



EDA and Feature engineering flight price prediction:

Colab:

https://colab.research.google.com/drive/1HLqXLLao2_QWuT-i7eIN9B9Vcz0jxN#scrollTo=lt5hzSv_w3Sa

Important from the lecture:

#suppose we have a column (in reality we don't have that dataset with us) where there is date of journey written as

DD/MM/YYYY, then to make it useful we will separate out the part of it

```
# df['Date'] = df['Date_of_Journey'].str.split('/').str[0]
```

```
# df['Month'] = df['Date_of_Journey'].str.split('/').str[1]
```

```
# df['Year'] = df['Date_of_Journey'].str.split('/').str[2]
```

But the above created columns - Date, month, year are in object (can be seen with df.info())

So, to convert them into integer we will use

```
# df['Date'] = df['Date'].astype(int)
```

```
# df['Month'] = df['Month'].astype(int)
```

```
# df['Year'] = df['Year'].astype(int)
```

now dropping that imaginary column

```
# df.drop('Date_of_Journey', axis=1, inplace=True)
```

now suppose we have one column named Arrival_hour where the data is in the form xx:xx [some have month too], then to make that useful we will use

```
# df['Arrival_hour'] = df['Arrival_Time'].str.split(':').str[0]
```

```
# df['Arrival_min'] = df['Arrival_Time'].str.split(':').str[1]
```