



Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Visualización de Datos de inmuebles en la ciudad de Valencia



Autores: Rubén Asensi Valls / Pablo García Pérez /
Manuel Martín Morante / Borja Rodrigo Garrido

Curso 2024 - 2025



Índice

1. Introducción	4
2. Las siete fases	5
2.1. Adquisición	5
2.2. Formateado	5
2.3. Filtrado	6
2.4. Minado	7
2.5. Representación	8
2.6. Refinado	9
2.7. Interacción	10
3. Resultados principales	12
3.1. Distribución de inmuebles por barrios	12
3.2. Precios promedio por metro cuadrado	12
3.3. Calidad y antigüedad de los inmuebles	12
3.4. Relación entre calidad y precio	12
3.5. Influencia de la proximidad al transporte público	12
3.6. Tendencias temporales	12
3.7. Clusters de barrios	13
4. Conclusiones	14



1. Introducción

En este proyecto hemos abordado el desafío de analizar el sector inmobiliario de Valencia a través de la visualización de datos, utilizando herramientas y metodologías aprendidas a lo largo de la asignatura. Nuestro objetivo principal ha sido explorar y responder a preguntas clave relacionadas con los precios, la calidad, la antigüedad y la distribución geográfica de los inmuebles en la ciudad, generando visualizaciones claras e interactivas que permitan una comprensión profunda de los datos.

El trabajo ha consistido en varias etapas, desde la adquisición y limpieza de los datos hasta su transformación en dashboards interactivos, en cuyo proceso hemos aplicado conceptos fundamentales de los tres bloques del curso. Este documento detalla las siete fases de la visualización de datos según lo aprendido en clase y describe cómo las hemos implementado en nuestro trabajo, justificando las decisiones tomadas en cada etapa.

A continuación, presentamos las fases de este proyecto, destacando los aprendizajes y los retos que hemos ido encontrando en la realización de cada una de ellas.

2. Las siete fases

2.1. Adquisición

La etapa de adquisición es el primer paso en cualquier proceso de visualización de datos. Consiste en recopilar los datos necesarios para abordar los objetivos que hemos propuesto inicialmente y garantizar que sean adecuados para el análisis. En esta fase, se debe considerar tanto el formato de los datos como su procedencia, asegurándonos de que sean fiables y estén actualizados. Además, es fundamental tener en cuenta la relevancia de los datos con respecto a los objetivos del proyecto, por lo que hemos priorizado aquellas fuentes que proporcionaban la información más completa y precisa. En concreto, hemos descargado los datos del siguiente repositorio de GitHub:

<https://github.com/paezha/idealista18>

El conjunto de datos mencionado fue publicado por la empresa Idealista y combinados con datos del Catastro de Valencia. Como nuestro objetivo principal es analizar el mercado inmobiliario en Valencia, hemos requerido de la obtención de datos relacionados con precios de inmuebles, características estructurales, ubicaciones geográficas... Estos datos cumplen con nuestras necesidades. Para ello, hemos recopilado información relacionada con los siguientes puntos:

1. **Datos de inmuebles en venta:** Hemos utilizado el archivo “valencia_sale.xlsx”, que contiene más de 33.000 registros detallados de propiedades, incluyendo variables como precio, superficie construida, número de habitaciones y antigüedad de los inmuebles. Estos datos nos permitieron, posteriormente, realizar análisis profundos sobre las características y tendencias del mercado.
2. **Datos geográficos:** Hemos incorporado el archivo “valencia_polygons.xlsx”, que contiene datos geográficos de los barrios de la ciudad y de los inmuebles de los mismos. Este conjunto de datos nos ha resultado crucial para realizar representaciones espaciales.

Hemos seleccionado cuidadosamente cada una de estas fuentes por su relevancia y su potencial para aportar valor al análisis. Todos los datos los hemos descargado estandarizados (Excel y RDA) para tener asegurada su compatibilidad con las herramientas que hemos utilizado en el proyecto.

Aunque no lo pueda parecer, esta fase inicial de recopilación ha sido crucial para sentar las bases del trabajo, ya que la selección de unos datos deficientes o incompletos podría haber comprometido nuestro resultado final.

2.2. Formateado

La fase de formateado consiste en organizar y estructurar los datos que hemos adquirido para que sean útiles en las etapas posteriores del análisis y la visualización. Este paso es esencial para garantizar que los datos tengan un formato coherente, con las unidades y

categorías adecuadas, y que estén listos para ser procesados sin generar errores. Además, en esta fase hemos renombrado variables para facilitar su interpretación y realizado ajustes para una mejor compatibilidad con las herramientas de análisis.

En nuestro caso, tras la adquisición de los datos, hemos realizado varias tareas de formateado para preparar los conjuntos de datos de acuerdo con las necesidades del proyecto:

1. **Conversión del tipo de ficheros:** Debido a que los datos originales estaban en formato “rda”, ha sido necesario convertirlos a un formato más adecuado para trabajar con ellos con las distintas herramientas utilizadas. Para ello, utilizamos R y algunas de sus librerías para transformarlos en archivos “csv”. En el caso de Tableau, los utilizamos en formato “xlsx”.
2. **Renombrado de columnas:** Para mejorar la comprensibilidad de las variables, renombramos las columnas de los datos geográficos y de los inmuebles. Por ejemplo, cambiamos LOCATIONNAME a NEIGHBORHOOD en el archivo de polígonos, y simplificamos nombres técnicos como UNITPRICE a PRICE_PER_M2.
3. **Transformaciones geométricas:** En los datos geográficos, hemos procesado las coordenadas de los polígonos utilizando la librería *shapely*, convirtiendo las cadenas de texto en objetos geométricos que pudieran ser interpretados con *Plotly* para realizar visualizaciones espaciales precisas.
4. **Estandarización de tipos de datos:** Convertimos las variables al tipo de dato adecuado para su análisis. Por ejemplo, asegurarnos de que las variables como PRICE y CONSTRUCTEDAREA estuvieran en formato numérico y convertir categorías como NEIGHBORHOOD y BUILDTYPE a texto categórico.
5. **Integración de datos:** Preparamos los diferentes conjuntos de datos para ser combinados posteriormente, lo que implica asegurar que los nombres de los barrios en los diferentes archivos coinciden en formato y estructura, eliminando inconsistencias en el texto y unificando criterios de nomenclatura.

Con el formateado hemos convertido datos crudos y heterogéneos en conjuntos organizados y listos para el análisis. Este paso ha sido fundamental para garantizar que todas las herramientas que hemos utilizado en las siguientes fases trabajaran de manera eficiente y sin errores.

2.3. Filtrado

La fase de filtrado consiste en seleccionar únicamente los datos relevantes para el análisis y eliminar aquellos que no aportan valor o que podrían generar ruido en los resultados. Este paso es crucial para enfocar el análisis en los objetivos planteados inicialmente, optimizar los recursos computacionales y mejorar la claridad de las visualizaciones. Un filtrado adecuado también nos permite reducir errores y evitar conclusiones erróneas debido a información irrelevante o incorrecta.

En nuestro proyecto, hemos llevado a cabo el filtrado teniendo en cuenta los objetivos de analizar el mercado inmobiliario en Valencia y responder a preguntas relacionadas con

precios, calidad y accesibilidad de las propiedades. A continuación, describimos las decisiones principales que hemos tomado durante esta etapa:

1. **Eliminación de registros con valores nulos o inconsistentes:** En el conjunto de datos de inmuebles, identificamos variables clave como PRICE, CONSTRUCTEDAREA y ROOMNUMBER que contenían algunos valores nulos. Eliminamos estos registros para garantizar la fiabilidad de los análisis y visualizaciones. Asimismo, descartamos propiedades con precios extremadamente bajos o altos, considerándolos como valores atípicos que podrían distorsionar los resultados.
2. **Selección geográfica:** Como nos enfocamos exclusivamente en los barrios de Valencia, hemos descartado los registros que no corresponden a esta área geográfica.
3. **Simplificación de categorías:** En el conjunto de datos de polígonos, eliminamos categorías que no aportaban información significativa para las visualizaciones. Por ejemplo, hemos descartado niveles de zonificación específicos (ZONELEVELID) que no estaban relacionados con nuestras preguntas de investigación.
4. **Enfoque en variables de interés:** Seleccionamos las columnas más relevantes para el análisis, como PRICE, AGE, QUALITY, DISTANCE_TO_CITY_CENTER y BUILDTYPE. Esto nos permite centrar el análisis en aspectos clave del mercado inmobiliario, evitando datos secundarios.
5. **Filtrado temporal:** Hemos optado por trabajar con los datos más recientes disponibles en los archivos para reflejar el estado actual del mercado inmobiliario, para que las conclusiones sean lo más relevantes y aplicables posibles. En concreto, del año 2018.

El filtrado nos ha ayudado a reducir el volumen de datos a un conjunto más manejable y enfocado, optimizando tanto el análisis como la presentación de resultados. Consideramos este paso de suma importancia, ya que nos permite enfocarnos de manera precisa y clara en las preguntas planteadas, eliminando cualquier información que pueda desviar la atención o complicar la interpretación.

2.4. Minado

La etapa de minado se centra en el análisis profundo de los datos mediante herramientas matemáticas, estadísticas y de minería de datos. En este paso tenemos como objetivo identificar patrones, tendencias y relaciones significativas en los datos, que nos sirvan como base para tomar decisiones en las fases posteriores, incluyendo la representación visual. En esta etapa se construye el conocimiento necesario para sustentar las posteriores visualizaciones. Las principales acciones que hemos realizado son:

1. **Cálculo de métricas descriptivas:** Analizamos indicadores como precios promedio por metro cuadrado (UNITPRICE_MEAN), antigüedad media (AGE_MEAN) y calidad promedio (QUALITY_MEAN) de los inmuebles en cada barrio. A partir de estas métricas obtenemos una visión general del comportamiento del mercado.

2. **Análisis de correlaciones:** Generamos matrices de correlación entre variables clave como PRICE, AGE, DISTANCE_TO_CITY_CENTER y QUALITY. Este análisis revela resultados como que los inmuebles con menor calidad del Catastro tienden a estar asociados a precios más altos.
3. **Análisis de patrones espaciales:** Evaluamos la proximidad de las propiedades a estaciones de metro utilizando las coordenadas geográficas, identificando tendencias como la concentración de inmuebles más caros en áreas cercanas al transporte público o el centro de la ciudad.
4. **Agrupamiento de datos:** Clasificamos los barrios según características como precio promedio, calidad y antigüedad en cuanto a sus inmuebles. Este análisis ha resultado en grupos homogéneos de barrios, facilitando la comparación entre zonas similares.
5. **Identificación de outliers:** Al detectar valores atípicos en variables como el precio y el tamaño de los inmuebles, hemos analizado dichos registros para decidir si debíamos excluirlos de las siguientes etapas o si reflejaban casos particulares relevantes para el análisis.

Como es obvio, el minado de datos ha sido fundamental para construir una base sólida de conocimiento sobre el mercado inmobiliario en Valencia. Los patrones y relaciones descubiertos en esta fase guían nuestras decisiones durante las etapas siguientes. Todo el conocimiento generado aquí se utiliza para preparar las representaciones gráficas y el dashboard.

2.5. Representación

La fase de representación consiste en seleccionar los modelo visual más adecuado para comunicar los resultados obtenidos en la fase de minado. En esta etapa, traducimos la información obtenida en gráficos básicos que facilitan la comprensión de los patrones y relaciones identificados. El objetivo es presentar los datos de manera clara y efectiva, permitiéndonos interpretar rápidamente las tendencias y los insights más relevantes.

Nosotros hemos seleccionado distintos modelos visuales para reflejar las características del sector inmobiliario en la ciudad. Cada elección ha estado guiada por los objetivos planteados inicialmente y los patrones descubiertos en la fase de minado. A continuación, detallamos las decisiones tomadas:

1. **Mapas temáticos:** Utilizamos mapas coropléticos para representar características como la antigüedad media (AGE_MEAN) y la calidad promedio (QUALITY_MEAN) de los inmuebles en cada barrio. Estos gráficos permiten visualizar la distribución geográfica de estas variables y destacar las diferencias entre los barrios.
2. **Gráficos de barras:** Para analizar el número total de inmuebles por barrio (REAL_ESTATE_TOTAL), empleamos gráficos de barras horizontales. Este enfoque facilita la comparación directa entre barrios y permitió identificar rápidamente las áreas con mayor actividad inmobiliaria. También lo utilizamos para representar las

correlaciones negativas o positivas entre las variables más correlacionadas con el precio de un inmueble.

3. **Diagramas de caja:** Para analizar la distribución de precios según los barrios utilizamos estos gráficos. En concreto, para los tres barrios con precios más caros y para los tres con precios más bajos.
4. **Datos agregados o totales:** Utilizamos datos agregados para proporcionar información relevante como las viviendas de nueva obra disponibles, el barrio con más inmuebles en venta, el barrio con los precios más caros...

La representación es un paso crucial para convertir los resultados del minado en un formato comprensible y accesible. Aunque las visualizaciones en esta etapa son iniciales, sientan las bases para los refinamientos y ajustes de interactividad, como filtros para seleccionar barrios o rangos de precios.

2.6. Refinado

El paso de refinado se centra en mejorar las representaciones gráficas básicas creadas en la etapa anterior. El objetivo es hacerlas más claras, estéticamente agradables y efectivas en la comunicación de la información. Esto incluye ajustes visuales como el uso de colores, tipografías, tamaños de fuente, y la organización de los elementos gráficos para garantizar una experiencia óptima para el usuario. Además, buscamos destacar los aspectos más importantes de los datos y minimizar distracciones.

En nuestro caso, hemos dedicado especial atención al refinamiento de las visualizaciones para garantizar que transmitieran los insights del mercado inmobiliario de manera clara y atractiva. A continuación, describimos las principales acciones realizadas durante esta fase:

1. **Selección y aplicación de paletas de colores:** Optamos por paletas cromáticas intuitivas que facilitan la interpretación de los datos. Por ejemplo, utilizamos colores cálidos como el amarillo y el naranja para representar precios elevados, y colores fríos como el azul para valores más bajos de precios. En los mapas coropléticos, las graduaciones de color permiten destacar las diferencias entre barrios de manera visualmente impactante y comprensible.
2. **Incorporación de etiquetas y leyendas:** Añadimos etiquetas descriptivas y leyendas claras en cada gráfico, asegurándonos de que los usuarios puedan entender rápidamente qué representa cada elemento visual.
3. **Jerarquía visual:** Establecemos una jerarquía visual en los dashboards, priorizando los gráficos más relevantes y ubicándolos en posiciones destacadas. Por ejemplo, los mapas temáticos los colocamos en la parte superior, ya que proporcionan una visión general clara del mercado. Utilizamos tamaños y estilos de fuente diferentes para los títulos, subtítulos y etiquetas, destacando los elementos clave sin saturar visualmente al usuario.
4. **Ajustes de escala y ejes:** Revisamos las escalas de los ejes en gráficos como los de barras para garantizar que sean coherentes y fáciles de interpretar. También hemos eliminado valores innecesarios en los ejes para reducir el ruido visual.

5. **Eliminación de elementos innecesarios:** Simplificamos los gráficos eliminando elementos que no aportan valor, como líneas de fondo excesivas o etiquetas redundantes.
6. **Consistencia visual:** Nos aseguramos de que todos los gráficos sigan un diseño coherente, utilizando una paleta de colores, fuentes y estilos similares en todo el dashboard, creando una experiencia uniforme, coherente y profesional.

Con estos ajustes, mejoramos significativamente la claridad y la accesibilidad de las visualizaciones, incrementando su impacto visual y comunicativo.

2.7. Interacción

La fase de interacción busca enriquecer las visualizaciones con funcionalidades que permitan al usuario explorar los datos de forma dinámica y personalizada. Este paso es fundamental para que las visualizaciones se conviertan en herramientas activas de análisis, ofreciendo flexibilidad y adaptabilidad según las necesidades de los usuarios. Las interacciones pueden incluir filtros, deslizadores, zoom, selección de subgrupos y otros elementos que faciliten una experiencia más inmersiva y comprensible. En nuestro dashboard, las características interactivas que hemos implementado son las siguientes:

1. **Filtros dinámicos:** Hemos incorporado filtros que permiten a los usuarios seleccionar inmuebles según diferentes criterios como el rango de precios, el número de habitaciones, la calidad y la antigüedad. Esto facilita la identificación de propiedades según necesidades específicas. Los filtros aplicados actualizan automáticamente los gráficos del dashboard indicados, garantizando que las visualizaciones permanezcan consistentes y relevantes.
2. **Mapas navegables:** En los mapas temáticos y de dispersión hemos incluido funcionalidades de zoom y desplazamiento, permitiendo a los usuarios explorar zonas específicas de Valencia con mayor detalle. También hemos implementado puntos interactivos en los mapas para mostrar información adicional al hacer clic en una ubicación, como el precio promedio por metro cuadrado o la proximidad a estaciones de metro.
3. **Selección de subgrupos:** Los usuarios pueden seleccionar subgrupos específicos de datos directamente en los gráficos, como inmuebles de un barrio concreto o propiedades con ciertas características. Esto permite realizar comparaciones directas entre diferentes categorías.
4. **Tooltips informativos:** En todos los gráficos interactivos, incluimos tooltips que muestran información detallada al pasar el cursor sobre un elemento.

La implementación de estas características interactivas transforma el dashboard en una herramienta poderosa donde no solo se pueden observar tendencias generales, sino también profundizar en detalles específicos según intereses concretos.

3. Resultados principales

3.1. Distribución de inmuebles por barrios

Benicalap es el barrio con mayor cantidad de inmuebles en venta, seguido por Campanar y L'Olivereta. Estos barrios muestran una oferta inmobiliaria considerablemente más amplia en comparación con otras zonas, reflejando su papel predominante en el mercado local.

3.2. Precios promedio por metro cuadrado

Los precios más altos por metro cuadrado se concentran en El Pla del Remei y Gran Vía, barrios caracterizados por su exclusividad y ubicación céntrica. En contraste, barrios periféricos como La Torre y Benimámet presentan precios significativamente más bajos.

3.3. Calidad y antigüedad de los inmuebles

Los inmuebles más antiguos se localizan predominantemente en barrios centrales como El Carmen, donde la antigüedad media supera los 60 años. Sin embargo, estos inmuebles también tienden a tener una calidad mayor según el catastro (a menor valor de calidad del catastro, la calidad del inmueble es mayor [ver gráfico de correlación]).

3.4. Relación entre calidad y precio

El análisis de correlaciones revela una relación negativa entre la calidad de los inmuebles según el catastro y su precio. Los inmuebles de menor calidad según el Catastro (es decir, una calidad mayor, i.e. 1 sería la calidad máxima)), ubicados en zonas privilegiadas, registran precios más elevados, mientras que las propiedades de menor calidad (mayor valor de la calidad del Catastro) tienden a ubicarse en áreas periféricas con precios más accesibles.

3.5. Influencia de la proximidad al transporte público

Las propiedades cercanas al centro de la ciudad muestran una clara tendencia a tener precios más altos. Esto es especialmente evidente en barrios como El Carme, Gran Vía, El Botànic...

3.7. Agrupaciones de barrios

Gracias a los análisis de agrupación de barrios podemos agrupar barrios en categorías según características similares. Por ejemplo:

- **Barrios exclusivos:** Gran Vía, Pla del Remei...
- **Barrios emergentes:** Sant Pau, Nou Campanar...
- **Barrios económicos:** Orriols, Torrefiel...

4. Conclusiones

La realización de este proyecto nos ha permitido comprender a fondo las dinámicas del sector inmobiliario en Valencia mediante el análisis y la visualización de datos. A través de cada fase, desde la adquisición hasta la interacción, hemos logrado desarrollar un dashboard interactivo que ofrece información clara y relevante sobre precios, calidad, antigüedad y accesibilidad de los inmuebles en la ciudad.

Este proyecto nos ha demostrado cómo con la visualización de datos se puede transformar conjuntos de datos masivos y complejos en herramientas comprensibles y útiles para la toma de decisiones. Además, el uso de tecnologías avanzadas como Tableau y Plotly nos ha permitido ofrecer un producto final tanto funcional como atractivo.

La experiencia adquirida a lo largo del proyecto nos ha permitido consolidar habilidades en el manejo de datos, la representación visual y el trabajo en equipo. Además, hemos superado retos como la integración de múltiples fuentes de datos y la creación de visualizaciones interactivas gracias a la colaboración y a la aplicación de los conceptos aprendidos durante la asignatura.

En resumen, con este trabajo no solo hemos cumplido con los objetivos planteados inicialmente, sino que también nos ha servido para desarrollar competencias clave en análisis de datos y visualización. Además, consideramos que el resultado final presentado es un recurso valioso para analizar el mercado inmobiliario de Valencia y ejemplifica el poder de la visualización para comunicar información compleja de manera efectiva.