0

   Delay2  Delay3
0    47.0     NaN
1     NaN     NaN
2    43.0    87.0
3     NaN     NaN
4    32.0     NaN
********************************************************************************
*
   FlightNumber              Airline          From_To  Delay1  Delay2  Delay3
0       10045.0               KLM(!)      London_paris    23.0    47.0     NaN
1       10055.0    <Air France> (12)      Madrid_milan     NaN     NaN     NaN
2       10065.0  (British Airways. )  London_stockholm    24.0    43.0    87.0
3       10075.0       12. Air France    Budapest_paris    13.0     NaN     NaN
4       10085.0          "Swiss Air"   Brussels_london    67.0    32.0     NaN
```

In [26]:
```
# Finally changed DataFrame looks like:

df
```

Out[26]:

	FlightNumber	Airline	From_To	Delay1	Delay2	Delay3
0	10045.0	KLM(!)	London_paris	23.0	47.0	NaN
1	10055.0	<Air France> (12)	Madrid_milan	NaN	NaN	NaN
2	10065.0	(British Airways.)	London_stockholm	24.0	43.0	87.0
3	10075.0	12. Air France	Budapest_paris	13.0	NaN	NaN
4	10085.0	"Swiss Air"	Brussels_london	67.0	32.0	NaN

In [ ]:
```
```