Problem Statement

It happens all the time: someone gives you data containing malformed strings, Python, lists and missing data. How do you tidy it up so you can get on with the analysis?
Take this monstrosity as the DataFrame to use in the following puzzles:
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"']})

1. Some values in the the FlightNumber column are missing. These numbers are meant to increase by 10 with each row so 10055 and 10075 need to be put in place. Fill in these missing numbers and make the column an integer column (instead of a float column).

Source Code:

```
import pandas as pd
import numpy as np
import numpy.ma.mrecords as mrecords
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"']})
df['FlightNumber'] = df['FlightNumber'].interpolate().astype(int)
df
```

Output Screenshot:

	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"

2. The From_To column would be better as two separate columns! Split each string on the underscore delimiter _ to give a new temporary DataFrame with the correct values. Assign the correct column names to this temporary DataFrame.

Source Code:

```
tDF = pd.DataFrame(df.From_To)
tDF['From']= df['From_To'].str.split('_').str[0]
tDF['To'] = df['From_To'][0:5].str.split('_').str[1]
tDF
```

Output Screenshot:

	From_To	From	To
0	LoNDon_paris	LoNDon	paris
1	MAdrid_miLAN	MAdrid	miLAN
2	londON_StockhOlm	londON	StockhOlm
3	Budapest_PaRis	Budapest	PaRis
4	Brussels_londOn	Brussels	londOn

3. Notice how the capitalisation of the city names is all mixed up in this temporary DataFrame. Standardise the strings so that only the first letter is uppercase (e.g. "londON" should become "London".)

Source Code:

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

Output Screenshot:

	From_To	From	To
0	LoNDon_paris	London	Paris
1	MAdrid_miLAN	Madrid	Milan
2	londON_StockhOlm	London	Stockholm
3	Budapest_PaRis	Budapest	Paris
4	Brussels_londOn	Brussels	London

4. Delete the From_To column from df and attach the temporary DataFrame from the previous questions.

Source Code:

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

Output Screenshot:

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

5. In the RecentDelays column, the values have been entered into the DataFrame as a list. We would like each first value in its own column, each second value in its own column, and so on. If there isn't an Nth value, the value should be NaN.
Expand the Series of lists into a DataFrame named delays, rename the columns delay_1, delay_2, etc. and replace the unwanted RecentDelays column in df with delays.

Source Code:

```
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'])
print(df)
```

Output Screenshot:

```
           From_To      From         To  Delay_1  Delay_2  Delay_3  \
0      LoNDon_paris    London      Paris     23.0     47.0      NaN
1      MAdrid_miLAN    Madrid      Milan      NaN      NaN      NaN
2  londON_StockhOlm    London  Stockholm     24.0     43.0     87.0
3    Budapest_PaRis  Budapest      Paris     13.0      NaN      NaN
4   Brussels_londOn  Brussels     London     67.0     32.0      NaN

   FlightNumber              Airline
0         10045               KLM(!)
1         10055    <Air France> (12)
2         10065  (British Airways. )
3         10075       12. Air France
4         10085          "Swiss Air"
```