

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Campus, Querétaro**



***Herramientas computacionales: el arte de la analítica  
(Gpo 201)***

**Actividad: Estadística de datos**

**Estudiantes:**

Karen Cebreros López

A01704254

**Profesor:**

Pedro Pérez

**Fecha de entrega:**

Miércoles 11 de mayo del 2022

IMPORTAR LIBRERÍAS Y SUBIR EL ARCHIVO "insurance.csv"

```
[3] import pandas as pd
import numpy as np
import random
```

```
[2] 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])))
```

Choose Files insurance.csv

- insurance.csv(text/csv) - 54289 bytes, last modified: 5/8/2022 - 100% done

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

MUESTRA LOS PRIMEROS 6 RENGLONES DEL CONJUNTO DE DATOS

```
df_insurance = pd.read_csv('insurance.csv')
df_insurance.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

+ Code

+ Text

TABLA RESUMEN CON LAS ESTADÍSTICAS GENERALES DE LAS VARIABLES NUMÉRICAS

```
df_insurance.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?

```
[8] df_insurance.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

DETERMINE SI EXISTE UNA CORRELACIÓN ENTRE EL "bmi" Y "charges"

```
[10] print('Correlación Pearson: ', df_insurance['bmi'].corr(df_insurance['charges'], method='pearson'))
```

Correlación Pearson: 0.19834096883362895

¿CUÁNTAS PERSONAS ASEGURADAS SON HOMBRES Y CUÁNTAS SON MUJERES?

```
[12] df_insurance['sex'].value_counts()
```

male 676  
female 662  
Name: sex, dtype: int64

¿CUÁNTOS HOMBRES Y MUJERES ASEGURADOS VIVEN EN CADA REGIÓN?

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

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

EN PROMEDIO, ¿QUIÉN PAGA MÁS DE CUOTA DE SEGURO? ¿LOS FUMADORES O LOS NO FUMADORES? MUÉSTRALO CON LOS DATOS

```
[15] df_insurance.groupby('smoker').mean()[['charges']]
```

	charges
smoker	
no	8434.268298
yes	32050.231832

¿CUÁLES SON LAS CUOTAS MÍNIMAS Y MÁXIMAS QUE LAS PERSONAS PAGAN DEPENDIENDO DEL GÉNERO Y DEL NÚMERO DE HIJOS?

```
df_insurance.groupby(['sex', 'children']).agg(['min', 'max'])[['charges']]
```

		charges	
		min	max
sex children			
female	0	1607.51010	63770.42801
	1	2201.09710	58571.07448
	2	2801.25880	47305.30500
	3	4234.92700	46661.44240
	4	4561.18850	36580.28216
	5	4687.79700	19023.26000
male	0	1121.87390	62592.87309
	1	1711.02680	51194.55914
	2	2304.00220	49577.66240
	3	3443.06400	60021.39897
	4	4504.66240	40182.24600
	5	4915.05985	14478.33015

¿CUÁL ES EL "bmi" PROMEDIO PARA HOMBRES Y MUJERES DEPENDIENDO DE LA REGIÓN EN LA QUE VIVEN Y SI SON FUMADORES?  
¿IMPACTA ESO EN LA TAFITA DEL SEGURO?

```
df_insurance.groupby(['sex', 'region', 'smoker']).mean()[['bmi', 'charges']]
```

			bmi	charges
sex region smoker				
female	northeast	no	29.777462	9640.426984
		yes	27.261724	28032.046398
	northwest	no	29.488704	8786.998679
		yes	28.296897	29670.824946
	southeast	no	32.780000	8440.205552
		yes	32.251389	33034.820716
male	southwest	no	30.050355	8234.091260
		yes	30.128571	31687.988430
	northeast	no	28.861760	8664.042222
		yes	29.560000	30926.252583
	northwest	no	28.930379	8320.689321
		yes	29.983966	30713.181419
	southeast	no	34.129552	7609.003587
		yes	33.650000	36029.839367
	southwest	no	31.019841	7778.905534
		yes	31.502703	32598.862854