**ESTADÍSTICA INFERENCIAL Y MUESTREO**

**INFORME ANÁLISIS INICAL DE DATOS**

**Neyfer Yesith Torres Vergara-U00177175**

En este informe exploraremos un análisis de una práctica aplicada utilizando herramientas fundamentales de Python, con el propósito de realizar un análisis descriptivo de datos reales relacionados con nacimientos de niños en La Dorada, Caldas.

**Objetivo:** Realizar una exploración y limpieza inicial del dataset, así como analizar la distribución y normalidad de las variables.

Para iniciar nuestro trabajo con el dataset, procedimos a cargarlo y a realizar una exploración profunda para entender a fondo su estructura y las características de sus variables, identificando tanto las categóricas como las numéricas. El siguiente paso crucial fue la limpieza de datos, donde nos dedicamos a manejar los valores faltantes y a identificar y corregir los valores atípicos, asegurando así que nuestro conjunto de datos fuera lo más confiable posible para análisis posteriores. Para visualizar y comprender mejor la distribución de la información, generamos gráficos de pastel y barras para las variables categóricas, y utilizamos histogramas y diagramas de caja para las variables numéricas. Como parte fundamental de este análisis inicial, evaluamos la normalidad de las distribuciones de las variables numéricas restantes a través de la aplicación de diversos tests estadísticos reconocidos, como Kolmogorov, Shapiro-Wilk, Anderson y Jarque-Bera, complementados con la creación de gráficos Q-Q, lo que nos proporcionó una clara idea de lo cerca que se encuentran estas distribuciones de un comportamiento normal. En retroceso, esta etapa inicial de análisis y preparación de datos ha sido esencial no solo para organizar y comprender el dataset, sino también para corregir problemas fundamentales y evaluar las propiedades de las variables, sentando una base sólida para continuar con estudios más detallados y ganar experiencia práctica para el futuro, un proceso en el que la asistencia y las capacidades de Gemini resultaron ser una herramienta de gran valor.

El estudio exploratorio del conjunto de datos sobre nacimientos en La Dorada, Santander, permitió obtener varios hallazgos relevantes:

**Resultado del análisis**

**Comprensión del Dataset**: Se reconocieron variables categóricas y numéricas, lo cual facilitó el uso de técnicas adecuadas para su análisis y representación.

**Manejo de Datos Incompletos**: Se detectaron registros con información faltante, los cuales fueron tratados adecuadamente según el tipo de análisis a realizar, garantizando así la consistencia de los resultados.

**Análisis de Variables Categóricas**: A través de gráficos de pastel y barras, se visualizaron las frecuencias de distintas categorías, evidenciando tendencias en variables como el sexo del recién nacido y el tipo de institución de salud.

**Distribución y Outliers en Variables Numéricas**: Se utilizaron histogramas y diagramas de caja para representar la distribución de variables cuantitativas como el peso al nacer o la edad de la madre. Asimismo, se identificaron valores atípicos, los cuales fueron eliminados aplicando el método del rango intercuartílico (IQR), con el fin de optimizar la calidad del análisis.

**Normalidad de los Datos**: Las pruebas estadísticas aplicadas (Kolmogorov-Smirnov, Shapiro-Wilk, Anderson-Darling y Jarque-Bera), junto con los gráficos Q-Q, mostraron que la mayoría de las variables numéricas no presentan una distribución normal, aspecto crucial a considerar en análisis estadísticos más avanzados.

**Conclusiones**

El conjunto de datos presentó valores faltantes y atípicos que fueron tratados para asegurar la calidad del análisis.

Las variables categóricas permitieron comprender la composición general de los registros.

Las variables numéricas mostraron diferentes comportamientos y, en su mayoría, no siguen una distribución normal.

Esta exploración inicial fue clave para dejar el dataset listo para futuros análisis más complejos.

**Bibliografía**

**https://www.datos.gov.co/Salud-y-Protecci-n-Social/Nacimientos-2021-2022/e7w8-zhsx/about\_data**