

Nombre: Hilda Beltrán

Matrícula: A01251916

```
from google.colab import drive
drive.mount('/content/drive')
```


Mounted at /content/drive

```
# Carga las librerías necesarias.
import pandas as pd
import numpy as np
```

```
# Carga el archivo insurance.csv
from google.colab import files
```

```
uploaded = files.upload()
```


```
for fn in uploaded.keys():
    print('User uploaded file "{name}" with length {length} bytes'.format(
        name=fn, length=len(uploaded[fn])))
```

  insurance.csv

- **insurance.csv**(text/csv) - 54289 bytes, last modified: 5/11/2022 - 100% done

Saving insurance.csv to insurance.csv  
User uploaded file "insurance.csv" with length 54289 bytes

```
# Carga el conjunto de datos al ambiente de Google Colab y muestra los primeros
# 6 renglones.
df = pd.read_csv('insurance.csv')
df.head(6)
```

	age	sex	bmi	children	smoker	region	charges	
0	19	female	27.900	0	yes	southwest	16884.92400	
1	18	male	33.770	1	no	southeast	1725.55230	
2	28	male	33.000	3	no	southeast	4449.46200	
3	33	male	22.705	0	no	northwest	21984.47061	
4	32	male	28.880	0	no	northwest	3866.85520	
5	31	female	25.740	0	no	southeast	3756.62160	

```
# Crea una tabla resumen con los estadísticas generales de las variables
```

```
# numéricas.
df1 = df[['age', 'bmi', 'children', 'charges']].copy()
df1.head(5)
```

	age	bmi	children	charges
0	19	27.900	0	16884.92400
1	18	33.770	1	1725.55230
2	28	33.000	3	4449.46200
3	33	22.705	0	21984.47061
4	32	28.880	0	3866.85520

```
df.describe()
```

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

```
# ¿Cómo se correlacionan las variables numéricas entre sí?
df.corr()
```

	age	bmi	children	charges
age	1.000000	0.109272	0.042469	0.299008
bmi	0.109272	1.000000	0.012759	0.198341
children	0.042469	0.012759	1.000000	0.067998
charges	0.299008	0.198341	0.067998	1.000000

```
# Determina si existe o no una correlación entre el índice de masa corporal
# (bmi) y el costo del seguro.
valcorr = df1['bmi'].corr(df1['charges'])
print(valcorr)
```

0.19834096883362895

```
# ¿Cuántas personas aseguradas son hombre y cuántas son mujeres?
```

```
def gen_asegurado():
    mujeres = 0
    hombres = 0
    for i in range (len(df.index)):
        if (df['sex'][i] == 'female'):
            mujeres += 1
        elif (df['sex'][i] == 'male'):
            hombres += 1
    return 'Mujeres: {} \nHombres: {}'.format(mujeres, hombres)
```

```
print(gen_asegurado())
```

Mujeres: 662

Hombres: 676

```
# ¿Cuántos hombres y mujeres asegurados viven en cada región?
```

```
pd.crosstab(df['region'], df['sex'])
```

	sex	female	male
region			
northeast		161	163
northwest		164	161
southeast		175	189
southwest		162	163

```
# En promedio, ¿quién paga más de cuota de seguro? ¿Los fumadores o los no
```

```
# fumadores? Muéstralo con los datos.
```

```
df.groupby(['smoker']).mean()
```

```
# Fumadores pagan más de cuota
```

	age	bmi	children	charges
smoker				
no	39.385338	30.651795	1.090226	8434.268298
yes	38.514599	30.708449	1.113139	32050.231832

```
# ¿Cuáles son las cuotas mínimas y máximas que las personan pagan dependiendo
```

```
# del género y del número de hijos?
```

```
df.groupby(['sex', 'children']).mean()[['charges']]
```

charges 

sex	children	
female	0	11905.714276
	1	12161.360414
	2	13941.317326
	3	13865.605066
	4	13937.674562
	5	9854.006419
male	0	12832.696736
	1	13273.522458
	2	16187.095325
	3	16789.167419
	4	13782.284829
	5	7931.658310

```
# ¿Cuál es el índice de masa corporal promedio para hombre y mujeres dependiendo
# región en la que viven y si son fumadores? ¿Impacta eso en la tarifa del
# seguro?
df.groupby(['region', 'smoker']).mean()[['bmi', 'charges']]
```

bmi charges 

region	smoker		
northeast	no	29.332082	9165.531672
	yes	28.565224	29673.536473
northwest	no	29.212678	8556.463715
	yes	29.140431	30192.003182
southeast	no	33.442418	8032.216309
	yes	33.096703	34844.996824
southwest	no	30.507865	8019.284513
	yes	31.005172	32269.063494

---

✓ 0 s completado a las 10:24

● ✕