



EXTRACCIÓN DE CONOCIMIEN TOS EN BASES DE DATOS

ING. LUIS ENRIQUE MASCOTE CANO



ELABORACIÓN DE GRÁFICAS

Lic. Ricardo

Hernández Martínez

Fecha de Entrega:

30/11/25

Introducción a la Visualización de Datos

Comprender el proceso de elaboración de gráficas mediante herramientas de visualización de datos y aplicarlas en un proyecto previamente desarrollado en el curso, integrando al menos tres tipos de visualizaciones e interpretando su aporte al análisis y extracción de conocimiento.

La visualización de datos es la representación gráfica de información con el fin de facilitar su comprensión, revelar patrones, tendencias, relaciones y apoyar la toma de decisiones. Gracias a la manera en que el cerebro humano procesa información, los gráficos permiten asimilar grandes volúmenes de datos más rápidamente que en tablas o textos.

En el contexto del análisis de datos, la visualización cumple un rol crucial porque:

- Permite detectar comportamientos atípicos.
- Ayuda a identificar correlaciones y tendencias.
- Facilita la comunicación de hallazgos a audiencias técnicas y no técnicas.
- Sirve como puente entre la exploración de datos y la extracción de conocimiento.

La visualización moderna se sustenta en tecnologías interactivas que permiten profundizar en los datos, filtrar información, y modificar vistas en tiempo real para obtener insights más precisos.

3. Importancia de la Visualización en la Extracción de Conocimiento

La extracción de conocimiento depende de la capacidad de entender patrones y estructuras dentro de grandes conjuntos de datos. Las gráficas facilitan:

- **Comprensión inmediata de información compleja.** Las relaciones abstractas se vuelven visibles.
- **Detección de tendencias emergentes.** La representación gráfica facilita observar cambios en el tiempo.
- **Identificación de relaciones entre variables.** Útil en análisis exploratorio y preparación previa al modelado.
- **Narrativa visual.** Permite comunicar hallazgos de forma clara, concisa e impactante.

En conjunto, la visualización se convierte en una herramienta clave para transformar datos en conocimiento útil.

4. Desarrollo del Proyecto de Visualización

Seleccione uno de los proyectos desarrollados en el curso (ej. clasificación, clustering, regresión, análisis estadístico, etc.) y compléméntelo con **tres tipos de gráficas**, tales como:

- Gráfica de dispersión (scatter plot)
- Histograma
- Gráfica de barras
- Boxplot
- Gráfica de líneas
- Heatmap

Las gráficas elegidas deben apoyar la interpretación del comportamiento de las variables o del desempeño del modelo.

4.1. Ejemplo de Implementación de Gráficas (Python + Matplotlib/Seaborn)

```
import seaborn as sns
import pandas as pd

df = pd.read_csv("datos.csv")

# Gráfica de dispersión\plt.figure(figsize=(6,4))
sns.scatterplot(data=df, x="feature1", y="feature2", hue="label")
plt.title("Relación entre Feature1 y Feature2")
plt.show()

# Histograma\plt.figure(figsize=(6,4))
sns.histplot(df["feature1"], kde=True)
plt.title("Distribución de Feature1")
plt.show()

# Boxplot\plt.figure(figsize=(6,4))
sns.boxplot(data=df, x="label", y="feature2")
plt.title("Distribución de Feature2 por Clase")
plt.show()
```

5. Interpretación de las Gráficas

Cada gráfica debe incluir una interpretación. Ejemplo:

5.1 Scatter Plot

Muestra la relación entre dos variables. Si los puntos forman agrupaciones claras, podría indicar la presencia de clusters o separabilidad entre clases.

5.2 Histograma

Permite observar la distribución de una variable. Una curva sesgada o concentrada revela asimetrías o falta de normalidad.

5.3 Boxplot

Útil para identificar valores atípicos y comparar distribuciones entre categorías.

6. Relación entre Visualización y Extracción de Conocimiento

La visualización no sólo presenta datos, sino que potencia el análisis al:

- Reducir la complejidad de la información.
- Facilitar la detección de patrones.
- Permitir comparar variables de manera inmediata.
- Ayudar a validar supuestos previos al modelado.

Por ejemplo, antes de entrenar un modelo de regresión, una gráfica de dispersión puede indicar si existe una relación lineal entre las variables. De igual forma, un heatmap de correlación ayuda a seleccionar variables relevantes

7. Conclusiones

- La visualización de datos es una herramienta indispensable para la comprensión de información y la extracción de conocimiento.
- Las técnicas visuales permiten identificar patrones, tendencias, correlaciones y anomalías que no serían evidentes en datos tabulares.
- Integrar diferentes tipos de gráficas en un proyecto mejora la interpretación y la comunicación de resultados.
- Con la ayuda de herramientas modernas es posible analizar grandes volúmenes de datos de forma interactiva, lo que abre nuevas posibilidades en el análisis exploratorio y la toma de decisiones.

8. Referencias

- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.
- Knaflitz, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Friendly, M. (2008). A brief history of data visualization. *Handbook of Data Visualization*, Springer.