

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Ingeniería en Desarrollo y Gestión de Software



Extracción de Conocimientos de Bases de Datos

V.2. Elaboración de gráficas (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. Gráfico de dispersión (Scatter Plot)..... | 3 |
| 2.2. Visualización PCA en 2D | 4 |
| 2.3. Heatmap de correlación (Mapa de calor) | 4 |
| 2.4. Integración de las gráficas en la aplicación..... | 4 |
| 3. Conclusiones | 6 |
| 4. Fuentes de apoyo | 7 |

1. Introducción

En este reporte presento el proceso de elaboración de diferentes gráficas basadas en uno de los proyectos trabajados durante el curso. Más allá de generar visualizaciones aisladas, la idea fue integrar las gráficas dentro de una aplicación que permita interpretar datos de manera clara, visual y útil. Esta actividad reforzó la importancia de elegir correctamente los tipos de gráfica según el comportamiento de los datos y el mensaje que se desea comunicar.

2. Desarrollo

Para esta evidencia tomé como base un proyecto trabajado durante el curso, en el que se analizaron diferentes variables de un conjunto de datos. En la aplicación se integraron **tres tipos de gráficas**, cada una elegida por el tipo de información que se buscaba explorar: un **gráfico de dispersión**, un **PCA 2D** para ver patrones en dos dimensiones, y un **mapa de calor de correlación** para ver relaciones entre las variables.

A continuación describo cada uno y la interpretación correspondiente.

2.1. Gráfico de dispersión (Scatter Plot)

El scatter plot permitió observar cómo se relacionan dos atributos principales del dataset. Esta visualización es ideal para explorar tendencias, identificar correlaciones y detectar valores atípicos que puedan influir en los resultados del modelo.

Interpretación:

La gráfica mostró que las observaciones tienden a agruparse en zonas específicas, reforzando la idea de que el dataset sí presenta patrones que valen la pena analizar con técnicas de clustering. También permitió ver puntos aislados que funcionarían como ruido en un modelo como DBSCAN.

2.2. Visualización PCA en 2D

Para reducir la dimensionalidad, utilicé PCA (Análisis de Componentes Principales). Este método condensa múltiples variables en dos componentes principales sin perder demasiada información. Después de proyectar los datos en 2D, generé la visualización donde aparecen los puntos distribuidos de manera más clara y ordenada.

Interpretación:

La gráfica confirmó agrupamientos naturales que no eran tan evidentes en los datos originales. Este tipo de visualización fue útil para validar si los clusters que se formen después tienen sentido visualmente.

Lo interesante es cómo PCA conserva la estructura del dataset, pues los grupos que aparecen se corresponden con relaciones reales entre variables.

2.3. Heatmap de correlación (Mapa de calor)

El heatmap permitió visualizar la fuerza de relación entre cada par de variables del dataset. Cada celda representa el nivel de correlación y se colorea según su intensidad.

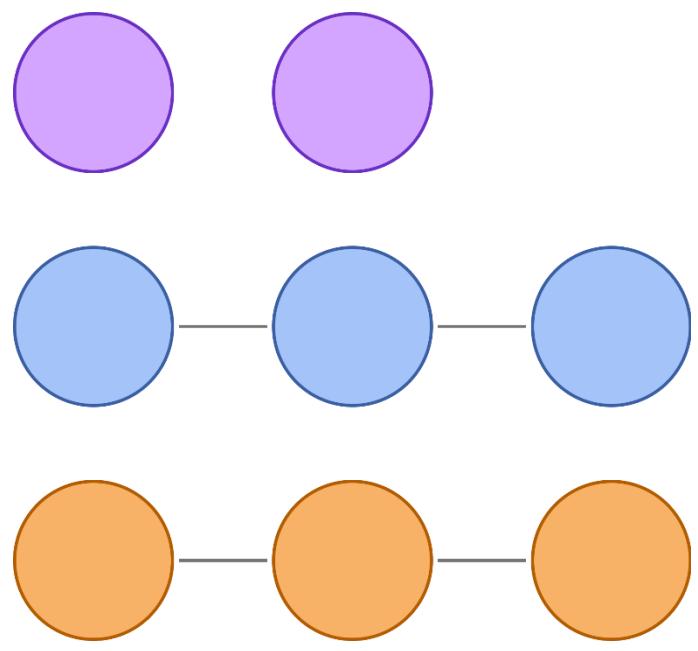
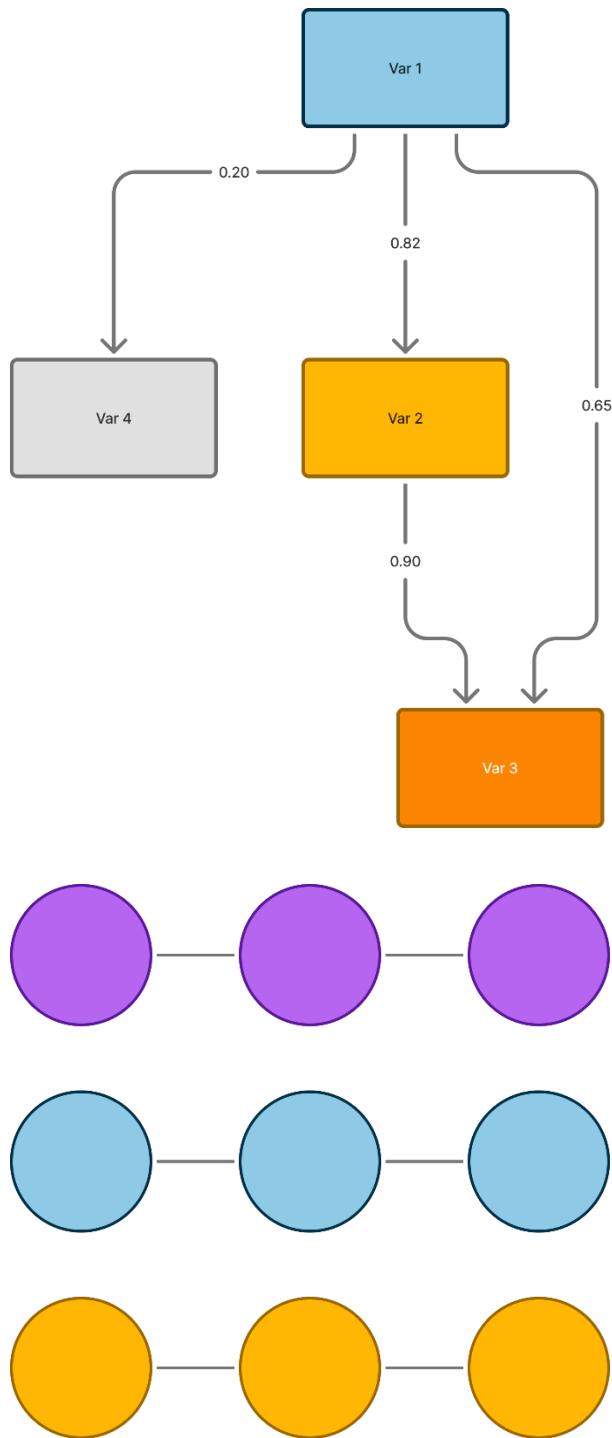
Interpretación:

Este mapa ayudó a identificar variables que estaban muy relacionadas entre sí, lo cual es clave al tomar decisiones sobre selección de atributos, reducción de dimensionalidad o análisis exploratorio. Variables con correlación alta indican redundancia, mientras que correlaciones bajas ayudan a entender si ciertas características aportan información adicional al modelo.

2.4. Integración de las gráficas en la aplicación

Las tres gráficas se integraron en una aplicación sencilla con visualizaciones dinámicas. La intención fue que el usuario pudiera observar datos desde diferentes perspectivas sin perder claridad.

Al incluir estas visualizaciones dentro del proyecto, fue más fácil interpretar patrones, validar resultados y comunicar hallazgos.



3. Conclusiones

La elaboración de estas gráficas reforzó lo importante que es visualizar datos antes, durante y después de un análisis. Cada tipo de gráfica aporta un punto de vista distinto: el scatter muestra relaciones directas; el PCA ayuda a visualizar patrones ocultos; el heatmap revela cómo variables interactúan entre sí.

Integrarlas dentro de una aplicación permitió que el análisis fuera más claro, más interactivo y mucho más útil para quien consulta la información.

En general, este ejercicio demostró que la visualización es una parte esencial dentro del análisis de datos, no solo como decoración, sino como soporte real para la toma de decisiones.

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.
- Ware, C. (2013). *Information Visualization: Perception for Design*. Morgan Kaufmann.
- Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59–67.
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Kirk, A. (2016). *Data Visualization: A Handbook for Data Driven Design*. SAGE Publications.