## Actividad - Estadística básica

· Nombre: Daniel Queijeiro Albo

Matrícula: A01710441

**Entregar:** Archivo PDF de la actividad, así como el archivo .ipynb en tu repositorio. **Nota:** Recuerda habrá una penalización de **50** puntos si la actividad fue entregada fuera de la fecha límite.

Carga el conjunto de datos insurance.csv (se encuentra en el repositorio de la clase) y realiza un análisis estadístico de las variables.

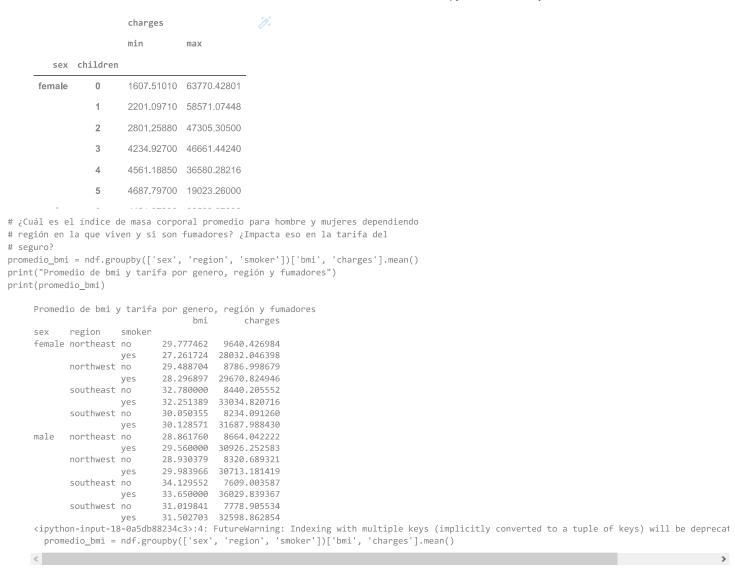
100	5428	39 100	54289	0 0		'	-::	
	age	sex	bmi	children	smoker	region	charges	7.
0	19	female	27.900	0	yes	southwest	16884.92400	
1	18	male	33.770	1	no	southeast	1725.55230	
2	28	male	33.000	3	no	southeast	4449.46200	
3	33	male	22.705	0	no	northwest	21984.47061	
4	32	male	28.880	0	no	northwest	3866.85520	
5	31	female	25.740	0	no	southeast	3756.62160	

El conjunto de datos contiene información demográfica sobre los asegurados en una compañía de seguros:

- age: Edad del asegurado principal
- sex: Género del asegurado. female o male
- bmi: Índice de masa corporal
- children: Número de hijos que estan cubiertos con la poliza.
- smoke: ¿El beneficiario fuma? (yes/no)
- region: ¿Dónde vive el beneficiario? Estos datos son de Estados Unidos. Regiones disponibles: northeast, southeast, southwest, northwest
- charges: Costo del seguro.

```
# Crea una tabla resumen con los estadísticas generales de las variables
# numéricas.
df.describe()
```

```
bmi
                                        children
                                                       charges
                    age
# ¿Cómo se correlacionan las varaibles numéricas entre sí?
df.corr()
                            bmi children
                                            charges
                   age
              1.000000 0.109272
                                  0.042469 0.299008
        age
              0.109272 1.000000
                                 0.012759 0.198341
       bmi
      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.
selected = df[['age', 'bmi', 'children', 'charges']]
print('Correlación pearson', selected['bmi'].corr(selected['charges'], method='pearson'))
print('Correlación spearman', selected['bmi'].corr(selected['charges'], method='spearman'))
print('Correlación kendall', selected['bmi'].corr(selected['charges'], method='kendall'))
     Correlación pearson 0.19834096883362895
    Correlación spearman 0.11939590358331145
    Correlación kendall 0.08252397079981415
# ¿Cuántas personas aseguradas son hombre y cuántas son mujeres?
ndf = df.dropna(subset=['charges'])
h = ndf.groupby('sex').get_group('male')
m = ndf.groupby('sex').get_group('female')
\label{eq:print('Hay', len(h), 'hombres asegurados y ', len(m), 'mujeres aseguradas.')}
    Hay 676 hombres asegurados y 662 mujeres aseguradas.
# ¿Cuántos hombres y mujeres asegurados viven en cada región?
group = ndf.groupby('region')
count = group['sex'].value_counts()
resultado = pd.DataFrame(count)
resultado.columns = ['Cantidad']
print(resultado)
                       Cantidad
     region
              sex
    northeast male
                            163
                           161
              female
    northwest female
                            164
              male
     southeast male
                            189
              female
                           175
     southwest male
                            163
                           162
              female
# En promedio, ¿quién paga más de cuota de seguro? ¿Los fumadores o los no
# fumadores? Muéstralo con los datos.
promedio = df.groupby('smoker')['charges'].mean()
print("La cuota promedio de fumadores es de ", promedio['yes'])
print("La cuota promedio de no fumadores es de ", promedio['no'])
     La cuota promedio de fumadores es de 32050.23183153285
    La cuota promedio de no fumadores es de 8434.268297856202
# ¿Cuáles son las cuotas mínimas y máximas que las personan pagan dependiendo
# del género y del número de hijos?
df.groupby(['sex','children']).agg(['min','max'])[['charges']]
```



Productos pagados de Colab - Cancela los contratos aquí