

STATISTICAL ESSENTIALS

Laboratorio Practico N°01:

Se entrega el siguiente set de datos de una muestra o grupo de personas, censun.csv, el cual es un dataset que cuenta con las principales características sociodemográficas y de localidad de las personas. Los censos son muy importantes para las poblaciones porque nos permiten estudiar, entender las principales características demográficas, sociales, económicas y demás de todos los habitantes de un país para así poder tomar acciones de prevención, precaución y diferentes políticas públicas. En este censo en particular se quería conocer el nivel económico de la población dónde la variable más relevante es el nivel de ingresos.

Se requiere que los ayude a identificar las principales características de las personas, así como poder solucionar los principales problemas que ocurren cuando se aborda un problema de análisis de datos.

- Las variables que se disponibilizan son:

Variables	Descripción de la Variable
id	ID de la persona o identificador único de ésta.
age	Edad de la persona.
employment	Tipo de empleo.
final.weight	Peso final en libras.
education	Nivel de educacion.
education.nun	Número de años de educación.
marital.status	Estado civil.
job	Rol que desempeña en su trabajo actual.
relationship	Relación o rol que ocupa en el hogar.
race	Raza o descendencia de la persona.
gender	Género de la persona.
capital.gain	Ganancia de capital que resulta de una disposición de un activo de capital.
capital.loss	Pérdida de capital que resulta de una pérdida de un activo de capital.
hours.per.wee	Horas por semana trabajadas.
native.country	País de nacimiento o nativo.
salary	Salario mediano que contestó a la hora de la entrevista.

Entregables del Laboratorio:

- 1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio bien definidos.
- 2.- ¿Qué tipo de variables se utilizan en el problema de negocio?
- 3.- *Realizar un análisis completo de gráficos multivariados de acuerdo a las tipologías de variables encontradas, es decir mínimamente 2 gráficos para variables cualitativas, 2 para variables cuantitativas.*
- 4.- *De lo antes explorado, cuáles son tus principales conclusiones. ¿Qué recomendación le darías a la empresa? ¿Qué solución adicional podrías plantear para mejorar el proceso actual?*
- 5.- *Evaluar la completitud de los datos, sacando indicadores para los datos antes de la imputación y después. Comente sus resultados. (Imputación datos).*
- 6.- *Evaluar los outliers o casos atípicos univariados para las variables cuantitativas. Muestre los principales gráficos para identificarlos y muéstrelos.*
- 7.- *Evaluar la colinealidad o multicolinealidad en las variables, identificando aquellas en las cuáles su coeficiente de asociación supere el punto de corte de ± 0.3 .*
- 8.- *Desarrolle las mejores transformaciones de variables para las variables cuantitativas, fundamente el porqué de las transformaciones y muestre los hallazgos más relevantes.*

