In [80]:
```
#Initialize#

import pandas as pd
import numpy as np
df = pd.DataFrame({'From_To': ['LoNDon_paris', 'MAdrid_miLAN', 'londON_StockhOlm','Buc
'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', '"

df
```

Out[80]:

	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 [81]:
```
#1.Some values in the the FlightNumber column are missing. These numbers are meant to

df['FlightNumber'] = df['FlightNumber'].interpolate().astype(int)
df
```

Out[81]:

	From_To	FlightNumber	RecentDelays	Airline
0	LoNDon_paris	10045	[23, 47]	KLM(!)
1	MAdrid_miLAN	10055	[]	<Air France> (12)
2	londON_StockhOlm	10065	[24, 43, 87]	(British Airways.)
3	Budapest_PaRis	10075	[13]	12. Air France
4	Brussels_londOn	10085	[67, 32]	"Swiss Air"

In [82]:
```
#2. The From_To column would be better as two separate columns! Split each string on t

tmpDF = pd.DataFrame(columns=['From','To'])
tmpDF[['From','To']] = df['From_To'].str.split('_', expand=True)
tmpDF
```

Out[82]:

	From	To
0	LoNDon	paris
1	MAdrid	miLAN
2	londON	StockhOlm
3	Budapest	PaRis
4	Brussels	londOn

In [83]:
```
#3. Notice how the capitalisation of the city names is all mixed up in this temporary

tmpDF['From'] = tmpDF.From.str.title()
tmpDF['To'] = tmpDF.To.str.title()
tmpDF
```

Out[83]:

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

In [84]:
```
#4. Delete the From_To column from df and attach the temporary DataFrame from the prev

df=pd.concat([tmpDF,df], axis=1)
df = df.drop('From_To', 1)
```

Out[84]:

	From	To	FlightNumber	RecentDelays	Airline
0	London	Paris	10045	[23, 47]	KLM(!)
1	Madrid	Milan	10055	[]	<Air France> (12)
2	London	Stockholm	10065	[24, 43, 87]	(British Airways.)
3	Budapest	Paris	10075	[13]	12. Air France
4	Brussels	London	10085	[67, 32]	"Swiss Air"

In [85]:
```
#5. In the RecentDelays column, the values have been entered into the DataFrame as a l

tDelay = pd.DataFrame(df.RecentDelays)
tDelay = pd.DataFrame(df['RecentDelays'].values.tolist())
tDelay.columns = ['Delay_1', 'Delay_2', 'Delay_3']
df = df.drop('RecentDelays', 1)
df.insert(3, "Delay_1", tDelay['Delay_1'])
df.insert(4, "Delay_2", tDelay['Delay_2'])
df.insert(5, "Delay_3", tDelay['Delay_3'])
df
```

Out[85]:

	From	To	FlightNumber	Delay_1	Delay_2	Delay_3	Airline
0	London	Paris	10045	23.0	47.0	NaN	KLM(!)
1	Madrid	Milan	10055	NaN	NaN	NaN	<Air France> (12)
2	London	Stockholm	10065	24.0	43.0	87.0	(British Airways.)
3	Budapest	Paris	10075	13.0	NaN	NaN	12. Air France
4	Brussels	London	10085	67.0	32.0	NaN	"Swiss Air"