

## Análisis de Datos Exploratorio.

El Análisis Exploratorio de Datos (EDA) es un proceso donde se utilizan una serie de herramientas y técnicas con el cual ayuda a comprender con mas facilidad el conjunto de datos que son utilizados. Una vez teniendo el conocimiento sobre el conjunto de datos como la extracción, la identificación, relación y las instancias son una de las cosas por la cual es importante como la frase de EDA cuando realiza algo seria cuando “la basura entra” y al acabar esa acción “la basura sale” una vez realizado eso muestra como si los datos eran poco entendibles al finalizarlo eso ya no es un problema ya que vuelve el conjunto de datos en perfectamente legibles.

El EDA tiene dos objetivos principales: el primero es la limpieza de los datos y el segundo es analizar las relaciones entre las variables para una mejor comprensión. Según Terence Shin en su artículo sobre la visualización de variables en EDA, existen tres componentes clave en el análisis exploratorio de datos: entender tus variables, limpiar tu conjunto de datos y analizar las relaciones entre las variables.

Una herramienta eficaz en EDA es la matriz de correlación, que permite visualizar rápidamente las relaciones entre múltiples variables. Por ejemplo, en un dataset de autos usados, se puede observar una correlación positiva entre el precio y el año del auto, y una correlación negativa entre el precio y el odómetro. Además, los scatterplots son muy útiles para identificar relaciones y outliers entre dos variables específicas, como el odómetro y el precio de un auto.

Una vez de comprender la lectura nos explica que el EDA no solo nos ayuda a limpiar los datos, sino también los transforma en información muy completa en tanto a la visualización de variables y el análisis de la relación, con ello nos asegura unos datos comprensibles y útiles para un futuro análisis.

**Para la realización del notebook** con la lectura pude observar que los valores que no son tan necesarios son 'Pclass', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin' con ello nos ayudara a como tendremos como mejor objetivo el saber con datos mas exactos y sin tantos datos innecesarios