

Documentación del uso de datos conforme al RGPD en el proyecto 2 del Máster en Big Data e Inteligencia Artificial de IMPELIA

1. Descripción del proyecto

El proyecto se centra en analizar datos de viviendas en Barcelona para entender patrones de precios y características que influyen en los mismos. Este enfoque fue elegido porque proporciona una base sólida para identificar correlaciones entre factores como el número de habitaciones, la ubicación y los precios por metro cuadrado. Utilizando modelos predictivos y técnicas de análisis descriptivo, se espera generar insights accionables que ayuden a los promotores, agentes y consultores a optimizar sus estrategias y anticiparse a tendencias futuras en el mercado inmobiliario. Además, esta metodología permite evaluar el impacto de factores externos, como cambios económicos o normativos, en los patrones de precios.

2. Finalidad del tratamiento de datos

La finalidad de este tratamiento es estudiar las características del mercado de viviendas para:

- Generar insights sobre el comportamiento de los precios, permitiendo identificar las variables que más influyen en su evolución.
- Diseñar herramientas predictivas que faciliten estimaciones precisas de precios basadas en variables como el número de habitaciones, el barrio y la presencia de ascensor o terraza.
- Apoyar a los actores del sector inmobiliario en la optimización de sus estrategias comerciales y de inversión, ayudándoles a tomar decisiones fundamentadas como la selección de ubicaciones para nuevas promociones o la mejora del posicionamiento de sus propiedades en el mercado.
- Proveer a analistas y consultores con una herramienta que permita entender las dinámicas de los precios en el contexto actual y proyectar tendencias futuras, integrando factores externos como cambios en la economía o nuevas regulaciones.

Esta finalidad es específica y está alineada con el RGPD, asegurando que los datos no se utilizarán para otros propósitos no relacionados.

3. Datos procesados

El dataset contiene las siguientes columnas:

- **price:** Precio de la vivienda (no es dato personal).

- **rooms:** Número de habitaciones (no es dato personal).
- **bathroom:** Número de baños (no es dato personal).
- **lift:** Presencia de ascensor (True/False, no es dato personal).
- **terrace:** Presencia de terraza (True/False, no es dato personal).
- **square_meters:** Metros cuadrados de la vivienda (no es dato personal).
- **real_state:** Tipo de inmueble (e.g., flat, no es dato personal).
- **neighborhood:** Barrio de la vivienda (puede considerarse semipersonal si se combina con otros datos).
- **square_meters_price:** Precio por metro cuadrado (no es dato personal).

No se procesan datos personales identificativos, ya que el dataset no incluye nombres, direcciones completas, números de teléfono ni otros identificadores directos.

4. Base jurídica

La base jurídica para el tratamiento de estos datos es el **interés legítimo**, ya que los datos no son personales y su análisis contribuye al entendimiento de patrones en el mercado inmobiliario. Esta base se justifica además en que el uso de datos anonimizados minimiza cualquier posible impacto en los derechos y libertades de las personas.

5. Medidas de protección y seguridad

- **Anonimización:** Aunque el dataset no contiene datos identificativos, las variables que podrían tener un rastro indirecto (como "neighborhood") se tratarán con cuidado para evitar la reidentificación mediante combinación con otros datasets.
- **Control de acceso:** El acceso al dataset está limitado a personas autorizadas dentro del equipo del proyecto, garantizando la confidencialidad.
- **Cifrado:** El dataset se almacena en dispositivos y servidores con cifrado activo (AES-256) para garantizar la seguridad en reposo y en tránsito.
- **Auditorías:** Se registran accesos y modificaciones en el dataset mediante sistemas de registro (logs) para garantizar la trazabilidad.
- **Pruebas de seguridad:** Periódicamente, se realizan pruebas de vulnerabilidad en los sistemas utilizados para el almacenamiento y procesamiento de datos.

6. Ética de la IA

El proyecto se ha diseñado teniendo en cuenta los principios de ética en la inteligencia artificial para garantizar:

- **Transparencia:** Los modelos utilizados están documentados de manera clara para que los resultados sean interpretables por los usuarios finales.
- **Imparcialidad:** Se han tomado medidas para mitigar posibles sesgos en los datos o en los algoritmos, asegurando que las predicciones sean justas y no discriminen a ningún grupo.
- **Responsabilidad:** El equipo ha implementado procesos de revisión interna para garantizar que el impacto de las predicciones sea positivo y no genere consecuencias negativas imprevistas.

Se realizarán auditorías periódicas para evaluar el desempeño ético del proyecto y tomar acciones correctivas en caso necesario.

7. Retención de datos

El dataset se conservará durante el tiempo necesario para completar el análisis y los modelos. Una vez finalizado el proyecto, los datos serán:

- **Eliminados:** De manera segura mediante herramientas de sobrescritura si no se requieren para nuevos análisis. Esto incluye la utilización de software especializado, basado en el estandar DoD 5220.22-M o NIST 800-88 que sobrescribe los datos una o varias veces para garantizar que no puedan ser recuperados.
- **Almacenados anonimizados:** Si es necesario retenerlos para investigaciones futuras, se asegurará que permanezcan anonimizados, utilizando técnicas como la eliminación de identificadores indirectos y la aplicación de algoritmos de anonimización avanzados.

8. Evaluación de Impacto en la Protección de Datos (EIPD)

Dado que no se procesan datos personales, este proyecto no requiere una EIPD conforme al artículo 35 del RGPD. Según la normativa, una EIPD es obligatoria cuando el tratamiento de datos personales implica un alto riesgo para los derechos y libertades de las personas, como en casos de decisiones automatizadas o tratamiento a gran escala de datos sensibles. En este proyecto, los datos procesados no permiten identificar directa o indirectamente a las personas, y se han implementado medidas de seguridad adecuadas que minimizan los riesgos potenciales.

No obstante, en caso de cambios en el alcance del proyecto o la inclusión de datos personales en el futuro, se reevaluará la necesidad de realizar una EIPD formal para garantizar el cumplimiento normativo.

9. Gobernanza de datos

El proyecto incorpora principios sólidos de gobernanza de datos para garantizar que:

- **Cumplimiento continuo:** El tratamiento de los datos se evalúa periódicamente para asegurar la conformidad con el RGPD y otras normativas aplicables.
- **Gestión responsable:** El equipo cuenta con formación en protección de datos y privacidad para manejar la información de manera ética y segura.
- **Auditorías internas:** Se realizan revisiones periódicas del manejo de datos para identificar y corregir posibles incumplimientos o áreas de mejora.

10. Transparencia

Se comunicará a los interesados (si aplica en el futuro) de manera clara:

- La finalidad del tratamiento.
- Los derechos que pueden ejercer.
- Las medidas de protección implementadas.

Esto se realizará a través de avisos de privacidad accesibles y un lenguaje claro para garantizar que los interesados comprendan cómo se gestionan sus datos.

11. Conclusión

El tratamiento de datos realizado en este proyecto cumple con los principios de privacidad y protección del RGPD. Se mantienen altos estándares de seguridad y se documenta cada paso para garantizar la conformidad con la normativa.

Además, el proyecto está diseñado para ser escalable en su cumplimiento, lo que permite una adaptación rápida en caso de nuevas regulaciones o requerimientos.