

DEFENSA TFG

MEJOR DISTRITO DE MADRID PARA INVERTIR

JUAN FERNÁNDEZ CEREZO

16/07/2024

1

INTRODUCCIÓN
Y OBJETIVOS

2

INGENIERÍA
DEL DATO

3

ANÁLISIS DEL
DATO

4

ANÁLISIS DE
NEGOCIOS

5

CONCLUSIÓN



Introducción y Objetivos del Estudio

1

Descripción del Problema

La compra de vivienda es una decisión crucial y compleja, especialmente en el contexto actual de precios históricos e incertidumbre en el mercado inmobiliario de Madrid

2

Motivación del Estudio

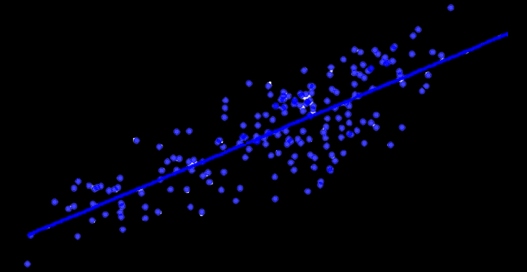
Proporcionar un análisis detallado que facilite la toma de decisiones informadas para inversores y compradores en el sector inmobiliario madrileño.

3

Objetivo Principal

Identificar el mejor distrito de Madrid para invertir en vivienda residencial, basándose en la revalorización esperada y considerando múltiples factores socioeconómicos.

Metodología y Herramientas



1

Herramientas utilizadas

- **Excel** para almacenamiento y cálculos básicos.
- **Python** con librerías como Pandas, Matplotlib y Scikit-Learn para análisis de datos y modelado.
- **PowerBI** para visualizaciones.

2

Proceso ETL

- **Extracción** de diferentes fuentes de datos como el Ayuntamiento de Madrid, Idealista y el INE.
- **Transformación** y limpieza de datos para su análisis.
- **Carga** en una base de datos unificada.

3

Modelos estadísticos

- Implementación de modelos de regresión LASSO, RIDGE y Random Forest Regressor para predecir la revalorización de viviendas en los distritos de Madrid.

Fase 1: Ingeniería del dato

Origen de los datos



- Datos demográficos

- 6 excels (1 por año)
- 21 hojas en cada Excel (1 por cada distrito)



- Precio medio venta/m²
- Precio medio alquiler/m²
- Numero viviendas disponibles venta y alquiler



- PIB anual
- PIB Per Cápita
- Valor Euribor



- Tipos de interés hipotecas (Variable/Total/Fijo)

FECHA INICIAL

2016

BBDD

FECHA FINAL

2023

Fase 1: Ingeniería del dato

Datos demográficos



- 1 Cambiar el formato de las variables y alternar sus posiciones.
- 2 Combinar todas las hojas de los diferentes excels en una misma hoja.
- 3 Eliminar las variables redundantes o innecesarias.
- 4 Sustituir los valores nulos mediante diferentes técnicas. No he observado datos dispersos (outliers).
- 5 Realizar un análisis estadístico de diferentes variables, además de gráficos.

Fase 1: Ingeniería del dato

Datos demográficos



BBDD inicial formada por:

- 237 variables, resumidas en:

- Características generales del distrito
- La población del distrito
- Indicadores económicos
- Indicadores de desempleo
- Educación
- Salud
- Calidad de vida
- Necesidades sociales
- Servicios sociales
- Vivienda
- Medio ambiente y residuos
- Seguridad
- Resultados elecciones locales
- Servicios y equipamientos municipales
- Participación ciudadana



BBDD final formada por:

- 45 variables, resumidas en:

- Características generales del distrito
- La población del distrito
- Indicadores económicos
- Indicadores de desempleo
- Educación
- Calidad de vida
- Vivienda
- Seguridad
- Servicios y equipamientos municipales

Fase 1: Ingeniería del dato (Datos viviendas - Idealista)

idealista

- Evaluación de los valores nulos y outliers.
- Corrección del formato de las variables. Por ejemplo, las fecha era en días y lo que he hecho es separar esa variable en 3: día, mes y año.
- Unificación de la BBDD demográfica con la de vivienda.

```
[16] datos_idealista.head()
```

	Localización	Año	Mes	Día	Fecha	Precio(€/m2)	Variación mensual	Variación trimestral	Variación anual
0	Arganzuela	2023	Octubre	1	2023-10-01	4510	0.005	0.013	0.047
1	Arganzuela	2023	Septiembre	1	2023-09-01	4485	0.000	0.021	0.052
2	Arganzuela	2023	Agosto	1	2023-08-01	4488	0.008	0.021	0.043
3	Arganzuela	2023	Julio	1	2023-07-01	4450	0.013	0.017	0.032
4	Arganzuela	2023	Junio	1	2023-06-01	4392	-0.001	0.008	0.028

He añadido 3
columnas separando
Fecha en Año, Mes y
Día para que sea
más fácil de analizar



Fase 1: Ingeniería del dato

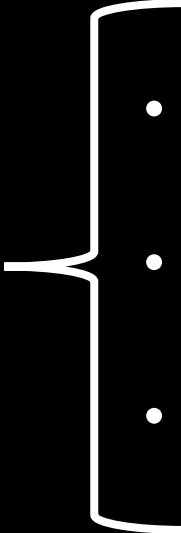
Datos económicos – Datosmacro + INE



- Evaluación de los valores nulos y outliers.
- Normalización de las variables
- Realización de una descripción de las principales medidas estadísticas.
- Unificación ambas BBDD que utilizaremos como contexto para entender el momento económico en el que se sitúa el mercado inmobiliario madrileño.

Fase 2: Análisis del dato (Modelos)

3 Modelos

- 
- Regresión LASSO
 - Regresión RIDGE
 - RANDOM FOREST

Fase 2: Análisis del dato (Métricas)

	RMSE	R ²	MAE
Regresión LASSO	703,88	0,56	339,46
Regresión RIDGE	280,69	0,93	179,08
RANDOM FOREST	280,69	0,65	287,35



Modelo con mejor rendimiento

- RMSE: Raíz del Error Cuadrático Medio
- R²: Coeficiente de determinación
- MAE: Error Absoluto Medio

Fase 3: Análisis de negocios

Preguntas objetivo y de investigación

Pregunta objetivo:

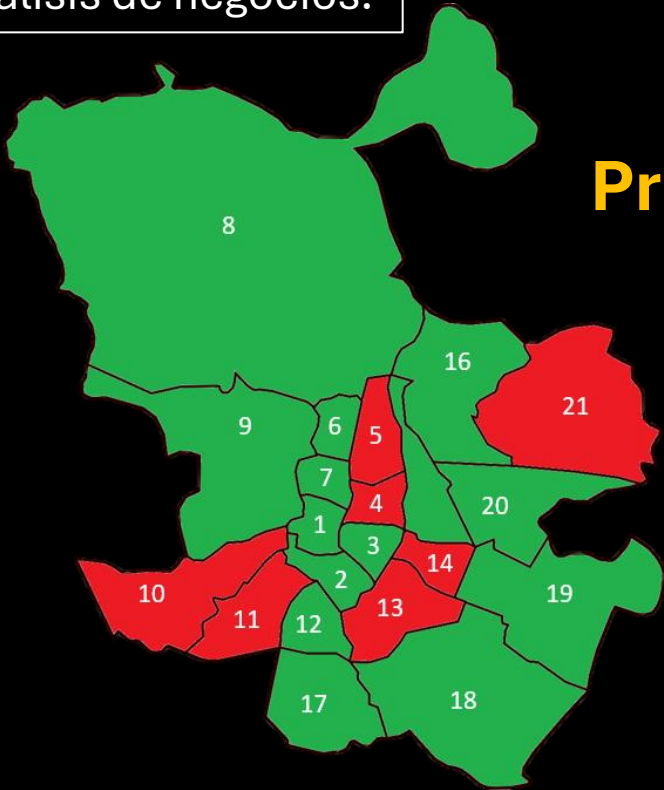
- ¿Cuál es el mejor distrito de la Comunidad de Madrid para invertir en vivienda residencial según la revalorización esperada?

Preguntas de investigación:

- ¿Cuál es la **tendencia histórica de precios de vivienda** en los diferentes distritos de Madrid?
- ¿Cuáles son las **variables más importantes** a la hora de revalorizarse un distrito?
- ¿Cómo influye la **demografía de cada distrito** en la revalorización de la vivienda?
- ¿Qué **indicadores económicos son relevantes** para prever la revalorización de las viviendas en los distintos distritos de Madrid?

Fase 3: Análisis de negocios

Pregunta objetivo: Distrito con mayor revalorización



Mapa revalorización positiva y negativa

1. Centro

2. Arganzuela

3. Retiro

4. Salamanca

5. Chamartín

6. Tetuán

7. Chamberí

8. Fuencarral-El Pardo

9. Moncloa-Aravaca

10. Latina

11. Carabanchel
12. Usera

13. Puente de Vallecas

14. Moratalaz

15. Ciudad Lineal

16. Hortaleza

17. Villaverde

18. Villa de Vallecas

19. Vicálvaro

20. San Blas

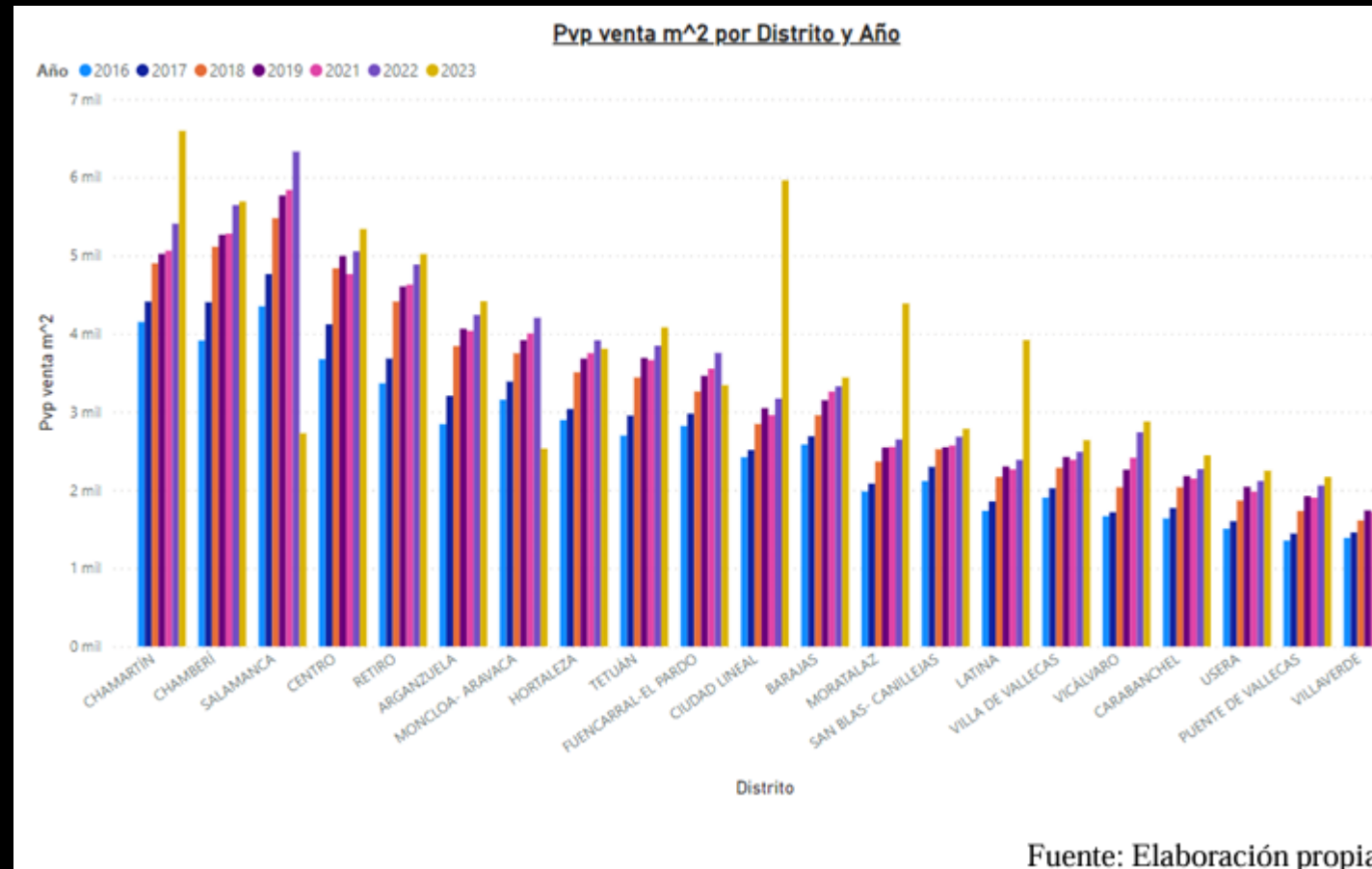
21. Barajas

Distrito	Precio Venta m ²	Revalorización (%)
Fuencarral-El Pardo	3437.00	3.88
Retiro	4307.72	3.46
Villaverde	1705.98	2.83
Barajas	3057	-0.50
Moratalaz	2631.39	-0.74
Puente de Vallecas	1777.34	-1.10

Top 3 mayores y menores distritos por revalorización

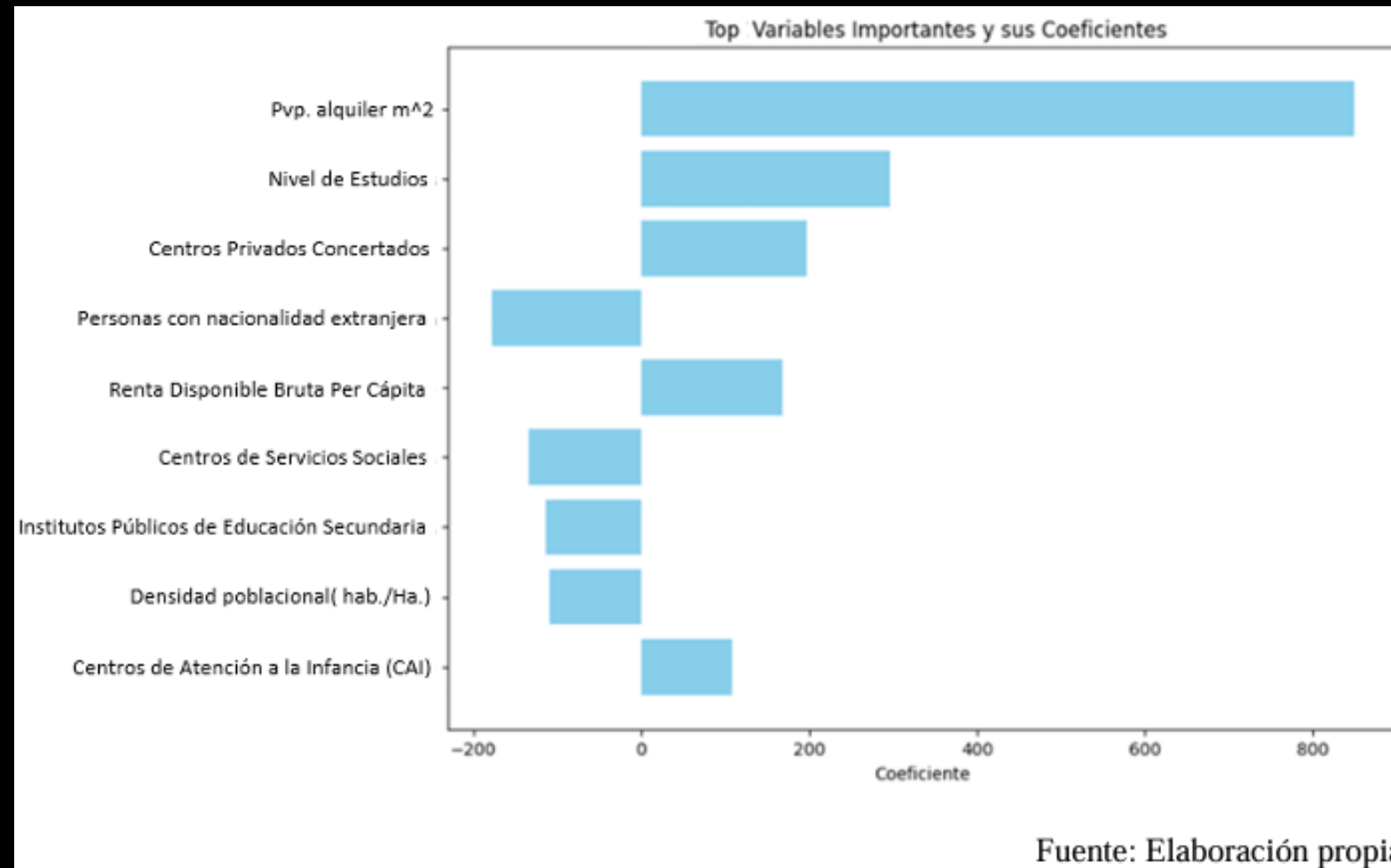
Fase 3: Análisis de negocios

Pregunta de investigación: tendencia histórica de precios



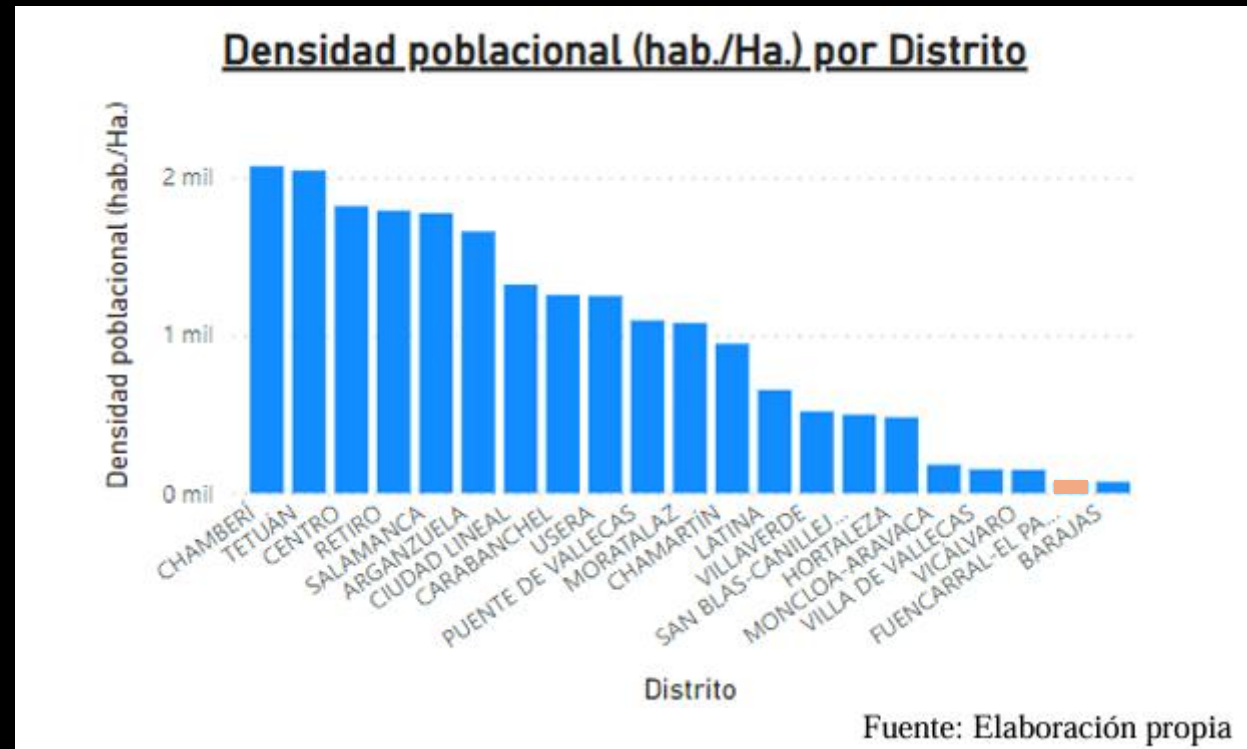
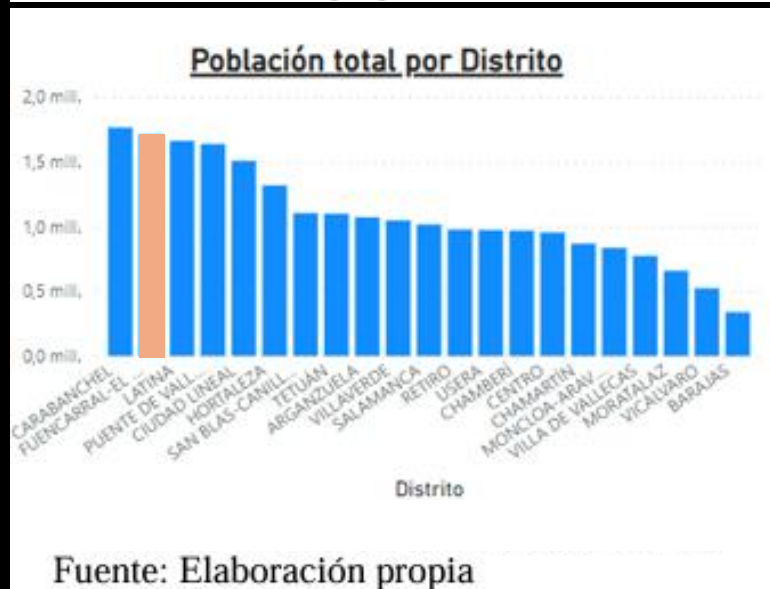
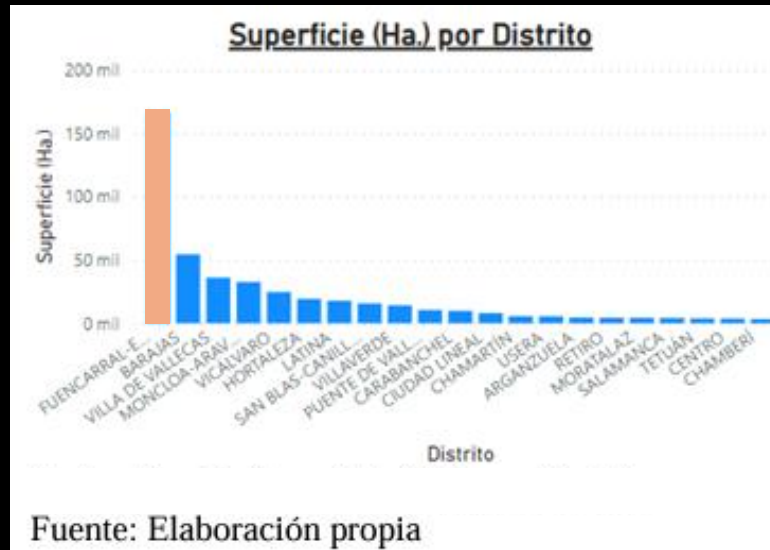
Fase 3: Análisis de negocios

Pregunta de investigación: Variables más importantes



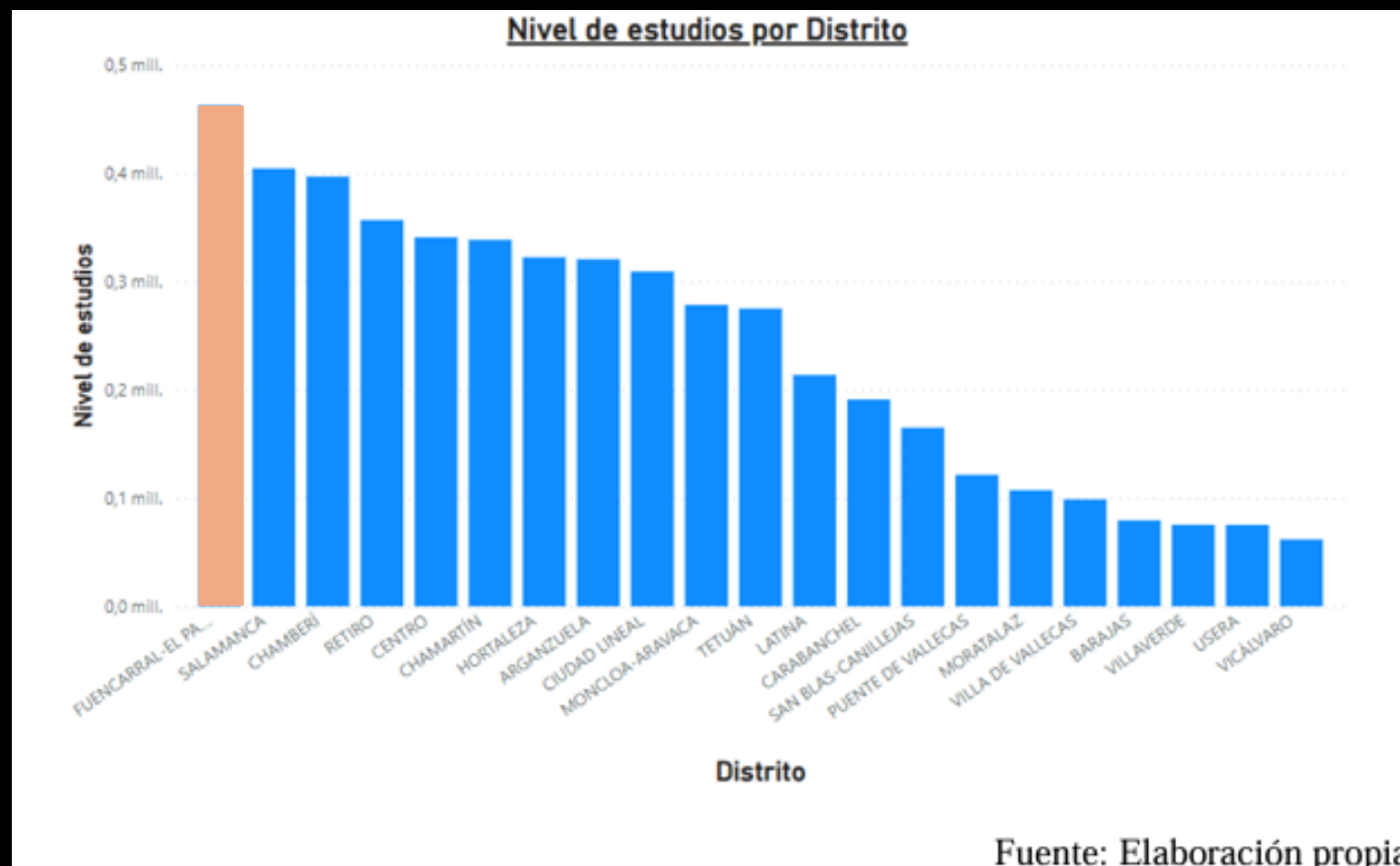
Fase 3: Análisis de negocios

Pregunta de investigación: Demografía de los distritos



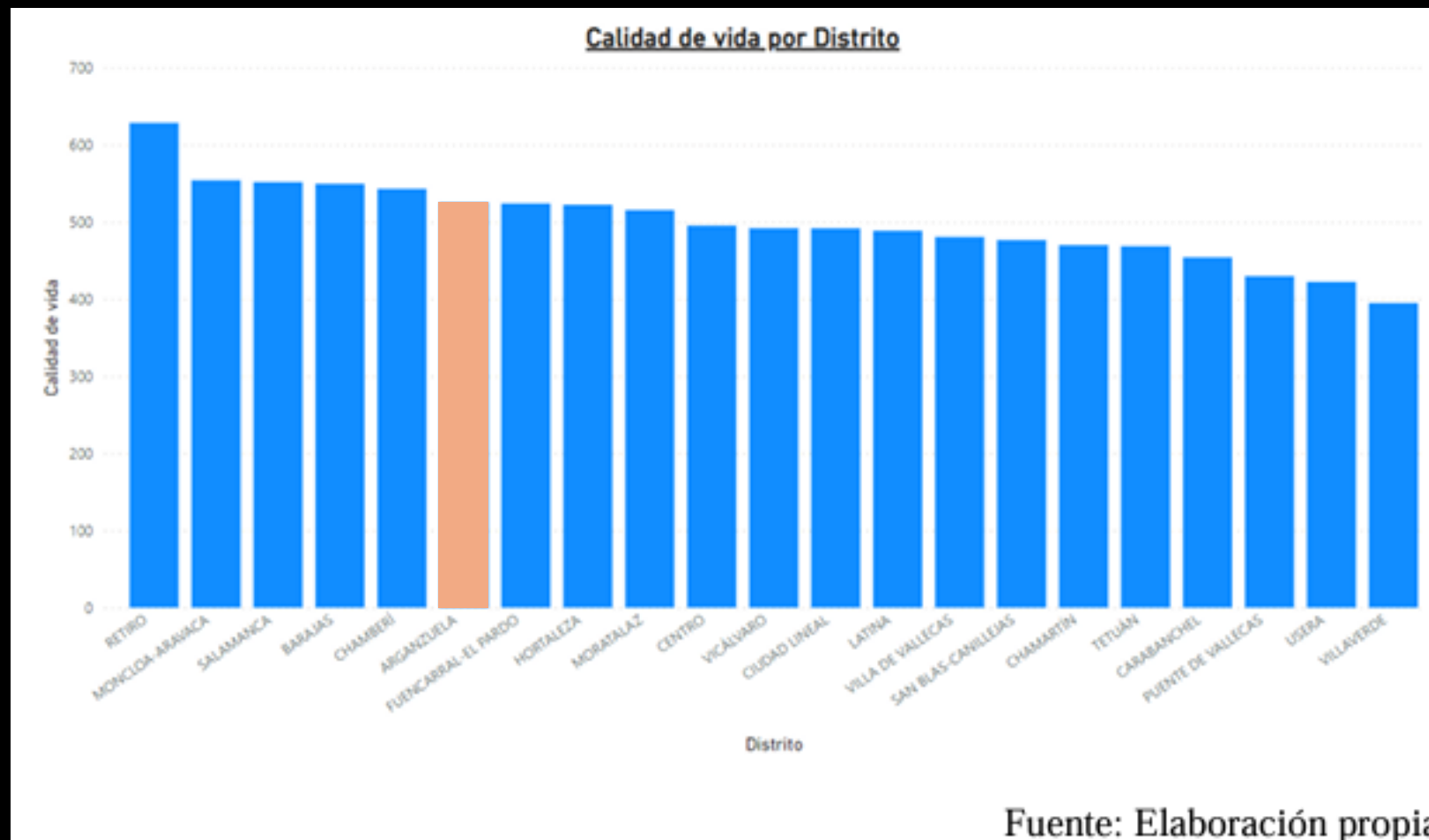
Fase 3: Análisis de negocios

Pregunta de investigación: Demografía de los distritos



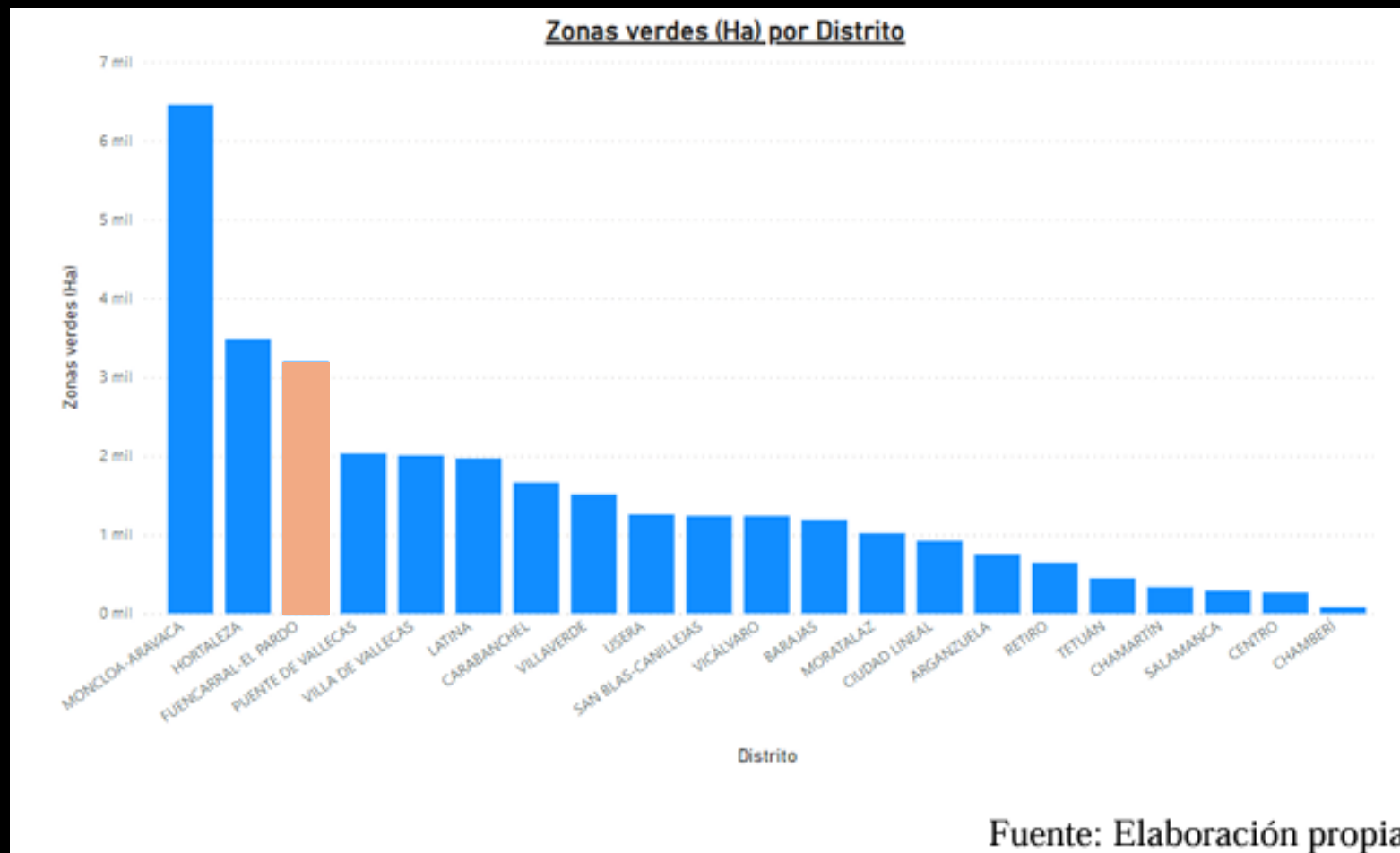
Fase 3: Análisis de negocios

Pregunta de investigación: Demografía de los distritos



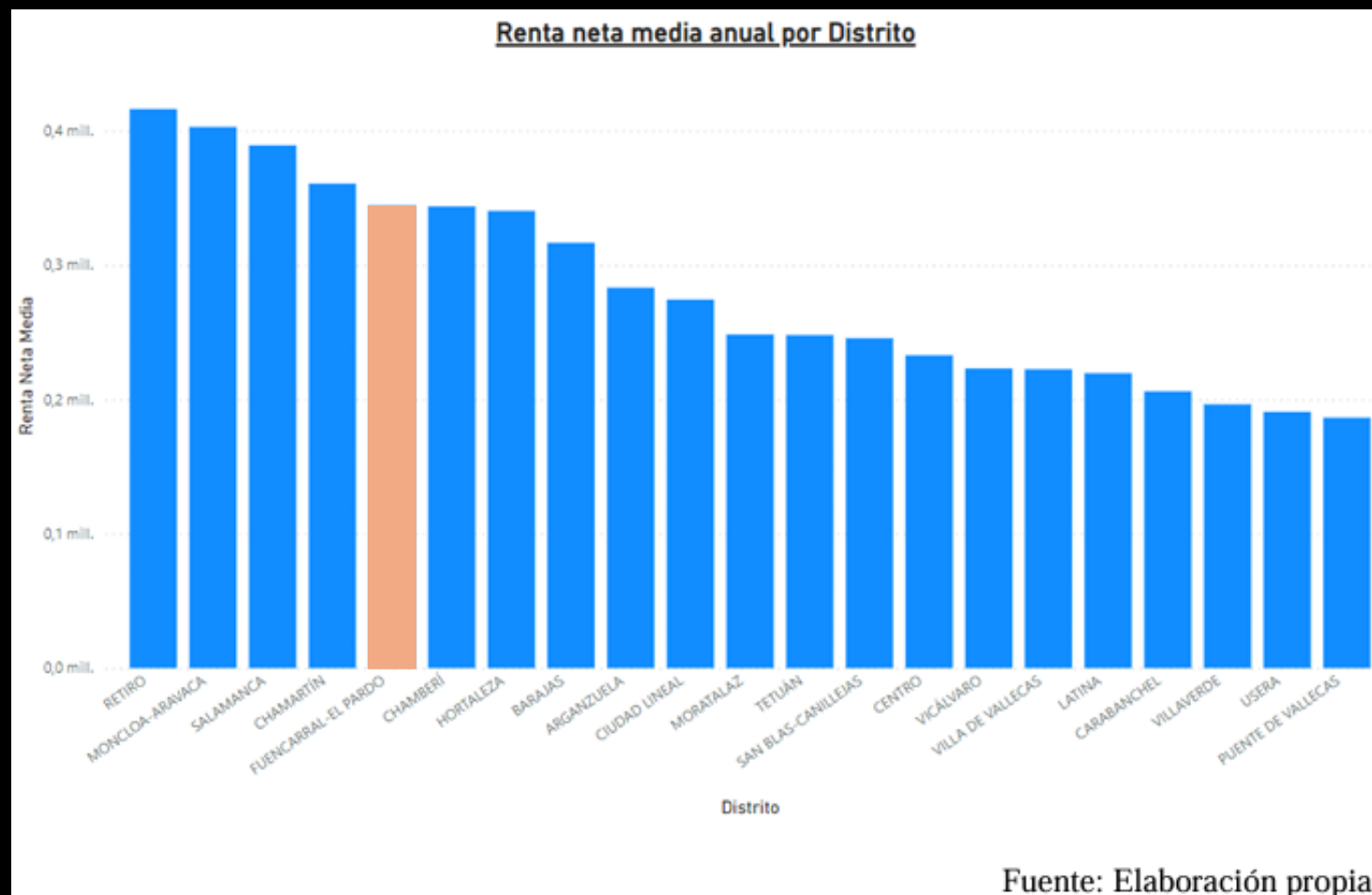
Fase 3: Análisis de negocios

Pregunta de investigación: Demografía de los distritos



Fase 3: Análisis de negocios

Pregunta de investigación: Indicadores económicos más relevantes



Resultados

Mejor Distrito para Invertir

Fuencarral-El Pardo se identifica como el distrito con mayor potencial de revalorización, con un 3.88% esperado.

Factores Clave

Alto nivel educativo, baja densidad poblacional, amplia zona verde y renta per cápita elevada contribuyen a su atractivo para inversión.

Desarrollo Futuro

El proyecto Madrid Nuevo Norte en Fuencarral-El Pardo refuerza su potencial de crecimiento y revalorización a largo plazo.



Conclusiones

1

Logro de Objetivos

El estudio ha cumplido su objetivo principal de identificar el mejor distrito para inversión inmobiliaria en Madrid, basándose en análisis de datos rigurosos.

2

Insights Valiosos

Se han proporcionado insights sobre tendencias históricas, variables influyentes y características de los distritos, útiles para inversores y compradores.

3

Guía para toma de decisiones informadas

El trabajo ofrece una base sólida para tomar decisiones de inversión inmobiliaria en Madrid, considerando múltiples factores y proyecciones futuras.



Gracias
