

TECNOLÓGICO DE MONTERREY Campus Querétaro

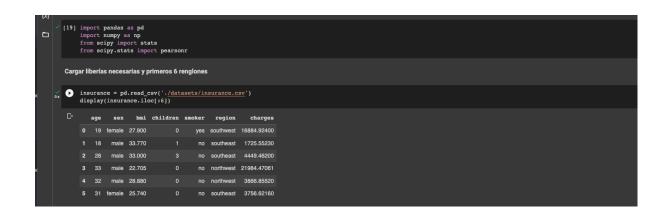
PROFESOR ENCARGADOS: Pedro Oscar Pérez Murueta

Semana Tec TC1002S

Actividad 6

Daniel Gutiérrez Gómez A01068056

FECHA DE ENTREGA: 22/03/2023



```
Crea una tabla resumen con los estadísticas generales de las variables numéricas
v [21] resumen = pd.DataFrame().assign(age = insurance["age"], bmi = insurance["bmi"], charges = insurance["charges"])
      display(resumen)
              age bmi charges
              19 27.900 16884.92400
         1 18 33.770 1725.55230
         2 28 33.000 4449.46200
         3 33 22.705 21984.47061
         4 32 28.880 3866.85520
        1333 50 30.970 10600.54830
              18 31.920 2205.98080
        1335 18 36.850 1629.83350
        1336 21 25.800 2007.94500
       1337 61 29.070 29141.36030
   ¿Cómo se correlacionan las varaibles numéricas entre sí? El coeficiente de correlación da una estimación de la relación entre dos variables. El
   coeficiente de correlación de Pearson se calcula multiplicando las desviaciones de la media de la variable 1 por las de la variable 2, dividiendo
   por el producto de las desviaciones estándar
```

```
[22] print('Correlación Pearson entre Edad e IMC: ', resumen.age.corr(resumen.bmi, method='pearson'))
    print('Correlación spearman entre Edad e IMC: ', resumen.age.corr(resumen.bmi, method='spearman'))
    print('Correlación kendall entre Edad y costo del seguro: ', resumen.age.corr(resumen.charges, method='pearson'))
    print('Correlación spearman entre Edad y costo del seguro: ', resumen.age.corr(resumen.charges, method='spearman'))
    print('Correlación kendall entre Edad y costo del seguro: ', resumen.age.corr(resumen.charges, method='spearman'))
    print('Correlación kendall entre IMC y costo del seguro: ', resumen.bmi.corr(resumen.charges, method='spearman'))
    print('Correlación spearman entre IMC y costo del seguro: ', resumen.bmi.corr(resumen.charges, method='spearman'))
    print('Correlación kendall entre IMC y costo del seguro: ', resumen.bmi.corr(resumen.charges, method='kendall'))

Correlación Pearson entre Edad e IMC: 0.1092718815485352
    Correlación spearman entre Edad e IMC: 0.10773603532023862
    Correlación kendall entre Edad y costo del seguro: 0.29900819333064754
    Correlación spearman entre Edad y costo del seguro: 0.5343921337718459
    Correlación kendall entre Edad y costo del seguro: 0.19834096883362895
    Correlación spearman entre IMC y costo del seguro: 0.19834096883362895
    Correlación spearman entre IMC y costo del seguro: 0.198399358331145
    Correlación kendall entre IMC y costo del seguro: 0.199399358331145
    Correlación kendall entre IMC y costo del seguro: 0.199399358331145
```

```
Determina si existe o no una correlación entre el índice de masa corporal

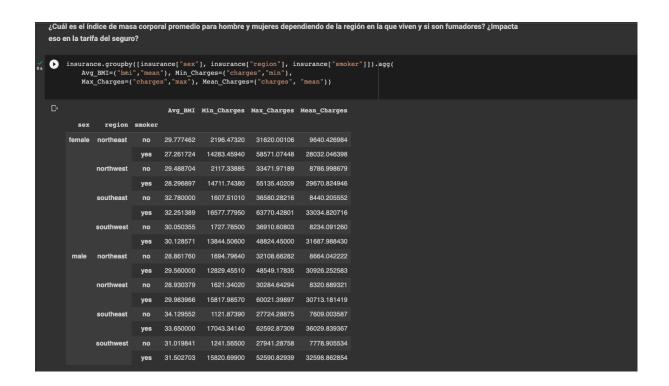
[23] r, p = stats.pearsonr(resumen.bmi, resumen.charges)
print(f*Correlación Pearson: r=(r), p-value={p}*)

Correlación Pearson: r=0.1983409688336288, p-value=2.459085535116766e-13

Como podemos ver no existe una relación valor de r=0.198 entre los valores del índice de la masa corporal y del costo del seguro, puesto que el valor de r es cercano a 0, ni -1 ni 1. Aunque correlación no significa causalidad, el hecho de que dos variables no estén correlacionadas no significa que no se afecten entre sí.

¿Cuántas personas aseguradas son hombre y cuántas son mujeres?

[24] print("Número de hombres aseguradas: ", (insurance["sex"] == "male").sum())
print("Número de mujeres aseguradas: 676
Número de hombres aseguradas: 662
```



Como podemos ver en la tabla agregada anterior, importa mucho la tarifa del seguro si eres fumador... Independientemente de si eres hombre o mujer y también independientemente del genero y de la región en la que vivas el fumar ya te incrementa demasiado la tarifa para pagar el seguro (ver columnas de Mean_charges, Max_charges y Min_charges)