

Visualización Avanzada y Automatización del Análisis de
Datos

Tema 4. Modelización y análisis exploratorio para visualización

Índice

Esquema

Ideas clave

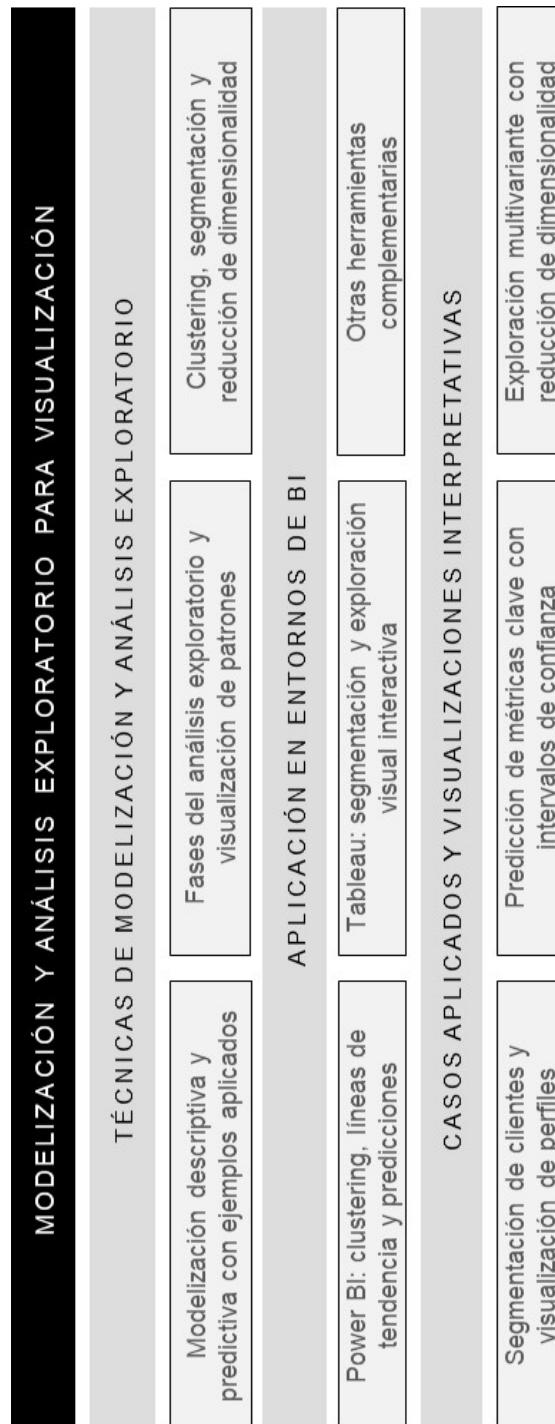
- 4.1. Introducción y objetivos
- 4.2. Técnicas básicas de modelización predictiva y descriptiva
- 4.3. Análisis exploratorio orientado a visualización
- 4.4. Agrupamiento, segmentación y reducción de dimensionalidad
- 4.5. Aplicación práctica de técnicas en entornos de BI
- 4.6. Casos aplicados con visualizaciones interpretativas
- 4.7. Resumen y conclusiones
- 4.8. Referencias bibliográficas

A fondo

Data Science Portfolio Project: Python y Tableau – PCA y Clustering

Learn Cluster Analysis in Power BI in 3 mins

Test



4.1. Introducción y objetivos

Detectar qué factores influyen en las ventas, anticipar tendencias de abandono de clientes, o descubrir segmentos con necesidades comunes. Estas preguntas, habituales en cualquier entorno profesional, encuentran respuesta gracias a la combinación de modelización de datos y análisis visual exploratorio. No se trata solo de predecir, sino de comprender qué sucede y por qué sucede, integrando técnicas analíticas con representaciones gráficas que faciliten la interpretación.

La capacidad de transformar grandes volúmenes de información en visualizaciones comprensibles permite validar hipótesis, identificar patrones ocultos y comunicar resultados con claridad. Desde la segmentación mediante *clustering* hasta la reducción de dimensionalidad con métodos como PCA o t-SNE, las organizaciones disponen hoy de un abanico de enfoques que potencian su inteligencia analítica y mejoran la toma de decisiones basada en datos.

En este tema se explorarán los principales métodos de modelización descriptiva y predictiva orientados a visualización, así como estrategias para analizar relaciones complejas de manera gráfica. Además, se presentarán ejemplos aplicados con Power BI y Tableau que permitirán al alumnado conocer las posibilidades reales de estas técnicas y desarrollar competencias prácticas para su implantación en proyectos de BI.

Al finalizar este tema, el alumnado será capaz de:

- ▶ Comprender los fundamentos de la modelización predictiva y descriptiva aplicados a la visualización de datos.
- ▶ Aplicar técnicas de análisis exploratorio para identificar patrones, relaciones y segmentos relevantes.
- ▶ Implementar procesos de agrupamiento y reducción de dimensionalidad en entornos profesionales de BI.
- ▶ Integrar visualizaciones interpretativas que acompañen el análisis y faciliten la toma de decisiones.
- ▶ Utilizar herramientas como Power BI y Tableau para desarrollar soluciones de exploración visual basadas en datos reales.

4.2. Técnicas básicas de modelización predictiva y descriptiva

La modelización estadística constituye un pilar esencial de la analítica avanzada. A través de estos métodos, es posible comprender cómo se relacionan distintas variables, identificar patrones de comportamiento y proyectar tendencias con un nivel de rigor que supera el análisis meramente descriptivo. Aplicadas de forma visual, estas técnicas contribuyen a que perfiles de negocio puedan interpretar resultados complejos sin requerir formación estadística especializada.

En entornos de BI, la capacidad de combinar modelización y visualización facilita no solo la validación de hipótesis, sino también la comunicación de hallazgos de forma intuitiva. Desde la predicción de indicadores clave, hasta la explicación de factores que impulsan ciertos resultados, la modelización predictiva y descriptiva amplía el alcance de la toma de decisiones informadas, siempre que se acompañe de representaciones gráficas claras y comprensibles.

Modelización descriptiva orientada a visualización

La modelización descriptiva se centra en explicar el comportamiento de los datos y determinar qué relaciones existen entre las variables de interés. Por ejemplo, un análisis de correlaciones puede desvelar si la satisfacción de los clientes depende más del tiempo de entrega o de la atención posventa. Estas relaciones, expresadas en métricas estadísticas, cobran vida cuando se representan mediante gráficos adecuados que evidencian patrones y dependencias.

Una técnica clásica de modelización descriptiva es la regresión lineal, que permite estimar el impacto de una o varias variables independientes sobre una variable dependiente. En contextos de BI, estas regresiones pueden visualizarse con diagramas de dispersión enriquecidos con líneas de tendencia y bandas de confianza, lo que facilita detectar la fuerza y el sentido de la relación. La claridad de la representación contribuye a que los usuarios valoren la solidez del modelo sin necesidad de consultar tablas numéricas complejas.

Otro ejemplo frecuente es el análisis de varianza (ANOVA), empleado para comparar grupos y evaluar si existen diferencias significativas entre ellos. Este método resulta especialmente útil en escenarios de segmentación de clientes o evaluación de campañas de marketing. La visualización mediante gráficos de cajas o diagramas de violín ayuda a ilustrar cómo se distribuyen los datos en cada grupo y qué nivel de variabilidad presentan, reforzando la comprensión del resultado.

Modelización predictiva y su valor interpretativo

La modelización predictiva permite estimar el comportamiento futuro de un fenómeno a partir de datos históricos. Entre las técnicas más utilizadas destacan la regresión logística, los árboles de decisión y los modelos de series temporales. Cada una de estas metodologías se adapta a diferentes tipos de problemas y proporciona resultados que pueden representarse de forma visual para simplificar su interpretación.

Por ejemplo, la regresión logística se aplica para predecir variables categóricas, como la probabilidad de que un cliente abandone un servicio. Las predicciones pueden plasmarse en gráficos de barras que muestren la probabilidad estimada de cada clase o en mapas de calor que relacionen las variables explicativas con el resultado previsto. Estos recursos visuales facilitan identificar umbrales críticos y comprender qué factores determinan la probabilidad final.

Los árboles de decisión, por su parte, ofrecen una representación jerárquica de las reglas que conducen a una predicción. Este enfoque es especialmente apreciado en entornos de negocio por su transparencia: cada división del árbol se interpreta como una condición lógica que explica cómo se llega al resultado. Los diagramas de árbol permiten seguir el razonamiento del modelo paso a paso y comunicar la lógica predictiva a audiencias no técnicas.

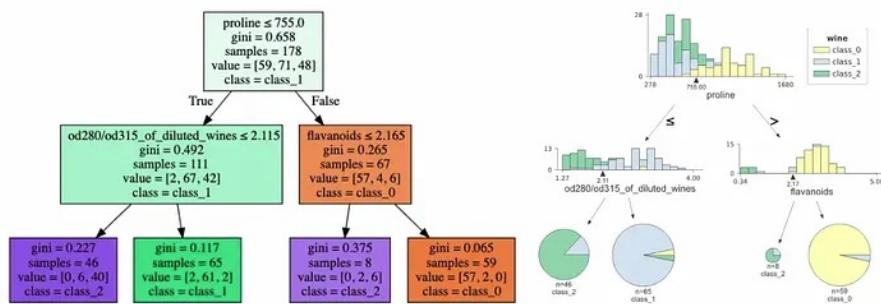


Figura 1. Visualización comparativa de un árbol de decisión con las bibliotecas scikit-learn (izquierda) y dtreeviz (derecha), mostrando la segmentación de clases y la distribución de variables predictoras.

Fuente: <https://medium.com/@akarabaev96/visualization-and-text-representation-of-a-decision-tree-ab52787b25c0>.

Ejemplos prácticos de técnicas básicas en entornos de BI

En la práctica, muchas herramientas de BI incorporan funciones de modelización básica que pueden configurarse sin programar código. Power BI, por ejemplo, permite realizar regresiones lineales y visualizar la relación entre variables con líneas de tendencia en gráficos de dispersión. Este tipo de recursos integrados agiliza la construcción de modelos exploratorios y la validación de hipótesis de negocio.

Tableau ofrece la posibilidad de añadir líneas de referencia y modelos de predicción sencillos que enriquecen los gráficos con componentes analíticos. Así, un responsable comercial puede superponer una previsión de ventas sobre los datos históricos y comprobar visualmente si los objetivos anuales son alcanzables. Estas funcionalidades refuerzan la capacidad de exploración autónoma y fomentan la cultura de toma de decisiones basada en datos.

Además de Power BI y Tableau, existen otras herramientas de propósito general, como Python o R que permiten construir modelos más sofisticados y exportar los resultados en formato visual. Por ejemplo, se pueden generar gráficos de importancia de variables o curvas ROC que muestren la capacidad predictiva de un modelo. La combinación de entornos especializados con plataformas de BI facilita integrar estos análisis en cuadros de mando interactivos.

4.3. Análisis exploratorio orientado a visualización

El análisis exploratorio de datos es el punto de partida de cualquier proceso analítico riguroso. Su objetivo principal no es confirmar hipótesis previas, sino descubrir patrones, relaciones y anomalías que pueden inspirar nuevas preguntas de negocio. Cuando este enfoque se combina con visualizaciones interactivas, se convierte en una herramienta potente para generar ideas e impulsar decisiones informadas.

A diferencia de los análisis confirmatorios, el carácter exploratorio prioriza la flexibilidad y la experimentación. La posibilidad de filtrar, agrupar y desglosar datos en tiempo real facilita identificar tendencias ocultas y conexiones inesperadas. Esta capacidad resulta especialmente valiosa en entornos de BI, donde la rapidez y la claridad visual marcan la diferencia en la adopción de las soluciones analíticas.

Principios y fases del análisis exploratorio

El análisis exploratorio se basa en algunos principios fundamentales: iteración, flexibilidad y enfoque visual. La iteración implica que el proceso no sigue una secuencia lineal, sino que requiere reformular preguntas y redefinir filtros conforme aparecen nuevos hallazgos. La flexibilidad permite al analista adaptar las herramientas y técnicas a la naturaleza cambiante de los datos y las necesidades de exploración.

Una fase inicial, imprescindible es la evaluación de la calidad de los datos. Antes de interpretar cualquier patrón, es necesario detectar valores ausentes, duplicados o inconsistentes que puedan sesgar los resultados. Esta tarea puede apoyarse en histogramas, gráficos de dispersión y tablas resumen que permitan examinar la distribución y la integridad de las variables principales.

A medida que se profundiza, el análisis exploratorio incorpora técnicas de reducción de complejidad, como la agrupación de categorías o la normalización de escalas. Estas prácticas facilitan la comparación entre segmentos y la identificación de relaciones significativas. El resultado final no es una única respuesta, sino un conjunto de hipótesis visuales que orientan fases posteriores de modelización y validación.

Visualización de patrones y relaciones

La representación gráfica es el núcleo del análisis exploratorio. Los gráficos de dispersión son una herramienta esencial para evaluar correlaciones y detectar concentraciones de datos que puedan indicar asociaciones relevantes. Superponer líneas de tendencia o zonas de densidad incrementa la capacidad interpretativa y permite priorizar áreas de interés.

Otro recurso habitual son los diagramas de caja y bigotes, que resumen la distribución de una variable y resaltan valores atípicos. En contextos de negocio, estas visualizaciones resultan útiles para identificar clientes con comportamientos extremos o productos con rendimientos inusuales. La detección temprana de estos casos puede ser clave para ajustar estrategias comerciales o priorizar intervenciones.

Las matrices de correlación completan el repertorio exploratorio al ofrecer una visión sintética de cómo se relacionan múltiples variables. Codificar la fuerza y el sentido de las relaciones mediante colores facilita que perfiles no técnicos comprendan rápidamente qué factores están más conectados. Este tipo de visualización es frecuente en informes ejecutivos por su capacidad de síntesis y claridad.

Herramientas y entornos de análisis exploratorio

Power BI y Tableau disponen de funciones específicas para el análisis exploratorio visual. En Power BI, las segmentaciones dinámicas y los paneles interactivos permiten filtrar datos y observar cómo cambian las visualizaciones en tiempo real. Esta interacción estimula la generación de hipótesis y facilita el debate entre distintos perfiles de usuario.

Tableau destaca por su enfoque orientado a la exploración intuitiva. Su funcionalidad de *Show Me* sugiere automáticamente representaciones adecuadas según el tipo de dato seleccionado. Además, la posibilidad de crear vistas detalladas con un solo clic agiliza la navegación entre niveles de información y fomenta la curiosidad analítica.

Para proyectos que requieren mayor sofisticación estadística, entornos como R o Python ofrecen bibliotecas especializadas en visualización exploratoria, como *ggplot2* o *seaborn*. Estas herramientas permiten construir gráficos complejos y personalizarlos al detalle, exportando los resultados para integrarlos en cuadros de mando o informes interactivos. La combinación de plataformas especializadas con soluciones de BI incrementa la potencia exploratoria de cualquier organización.

4.4. Agrupamiento, segmentación y reducción de dimensionalidad

En proyectos de análisis avanzados, la simple visualización de métricas agregadas no es suficiente para comprender los patrones que subyacen en los datos. Muchas veces, la clave reside en identificar segmentos de comportamiento similar o en reducir la complejidad de conjuntos con decenas de variables. Las técnicas de agrupamiento y reducción de dimensionalidad permiten descubrir estructuras latentes y simplificar la información de forma que sea comprensible y accionable.

Estas metodologías no sólo aportan valor en fases exploratorias, sino que también resultan esenciales para construir modelos predictivos más precisos y para diseñar visualizaciones que transmitan información sin saturar al usuario. La combinación de segmentación, proyección en espacios reducidos y representación gráfica de los resultados es una de las prácticas más eficaces para comunicar hallazgos complejos en entornos de BI.

Técnicas de clustering y segmentación visual

El clustering consiste en agrupar observaciones que comparten características similares. Una de las técnicas más conocidas es *k-means*, que asigna cada punto de datos al grupo más cercano según su distancia al centroide. Este método resulta útil para segmentar clientes, productos o regiones con comportamientos homogéneos. Visualizar los clústeres con colores distintos facilita interpretar las diferencias y entender qué variables definen cada segmento.

Otro enfoque de segmentación es el clustering jerárquico, que construye un árbol de relaciones (dendrograma) mostrando cómo se agrupan los elementos progresivamente. Esta técnica permite determinar el número óptimo de segmentos en función de la distancia entre grupos. Su representación visual en forma de árbol facilita explicar la estructura de la segmentación a audiencias no técnicas.

Además, algunos entornos de BI, como Power BI y Tableau, incorporan funciones automáticas de clustering que generan segmentaciones con unos pocos clics. Estas funcionalidades aceleran la exploración inicial y ofrecen un punto de partida para análisis más profundos. Es importante, sin embargo, validar siempre los resultados con métricas de cohesión interna y criterio experto antes de tomar decisiones estratégicas.

Métodos de reducción de dimensionalidad (PCA, t-SNE)

La reducción de dimensionalidad permite simplificar conjuntos de datos con muchas variables manteniendo la información más relevante. El método más utilizado es el Análisis de Componentes Principales (PCA), que transforma las variables originales en un nuevo conjunto de componentes que capturan la mayor parte de la varianza. Esta técnica facilita representar datos multivariantes en gráficos bidimensionales donde se visualizan tendencias y agrupaciones.

Por su parte, *t-SNE* (*t-distributed Stochastic Neighbor Embedding*) es un método no lineal que proyecta datos en espacios de baja dimensión preservando las relaciones de proximidad local. Es especialmente eficaz para visualizar clústeres en problemas de alta complejidad, como segmentación de clientes basada en comportamiento digital o análisis de textos. Aunque su interpretación numérica es menos directa que la del PCA, su capacidad de descubrimiento visual es muy apreciada.

En entornos profesionales, es habitual utilizar la reducción de dimensionalidad como paso previo a la visualización o al clustering. Por ejemplo, tras aplicar PCA, se pueden representar los dos primeros componentes en un diagrama de dispersión y colorear los puntos según el segmento asignado. Esta práctica facilita entender la estructura global de los datos y comunicarla de forma intuitiva.

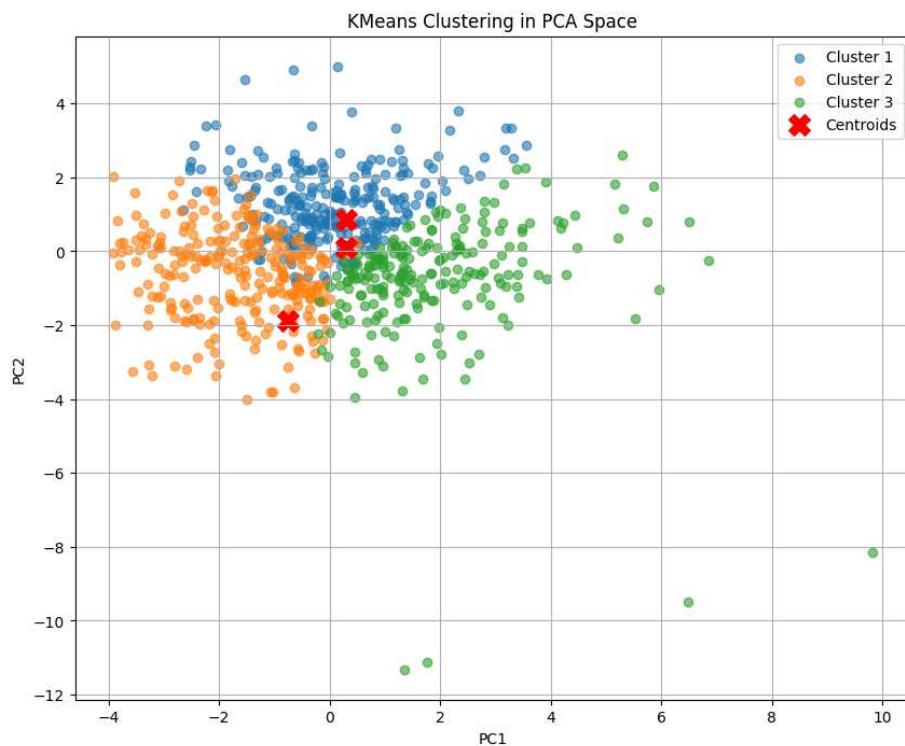


Figura 2. Distribución de observaciones en el espacio de componentes principales (PCA) con agrupación k-means en tres clústeres y visualización de centroides. Fuente: <https://blog.gopenai.com/pca-and-clustering-in-sports-analytics-4e7d92972f5c>.

Interpretación visual de resultados y casos de uso

La representación gráfica de los resultados es esencial para que la reducción de dimensionalidad y el clustering tengan impacto en la toma de decisiones. Un ejemplo frecuente es la matriz de calor que muestra cómo contribuye cada variable original a los componentes principales. Esta visualización ayuda a comprender qué factores explican la mayor parte de la varianza y orienta futuras hipótesis de negocio.

Los gráficos de dispersión con códigos de color son otra herramienta clave. Al mostrar los puntos proyectados en dos dimensiones y diferenciarlos por segmento, se genera una vista clara de cómo se distribuyen los grupos. Esta representación es especialmente eficaz en presentaciones ejecutivas, ya que permite identificar con rapidez las diferencias entre perfiles y discutir estrategias específicas.

Por último, es recomendable acompañar estos gráficos de indicadores de calidad del modelo, como la varianza explicada o el coeficiente de silueta en el caso del clustering. Incluir estas métricas en las visualizaciones refuerza la credibilidad del análisis y ayuda a evaluar si la segmentación o la reducción de dimensionalidad son suficientemente robustas para basar decisiones estratégicas en ellas.

4.5. Aplicación práctica de técnicas en entornos de BI

La utilidad real de la modelización y el análisis exploratorio depende de su capacidad para integrarse en herramientas que formen parte del flujo habitual de trabajo. Hoy en día, plataformas como Power BI y Tableau permiten aplicar técnicas de segmentación, reducción de dimensionalidad y análisis predictivo de manera accesible, con interfaces visuales que simplifican la interpretación de los resultados.

Este enfoque democratiza el uso de métodos avanzados, acercándolos a perfiles profesionales que no tienen formación técnica especializada. La posibilidad de combinar procesos analíticos con paneles interactivos incrementa la adopción de estas técnicas y refuerza la cultura de toma de decisiones basada en datos.

Power BI: Clustering, segmentación y reducción de dimensionalidad

Power BI incluye funcionalidades de clustering que permiten agrupar datos automáticamente en segmentos con características similares. Esta función se activa desde el panel de *Análisis* en un gráfico de dispersión, donde el usuario puede definir el número de clústeres o dejar que el algoritmo determine la mejor agrupación. La visualización se actualiza en tiempo real, mostrando cada segmento con un color diferente y etiquetas que identifican su centroide.

Annual Income (k\$) and Spending Score (1-100)

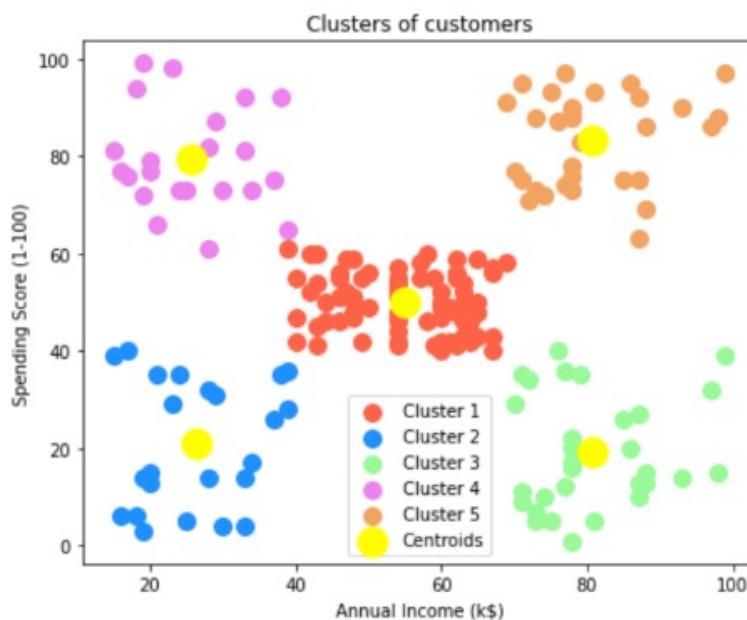


Figura 3. Segmentación de clientes en cinco clústeres según ingresos anuales y nivel de gasto, con visualización de centroides de cada grupo. Fuente: <https://yannawut.medium.com/clustering-in-power-bi-e1c7fd5700e>.

La herramienta también facilita la creación de líneas de tendencia, que son una forma sencilla de aplicar regresiones lineales para observar relaciones entre variables. Estas líneas pueden configurarse para mostrar intervalos de confianza, resaltando la fiabilidad de la predicción visual. Esta característica resulta muy útil para validar hipótesis de negocio sin necesidad de recurrir a modelos estadísticos externos.

En proyectos que requieren reducción de dimensionalidad, Power BI permite importar datos previamente procesados en R o Python. Por ejemplo, tras realizar un PCA en R, el usuario puede cargar los componentes principales como columnas adicionales y representarlas en un diagrama de dispersión. Este enfoque híbrido aprovecha la potencia de entornos especializados sin renunciar a la interactividad de Power BI.

Tableau: Modelización básica y análisis exploratorio

Tableau ofrece múltiples opciones para enriquecer visualizaciones con modelos predictivos sencillos. Desde el menú *Análisis*, es posible agregar tendencias lineales, cuadráticas o exponenciales que se ajustan automáticamente a los datos. Estas líneas permiten comprobar de un vistazo si existe una relación consistente entre variables y comunicarla de forma intuitiva a otros usuarios.

El clustering en Tableau se implementa arrastrando la funcionalidad de *Clúster* sobre una hoja de trabajo. La herramienta calcula automáticamente los grupos y los colorea en el gráfico de dispersión o en otros tipos de visualización. Además, genera un panel de resumen que muestra estadísticas clave de cada segmento, como el tamaño o la media de las principales variables.

Para explorar relaciones complejas, Tableau permite crear vistas combinadas con múltiples dimensiones y medidas. Por ejemplo, se puede construir una matriz de dispersión con filtros interactivos que segmenten los datos según región o período temporal. Este tipo de representaciones facilita el descubrimiento de patrones y la generación de hipótesis en fases tempranas del análisis exploratorio.

Otras herramientas: aplicaciones sencillas

Aunque Power BI y Tableau son las opciones más extendidas en entornos de BI, existen otras plataformas que permiten incorporar análisis exploratorio y modelización de forma visual. Qlik Sense, por ejemplo, integra funciones de clustering y análisis de tendencias con un enfoque basado en búsquedas asociativas que facilita la exploración de grandes volúmenes de datos.

Algunas soluciones más especializadas, como Alteryx, combinan flujos de procesamiento de datos con herramientas de modelización predictiva y generación de gráficos. Este enfoque resulta útil en proyectos que requieren preparar datos complejos antes de integrarlos en paneles visuales. La posibilidad de automatizar procesos y documentar cada paso aporta transparencia y facilita la reproducibilidad del análisis.

Por último, en casos donde se dispone de entornos de desarrollo como RStudio o notebooks de Python, es posible crear modelos avanzados y exportar gráficos interactivos en formatos web (por ejemplo, con *plotly* o *ggplotly*). Estas visualizaciones pueden integrarse posteriormente en informes dinámicos o en plataformas de BI, ampliando el alcance del análisis sin perder flexibilidad técnica.

4.6. Casos aplicados con visualizaciones interpretativas

La mejor forma de comprender el potencial de la modelización y el análisis exploratorio, es a través de ejemplos prácticos que muestran cómo estas técnicas se aplican en situaciones reales. Cada caso ilustra un reto habitual en entornos de BI y demuestra cómo la combinación de métodos estadísticos y visualización avanzada facilita tomar decisiones informadas.

En este apartado se presentan tres escenarios concretos: segmentación de clientes, predicción de métricas clave y exploración visual de datos multivariantes. Cada uno incorpora gráficos interpretativos que transforman resultados técnicos en conclusiones accesibles para perfiles no especializados.

Caso 1. Segmentación de clientes con visualización de perfiles

Una compañía de telecomunicaciones desea identificar grupos de clientes con patrones de consumo similares para personalizar sus ofertas. Mediante clustering *k-means*, se crean cuatro segmentos basados en variables como frecuencia de uso, gasto mensual y antigüedad en la compañía. El modelo identifica un clúster con alta rentabilidad y baja fidelidad, prioritario para acciones de retención.

La visualización principal consiste en un diagrama de dispersión donde cada punto representa un cliente, coloreado según su clúster. Este gráfico permite identificar de inmediato qué grupos son más homogéneos y cuáles presentan mayor dispersión interna. Además, un panel complementario muestra los promedios de cada variable por segmento, facilitando la comparación y la interpretación.

Para presentar los resultados a la dirección comercial, se diseña un dashboard que combina los gráficos de dispersión, tablas de resumen y filtros interactivos. De este modo, los responsables pueden explorar cada clúster, conocer sus características y diseñar campañas específicas. La claridad visual refuerza la confianza en el análisis y mejora la toma de decisiones.

Caso 2. Predicción de métricas clave con representación visual

Una empresa del sector retail necesita anticipar el volumen de ventas mensual con el fin de optimizar inventario y planificación logística. Utilizando un modelo de regresión lineal múltiple, se incorporan variables como campañas de marketing, precios promedio y estacionalidad. El modelo muestra que las promociones tienen un efecto positivo significativo, mientras que la estacionalidad explica parte de la variación.

El dashboard resultante presenta una serie temporal con los datos históricos de ventas y la predicción para los seis meses siguientes. Las áreas de confianza alrededor de la línea de proyección visualizan la incertidumbre asociada a la estimación, ayudando a contextualizar los riesgos. Este recurso facilita que los directivos comprendan tanto la tendencia central como la variabilidad potencial.

Además, se incluyen gráficos de barras que muestran el peso relativo de cada variable independiente sobre la predicción. Este enfoque visual permite priorizar los factores más influyentes y tomar decisiones proactivas en base a los resultados. La integración del modelo en una plataforma interactiva asegura que los análisis se actualicen automáticamente con nuevos datos.

Caso 3. Exploración visual de datos multivariantes

Una organización sin ánimo de lucro dispone de un conjunto de datos con más de 20 indicadores sociales, económicos y sanitarios de distintos municipios. El objetivo es identificar patrones de vulnerabilidad y orientar recursos hacia las zonas con mayores carencias. La complejidad de la información hace necesario aplicar reducción de dimensionalidad antes de interpretarla.

Mediante un análisis de componentes principales (PCA), se proyectan los datos en dos dimensiones que capturan el 75% de la varianza. Cada municipio se representa como un punto en el plano, permitiendo identificar conglomerados con perfiles similares. Los puntos se colorean según el nivel de renta media, revelando tendencias geográficas de exclusión social.

El informe final incluye un gráfico de dispersión con etiquetas dinámicas que muestran los indicadores principales de cada municipio al pasar el cursor. Este recurso interactivo facilita a los responsables de programa explorar los datos con autonomía y formular hipótesis sobre las prioridades de intervención. La claridad visual convierte un conjunto complejo en una base de conocimiento accesible.

4.7. Resumen y conclusiones

A lo largo de este tema se ha revisado el papel fundamental que desempeñan la modelización y el análisis exploratorio en los procesos de visualización avanzada. Estas técnicas permiten no sólo describir qué ha ocurrido, sino también comprender por qué suceden determinados fenómenos y anticipar escenarios futuros. La capacidad de integrar métodos predictivos y descriptivos en entornos visuales facilita que los hallazgos se conviertan en acciones concretas.

Se han presentado técnicas de segmentación y reducción de dimensionalidad que ayudan a simplificar conjuntos de datos complejos y a identificar patrones latentes. El uso de gráficos adecuados, como diagramas de dispersión coloreados por clúster o series temporales con intervalos de confianza, potencia la comprensión de estos análisis por parte de perfiles no técnicos. Además, la exploración interactiva contribuye a una toma de decisiones más ágil y fundamentada.

Por último, se han revisado ejemplos prácticos y funcionalidades específicas de herramientas como Power BI y Tableau, que permiten democratizar el uso de estos métodos en entornos profesionales. La combinación de técnicas estadísticas y representaciones visuales avanzadas constituye una de las competencias más relevantes en proyectos de BI orientados a crear valor a partir de los datos.

4.8. Referencias bibliográficas

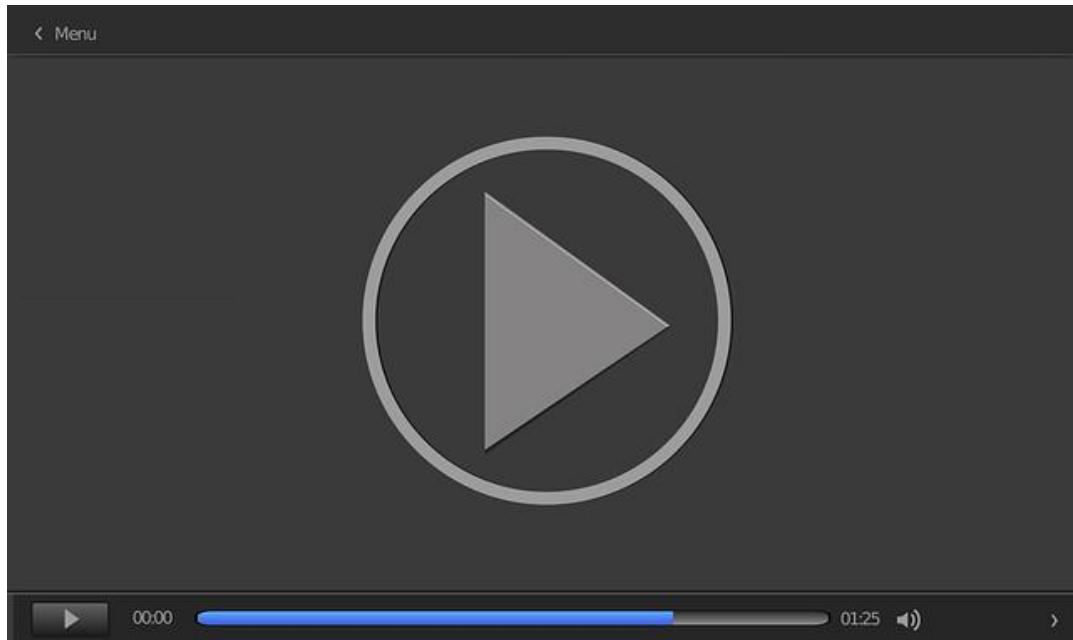
James, G., Witten, D., Hastie, T. y Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer. <https://www.statlearning.com/>

McKinney, W. (2023). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter*. O'Reilly Media. <https://wesmckinney.com/book/>

Wickham, H., Cetinkaya-Rundel, M. y Grolemund, G. (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (2nd ed.). O'Reilly Media. <https://r4ds.hadley.nz/>

Data Science Portfolio Project: Python y Tableau – PCA y Clustering

Zhao, A. (2024, marzo). *Data Science Portfolio Project: Python y Tableau – PCA y Clustering* [Vídeo]. YouTube. Recuperado de <https://www.youtube.com/watch?v=MSzZN3xighQ>



Accede al vídeo:

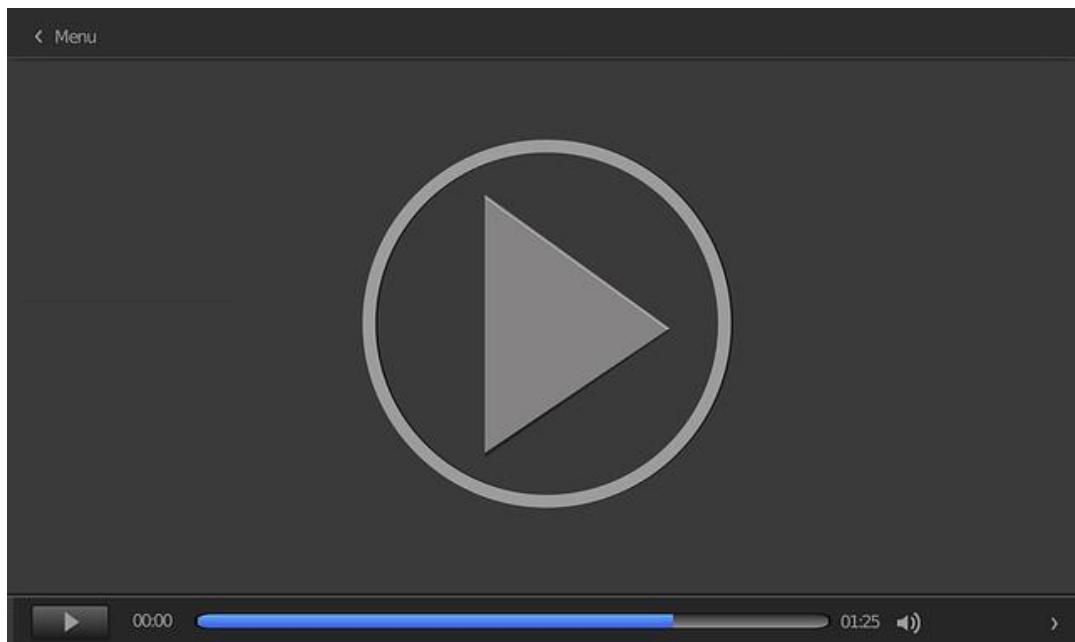
<https://www.youtube.com/embed/MSzZN3xighQ>

Vídeo práctico (~30 min) en el que Alice Zhao muestra cómo aplicar PCA y clustering con Python, y luego visualizar los resultados en Tableau. Incluye el dataset y el workbook para seguir el proyecto paso a paso, ideal para replicar en clase y reforzar la parte de reducción de dimensionalidad y segmentación visual.

Learn Cluster Analysis in Power BI in 3 mins

Data & Analytics Team. (2021). *Learn Cluster Analysis in Power BI in 3 mins* [Vídeo].

YouTube. Recuperado de <https://www.youtube.com/watch?v=SOJp1vDutIE>



Accede al vídeo:

<https://www.youtube.com/embed/SOJp1vDutIE>

Vídeo breve y directo (~3 min) que muestra cómo activar y configurar clustering en Power BI. El ejemplo utiliza un scatter plot interactivo donde se generan segmentos automáticamente, ideal como demo corta para introducir el concepto en clase.

1. ¿Qué propósito principal tiene la modelización descriptiva en un proyecto de visualización?

 - A. Predecir valores futuros automáticamente
 - B. Explicar relaciones entre variables y patrones históricos
 - C. Reducir el volumen de datos sin pérdida de información
 - D. Asignar probabilidades de clasificación

2. ¿Qué técnica es más adecuada para proyectar datos multivariantes en dos dimensiones manteniendo la varianza?

 - A. Regresión logística
 - B. Árboles de decisión
 - C. Análisis de Componentes Principales (PCA)
 - D. Clustering jerárquico

3. ¿Qué ventaja aporta la representación visual de un árbol de decisión?

 - A. Incrementa la precisión predictiva.
 - B. Permite reducir automáticamente la dimensionalidad.
 - C. Facilita la interpretación de reglas lógicas y condiciones de decisión.
 - D. Sustituye la validación estadística del modelo.

4. ¿Cuál de las siguientes afirmaciones sobre t-SNE es correcta?

 - A. Es una técnica lineal que maximiza la varianza.
 - B. Es un método no lineal que preserva relaciones locales.
 - C. Sustituye la necesidad de clustering previo.
 - D. Utiliza componentes principales como base del cálculo.

5. ¿Qué funcionalidad ofrece Power BI para segmentar datos automáticamente en un gráfico?
 - A. Creación de jerarquías de campos
 - B. Inserción de líneas de predicción
 - C. Agrupación por clustering con colores diferenciados
 - D. Generación de mapas geográficos
6. ¿Qué elemento no es propio del análisis exploratorio?
 - A. Confirmar hipótesis formuladas previamente
 - B. Identificar patrones inesperados
 - C. Iterar preguntas conforme avanza el análisis
 - D. Visualizar relaciones entre variables
7. ¿Qué métrica se utiliza habitualmente para valorar la calidad de un clustering?
 - A. R-cuadrado
 - B. Coeficiente de silueta
 - C. AUC-ROC
 - D. Error cuadrático medio
8. ¿Qué recurso visual facilita entender la contribución de variables en PCA?
 - A. Árbol de decisión
 - B. Serie temporal
 - C. Matriz de calor con pesos de cada variable en los componentes
 - D. Diagrama de barras apiladas

- 9.** ¿Qué diferencia principal tiene la regresión logística respecto a la lineal?
- A. Predice valores continuos
 - B. Es un método no supervisado
 - C. Predice variables categóricas mediante probabilidad
 - D. No requiere variables independientes
- 10.** ¿Qué herramienta destaca por su capacidad de sugerir automáticamente tipos de visualización?
- A. Power Query
 - B. Tableau (función Show Me)
 - C. Excel
 - D. DAX Studio