

Grado en

# Business Data Analytics

Informe final

Reto 0: Agintzari

Curso: 1º 2024-2025

Equipo: Verde claro

URL del repositorio en Github: [Reto-0-verde-claro/24-25\\_1\\_R0\\_VerdeClaro.docx at c2258fc6d972653f8a3fc9d8187cffbd506da9ac · verdeclaro/Reto-0-verde-claro \(github.com\)](https://github.com/verdeclaro/Reto-0-verde-claro)

Autores:

- Uxue Duran
- Aiala Etxebarrieta
- Eider Gana
- Peio Garate
- Imanol Eiguren
- Ander Duñabeitia

## **ÍNDICE**

<b>1. INTRODUCCIÓN.....</b>	<b>6</b>
<b>2. IDENTIFICAR LA PROBLEMÁTICA.....</b>	<b>7</b>
<b>3. RECOGER Y ALMACENAR LOS DATOS.....</b>	<b>8</b>
<b>3.1 Fuentes de datos utilizadas.....</b>	<b>8</b>
<b>3.2 Procesamiento de los datos.....</b>	<b>8</b>
<b>4. ANALIZAR Y MODELAR LOS DATOS.....</b>	<b>10</b>
<b>4.1 Análisis descriptivos.....</b>	<b>10</b>
<b>4.2 Modelar los datos y Visualizar los resultados obtenidos.....</b>	<b>12</b>
<b>4.3 Conclusiones.....</b>	<b>18</b>
<b>5. TRANSFORMAR LOS NEGOCIOS.....</b>	<b>19</b>
<b>6. IMPLICACIONES LEGALES Y ÉTICAS.....</b>	<b>20</b>
<b>GLOSARIO DE TERMINOLOGÍA.....</b>	<b>23</b>
<b>GLOSARIO DE SIGLAS.....</b>	<b>26</b>
<b>ANEXO: Resumen del Proyecto.....</b>	<b>27</b>
<b>BIBLIOGRAFÍA:.....</b>	<b>28</b>

## ÍNDICE DE FIGURAS

FIGURA 1: POBLACIÓN-EXPEDIENTES.....	12
FIGURA 2: EXPEDIENTES POR CADA 1000 HABITANTES.....	12
FIGURA 3: RENTA CAPITAL MOBILIARIO (EXPEDIENTE POR 1000 HABITANTES).	13
FIGURA 4: EXPEDIENTES POR CADA 1000 HABITANTES.....	13
FIGURA 5: POBLACIÓN.....	14
FIGURA 6: RENTA CAPITAL MOBILIARIO.....	15

## **ÍNDICE DE TABLAS**

<b>TABLA 1: VARIABLES ESTADÍSTICAS DESCRIPTIVAS.....</b>	<b>9</b>
--	----------

<b>TABLA 2: Predicción de los indicadores con mayor correlación respecto al número de expedientes.....</b>	<b>16</b>
--	-----------

<b>TABLA 3: Predicciones del número de expedientes.....</b>	<b>16</b>
---	-----------

<b>TABLA 4: Predicciones del número de expedientes hombres y mujeres.....</b>	<b>16</b>
---	-----------

## 1. INTRODUCCIÓN

En una sociedad en constante evolución, las necesidades de la población cambian rápidamente debido a factores como la demografía y la economía. Sin embargo, nuestros servicios públicos no siempre responden de la manera adecuada a las necesidades actuales. Y es por eso que la cooperativa Agintzari, que se centra en la promoción y el desarrollo de la lengua y cultura vasca, nos ha propuesto un reto centrado en pronosticar las necesidades de los usuarios.

El objetivo principal de Agintzari es claro y se centra en ofrecer ayuda a la comunidad vasca en diversas áreas fundamentales, como la educación, la cultura y el bienestar social. La cooperativa no se limita a ofrecer servicios; su misión es crear un impacto positivo y duradero en la vida de las personas que forman parte de esta comunidad. Para lograrlo, se dedica a ofrecer una amplia gama de servicios, que incluyen formación especializada, asesoramiento personalizado y el desarrollo de proyectos culturales innovadores.

Asimismo, nuestro objetivo ante este reto es predecir los datos de los próximos tres años en el distrito 04 (Begoña, Bolueta, Santutxu) para que los servicios sociales y servicios públicos se conviertan en un sistema adaptativo basado en datos actuales y coordinados, con suficiente personal técnico capacitado para responder efectivamente a las demandas sociales. Al predecir la tendencia de los distintos indicadores obtenidos para los próximos tres años, buscamos anticipar el número de expedientes que se abrirán en el futuro. Esto garantiza la prestación de servicios de alta calidad adaptados a las necesidades cambiantes de la población y optimiza el apoyo prestado a las comunidades.

## **2. IDENTIFICAR LA PROBLEMÁTICA**

La problemática de este reto refleja el hecho de que los servicios sociales y públicos no se adaptan adecuadamente a los cambios de las necesidades de la población, las cuales están influenciadas por factores como la demografía, la economía y las condiciones de vivienda. El principal problema es la desconexión entre la evolución de las necesidades de las personas y la capacidad de los servicios públicos para ajustarse a ellas. Esto se debe en parte, a que las administraciones no planifican basándose en datos actualizados, ni coordinan correctamente los diferentes sectores como la salud, empleo, vivienda y educación para ofrecer una respuesta integral.

Además los equipos técnicos que trabajan en estos servicios sufren de un desajuste entre el número de profesionales disponibles y las competencias necesarias para tratar a la población y las nuevas demandas. La ausencia de datos públicos o la falta de comparabilidad dificultan el análisis y la planificación efectiva. Sin información clara, es difícil identificar tendencias y prevenir la evolución de las necesidades sociales. Esto no solo afecta la calidad del servicio, sino que también puede generar frustración tanto en los profesionales como en los usuarios.

La finalidad de este reto es analizar y, prever la evolución de las variables demográficas, económicas y residenciales estudiadas que inciden en las necesidades sociales, para poder disponer de los conocimientos necesarios para rediseñar y mejorar los Servicios Sociales de manera que puedan ofrecer respuestas adecuadas, eficientes y eficaces a las futuras demandas. Es decir, se trata de planificar y coordinar la prestación del servicio, así como de disponer de una plantilla técnica en número, formación y organización para que puedan prestar asistencia de calidad a las necesidades sociales detectadas. Asimismo, se han seleccionado dos tipos de servicio, el 1.3 y el 2.7.3, por la fuerte incidencia que tienen en la Cartera Servicios Sociales, para adaptar y mejorar.

### 3. RECOGER Y ALMACENAR LOS DATOS

#### 3.1 Fuentes de datos utilizadas

A la hora de recoger y almacenar datos hemos utilizado bases de datos de Trebatu, para acceder a datos sobre expedientes abiertos en Gipuzkoa. Además, para poder añadir distintos indicadores a cada municipio de Gipuzkoa hemos usado fuentes como EUSTAT o INE. Con estas hemos conseguido distintos indicadores, así como, el PIB de cada municipio, la tasa de paro, tasa de inmigración, población, poblaciones por rangos de edades, etc.

#### 3.2 Procesamiento de los datos

Para comenzar con este reto, empezamos recogiendo datos de Gipuzkoa, ya que en Gipuzkoa podríamos clasificar los datos por municipios, lo cual es más accesible a la hora de la recopilación de datos, ya que es fácil encontrar datos por municipios, pero no por distritos, como es el caso de la zona que tenemos que analizar.

Para esto, elegimos un año para analizar, qué es el año 2019. Elegimos este año ya que pensamos que era una buena idea omitir los años covid y post covid. Esto es porque queríamos ver las relaciones entre los distintos indicadores y el número de expedientes abiertos por municipios sin el afecto de este suceso pandémico, dado que cada vez es un tema menos influyente, y no creemos que los servicios sociales que se vayan a necesitar en los próximos años vayan a ser los mismos que se han necesitado en estos últimos cuatro años.

Otro punto a tener en cuenta es que filtramos los datos con el año 2019, pero respecto a los expedientes sólo aplicamos este filtro a la fecha de inicio, no a la fecha final ni a la fecha de derivación.

A continuación, para el procesamiento de los datos, hemos utilizado Excel para limpiar los datos de los que disponemos de Gipuzkoa, ya que es una herramienta útil debido a su capacidad para gestionar, organizar y analizar grandes volúmenes de información. Tras limpiar los datos que teníamos de Gipuzkoa, gracias a sus funciones de filtrado, fórmulas avanzadas y gráficos, nos quedamos con un conjunto de datos que incluía distintos indicadores por municipio, los cuales los sacamos de los datos de Trebatu.

Estos son los indicadores:

- Número de expedientes abiertos en 2019
- Mujeres
- Hombres
- Nacionales
- Extranjeros
- Doble nacionalidad

Posteriormente, empezamos a recopilar datos de indicadores adicionales, así como:

- Indicadores económicos
  - PIB
  - Renta
- Indicadores demográficos
  - Tasa de inmigración
  - Tasa de paro

Además de todos estos indicadores obtenidos, para que los gráficos que realizamos posteriormente fueran más precisos, transformamos el indicador del número de expedientes, por el número de expedientes por cada 1000 habitantes. También convertimos ciertos indicadores en porcentajes, como por ejemplo, el porcentaje de hombres, mujeres, nacionales o extranjeros.

Respecto a Bizkaia, seguimos los mismos pasos de recogida y limpieza de datos que en Gipuzkoa, pero esta vez solo incluimos los indicadores que resultaron más relevantes en relación al número de expedientes.



#### 4. ANALIZAR Y MODELAR LOS DATOS

##### 4.1 Análisis descriptivos

**TABLA 1: VARIABLES ESTADÍSTICAS DESCRIPTIVAS**

5/10/2024	AÑO	EXPEDIENTES	DESVIACIÓN ESTÁNDAR (EXPEDIENTES)	MEDIANA	CUARTIL 1	CUARTIL 3	RANGO INTERCU ARTÍLIC O	RANGO (EXPEDI ENTES)	MEDIA DE EXPEDIE NTES ANUAL
SANTUTXU-B OLUETA	2016	103	31,84336666	128,5	116,5	133,5	17	109	128
	2017	147	31,84336666	128,5	116,5	133,5	17	109	128
	2018	129	31,84336666	128,5	116,5	133,5	17	109	128
	2019	128	31,84336666	128,5	116,5	133,5	17	109	128
	2020	81	31,84336666	128,5	116,5	133,5	17	109	128
	2021	190	31,84336666	128,5	116,5	133,5	17	109	128
	2022	125	31,84336666	128,5	116,5	133,5	17	109	128
	2023	121	31,84336666	128,5	116,5	133,5	17	109	128
SUMA TOTAL		507							512
BARAKALDO	2014	61	77,80145671	203,5	131,75	224,25	92,5	92,5	178,6666 667
	2015	108	77,80145671	203,5	131,75	224,25	92,5	92,5	178,6666 667
	2016	203	77,80145671	203,5	131,75	224,25	92,5	92,5	178,6666 667
	2017	265	77,80145671	203,5	131,75	224,25	92,5	92,5	178,6666 667
	2018	204	77,80145671	203,5	131,75	224,25	92,5	92,5	178,6666 667
	2019	231	77,80145671	203,5	131,75	224,25	92,5	92,5	178,6666 667
SUMA TOTAL		1072							1072
GETXO	2014	14	12,83744523	17	12,5	29	16,5	16,5	20
	2015	4	12,83744523	17	12,5	29	16,5	16,5	20
	2016	12	12,83744523	17	12,5	29	16,5	16,5	20
	2017	38	12,83744523	17	12,5	29	16,5	16,5	20
	2018	20	12,83744523	17	12,5	29	16,5	16,5	20
	2019	32	12,83744523	17	12,5	29	16,5	16,5	20
SUMA TOTAL		120							120

Como parte de este proyecto, se han utilizado varias medidas estadísticas para facilitar la comparación y la interpretación de los datos obtenidos previamente, entre esas medidas estadísticas están, la media, mediana, desviación estándar, el rango, los cuartiles y el rango intercuartílico.

La media es una medida de tendencia central que se consigue al sumar todos los valores de un conjunto de datos y dividiéndolos por el total de valores. Representa el promedio de los datos y es útil para conseguir el valor central de una distribución. La media nos ha permitido entender cuál es el valor promedio de las variables analizadas. Por ejemplo, si se analiza el número de expedientes promedio en diferentes distritos, la media nos ha ayudado a comparar cuál es el distrito con más o menos expedientes. Es muy fácil de usar e interpretar, además tiene en cuenta todos los valores del conjunto, por lo que es una medida completa.

La mediana es otra medida de tendencia central que divide el conjunto de datos en dos partes iguales, la mediana es el valor que se encuentra en el medio cuando los valores están ordenados de menor a mayor o viceversa. También es muy sencilla de calcular ya que si el número de observaciones es impar, la mediana es el valor central, y en el caso de que el número de observaciones sea par, se calcula haciendo la media entre los dos valores centrales. La mediana es útil al trabajar con distribuciones asimétricas o con datos que tienen valores extremos, ya que no es influenciada por estos. En nuestro caso, la mediana nos ha ayudado ya que muestra el punto medio y, por lo tanto, evita distorsiones causadas por valores muy altos o bajos. Una de sus ventajas es que no es afectada por outliers.

La desviación estándar mide la dispersión de un conjunto de datos con respecto a su media. Es decir, indica cuánto se alejan los valores individuales de la media del conjunto. Una desviación estándar baja sugiere que los valores están más cerca de la media, mientras que una desviación alta indica una variabilidad mayor. Es útil para evaluar la consistencia de los datos y la variabilidad en la información. Y no como en la media, la desviación estándar si que es sensible a valores extremos, lo que puede incrementar artificialmente la variabilidad percibida.

El rango es la diferencia entre el valor máximo y el valor mínimo en un conjunto de datos, simplemente proporciona una medida simple de la extensión de los datos. Se utiliza para dar una idea rápida de la variabilidad total en un conjunto de datos. A nosotros nos ha servido para analizar las edades de personas con expedientes abiertos, indicándonos la amplitud de la distribución de sus edades. Esta medida también es muy fácil de calcular, y útil para comparar la extensión de dos o más conjuntos de datos.

Los cuartiles son tres, primero, segundo y tercero. El primero delimita el 25% inferior del conjunto de datos, el segundo divide los datos en dos partes iguales al igual que la mediana, y el tercero delimita el 25% superior de los datos. Los cuartiles ayudan a identificar la distribución de los datos y también identificar puntos de referencia, como el 25% o el 75% de los valores. Son útiles para evaluar la dispersión de datos y no están afectados por valores extremos.

La última de las medidas estadísticas utilizadas ha sido el rango intercuartílico, el cual es la diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1), este mide la dispersión de la parte central de los datos sin ser influenciado por los extremos. Esta medida nos ha permitido detectar la concentración de datos en la parte central de la distribución. Ha sido útil para resumir la variabilidad interna sin incluir valores atípicos que distorsionan la medida, proporcionando una visión clara de la dispersión del centro de la distribución, por ejemplo, en Barakaldo, el rango intercuartílico es de 92,5, lo que indica una dispersión mayor que en Santutxu-Bolueta.

El uso de estas medidas estadísticas en la comparación de expedientes y población entre barrios o distritos nos ha permitido una evaluación más profunda y detallada de la información. Cada medida nos ha ofrecido una perspectiva distinta sobre los datos, desde la tendencia central hasta la dispersión y variabilidad, ayudando así a identificar patrones y tomar decisiones en el análisis.

## 4.2 Modelar los datos y Visualizar los resultados obtenidos

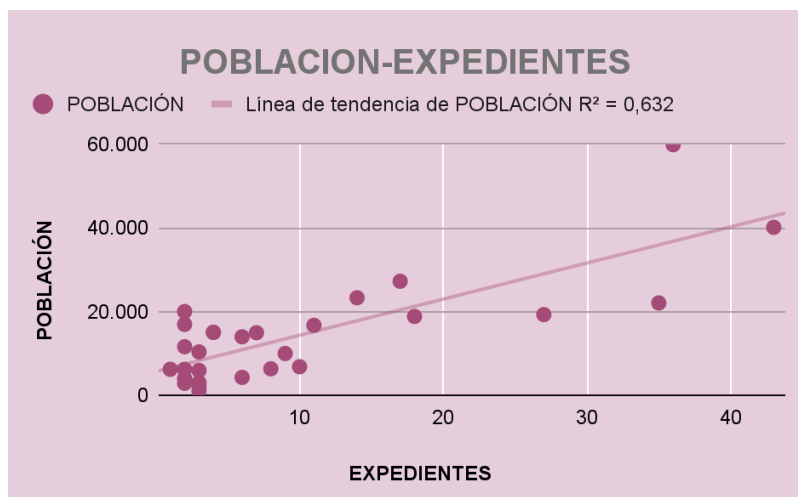
Para poder facilitar la interpretación de los datos, y así ver las distintas relaciones de cada indicador respecto al número de expedientes abiertos, creamos un gráfico de dispersión en excel por cada indicador en relación al número de expedientes por cada 1000 habitantes. A estos gráficos le añadimos la línea de tendencia y así poder ver cuánta relación había entre cada indicador y el número de expedientes, para luego quedarnos con los indicadores que más correlación tuvieran y hacer uso de esos en el análisis de nuestro distrito de Bizkaia.

Al finalizar con este proceso, los indicadores que resultaron relevantes en Gipuzkoa en relación con el número de expedientes los utilizamos para Bizkaia. Sin embargo, en Bizkaia el estudio se basa en analizar estos indicadores a medida que van pasando los años. Es decir, cómo han ido evolucionando esos indicadores a medida que han ido pasando los años. A parte de estos indicadores, también hemos sacado otros datos, que son: el %de mujer con expediente, el %de hombre con expediente, el % de edad de <19 años con expediente, el % de edad de 20-64 años con expediente y el % de edad de >65 años con expediente. Estos datos nos van a ser muy útiles a la hora de hacer la predicción, debido a que nos permitirán especificar mucho más la predicción.

Con el análisis realizado con los datos de Gipuzkoa nos dimos cuenta que las variables que más relación tienen con el aumento o disminución del número de expedientes son las siguientes:

- El primer indicador que nos indica una clara relación respecto al número de expedientes es la población.

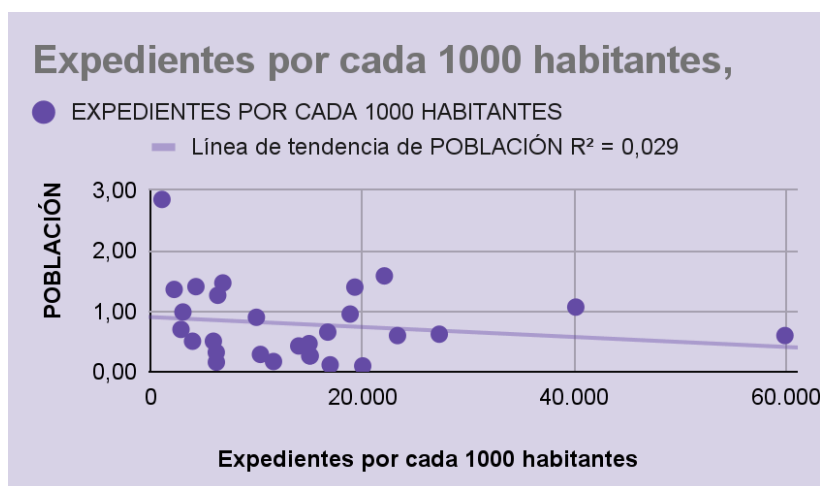
**FIGURA 1: POBLACIÓN-EXPEDIENTES**



En este gráfico de dispersión en el que en el eje x está colocado el número de expedientes abiertos, mientras que en el eje y se encuentra el número de población; nos indica que mientras el número de habitantes de un lugar aumente, en ese lugar se abrirán más expedientes.

Sin embargo, en proporción, es decir, si nos fijamos en el número de expedientes por cada 1000 habitantes, cuanto más alta sea la población el número de expedientes abiertos por cada 1000 disminuye; como nos indica el siguiente gráfico:

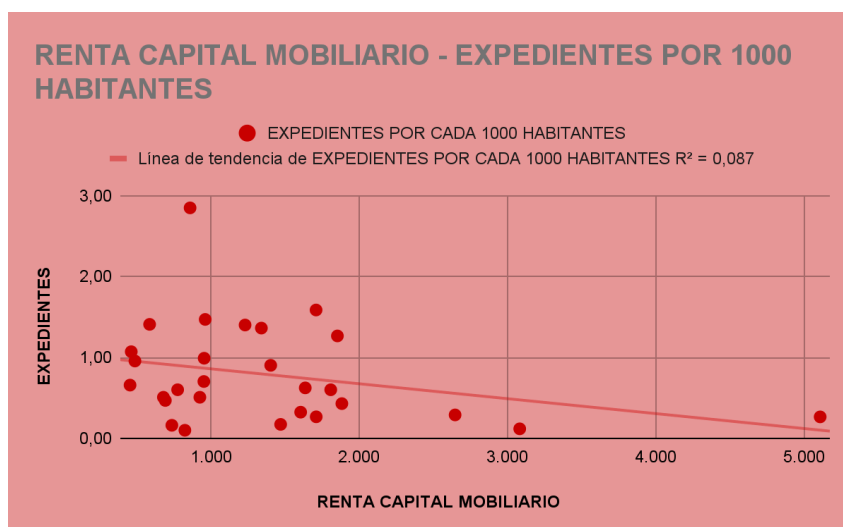
**FIGURA 2: EXPEDIENTES POR CADA 1000 HABITANTES**



En este gráfico vemos como en el eje X está el número de población, mientras que en el eje Y se coloca el número de expedientes por cada 1000 habitantes. Cuya relación, como bien hemos mencionado anteriormente, se basa en que mientras más alta sea la población el número de expedientes por cada 1000 habitantes más bajo es.

- Otro indicador que nos indica una clara relación con los expedientes abiertos es la **RENTA CAPITAL MOBILIARIO**:

**FIGURA 3: RENTA CAPITAL MOBILIARIO (EXPEDIENTE POR 1000 HABITANTES)**

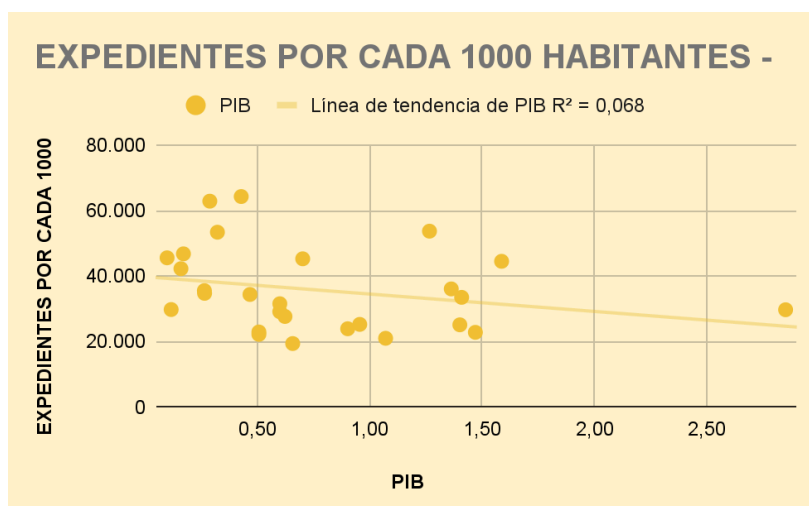


En este gráfico de dispersión vemos como en el eje X se coloca la renta del capital inmobiliario mientras que en el eje Y se coloca el número de expedientes por cada 1000 habitantes.

Se observa como su tendencia es negativa puesto que a medida que el capital inmobiliario aumenta, los expedientes por cada 1000 habitantes disminuyen.

- Por último, la última variable que indica otra relación respecto al **NÚMERO DE EXPEDIENTES ES EL PIB**:

**FIGURA 4: EXPEDIENTES POR CADA 1000 HABITANTES**



En este gráfico en el eje X se encuentra el valor del PIB mientras que en el eje Y se encuentra el valor de los expedientes por cada 1000 habitantes.

En el gráfico se observa que mientras más alto es el PIB el número de expedientes por cada 1000 habitantes es menor; ya que tiene una tendencia descendente.

No obstante, esta variable del PIB no la hemos podido utilizar para hacer el estudio en nuestro distrito, ya que los datos anuales del PIB por distritos en Bilbao no son públicos.

Como hemos nombrado anteriormente, estas variables las hemos utilizado en nuestro distrito para ver cómo han ido evolucionando a medida que ha ido pasando el tiempo:

- **POBLACIÓN:**

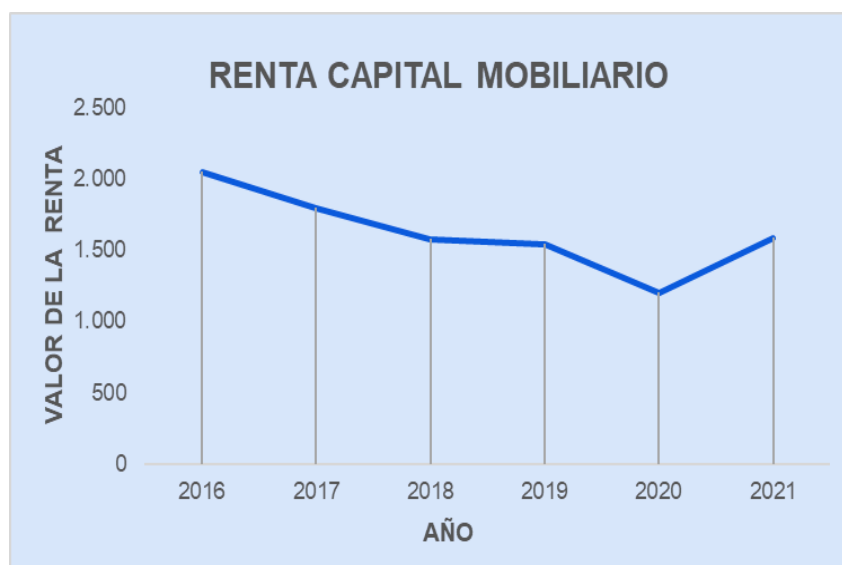
**FIGURA 5: POBLACIÓN**



En este gráfico lineal vemos la evolución de la población en Santutxu-Boluda. Se ve que en general tiene una tendencia ascendente. Esto es un buen dato para la hora de hacer la predicción.

- **RENTA CAPITAL MOBILIARIO:**

**FIGURA 6: RENTA CAPITAL MOBILIARIO**



Aquí vemos un gráfico lineal que indica la evolución de la renta del capital mobiliario los últimos años en Santutxu-Bolueta. Vemos que el año 2020 tuvo una disminución notoria a causa de la pandemia. Aun así, en general, tiene una tendencia descendente.

A continuación, después de visualizar gráficamente qué indicadores tienen más correlación con el número de expedientes, pasamos a crear modelos mediante regresiones lineales en R. Después de intentar varias regresiones lineales simples, en las cuales la y era el número de expedientes y la x eran los distintos indicadores de los que disponíamos, nos quedamos con los dos indicadores que resultaron tener más alto el coeficiente de correlación, dado que eran los que nos iban a llevar a una mejor predicción del número de expedientes que se abrirán en el futuro. Estos indicadores fueron el número de extranjeros y la renta del capital mobiliario. Con dichos indicadores, creamos modelos de regresiones lineales simples, donde la y eran los indicadores (extranjeros y renta capital mobiliario) y la x era el año (2024, 2025, y 2026). Es decir, predijimos esos indicadores en función del año, para así poder ver la evolución de estos en el futuro. Después de crear los modelos, los aplicamos para predecir esos indicadores en los años siguientes.

Para saber interpretar las fórmulas primero debemos saber lo siguiente:

y: Variable dependiente o resultado que estamos tratando de predecir o explicar.

$\beta_0$ : Es el coeficiente de la primera variable independiente.

$\beta_1$ : Es el coeficiente de la segunda variable independiente, representa la pendiente indicando cuánto cambia y cuando  $\beta_1$  cambia en una unidad.

$\beta_3$ : Es el coeficiente de la tercera variable independiente, representa la pendiente indicando cuánto cambia y cuando  $\beta_2$  cambia en una unidad.

$$y = -401820.1091\beta_0 + 200.5212\beta_1$$

$$y = 214204.3091\beta_0 - 105.3879\beta_1$$

**TABLA 2: Predicción de los indicadores con mayor correlación respecto al número de expedientes.**

AÑO	RENTA CAPITAL MOBILIARIO	EXTRANJEROS
2024	875	4080
2025	782	4258
2026	583	4636

Posteriormente, predijimos el número de expedientes que se abrirán en el futuro mediante una regresión lineal múltiple, donde la y era el número de expedientes y la x eran los dos indicadores. Después de aplicar ese modelo para predecir el número de expedientes, las predicciones obtenidas fueron las siguientes.

$$y = -358.10716128\beta_0 + 0.16700851\beta_1 + 0.08290562\beta_2$$

**TABLA 3: Predicciones del número de expedientes**

AÑO	Nº DE EXPEDIENTES
2024	126
2025	125
2026	123

Por último, realizamos la predicción de expedientes en los próximos tres años, pero separando las predicciones en los expedientes de hombre y mujeres. Para ello, seguimos el mismo procedimiento, pero esta vez, haciendo dos regresiones lineales múltiples, siendo la y el número de expedientes hombres y mujeres, y la x los indicadores. Aplicamos los modelos para predecir el número de expedientes hombres y mujeres por separado, y obtuvimos estos resultados.

$$y = -174.32946933\beta_0 + 0.08259967\beta_1 + 0.03758969\beta_2$$

$$y = -196.39153804\beta_0 + 0.08890934\beta_1 + 0.04726208\beta_2$$

**TABLA 4: Predicciones del número de expedientes hombres y mujeres**

AÑO	Nº EXPEDIENTES HOMBRES	Nº EXPEDIENTES MUJERES
2024	52	74
2025	51	74
2026	48	75



### 4.3 Conclusiones

El análisis de los datos en el distrito de Santutxu-Bolueta reveló algunos patrones importantes que nos ayudan a entender cómo ciertos factores están relacionados con el número de expedientes abiertos. En primer lugar, se encontró que, a medida que la población crece, el número total de expedientes también aumenta. Sin embargo, si lo miramos por cada 1000 habitantes, el número de expedientes disminuye en las áreas más pobladas, lo que sugiere que estas zonas podrían estar manejando mejor los servicios sociales.

Además, tanto la renta del capital mobiliario como el PIB mostraron una relación inversa con el número de expedientes por cada 1000 habitantes. Es decir, en las zonas donde estos indicadores económicos son más altos, la necesidad de abrir nuevos expedientes es menor, lo que indica que un mejor entorno económico puede reducir la demanda de servicios sociales.

En cuanto a la evolución de los indicadores en el tiempo, se pudo observar que la población en Santutxu-Bolueta ha mostrado una tendencia ascendente, lo que es útil para prever futuras demandas. Por otro lado, la renta del capital mobiliario mostró una ligera disminución tras la pandemia, lo que podría influir en un aumento de expedientes a corto plazo.

Finalmente, en el análisis de R podemos reafirmar que los indicadores que hemos elegido tienen correlación y podemos ver los patrones de forma más clara.

Estos patrones, junto con el análisis de las tendencias futuras, permitirán prever la demanda de trabajo y mejorar la distribución de recursos en los próximos años en el distrito.

## 5. TRANSFORMAR LOS NEGOCIOS

El estudio de las variables socioeconómicas de Bizkaia y Gipuzkoa, con la información suministrada por Agintzari, facilita la creación de modelos predictivos que evalúan la cantidad de intervenciones requeridas en los próximos años. Estas herramientas de análisis, que emplean métodos como la regresión logística y el estudio de series de tiempo, no solo ayudan a la predicción del número de casos, sino que además reconocen los tipos problemas particulares que podrían ocurrir en cada distrito o municipio.

Gracias a estas predicciones, Agintzari puede administrar de forma más eficaz los recursos y el reparto de su personal. Por ejemplo, si los modelos señalan un incremento en la necesidad de programas de intervención en zonas urbanas con elevados índices de inmigración, desocupación juvenil o incluso desempleo, la cooperativa experimenta un aumento en la demanda de dichos programas.

Este método predictivo también simplifica la organización de los gastos y la distribución de recursos, disminuyendo así el peligro de la sobrecarga en ciertos sectores. Por lo tanto, se logra mejorar la eficiencia en las operaciones y se asegura un uso más estratégico de los recursos existentes, incrementando la eficacia de las operaciones.

Desde un punto de vista empresarial, la aplicación de modelos predictivos no solo optimiza el desarrollo empresarial, sino que también destaca la calidad del servicio que brinda Agintzari. La habilidad para prever dificultades y ajustar los recursos a las demandas futuras fortalece la percepción de la cooperativa como un servicio rápido y efectivo en la determinación de problemas sociales. Lo que a su vez potencia la sostenibilidad y la capacidad de respuesta del modelo de negocio. En consecuencia, se garantiza un incremento en la satisfacción de los beneficiarios y se produce un efecto positivo en las comunidades beneficiadas.

## 6. IMPLICACIONES LEGALES Y ÉTICAS

La ética es uno de los aspectos clave en este reto con Agintzari, una cooperativa de iniciativa social. Esto es porque las decisiones que toman este tipo de organizaciones no solo afectan a la propia empresa, sino que también al bienestar de personas y comunidades necesitadas. Hoy en día, cada vez más decisiones se toman a través del análisis de datos, basándose en modelos predictivos creados por algoritmos. Por este motivo, es imprescindible que estos modelos se diseñen y usen de manera justa y responsable. Si los algoritmos no se supervisan adecuadamente, pueden heredar sesgos, discriminar a ciertos grupos de personas o priorizar resultados que puede que no sean justos. Así pues, hacer una reflexión ética en cada fase del reto no solo es algo necesario desde el punto de vista moral, sino que también es esencial para asegurar que las soluciones tecnológicas basadas en datos ayuden al bienestar de la sociedad y no creen nuevas desigualdades.

El hecho de que se puede explicar a una persona o a un grupo de personas a través de los datos, crea una gran cuestión ética, ya que convierte a los humanos en números y patrones, y además ignora factores personales e individuales que pueden ser necesarios para entender la realidad de cada uno. Así pues, dos personas que tienen los mismos datos pueden tener circunstancias de vida completamente diferentes, ya sea por su entorno cultural, socioeconómico o por otros aspectos personales que no se pueden captar en un análisis cuantitativo. Esta situación, puede llevar a soluciones que no satisfacen realmente las necesidades de los clientes. Por eso, es esencial que cualquier análisis de datos no se base sólo en números, sino que también se centre en un entendimiento de las circunstancias y experiencias individuales, asegurando que las decisiones tomadas a partir de estos datos sean justas y respeten la dignidad humana.

En general, hacer pública la información sobre los expedientes de intervención social controlados por la Administración Pública es favorable para justificar cualquier decisión, acción o resultado de las empresas, y así conseguir que estas generen confianza. No obstante, al revelar datos sensibles relacionados con aquellos que han recibido ayuda social, esa revelación puede afectar negativamente a su seguridad y provocarles discriminación o estigmatización. Una buena solución a este dilema es proteger la confidencialidad de las personas haciendo públicos solo datos anónimos o generales, que no expongan a nadie.

La utilización de datos socioeconómicos y sociodemográficos en la planificación de este tipo de servicios que ofrece Agintzari, puede conllevar varios riesgos éticos, que si no se solucionan correctamente pueden excluir a ciertos grupos sociales o generar discriminación.

Por un lado, existe el riesgo de que los datos socioeconómicos y sociodemográficos estén sesgados de varias maneras. Por ejemplo, puede darse el caso de que estemos jugando con datos incompletos, ya que hay poblaciones con menor accesibilidad y que por eso no salen representadas en las bases de datos. Y esto les podría llevar a recibir menos recursos o servicios que otras poblaciones. Además de esto, las diferentes interpretaciones que se le puede dar a los datos socioeconómicos o sociodemográficos, así como los ingresos o el nivel educativo, pueden crear distintos

estereotipos al asumir que ciertos grupos “necesitan” ciertos tipos de servicios, lo que puede llegar a ser una planificación discriminatoria.

Por otro lado, otro riesgo que existe a la hora de utilizar este tipo de datos en la planificación de servicios, es que las desigualdades pueden incrementar si utilizamos estos datos para crear modelos o algoritmos. Esto puede pasar si los algoritmos se crean con datos no actuales, que pueden reflejar ciertas desigualdades que hoy en día no se dan. Además, las planificaciones que se toman a base de estos modelos, no tienen en cuenta las complejidades humanas y sociales.

Las administraciones necesitan prever las necesidades futuras de la población y para eso se basan en datos. Dado que existen ciertos riesgos que hemos comentado anteriormente, hay que tener responsabilidades a la hora de hacer uso de estos datos. En primer lugar, las administraciones tienen la responsabilidad de anticipar el bienestar de la sociedad, lo que incluye el hecho de prever los servicios que serán necesarios en el futuro, así como la salud, la educación, el empleo, etc. Para eso, deben ser capaces de prever cambios demográficos, económicos y ambientales, y así poder garantizar que los servicios se ajusten a estas transformaciones. Así pues, el análisis de datos permite a las administraciones anticiparse a estos cambios sociales.

Por otra parte, las administraciones también tienen la responsabilidad de asegurarse de que están trabajando con datos de calidad, ya que sino los modelos predictivos podrían estar sesgados y se podrían tomar decisiones incorrectas. Por ejemplo, si a la hora de predecir el número de viviendas que harían falta en un futuro solo se usan datos que reflejan ciertas zonas urbanas, podrían ignorarse las necesidades de zonas rurales.

En relación a los Objetivos de Desarrollo Sostenible (ODS), hay varios que podrían estar implicados en este reto. Cada uno de los ODS que están implicados, podrían estar afectando de forma positiva, en caso de estar trabajando con datos de calidad, ya que nos ayudará a hacer predicciones útiles. Sin embargo, también pueden afectar de forma negativa si los datos con los que se trabajan están sesgados, puesto que las predicciones podrían marginar a ciertos grupos, lo que iría en contra de los ODS.

A continuación algunos ejemplos de los ODS que están relacionados con el reto. El ODS 1, el fin de la pobreza. El objetivo de este es poner fin a la pobreza en todas sus formas y en todo el mundo. Este ODS está implicado en el reto ya que uno de los objetivos de los servicios sociales también es reducir la pobreza, y para eso garantizan recursos básicos como la vivienda, alimentación, salud y educación a personas necesitadas.

Otro ODS implicado sería el tres, salud y bienestar. El objetivo de este es asegurar una vida sana y promover el bienestar de todas las personas a todas las edades. La implicación de este ODS en el reto viene a cuenta de que los servicios sociales también incluyen atención médica o apoyo a la salud mental, lo cual es esencial para cumplir con este ODS. Para poder prever este tipo de servicios del futuro es necesario hacer un análisis de datos médicos para así anticiparse a problemas de salud mental o futuras pandemias.

Teniendo en cuenta todos los aspectos éticos tratados anteriormente, es evidente que la ética es esencial a la hora de planificar y ofrecer servicios sociales. Por esto, una buena sugerencia a tener en cuenta sería la creación de un comité ético, el cual este encargado de evaluar las implicaciones del uso de datos en los servicios sociales, garantizando así que las decisiones tomadas sean justas para todos. Y también de identificar posibles riesgos para poder prevenir la exclusión de grupos o desigualdades.

## GLOSARIO DE TERMINOLOGÍA

TERMINOLOGÍA	DEFINICIÓN
<b>Distrito 04</b>	Área geográfica específica mencionada en el texto, que incluye Begoña, Bolueta y Santutxu.
<b>Cooperativa</b>	Organización formada por un grupo de personas que se asocian para satisfacer necesidades comunes, en este caso, la promoción de la lengua y cultura vasca.
<b>Tendencia de indicadores</b>	Análisis de datos que muestra cómo ciertos factores están cambiando a lo largo del tiempo, lo que ayuda a prever futuros comportamientos o necesidades.
<b>Demografía</b>	Estudio de las poblaciones humanas, incluyendo su tamaño, distribución y cambios a lo largo del tiempo.
<b>Comparabilidad</b>	Capacidad de comparar datos o información de diferentes fuentes o períodos para identificar tendencias y patrones.
<b>Cartera de Servicios Sociales</b>	Conjunto de servicios que se ofrecen a la población en el ámbito social.
<b>Plantilla técnica</b>	Conjunto de profesionales que trabajan en un área específica, en este caso, en servicios sociales.
<b>Tipos de servicio (1.3 y 2.7.3)</b>	Referencias específicas a categorías de servicios dentro de la Cartera de Servicios Sociales que se consideran prioritarios para la adaptación y mejora.
<b>Base de datos Trebatu</b>	Sistema de gestión de datos que proporciona información sobre expedientes abiertos en Gipuzkoa.
<b>Indicadores</b>	Medidas o métricas que se utilizan para evaluar y comparar diferentes aspectos de una población o situación, como el PIB, la tasa de paro, etc.
<b>Medidas estadísticas</b>	Herramientas matemáticas utilizadas para resumir, analizar e interpretar datos, facilitando la comparación y la

	comprensión de la información.
<b>Media</b>	Medida de tendencia central que se obtiene sumando todos los valores de un conjunto de datos y dividiéndolos por el número total de valores. Representa el promedio de los datos.
<b>Mediana</b>	Medida de tendencia central que divide un conjunto de datos en dos partes iguales. Es el valor que se encuentra en el medio cuando los datos están ordenados.
<b>Desviación estándar</b>	Medida que indica la dispersión de un conjunto de datos respecto a su media. Mide cuánto se alejan los valores individuales de la media.
<b>Rango</b>	Diferencia entre el valor máximo y el valor mínimo en un conjunto de datos. Proporciona una medida simple de la extensión de los datos.
<b>Cuartiles</b>	Valores que dividen un conjunto de datos en cuatro partes iguales. El primer cuartil (Q1) delimita el 25% inferior, el segundo cuartil (Q2) es la mediana, y el tercer cuartil (Q3) delimita el 25% superior.
<b>Rango intercuartílico</b>	Diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1). Mide la dispersión de la parte central de los datos, excluyendo los valores extremos.
<b>Tendencia central</b>	Diferencia entre el tercer cuartil (Q3) y el primer cuartil (Q1). Mide la dispersión de la parte central de los datos, excluyendo los valores extremos.
<b>Dispersión</b>	Grado en que los valores de un conjunto de datos se distribuyen o se separan entre sí. Se mide a través de la desviación estándar, el rango y el rango intercuartílico.
<b>Outliers</b>	Datos que se encuentran significativamente alejados del resto de los valores en un conjunto de datos. Pueden influir en las medidas de tendencia central y dispersión.
<b>Línea de tendencia</b>	Línea que se ajusta a los datos en un gráfico de dispersión, que ayuda a

	visualizar la dirección general de la relación entre las variables.
<b>Algoritmos</b>	Conjuntos de instrucciones o reglas definidas que permiten resolver problemas o realizar tareas específicas en el análisis de datos.
<b>Sesgos</b>	Tendencias que pueden influir en la interpretación de datos y generar decisiones injustas o discriminatorias.
<b>Estigmatización</b>	Proceso de desvalorización social de un individuo o grupo, que puede resultar de la revelación de datos sensibles.
<b>Renta del Capital Mobiliario</b>	La renta de capital mobiliario se refiere a los ingresos generados por bienes muebles, como dividendos de acciones, intereses de cuentas de ahorro, o alquiler de bienes de consumo.



## GLOSARIO DE SIGLAS

SIGLA	DEFINICIÓN
EUSTAT	Instituto Vasco de Estadística
INE	Instituto Nacional de Estadística
PIB	Producto Interno Bruto
ODS	Objetivos de Desarrollo Sostenible

## **ANEXO: Resumen del Proyecto**

El principal problema radica en que los servicios sociales no se ajustan adecuadamente a las cambiantes necesidades de la población, debido a una planificación no basada en datos actualizados y a una falta de coordinación entre sectores como salud, empleo y vivienda. Además, hay un desajuste entre el número de profesionales y las competencias necesarias, complicando la respuesta a las demandas sociales emergentes.

### **Recopilación y procesamiento de datos:**

Se han utilizado bases de datos de Trebatu, EUSTAT e INE para recopilar información de los municipios de Gipuzkoa y Bizkaia, como PIB, tasa de paro, inmigración, población, etc. Elegimos el año 2019 para omitir los efectos de la pandemia. Los datos los procesamos en Excel, aplicando filtros y transformaciones como la conversión de expedientes por cada 1000 habitantes y el uso de porcentajes para analizar variables.

### **Análisis descriptivo y modelado de datos:**

Se utilizaron medidas estadísticas como media, mediana y desviación estándar para comparar indicadores. El análisis mostró que un mayor número de habitantes se correlaciona con más expedientes abiertos, pero menos por cada 1000 habitantes en zonas más pobladas. La renta y el PIB también mostraron una relación inversa con el número de expedientes, indicando que un entorno económico mejor reduce la demanda de servicios sociales.

### **Transformación del negocio:**

El uso de modelos predictivos basados en datos socioeconómicos facilita a Agintzari anticipar la demanda de servicios sociales y optimizar la distribución de recursos. Esto incrementa la eficacia y la calidad del servicio, mejorando la percepción de la cooperativa como una organización que responde a las necesidades sociales de forma efectiva.

### **Implicaciones legales y éticas:**

El uso de datos en modelos predictivos conlleva riesgos éticos, como el sesgo y la discriminación. Es crucial diseñar los modelos de manera justa, supervisar los algoritmos y garantizar la confidencialidad de los datos personales. La planificación debe basarse en datos de calidad para evitar exclusión de grupos y estereotipos.

## **BIBLIOGRAFÍA:**

[BILBAO OPEN DATA. EDADES](#)

[EUSTAT. POBLACIÓN](#)