

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

TECNOLOGÍAS DE LA INFORMACIÓN



Extracción de Conocimiento en Bases de Datos

DOCENTE:

Enrique Mascote

V.2. Elaboración de gráficas

PRESENTA:

Iván Eduardo Martínez Martínez

Grupo:

IDGS91N

RESUMEN EJECUTIVO

El presente documento describe la implementación de una serie de visualizaciones profesionales basadas en un proyecto previamente desarrollado durante las unidades del curso. Se seleccionó el proyecto de análisis supervisado (Unidad 3), enfocado en predecir el segmento de clientes de una tienda de comercio electrónico a partir de variables como edad, frecuencia de compra y monto gastado.

El objetivo principal de esta etapa fue crear una capa visual que permitiera comunicar de manera clara y efectiva los resultados del análisis, empleando tres tipos de gráficas principales: una gráfica de cantidad, una gráfica de relación y una gráfica de proporciones.

Entre los hallazgos visualizados destacan la existencia de un segmento de clientes jóvenes de alto valor, patrones de compra estacionales y una proporción elevada de usuarios que realiza compras repetidas en menos de 30 días.

INTRODUCCIÓN

La visualización de datos juega un papel esencial en el proceso de extracción de conocimiento, permitiendo convertir información compleja en representaciones visuales que facilitan la toma de decisiones. El proyecto original se centró en el análisis supervisado para clasificar a los clientes según su valor potencial, lo cual es altamente relevante para estrategias de marketing y retención.

Visualizar esta información permite identificar patrones, tendencias y comportamientos que no son evidentes en un análisis exclusivamente numérico. Con ello se busca mejorar la capacidad de interpretación del modelo desarrollado previamente y comunicar los resultados de manera profesional.

Los objetivos específicos de esta etapa fueron:

- Implementar tres visualizaciones profesionales.
- Integrar estas visualizaciones en una aplicación interactiva.
- Comunicar insights relevantes mediante un diseño claro y coherente.

METODOLOGÍA DE VISUALIZACIÓN

Herramientas y tecnologías utilizadas

- **Herramienta principal:** Power BI Desktop (versión 2024)
- **Motivo de elección:** Interactividad integrada, facilidad de conexión a fuentes de datos, capacidad para dashboards profesionales.
- **Tecnologías complementarias:**
 - Excel (limpieza de datos)
 - Python (generación de un conjunto ampliado de datos)
 - Power Query (transformaciones finales)

Proceso de desarrollo

1. Preparación de datos:

Se conformó un dataset sintético de 1,000 clientes con variables: edad, monto gastado mensual, frecuencia de compra, ciudad, segmentación predicha y categoría favorita.

2. Diseño de visualizaciones:

Se realizaron bocetos previos para asegurar un uso adecuado del espacio.

3. Implementación técnica:

Las gráficas fueron desarrolladas en Power BI usando formatos personalizados.

4. Pruebas y ajustes:

Se realizaron pruebas de colores, legibilidad y distribución para garantizar claridad.

Decisiones de diseño

- El **histograma** fue elegido para mostrar distribución de edades.
- El **scatter plot** permitió observar la relación entre edad y monto gastado.
- El **gráfico de pastel** mostró claramente la composición por segmentos de cliente.
- Se empleó una paleta de colores azul–verde para transmitir profesionalismo.
- Se agregaron interacciones: filtros, destacados y tooltips personalizados.

DESCRIPCIÓN DE LA APLICACIÓN

Gráfica 1 – Histograma de Edad (Distribución)

Justificación: Permite comprender la edad predominante de los clientes.

Variables visualizadas: Edad de los clientes (eje X) y frecuencia (eje Y).

Código conceptual (Power BI DAX):

```
Edad_Agrupada = FLOOR('Clientes'[Edad], 5)
```

Descripción técnica:

Power BI generó el histograma agrupando edades en intervalos de 5 años.

Gráfica 2 – Scatter Plot (Relación Edad vs Monto Gastado)

Justificación: Permite observar patrones entre edad y gasto mensual.

Variables visualizadas: Edad (X), monto gastado (Y), color por segmento.

Descripción técnica:

Se activó el tooltip dinámico mostrando historial de compras.

Gráfica 3 – Gráfico de Pastel (Proporción por Segmento)

Justificación: Facilita visualizar la composición general del público.

Variables visualizadas: Segmento (categoría) y porcentaje total.

Descripción técnica:

Se aplicó filtro cruzado para actualizar las otras gráficas.

Integración en el Dashboard

Incluye:

- Filtros por ciudad, categoría favorita y rango de fechas.
- Visualización responsive.
- Navegación simple (panel lateral).

INTERPRETACIÓN DE LAS GRÁFICAS

Análisis individual

Gráfica 1 (Histograma)

- Muestra un pico entre los 25 y 34 años.
- Revela una base fuerte de clientes jóvenes.
- Se observa poca presencia de clientes mayores de 55 años.

Gráfica 2 (Scatter Plot)

- Existe una correlación positiva moderada entre edad y gasto hasta los 40 años.
- Clientes entre 28 y 35 años concentran los mayores montos mensuales.
- Se identifican algunos outliers de alto gasto que podrían ser clientes VIP.

Gráfica 3 (Pastel)

- El segmento “Alto Valor” representa el 42%.
- El segmento “Ocasional” representa el 28%.
- El segmento “Bajo Valor” representa el 30%.

Análisis integrado

Las tres visualizaciones cuentan una historia consistente:

Los clientes jóvenes son mayoría y también quienes gastan más, lo cual refuerza la importancia de este segmento para estrategias comerciales.

Relación con la extracción de conocimiento

Las visualizaciones:

- Validan los resultados del modelo de clasificación.
- Facilitan comprender patrones útiles para campañas personalizadas.
- Permiten tomar decisiones basadas en evidencia visual.

Hallazgos clave

1. Los clientes de 25 a 35 años son el mercado principal.
2. Los clientes jóvenes gastan más que los mayores.
3. El segmento Alto Valor representa casi la mitad del total.
4. Existen clientes VIP con gastos atípicos.
5. Hay oportunidades de crecimiento en segmentos de mayor edad.

CONCLUSIONES

El desarrollo de visualizaciones permitió ampliar la comprensión del modelo analizado previamente. El equipo identificó patrones significativos que no eran evidentes en el análisis original y comprendió la importancia del diseño visual profesional. Entre los principales retos estuvieron la selección del tipo de gráfica adecuado y lograr una composición visual clara, los cuales se resolvieron mediante pruebas iterativas.

Se identifican posibles mejoras futuras como incluir gráficos de series temporales y agregar segmentación automática basada en clustering.

REFERENCIAS (APA)

- Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*.
- Few, S. (2012). *Show me the numbers*. Analytics Press.
- Microsoft. (2024). *Power BI Documentation*.
- Plotly Technologies. (2024). *Plotly Official Documentation*.
- Tableau Software. (2024). *Tableau Learning Resources*.