

Análisis de datos

Nivel Básico – Explorador

Misión 2

Comparación de diferentes métodos de análisis en un dataset.

1. Análisis Descriptivo

Descripción:

- **Objetivo:** Resumir y describir las características del dataset.
- **Métodos:** Estadísticas descriptivas (media, mediana, moda, desviación estándar), gráficos (histogramas, gráficos de dispersión).

Pros:

- Rápido y fácil de implementar.
- Proporciona una visión general clara de los datos.

Contras:

- No revela relaciones causales ni patrones complejos.
- Puede ser insuficiente para análisis profundos.

Ejemplo: Analizar la distribución de edades en un dataset de clientes.

2. Análisis Exploratorio de Datos (EDA)

Descripción:

- **Objetivo:** Explorar los datos para identificar patrones, anomalías y relaciones.
- **Métodos:** Visualizaciones (mapas de calor, gráficos de caja), análisis de correlación, reducción de dimensionalidad (PCA).

Pros:

- Ayuda a comprender la estructura subyacente de los datos.
- Facilita la detección de problemas con los datos (valores atípicos, datos faltantes).

Contras:

- Puede ser intensivo en tiempo y computación.
- Los resultados pueden ser interpretativos y subjetivos.

Ejemplo: Usar gráficos de dispersión para encontrar relaciones entre variables.

3. Análisis Predictivo

Descripción:

- **Objetivo:** Predecir valores futuros basados en datos históricos.
- **Métodos:** Regresión lineal, regresión logística, árboles de decisión, redes neuronales.

Pros:

- Útil para hacer predicciones y tomar decisiones basadas en datos.
- Puede manejar datos complejos y no lineales.

Contras:

- Requiere una cantidad significativa de datos para entrenar modelos precisos.
- Los modelos pueden ser complejos y difíciles de interpretar.

Ejemplo: Predecir la probabilidad de que un cliente realice una compra basada en su historial de compras.

4. Análisis Inferencial

Descripción:

- **Objetivo:** Hacer inferencias y generalizaciones sobre una población basada en una muestra.
- **Métodos:** Pruebas de hipótesis, intervalos de confianza, análisis de varianza (ANOVA).

Pros:

- Permite hacer afirmaciones sobre una población sin tener que analizar todos los datos.
- Basado en teoría estadística sólida.

Contras:

- Asume que los datos de la muestra son representativos de la población.
- Puede ser afectado por errores de muestreo y sesgos.

Ejemplo: Determinar si hay diferencias significativas en las ventas entre diferentes regiones usando ANOVA.

5. Análisis Causal

Descripción:

- **Objetivo:** Identificar relaciones causales entre variables.
- **Métodos:** Experimentos controlados, modelos de variables instrumentales, técnicas de causalidad como Granger.

Pros:

- Permite establecer relaciones de causa y efecto.
- Proporciona una comprensión profunda de los factores que afectan a las variables.

Contras:

- Puede ser difícil de implementar sin experimentos controlados.
- Requiere datos adecuados y métodos estadísticos complejos.

Ejemplo: Evaluar el impacto de una campaña de marketing en las ventas utilizando un diseño experimental.

Comparación General

Método	Pros	Contras	Uso Común
Análisis Descriptivo	Rápido, fácil de implementar	Superficial, no revela relaciones complejas	Resumen general de datos
EDA	Identifica patrones, problemas y relaciones	Intensivo en tiempo y computación, subjetivo	Exploración de datos inicial
Análisis Predictivo	Hace predicciones, maneja datos complejos	Requiere muchos datos, modelos complejos	Predicción de valores futuros
Análisis Inferencial	Generaliza sobre poblaciones, basado en teoría	Supone representatividad, errores de muestreo	Generalización y pruebas de hipótesis
Análisis Causal	Establece relaciones causa-efecto	Difícil sin experimentos, requiere datos adecuados	Comprensión profunda de relaciones causales