

### 3. Adición de nuevas variables (columnas)

June 19, 2022

## 1 Transofrmación de datos

### 1.0.1 Adición de nuevas variables

Carga de librerías

```
[2]: import pandas as pd
```

Importado de datos

```
[2]: df = pd.read_csv("nycflights.csv")
      #df.info()
      #df.head()
```

```
[2]:
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	\
0	2013	6	30	940	15	1216	-4	VX	N626VA	
1	2013	5	7	1657	-3	2104	10	DL	N3760C	
2	2013	12	8	859	-1	1238	11	DL	N712TW	
3	2013	5	14	1841	-4	2122	-34	DL	N914DL	
4	2013	7	21	1102	-3	1230	-8	9E	N823AY	

  

	flight	origin	dest	air_time	distance	hour	minute
0	407	JFK	LAX	313	2475	9	40
1	329	JFK	SJU	216	1598	16	57
2	422	JFK	LAX	376	2475	8	59
3	2391	JFK	TPA	135	1005	18	41
4	3652	LGA	ORF	50	296	11	2

Creación de nuevas columnas con un valor independiente

```
[15]: df = pd.DataFrame({'Name'      : ['Carlos', 'Andrés', np.nan, 'Santiago', 'Fernando', "Marcelo", np.nan],
                        'Age'       : [23, 24, 24, 25, None, 27, None],
                        'University': ['AA', pd.NA, 'BB', None, 'CC', 'EE', pd.NA]})
      # Adición de una columna
      df["Sex"] = "Male"
      df.head()
```

```
[15]:
```

	Name	Age	University	Sex
0	Carlos	23.0	AA	Male
1	Andrés	24.0	<NA>	Male
2	NaN	24.0	BB	Male
3	Santiago	25.0	None	Male
4	Fernand0	NaN	CC	Male

```
[11]: df = pd.DataFrame({'Name'      : ['Carlos', 'Andrés', np.nan, 'Santiago', np.
    ↪ "Fernand0", "Marcelo", np.nan],
    'Age'      : [23, 24, 24, 25, None, 27, None],
    'University': ['AA', pd.NA, 'BB', None, 'CC', 'EE', pd.NA]})

# Adición de múltiples columnas independientes
df[["Sex", "Grade"]] = ["Male", 75]
df.head()
```

```
[11]:
```

	Name	Age	University	Sex	Grade
0	Carlos	23.0	AA	Male	75
1	Andrés	24.0	<NA>	Male	75
2	NaN	24.0	BB	Male	75
3	Santiago	25.0	None	Male	75
4	Fernand0	NaN	CC	Male	75

```
[14]: df = pd.DataFrame({'Name'      : ['Carlos', 'Andrés', np.nan, 'Santiago', np.
    ↪ "Fernand0", "Marcelo"],
    'Age'      : [23, 24, 24, 25, None, 27],
    'University': ['AA', pd.NA, 'BB', None, 'CC', 'EE']})

# Adición de una nueva columna con un valor para cada fila
# La lista debe tantos elementos como filas en el dataframe
df["Sex"] = ["Male", 1, "Female", "Female", 0, "Female"]
df.head(8)
```

```
[14]:
```

	Name	Age	University	Sex
0	Carlos	23.0	AA	Male
1	Andrés	24.0	<NA>	1
2	NaN	24.0	BB	Female
3	Santiago	25.0	None	Female
4	Fernand0	NaN	CC	0
5	Marcelo	27.0	EE	Female

```
[24]: df = pd.DataFrame({'Name'      : ['Carlos', 'Andrés', np.nan, 'Santiago', np.
    ↪ "Fernand0", "Marcelo"],
    'Age'      : [23, 24, 24, 25, None, 27],
    'University': ['AA', pd.NA, 'BB', None, 'CC', 'EE']})

# El método insert modifica el dataframe original
```

```
df.insert(loc = 0,
          column = 'New_column',
          value= [20,40,60,80,100,120])
df.head(6)
```

```
[24]:
```

	New_column	Name	Age	University
0	20	Carlos	23.0	AA
1	40	Andrés	24.0	<NA>
2	60	NaN	24.0	BB
3	80	Santiago	25.0	None
4	100	Fernand0	NaN	CC
5	120	Marcelo	27.0	EE

```
[25]: df = pd.DataFrame({'Name'      : ['Carlos', 'Andrés', np.nan, 'Santiago',
↪ "Fernand0", "Marcelo"],
                        'Age'       : [23, 24, 24, 25, None, 27],
                        'University' : ['AA', pd.NA, 'BB', None, 'CC', 'EE']})

# El método loc modifica el dataframe original
df.loc[:, 'new_column'] = 99
df.head(6)
```

```
[25]:
```

	Name	Age	University	new_column
0	Carlos	23.0	AA	99
1	Andrés	24.0	<NA>	99
2	NaN	24.0	BB	99
3	Santiago	25.0	None	99
4	Fernand0	NaN	CC	99
5	Marcelo	27.0	EE	99

### Creación de columnas calculadas

```
[26]: df = pd.DataFrame({'Name'      : ['Carlos', 'Andrés', np.nan, 'Santiago',
↪ "Fernand0", "Marcelo"],
                        'Age'       : [23, 24, 24, 25, None, 27],
                        'University' : ['AA', pd.NA, 'BB', None, 'CC', 'EE']})
```

```
[27]: # Adición de una columna
df1 = df.copy()
df1['New_column'] = df1['Age']*2
df1.head()
```

```
[27]:
```

	Name	Age	University	New_column
0	Carlos	23.0	AA	46.0
1	Andrés	24.0	<NA>	48.0
2	NaN	24.0	BB	48.0
3	Santiago	25.0	None	50.0

4	Fernand0	NaN	CC	NaN
---	----------	-----	----	-----

```
[34]: x = df1['Age']*2
df1 = df.copy()
df1.insert(loc = len(df1.columns.tolist()),
           column = 'New_column',
           value = x)
df1.head(6)
```

```
[34]:
```

	Name	Age	University	New_column
0	Carlos	23.0	AA	46.0
1	Andrés	24.0	<NA>	48.0
2	NaN	24.0	BB	48.0
3	Santiago	25.0	None	50.0
4	Fernand0	NaN	CC	NaN
5	Marcelo	27.0	EE	54.0