

Documento de Procesos:

Para el desarrollo de este proyecto de Análisis Exploratorio de Datos (EDA) en el área de Ingeniería de Sistemas de la Universidad del Sinú, se siguieron los siguientes procesos:

Recolección de Datos:

Se recopiló información de diferentes fuentes institucionales, como registros académicos y bases de datos, para obtener los datos relevantes para el análisis.

Los datos recolectados incluyeron variables como ID del estudiante, promedio de calificaciones, edad y número de cursos tomados.

Preparación de Datos:

Se realizó un proceso de limpieza y estandarización de los datos, lo que incluyó la identificación y manejo de valores atípicos (outliers) y datos faltantes.

Los valores atípicos en los promedios de calificaciones fueron eliminados, ya que no representaban la variabilidad natural de los datos y distorsionaban el análisis.

Los datos faltantes en la columna de edad también fueron eliminados, ya que su ausencia no afectaba de manera significativa el análisis general.

Análisis Estadístico:

Se aplicaron técnicas de análisis estadístico descriptivo, como el cálculo de medidas de tendencia central (media, mediana), medidas de dispersión (desviación estándar) y rango, para cada una de las variables.

Esta información estadística permitió comprender mejor la distribución y características de los datos.

Visualización de Datos:

Se utilizó la herramienta Plotly para crear visualizaciones interactivas y dinámicas de los datos, como histogramas, gráficos de dispersión y matrices de correlación.

Estas visualizaciones facilitaron la identificación de patrones, tendencias y relaciones entre las variables analizadas.

Interpretación de Resultados:

A partir de los análisis estadísticos y las visualizaciones generadas, se interpretaron los hallazgos clave, como la identificación de outliers, la distribución de la edad y el número de cursos de los estudiantes, y la correlación entre las variables.

Se resaltaron los insights más relevantes que podrían informar la toma de decisiones y la implementación de estrategias para mejorar el rendimiento académico de los estudiantes de Ingeniería de Sistemas.

Documentación y Reporte:

Se documentó de manera detallada todo el proceso, incluyendo la metodología utilizada, los hallazgos clave, las conclusiones y recomendaciones.

Se generó un reporte final que presenta de manera clara y concisa los resultados del análisis exploratorio de datos.

Puntos de Hallazgos:

Problemas Encontrados:

- Presencia de valores atípicos (outliers) en los promedios de calificaciones de los estudiantes, superando el rango esperado de 0 a 5.
- Algunos datos faltantes en la columna de edad de los estudiantes.
- Débil correlación entre las variables analizadas (ID del estudiante, promedio de calificaciones, edad y número de cursos).

Soluciones Implementadas:

- Eliminación de los valores atípicos (outliers) en los promedios de calificaciones, ya que distorsionaban el análisis y no representaban la variabilidad natural de los datos.
- Eliminación de los datos faltantes en la columna de edad, ya que su ausencia no afectaba significativamente el análisis general.
- Análisis de la matriz de correlación para comprender la fuerza de las relaciones entre las variables y enfocar el análisis en aquellas que presenten una correlación más significativa.

Lecciones Aprendidas:

- La importancia de realizar un análisis exploratorio de datos exhaustivo, incluyendo la identificación y manejo adecuado de valores atípicos y datos faltantes, para asegurar la calidad y confiabilidad de los resultados.
- La necesidad de comprender las relaciones y patrones presentes en los datos, más allá de la simple aplicación de técnicas estadísticas, para generar insights valiosos y tomar decisiones informadas.
- La utilidad de herramientas de visualización, como Plotly, para facilitar la interpretación de los datos y comunicar de manera efectiva los hallazgos del análisis.