

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Ingeniería en Desarrollo y Gestión de Software



Extracción de Conocimientos de Bases de Datos

V.1. Reporte de investigación de técnicas de visualización (50%)

IDGS91N

PRESENTAN:

Giselle Cantú Chávez

NOMBRE DEL DOCENTE:

Ing. Luis Enrique Mascote Cano

Chihuahua, Chih., 29 de noviembre de 2025

Índice

1. Introducción.....	3
2. Desarrollo	3
2.1. Importancia de la visualización en análisis de datos	3
2.2. Técnicas de visualización más utilizadas.....	3
1) Gráficos de barras.....	3
2) Gráficos de líneas.....	4
3) Mapas de calor (heatmaps).....	4
4) Diagramas de dispersión (scatter plots).....	5
5) Histogramas.....	5
6) Diagramas de cajas (boxplots)	5
7) Gráficos avanzados: PCA, clusters y redes.....	6
2.3. Relación con la extracción del conocimiento.....	6
2.4. Esquema y diagrama sugerido	7
3. Conclusiones	8
4. Fuentes de apoyo	8

1. Introducción

La visualización de información se ha convertido en una parte esencial dentro del proceso de extracción de conocimiento, porque permite traducir datos complejos en mensajes claros. Cuando los datos se representan de manera visual, se vuelven más fáciles de interpretar y conectar con decisiones reales. En este reporte exploró técnicas de visualización ampliamente utilizadas en análisis de datos, ciencia aplicada y entornos ejecutivos, explicando su utilidad, cuándo conviene usarlas y cómo se relacionan directamente con la identificación de patrones y el descubrimiento de conocimiento en un proyecto analítico.

2. Desarrollo

2.1. Importancia de la visualización en análisis de datos

Visualizar datos no es solo “hacer gráficas”; es construir representaciones que revelen historias, tendencias, anomalías y decisiones posibles. La visualización actúa como un puente entre el análisis técnico y la interpretación humana, especialmente cuando se trabaja con grandes volúmenes de información. Sin ella, entender patrones, detectar errores o comunicar hallazgos sería un proceso lento e impreciso.

2.2. Técnicas de visualización más utilizadas

1) Gráficos de barras

Sirven para comparar categorías o mostrar variaciones discretas. Son fuertes en escenarios donde se necesita ver diferencias claras entre grupos: ventas por región, productos más vendidos, clasificación por niveles, etc.

Ventajas:

- Fáciles de leer.
- Adecuados cuando se necesitan comparaciones directas.

Limitaciones:

- No representan bien cambios temporales.
- Pueden distorsionarse si se saturan de colores o categorías.

2) Gráficos de líneas

Se usan para analizar tendencias a lo largo del tiempo. Representan cómo un indicador evoluciona: crecimiento mensual, comportamiento histórico, fluctuaciones de mediciones, entre otros.

Relación con extracción de conocimiento:

Ayudan a identificar ciclos, anomalías, puntos de inflexión y patrones estacionales.

3) Mapas de calor (heatmaps)

Visualizan matrices o datos en escala de colores. Son útiles en correlaciones, análisis de similitud o patrones de intensidad.

Fortaleza principal: permiten identificar relaciones fuertes o débiles entre variables de manera inmediata.

4) Diagramas de dispersión (scatter plots)

Permiten observar relaciones entre dos variables, detectar patrones, linealidades, grupos y posibles anomalías.

Caso típico:

Analizar si existe correlación entre ingreso y gasto, entre edad y consumo, entre tamaño y productividad, etc.

5) Histogramas

Muestran la distribución de una variable: frecuencias, sesgos, dispersión, extremos y forma general. Son esenciales para explorar datos antes de modelar, porque permiten ver si los datos cumplen supuestos o necesitan transformaciones.

6) Diagramas de cajas (boxplots)

Se usan para evaluar la variabilidad, los valores atípicos y el comportamiento de una variable dentro de diferentes categorías.

Valor en extracción de conocimiento:

Revela asimetrías, grupos anómalos y discrepancias que pueden influir en modelos predictivos o de agrupación.

7) Gráficos avanzados: PCA, clusters y redes

- **Visualización con PCA**

Reduce dimensionalidad y permite proyectar datos complejos en 2D o 3D.

Es fundamental para explorar patrones ocultos y detectar agrupamientos naturales antes de aplicar modelos.

- **Visualización de clustering**

Cuando un algoritmo como K-Means, DBSCAN o jerárquico forma grupos, la visualización permite validar si realmente existen sectores diferenciados o si el modelo está forzando separaciones.

- **Grafos y redes**

Permiten representar relaciones de proximidad o interacción entre elementos: redes sociales, nodos conectados, sistemas distribuidos, etc.

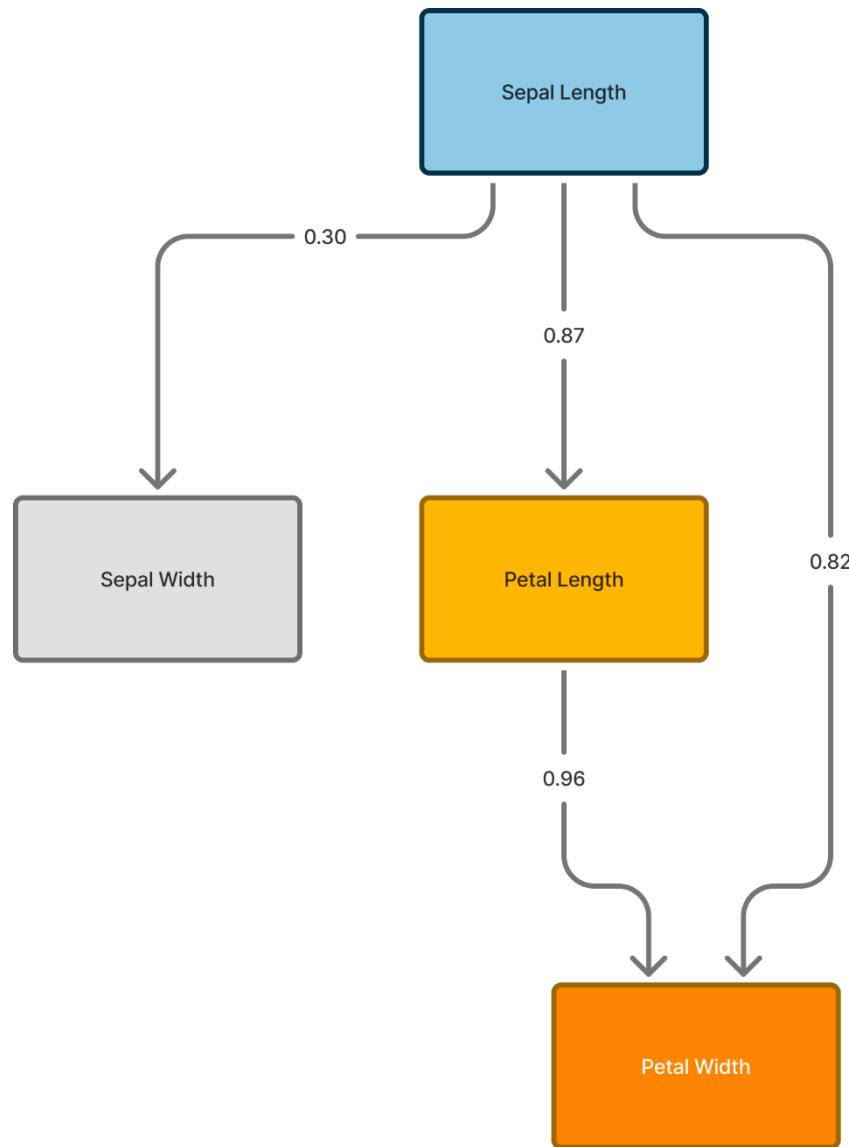
2.3. Relación con la extracción del conocimiento

La visualización juega un papel directo en el proceso KDD (Knowledge Discovery in Databases) por varias razones:

- **Apoya la comprensión inicial del dataset** mediante análisis exploratorio.
- **Ayuda a seleccionar técnicas de minería** basándose en patrones visibles.
- **Valida resultados** comparando lo que el modelo dice contra lo que se ve en los datos.
- **Facilita comunicar hallazgos** a un público no técnico.
- **Accelera la toma de decisiones**, ya que las visualizaciones reducen ambigüedades.

En pocas palabras, visualizamos para entender, para validar y para comunicar. Sin visualización, el conocimiento no se comparte ni se aprovecha correctamente.

2.4. Esquema y diagrama sugerido



3. Conclusiones

Las técnicas de visualización no solo representan datos: permiten entender, cuestionar y validar lo que ocurre dentro de un sistema. Cada técnica aporta una perspectiva distinta: unas destacan comparaciones, otras explican distribuciones, otras muestran tendencias, relaciones o estructuras complejas. Para extraer conocimiento de forma efectiva, no basta con aplicar algoritmos; es necesario ver lo que los datos están diciendo a través de gráficos claros, interpretables y adecuados al contexto.

En investigaciones aplicadas, un buen gráfico puede revelar patrones que jamás se verían en una tabla o al revisar columnas aisladas. Por eso, la visualización es un componente inseparable del análisis y una herramienta crítica para comunicar hallazgos en ambientes académicos, ejecutivos y científicos.

4. Fuentes de apoyo

- Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.
- Knafllic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Ware, C. (2013). *Information Visualization: Perception for Design*. Morgan Kaufmann.
- Zhou, M. X., & Chen, M. (2020). *Human-Centered Data Visualization*. MIT Press.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press.
- Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59–67.