

Treball de Fi de Grau/Màster

## **Grau en Enginyeria de Tecnologies Industrials**

# **Ciencia de datos para el análisis de la percepción ciudadana sobre la ciudad de Barcelona (CS)**

### **MEMORIA**

**Autor/a:** Adrià Núñez Fernández

**Director/a:** Luis José Talavera Mendez

**Convocatoria:** 04/2023



Escola Tècnica Superior  
d'Enginyeria Industrial de Barcelona





## Resumen

El estudio de fin de grado que el lector se dispone a leer se centra en la utilización de la ciencia de datos para analizar la percepción ciudadana sobre la ciudad de Barcelona y la obtención de relaciones entre los datos de los habitantes encuestados.

Con tal de desarrollar el trabajo se han utilizado la minería de datos, análisis estadísticos exploratorios y análisis de Clustering.

En esta memoria se presenta la documentación teórica relacionada con el estudio, los archivos de datos utilizados, los métodos estadísticos realizados, los resultados obtenidos y las conclusiones extraídas.

Todos los procesos analíticos de este trabajo han sido realizados con el lenguaje de programación Python con las librerías de Pandas y Kmodes.

## Resum

L'estudi de fi de grau que el lector es disposa a llegir es centra en la utilització de la ciència de dades per a analitzar la percepció ciutadana sobre la ciutat de Barcelona i l'obtenció de relacions entre les dades dels habitants enquestats.

Per tal de desenvolupar el treball s'han utilitzat la mineria de dades, anàlisis estadístiques exploratòries i anàlisis de Clustering.

En aquesta memòria es presenta la documentació teòrica relacionada amb l'estudi, els arxius de dades utilitzades, els mètodes estadístics realitzats, els resultats obtinguts i les conclusions extretes.

Tots els processos analítics d'aquest treball han estat realitzats amb el llenguatge de programació Python amb les llibreries de Pandas i Kmodes.

## Abstract

The final-degree project that the reader is about to read focuses on the use of data science to analyse citizen perception of the city of Barcelona and to obtain relationships in the data of the inhabitants surveyed.

In order to develop the project, data mining, exploratory statistical analysis and clustering analysis have been used.

This report presents the theoretical documentation related to the study, the data files used, the statistical methods used, the results obtained and the conclusions drawn.

All the analytical processes in this work have been done using the Python programming language with the Pandas and Kmodes libraries.



# Índice

Resumen.....	3
Resum.....	4
Abstract .....	5
Índice.....	7
Glosario y Nomenclatura .....	8
Lista de figuras.....	9
1. Prefacio .....	11
2. Introducción .....	12
2.1. Motivación .....	12
2.2. Alcance del trabajo.....	13
2.3. Objetivos del trabajo .....	13
3. Fundamentos teóricos.....	14
3.1. Python .....	14
3.2 Pandas .....	15
3.3 Entorno de trabajo .....	15
3.4 Exploratory Data Analysis .....	17
3.5 Clustering .....	17
4. Datos iniciales .....	21
4.1 Encuesta de opinión .....	21
4.2 Datos de inseguridad .....	23
4.3 Datos de limpieza.....	23
4.4 Datos de vivienda.....	24
5. Metodología y resultados .....	25
5.1 Encuesta de opinión .....	25
5.2 Estudio de inseguridad .....	34
5.3 Estudio de limpieza .....	39
5.4 Estudio de vivienda .....	42
5.5 Clustering .....	46
6. Planificación.....	62
7. Estudio económico .....	63
8. Estudio ambiental.....	64
9. Estudio social y de igualdad de género .....	65
10. CONCLUSIONES .....	66
11. Bibliografía .....	68

## Glosario y Nomenclatura

Csv: archivo de texto en el cual los caracteres están separados por comas, haciendo una especie de tabla en filas y columnas.

Python: lenguaje de programación utilizado para desarrollar el proyecto.

Jupyter notebook: entorno interactivo basado en la web utilizado para desarrollar el código Python

Clustering: técnica de análisis estadístico utilizada para definir grupos similares dentro de un conjunto de datos.

K-Modes: método de desarrollo de clustering utilizado en este trabajo.

Clúster: grupo definido mediante clustering.

## Lista de figuras

Figura 1: Agrupación K-Means.

Figura 2: Fórmula K-Means.

Figura 3: Fichero csv “2022\_r22032\_servicios\_municipales\_evolucio2022\_bd\_csvtext\_v1\_0.csv”.

Figura 4: Dataframe de datos de opinión.

Figura 5: Fichero csv “Fets\_penals\_coneguts\_fets\_coneguts\_resolts\_i\_detencions.csv”.

Figura 6: Fichero csv “estat\_espais\_prioritaris\_2019.csv”

Figura 7: Fichero csv “2022\_comp\_vend\_preu\_trim.csv”

Figura 8: Fichero csv “2019\_atles\_renda\_bruta\_persona.csv”

Figura 9: Cuenta de NOM\_DISTRICTE

Figura 10: Gráfico de barras de NOM\_DISTRICTE

Figura 11: Cuenta de SEXE

Figura 12: Gráfico de barras de SEXE

Figura 13: Cuenta de EDAT1889\_1A6

Figura 14: Gráfico de barras de EDAT1889\_1A6

Figura 15: Cuenta de SITLABORAL\_1A9

Figura 16: Gráfico de barras de SITLABORAL\_1A9

Figura 17: Cuenta de SATISF\_RES\_BCN\_1A4

Figura 18: Gráfico de barras de SATSF\_RES\_BCN1A4

Figura 19: Cuenta de PROB\_BCN\_AGR

Figura 20: Gráfico de barras de PROB\_BCN\_AGR

Figura 21: Cuenta de US30D\_XARXES\_Q01

Figura 22: Gráfico de barras de US30D\_XARXES\_Q01

Figura 23: Cuenta de IDEO\_1A8

Figura 24: Gráfico de barras de IDEO\_1A8

Figura 25: Gráfico de barras de ESTUDIS\_1A6

Figura 26: Gráfico de barras de ESTUDIS\_1A6

Figura 27: Cuenta de LLAR\_INGRESSOS\_1A10

Figura 28: Gráfico de barras de LLAR\_INGRESSOS\_1A10

Figura 29: Porcentaje primer problema inseguridad

Figura 30: Gráfico de barras porcentaje primer problema inseguridad

Figura 31: Índice de robos

Figura 32: Gráfico de barras índice de robos

Figura 33: Índice de lesiones

Figura 34: Gráfico de barras de índice de lesiones

Figura 35: Índice de delitos

Figura 36: Gráfica de barras del índice de delitos

Figura 37: Porcentaje primer problema limpieza

Figura 38: Gráfico de barras porcentaje primer problema limpieza

Figura 39: Cantidad de basura

Figura 40: Gráfico de barras de cantidad de basura

Figura 41: Porcentaje primer problema vivienda

Figura 42: Gráfico de barras primer problema vivienda

Figura 43: Precio operaciones compra-venta

Figura 44: Gráfico de barras de precio operaciones compra-venta

Figura 45: Renta por cápita

Figura 46: Gráfico de barras de renta por cápita

Figura 47: Gráfico número óptimo de clusters

Figura 48: Clusters

Figura 49: Cuenta de NOM\_DISTRICTE de Cluster 1

Figura 50: Cuenta de SEXE de Cluster 1

Figura 51: Cuenta de EDAT1899\_1A6 de Cluster 1

- Figura 52: Cuenta de SATISF\_RES\_BCN de Cluster 1  
Figura 53: Cuenta de PROB\_BCN\_AGR de Cluster 1  
Figura 54: Cuenta de IDEO\_1A8 de Cluster 1  
Figura 55: Cuenta de ESTUDIS\_1A6 de Cluster 1  
Figura 56: Cuenta de LLAR\_INGRESSOS\_1A10 de Cluster 1  
Figura 57: Cuenta de NOM\_DISTRICTE de Cluster 2  
Figura 58: Cuenta de SEXE de Cluster 2  
Figura 59: Cuenta de EDAT1899\_1A6 de Cluster 2  
Figura 60: Cuenta de SATISF\_RES\_BCN de Cluster 2  
Figura 61: Cuenta de PROB\_BCN\_AGR de Cluster 2  
Figura 62: Cuenta de IDEO\_1A8 de Cluster 2  
Figura 63: Cuenta de ESTUDIS\_1A6 de Cluster 2  
Figura 64: Cuenta de LLAR\_INGRESSOS\_1<sup>a</sup>10 de Cluster 2  
Figura 65: Cuenta de NOM\_DISTRICTE de Cluster 3  
Figura 66: Cuenta de SEXE de Cluster 3  
Figura 67: Cuenta de EDAT1899\_1<sup>a</sup>6 de Cluster 3  
Figura 68: Cuenta de SATISF\_RES\_BCN de Cluster 3  
Figura 69: Cuenta de PROB\_BCN\_AGR de Cluster 3  
Figura 70: Cuenta de IDEO\_1A8 de Cluster 3  
Figura 71: Cuenta de ESTUDIS\_1<sup>a</sup>6 de Cluster 3  
Figura 72: Cuenta de LLAR\_INGRESSOS\_1A10 de Cluster 3  
Figura 73: Cuenta de NOM\_DISTRICTE de Cluster 4  
Figura 74: Cuenta de SEXE de Cluster 4  
Figura 75: Cuenta de EDAT1899\_1<sup>a</sup>6 de Cluster 4  
Figura 76: Cuenta de SATISF\_RES\_BCN de Cluster 4  
Figura 77: Cuenta de PROB\_BCN\_AGR de Cluster 4  
Figura 78: Cuenta de IDEO\_1<sup>a</sup>8 de Cluster 4  
Figura 79: Cuenta de ESTUDIS\_1<sup>a</sup>6 de Cluster 4  
Figura 80: Cuenta de LLAR\_INGRESSOS\_1<sup>a</sup>10 de Cluster 4  
Figura 81: Cuenta de NOM\_DISTRICTE de Cluster 5  
Figura 82: Cuenta de SEXE de Cluster 5  
Figura 83: Cuenta de EDAT1899\_1<sup>a</sup>6 de Cluster 5  
Figura 84: Cuenta de SATISF\_RES\_BCN de Cluster 5  
Figura 85: Cuenta de PROB\_BCN\_AGR de Cluster 5  
Figura 86: Cuenta de IDEO\_1A8 de Cluster 5  
Figura 87: Cuenta de ESTUDIS\_1A6 de Cluster 5  
Figura 88: Cuenta de LLAR\_INGRESSOS\_1A10 de Cluster 5  
Figura 89: Diagrama de Gant

# 1. Prefacio

El presente trabajo consta de los siguientes análisis:

- Una exploración de los datos obtenidos de la encuesta de satisfacción realizada a los habitantes de Barcelona para comprender lo que nos indica cada variable tomada en cuenta y la conclusión que se obtiene de los valores de cada una.
- Un análisis exploratorio de la inseguridad en los diferentes distritos de Barcelona, problema que más encuestados consideran como el más importante de vivir en la ciudad.
- Un análisis exploratorio de la limpieza en los diferentes distritos de Barcelona, segundo problema que más encuestados consideran como el más importante de vivir en la ciudad.
- Un análisis exploratorio del acceso a la vivienda en los diferentes distritos de Barcelona, tercer problema que más encuestados consideran como el más importante de vivir en la ciudad.
- Un análisis de Clustering con tal de dividir la población encuestada en varios grupos de características similares para poder comprender toda la información que nos muestran los datos de la encuesta de opinión de Barcelona y relacionarla con los análisis previos.

## 2. Introducción

El Ayuntamiento de Barcelona realiza cada año una encuesta de satisfacción y opinión a los habitantes residentes en la ciudad y publica los datos en el portal de Open Data de Barcelona. Esta encuesta consta de muchas preguntas realizadas a los encuestados y se sube en un archivo csv. En este portal también se encuentran muchos archivos csv que contienen datos relativos a servicios y otros aspectos de la ciudad.

En este trabajo se realizarán análisis estadísticos a partir de los datos publicados de las encuestas de opinión y otros archivos de datos presentes en el portal Open Data para extraer la máxima información posible de ellos.

Dicho estudio será de utilidad para la población de Barcelona en tanto que les permitirá identificar los principales problemas a resolver, y sobre todo para el Ayuntamiento y otros organismos de gestión, dándoles una información que pueden utilizar para mejorar su gestión.

Este estudio también proveerá información útil a partir de la cual realizar estudios políticos y de gestión pública.

### 2.1. Motivación

La decisión de realizar este trabajo ha comenzado a partir de la motivación de aprender a realizar estudios estadísticos en Python y aprender a trabajar mejor con dicho lenguaje de programación, descubriendo más sobre la estadística de análisis y sobre las herramientas que se disponen para realizarla con Python.

Se eligió el tema sobre el cual realizar el análisis a partir del descubrimiento de los archivos de datos publicados por el ayuntamiento de Barcelona y la curiosidad generada por ellos a raíz también de las próximas elecciones del Ayuntamiento de Barcelona y el clima socio-político general.

El Ayuntamiento de Barcelona dispone de archivos de datos en el depósito de datos digitales abierto al público Open Data BCN. Sin embargo solo se permite visualizar estos archivos csv. La excepción es el caso de los datos de la encuesta de opinión a los habitantes de Barcelona, la cual se puede visualizar de forma gráfica en el mismo portal. Sin embargo dicha herramienta digital solamente permite cruzar dos valores de cada variable. Tampoco permite comparar los datos de diferentes archivos entre sí. De esta forma, resulta muy complicado obtener conclusiones. Se decide realizar el análisis que se presenta en este trabajo para poder obtener mayores conclusiones.

## 2.2. Alcance del trabajo

El estudio se realizará solamente a partir de los archivos de datos csv publicados por el Ayuntamiento de Barcelona en el portal Open Data BCN y un archivo de datos publicado por los Mossos d'Esquadra. Además, se tendrán en cuenta solamente los datos de años más próximos al año de publicación de este trabajo, el 2023.

## 2.3. Objetivos del trabajo

El objetivo general del trabajo es la extracción de conclusiones estadísticas de la percepción ciudadana sobre la ciudad de Barcelona.

Los objetivos particulares son los siguientes:

- Búsqueda y obtención de los archivos de datos necesarios.
- Estudio bibliográfico de las técnicas disponibles para realizar estudios estadísticos en Python.
- Exploración de los datos disponibles y comprensión de lo que estos datos indican.
- Realizar análisis de datos exploratorios de los principales problemas de la población de Barcelona obteniendo patrones y averiguando información importante que estos datos puedan aportar.
- Distinguir grupos de características similares en la población de Barcelona con respecto a su opinión de vivir en ella.

## 3. Fundamentos teóricos

En este apartado se aporta la información teórica necesaria para realizar el trabajo. Se explica el código de programación que se usa para realizar los análisis, la librería del lenguaje más utilizada, el entorno de trabajo, y las técnicas estadísticas que se utilizan.

### 3.1. Python

Python es un lenguaje de programación de alto nivel creado con el objetivo principal de tener un código fácilmente legible. Es un lenguaje muy utilizado para el desarrollo de todo tipo de programas, Instagram, Spotify y Netflix son ejemplos de aplicaciones famosas que contienen código Python.

Al contrario que otros lenguajes como por ejemplo Java o .net, Python es interprete, lo cual significa que no es necesario compilarlo para ejecutar sus funciones, si no que es posible ejecutar directamente dichas funciones mediante un programa interpretador.

Python es además un lenguaje multiplataforma de código abierto y gratuito, lo cual permite una amplia utilización en todo el mundo y en ambientes de todo tipo.

Todas estas características hacen que el lenguaje de programación Python se haya convertido en uno de los más populares. Además, es un código especialmente útil para trabajar en los ámbitos de inteligencia artificial, Big Data, machine learning y data science, campos cada vez más necesarios y presentes en empresas y la sociedad en general. Será para trabajar dentro de estos ámbitos para lo que se utiliza Python en este proyecto.

Gracias a la facilidad con que Python permite limpiar y organizar los datos, crear algoritmos de aprendizaje automático y cruzar los datos entre sí mediante motores numéricos como Pandas y Scikit-learn, que son dos de las muchas bibliotecas de funciones diferentes que posee Python, es un lenguaje de programación especialmente adecuado para analizar grandes cantidades de datos y transformarlos en información útil para su exploración y la predicción de futuras tendencias utilizándose por lo tanto en la actualidad en muchísimas empresas dedicadas al sector de data analytics, tanto directamente como integrado dentro de softwares comerciales de análisis de datos.

Sin embargo, y a pesar de que en este programa se utilizará Python para las funciones descritas anteriormente, también es utilizado para muchas otras aplicaciones.

Python también se utiliza para la base de datos distribuida Blockchain, la base sobre la que se sustentan las criptomonedas, un tema popular en la actualidad. La sencillez, versatilidad y rapidez de éste lenguaje de programación permite a los desarrolladores crear cadenas de bloques a partir de pocas líneas de código.

También se puede utilizar Python para el desarrollo web. Permite crear webs mediante códigos más cortos y rápidos de hacer que otros lenguajes, además de más claros para el entendimiento. Las amplias bibliotecas de las que dispone Python son siempre de ayuda, siendo Django uno de los frameworks más utilizados para la creación de webs en la actualidad. Relacionado con este aspecto de la utilización de Python también destaca el scraping, la obtención de datos de todo tipo de webs, método utilizado por muchísimas empresas.

Otro campo para el cual se utiliza Python es en el desarrollo gráfico. Existen juegos éxitos en ventas cuyo motor gráfico ha sido desarrollado por dicho lenguaje de programación. La principal característica de Python para esta utilidad vuelve a ser la gran cantidad de herramientas, siendo PyGame, Blender y Arcade unas de las bibliotecas más utilizadas en este ámbito.

## 3.2 Pandas

Pandas es una de las muchas librerías de código abierto de las que dispone Python. Su nombre proviene del término “Panel Data” término de econometría que designa datos que combinan una dimensión temporal con otra dimensión transversal.

Se creó con el objetivo de juntar en una sola librería todas las herramientas necesarias para trabajar realizando análisis de datos, como por ejemplo adquirir los datos de un archivo, manipularlos y prepararlos, modelarlos y analizarlos.

Las dos principales estructuras de la librería Pandas son las Series, objetos unidimensionales capaces de almacenar cualquier tipo de dato, siendo similares a una columna de una tabla, y las DataFrames, objetos bidimensionales que pueden almacenar también todo tipo de dato, consistentes en una estructura de series componiendo una tabla con filas y columnas ordenadas.

La librería Pandas permite también realizar muchas funciones como pueden ser curvas, gráficas, diagramas e historiogramas.

## 3.3 Entorno de trabajo

### Interprete de Python

La forma básica de utilizar el lenguaje Python es mediante su propia consola presente en los sistemas operativos comerciales. Es un intérprete interactivo que permite ejecutar

instrucciones y ver el resultado de la comanda inmediatamente. No se suele utilizar dicho interprete para trabajar con el programa pero resulta útil para testear pequeñas partes del código y para el aprendizaje.

### **Entornos de desarrollo integrado**

Los entornos de desarrollo integrado, o IDE en inglés, son un tipo de aplicaciones diseñadas con tal de mejorar la productividad a la hora de diseñar código. Se disponen de muchos IDEs diferentes para trabajar con Python. Ofrecen diferentes beneficios respecto al interprete usual de Python, como autocompletado, información de errores, interfaz cómoda, etc.

### **Entornos interactivos basados en la web**

Un ejemplo de entorno interactivo basado en la web es el utilizado para desarrollar el código de este trabajo, Jupyter Notebook. Este entorno de desarrollo consiste en una aplicación web a la que se accede a través de cualquier navegador de internet comercial. En concreto, esta aplicación facilita la creación de cuadernos (notebooks) compuestos por celdas. En estas celdas podemos desarrollar nuestro código Python e ir las ejecutando una a una.

Dicho entorno es altamente utilizado en data science y machine learning, ya que permite visualizar gráficos y disponer del código que los ha generado en el mismo documento.

### **Jupyter Notebook**

Jupyter Notebook, como se ha comentado en el apartado anterior, es el entorno interactivo basado en la web que se ha utilizado para desarrollar el código del presente trabajo. Es un software libre basado consecuentemente en componentes libres, por lo cual está secundado, desarrollado y mantenido por una amplia comunidad. Jupyter Notebook sigue un esquema y una lista ordenada de celdas de entrada y de salida. Estas celdas albergan, entre otras cosas, código, texto (en formato Markdown), fórmulas matemáticas, ecuaciones, y contenido multimedia (Rich Mediana). Los documentos creados en Jupyter pueden exportarse, entre otros formatos, a HTML, PDF, Markdown o Python. Los dos componentes principales de Jupyter Notebook son un conjunto de núcleos (Interpreter) y los Dashboard. Cada núcleo es un motor de ejecución para un lenguaje que se encarga de procesar las solicitudes y devolver las respuestas apropiadas. El núcleo por defecto es Python. La principal ventaja que ofrece este entorno es la creación rápida de prototipos y la facil visualización de los resultados.

### 3.4 Exploratory Data Analysis

El análisis exploratorio de datos, EDA en inglés, se utiliza para analizar conjuntos de datos con tal de comprender la información que se presenta en ellos, observando las principales características, patrones sencillos, compararlos entre sí, identificar errores presentes en los datos, etc.

Existen cuatro tipos principales de métodos:

- Cálculos univariantes. Se utilizan para comprender los datos y identificar patrones presentes en ellos.
- Gráficos univariantes. Sirven para comprender los datos y identificar patrones presentes en ellos de una forma más visual. Hay muchos tipos de gráficos, como por ejemplo los gráficos de barras y los historiogramas
- Cálculos bivariantes. Se utilizan para establecer relaciones estadísticas entre datos
- Gráficos bivariantes. Se utilizan para establecer relaciones estadísticas entre datos de forma más visual.

En este trabajo se utiliza el análisis de datos exploratorio mediante los métodos univariantes para comprender los datos de los que disponemos, establecer patrones, y compararlos entre sí.

### 3.5 Clustering

El clustering es un método no supervisado de machine-learning . Su objetivo es definir, de forma automática, grupos de datos con aspectos en común a partir de un conjunto de elementos de diferentes características o propiedades pero que comparten ciertas similitudes. Esto significa que no es necesario partir de ninguna hipótesis de correlación inicial, si no que es posible descubrir relaciones que no se habían planteado en un inicio y que no habían sido investigadas mediante el análisis exploratorio de datos.

Algunos ejemplos de la utilidad del clustering pueden ser la segmentación de clientes en grupos, determinar patrones de comportamiento de un grupo de personas, patrones climáticos, agrupar artículos o noticias por temas...

En este proyecto se utilizará para determinar grupos de personas encuestadas para descubrir posibles relaciones entre algunas características y la opinión personal que tienen respecto al problema principal de vivir en Barcelona.

El algoritmo de clustering más utilizado es el K-Means, el cual agrupa los datos a partir de

la definición de un centroide y minimiza la distancia cuadrática entre cada dato y el centroide de su grupo mediante iteraciones.

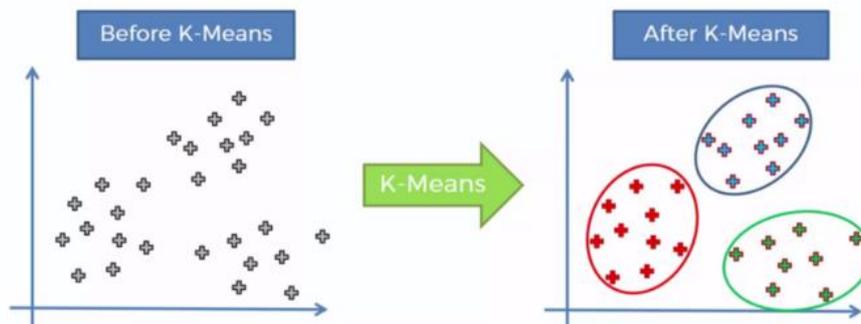


Figura 1: Agrupación K-Means

Este método sigue los siguientes pasos:

1. Escoger las variables que se desea usar para agrupar los datos
2. Determinar una cantidad de grupos  $k$  y establecer un centroide en cada grupo
3. Asignar los datos más cercanos a cada centroide a su grupo
4. Calcular la distancia cuadrática entre cada dato y su centroide
5. Establecer nuevos grupos con sus respectivos centroides, volver a calcular la distancia cuadrática entre ellos y comprobar si crece o decrece.
6. Iterar el proceso anterior y finalizar obteniendo los grupos donde menor distancia cuadrática existe entre los datos de cada grupo y su centroide.

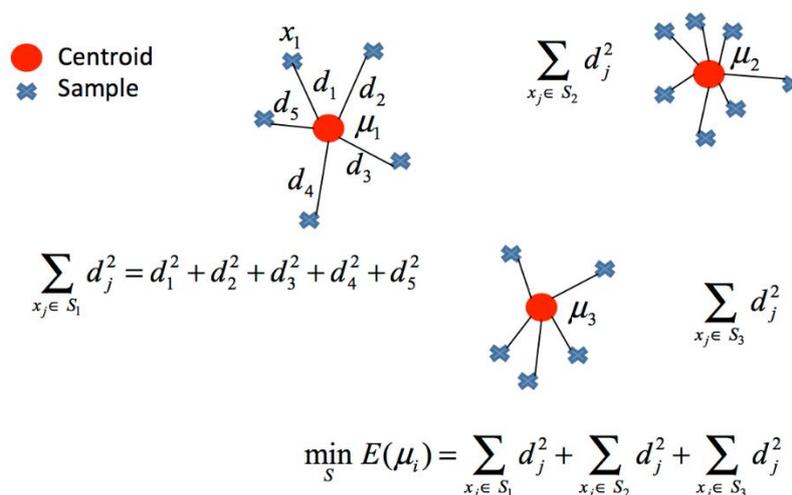


Figura 2: Fórmula K-Means

Para identificar el número de clústers óptimo, el método más utilizado es el de Elbows. Este método comienza creando un grupo y calcula la suma de las distancias al cuadrado entre los diferentes objetos y el centroide. Continúa creando dos grupos y calculando la suma de las distancias al cuadrado entre los diferentes objetos y sus centroides. Cuantos más grupos se crean, menores son las distancias entre los objetos y los centroides. Continúa iterando hasta que se tienen muestras hasta un número elegido de grupos. La cantidad óptima de clústers se obtiene cuando la suma de las distancias al cuadrado entre los objetos y sus centroides comienza a decrecer levemente al añadir nuevos grupos.

Sin embargo, el algoritmo K-Means solamente se puede utilizar para analizar un conjunto de datos cuyas variables sean numéricas, puesto que se debe calcular la distancia entre los centroides y cada uno de los puntos.

Para poder realizar el análisis necesario en nuestro trabajo se utiliza el algoritmo K-Modes, que funciona para conjuntos de datos que contienen variables categóricas, como es el caso de los datos de los cuales se disponen.

El método del K-modes se diferencia del K-Means en que, en vez de calcular la distancia entre puntos, pues al no haber valores numéricos esto no es posible, calcula la moda de cada variable de cada objeto y las compara con las modas de las variables de los centroides.

El objetivo de este algoritmo es minimizar cuantas modas de cada objeto son diferentes respecto a las modas de los centroides de sus clusters. De la misma forma que en el método k-means, itera diferentes grupos con diferentes centroides para averiguar qué agrupación minimiza estas diferencias de las modas entre los objetos de cada grupo.

Respecto al método de los Elbows para identificar el número de clústers óptimos, también se realiza calculando la cantidad de modas de cada objeto que varían respecto de las de sus centroides en cada iteración, y se eligen el número de centroides a partir del cual la diferencia entre las modas entre iteración y iteración comienza a decrecer en menor cantidad.

## 4. Datos iniciales

En los siguientes puntos se describen los datos iniciales a partir de los cuales se realizan los estudios presentes en el trabajo.

### 4.1 Encuesta de opinión

Los datos iniciales relativos a la encuesta de opinión parten del archivo csv separado por comas “2022\_r22032\_servicios\_municipales\_evolutio2022\_bd\_csvtext\_v1\_0.csv”, y es descargado del Servicio de datos abiertos del Ayuntamiento de Barcelona Datos abiertos BCN, consistente a un registro de todas las encuestas realizadas a los habitantes.

Este archivo csv se compone de 271 columnas y 154489 filas. Las columnas corresponden a las preguntas realizadas en la encuesta y las filas a las respuestas de cada ciudadano. Es decir, cada columna contiene una pregunta de la encuesta y las diferentes respuestas y en cada fila toda la información de un solo habitante.

La siguiente figura consiste en la lectura del csv inicial mediante Python y Jupyter Notebook ya convertido en un dataframe de Python:

ANY	NOM_DISTRICTE	NOM_BARRI	NOM_GRANBARRI	DATA_ENG	SEXE	EDAT1899_1A7	EDAT1899_1A8	NACIO_ACT_UE	NACIO_CANVI	PMI_PROF_ACT_B	PMI_PROF_ABANS	PMI_PROF_ABANS_A	PMI_PROF_ABANS_B	PMI_PROF_ABANS_C	PMI_PROF	NT
0	2022	GRACIA	LA VILA DE GRACIA	LA VILA DE GRACIA	2022-02-17	DONA	DE 35 A 44 ANYS	DE 35 A 44 ANYS	ESpanyola	SÍ, SEMPRE HA TINGUT NACIONALITAT ESPANYOLA	PROFESIONAL / TÈCNIC/A / COMANDAMENT INTERMEDI	NaN	NaN	NaN	NaN	PROFESIONAL / TÈCNIC/A / COMANDAMENT INTERMEDI
1	2022	SANT MARTÍ	EL CLOT	EL CLOT I EL CAMP DE L'ARPA	2022-02-17	HOME	DE 65 A 74 ANYS	DE 65 ANYS I MÉS	ESpanyola	SÍ, SEMPRE HA TINGUT NACIONALITAT ESPANYOLA	NaN	PERSONAL ADMINISTRATIU D'OFICINA, SERVEIS, COM...	PER COMPTE D'ALTRI (COM A ASSALARIAT/ADA)	NaN	PERSONAL ADMINISTRATIU D'OFICINA, SERVEIS, COM...	PERSONAL ADMINISTRATIU D'OFICINA, SERVEIS, COM...
2	2022	SANT MARTÍ	EL CLOT	EL CLOT I EL CAMP DE L'ARPA	2022-02-17	HOME	DE 18 A 24 ANYS	DE 18 A 24 ANYS	ESpanyola	NO, ABANS TENIA UNA ALTRA NACIONALITAT	PERSONAL DEL SECTOR DE RESTAURACIÓ, COMERÇ, S...	NaN	NaN	NaN	NaN	PERSONAL DEL SECTOR DE RESTAURACIÓ, COMERÇ, S...
3	2022	SANT MARTÍ	EL CLOT	EL CLOT I EL CAMP DE L'ARPA	2022-02-17	HOME	DE 75 ANYS I MÉS	DE 65 ANYS I MÉS	ESpanyola	SÍ, SEMPRE HA TINGUT NACIONALITAT ESPANYOLA	NaN	PERSONAL ADMINISTRATIU D'OFICINA, SERVEIS, COM...	PER COMPTE D'ALTRI (COM A ASSALARIAT/ADA)	NaN	PERSONAL ADMINISTRATIU D'OFICINA, SERVEIS, COM...	PERSONAL ADMINISTRATIU D'OFICINA, SERVEIS, COM...
4	2022	SANT MARTÍ	EL CLOT	EL CLOT I EL CAMP DE L'ARPA	2022-02-17	HOME	DE 35 A 44 ANYS	DE 35 A 44 ANYS	ESpanyola	NO, ABANS TENIA UNA ALTRA NACIONALITAT	PERSONAL OBRER DE LA INDÚSTRIA I LA CONSTRUCCIÓ	NaN	NaN	NaN	NaN	PERSONAL OBRER DE LA INDÚSTRIA I LA CONSTRUCCIÓ
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
154483	2020	SANT MARTÍ	EL BESÓS I EL MARESME	EL BESÓS, EL MARESME I PROVENÇALS	2020-12-02	HOME	DE 35 A 44 ANYS	DE 35 A 44 ANYS	RESTA PAÏSOS	NaN	NaN	NaN	NaN	NaN	NaN	ESTUDIA
154484	2020	SANT MARTÍ	EL BESÓS I EL MARESME	EL BESÓS, EL MARESME I PROVENÇALS	2020-12-02	DONA	DE 18 A 24 ANYS	DE 18 A 24 ANYS	ESpanyola	SÍ, SEMPRE HA TINGUT NACIONALITAT ESPANYOLA	PERSONAL OBRER DE LA INDÚSTRIA I LA CONSTRUCCIÓ	NaN	NaN	NaN	NaN	PERSONAL OBRER DE LA INDÚSTRIA I LA CONSTRUCCIÓ
154485	2020	SANT MARTÍ	EL BESÓS I EL MARESME	EL BESÓS, EL MARESME I PROVENÇALS	2020-12-02	HOME	DE 25 A 34 ANYS	DE 25 A 34 ANYS	RESTA PAÏSOS	NaN	NaN	PERSONAL DEL SECTOR DE RESTAURACIÓ, COMERÇ, SE...	PER COMPTE D'ALTRI (COM A ASSALARIAT/ADA)	NaN	PERSONAL DEL SECTOR DE RESTAURACIÓ, COMERÇ, S...	PERSONAL DEL SECTOR DE RESTAURACIÓ, COMERÇ, S...
154486	2020	SANT MARTÍ	EL BESÓS I EL MARESME	EL BESÓS, EL MARESME I PROVENÇALS	2020-12-02	DONA	DE 75 ANYS I MÉS	DE 65 ANYS I MÉS	ESpanyola	SÍ, SEMPRE HA TINGUT NACIONALITAT ESPANYOLA	PERSONAL OBRER DE LA INDÚSTRIA I LA CONSTRUCCIÓ	NaN	NaN	NaN	NaN	PERSONAL OBRER DE LA INDÚSTRIA I LA CONSTRUCCIÓ
154487	2020	SANT MARTÍ	EL BESÓS I EL MARESME	EL BESÓS, EL MARESME I PROVENÇALS	2020-12-02	DONA	DE 45 A 54 ANYS	DE 45 A 54 ANYS	RESTA PAÏSOS	NaN	NaN	AUTÒNOM/A / EMPRESARIA/AMB'S O MENYS TREBALLA...	TREBALLAVA PER COMPTE PROPÍ (COM AUTÒNOM/A O A...)	NO TENIA TREBALLADORS/ES (AUTÒNOM/A)	NaN	AUTÒNOM/A / EMPRESARIA/AMB'S O MENYS TREBALLA...

154488 rows x 271 columns

Figura 3: Fichero csv “2022\_r22032\_servicios\_municipales\_evolutio2022\_bd\_csvtext\_v1\_0.csv”

Las columnas que se utilizarán para realizar el estudio serán las correspondientes a las siguientes preguntas realizadas en la encuesta, siendo cada pregunta numerada una columna y presentadas en el orden deseado:

- 1) Año
- 2) Distrito municipal
- 3) Sexo
- 4) Edad
- 5) Situación laboral
- 6) Problema que mas afecta personalmente al encuestado
- 7) En qué redes sociales participa
- 9) Ideología política
- 10) Nivel de estudios acabados
- 11) Grado de satisfacción de vivir en Barcelona
- 12) Volumen de ingresos mensuales netos de su familia según tramos

Así pues, se procede a conservar solamente dichas columnas del dataframe, que corresponden a las 0, 1, 5, 7, 12, 25, 27, 188, 203, 227 i 200 según su índice numérico.

También se conservan solamente las filas del archivo que corresponden a encuestas realizadas el último año del estudio, el 2022, con tal de asegurar que las conclusiones extraídas se encuentra vigentes en la fecha de realización del presente trabajo. Para esto se eliminan las filas del dataframe cuyo valor en la primera columna difiere de 2022.

Una vez realizado dicho tratamiento se obtiene el dataframe con la información deseada, que resulta ser de 11 columnas y 6.000 filas. Esto significa que se tiene la información de 6.000 encuestas realizadas en 2022. En la siguiente figura se muestra el dataframe a partir del cual se realizará el estudio.

ANY	NOM_DISTRICTE	SEXE	EDAT1899_1A8	SITLABORAL_1A9	SATISF_RES_BCN_1A4	PROB_BCN_AGR	US30D_XARXES_Q01	IDEO_1A8	ESTUDIS_1A8	LLAR_INGRESSOS_1A10
0	2022	GRÀCIA	DONA	DE 35 A 44 ANYS	TREBALLA PER COMPTE D'ALTRI (COM ASSALARIAT/ADA)	BASTANT SATISFETA (7 A 8)	CONTAMINACIÓ I MEDI AMBIENT	TELEGRAM	ESQUERRA	POST UNIVERSITARIS DE 2.001 A 2.500 EUROS
1	2022	SANT MARTÍ	HOME	DE 65 ANYS I MES	COBRA JUBILACIÓ, PREJUBILACIÓ O PENSIO / INCAP...	POC/GENS SATISFETA (0 A 4)	CONGESTIÓ DE TRANSIT	INSTAGRAM	ESQUERRA	SECUNDARIS GENERALS DE 3.001 A 5.000 EUROS
2	2022	SANT MARTÍ	HOME	DE 18 A 24 ANYS	ESTUDIA	BASTANT SATISFETA (7 A 8)	INSEGURETAT	INSTAGRAM	CENTRE ESQUERRA	SECUNDARIS GENERALS DE 500 A 1.000 EUROS
3	2022	SANT MARTÍ	HOME	DE 65 ANYS I MES	COBRA JUBILACIÓ, PREJUBILACIÓ O PENSIO / INCAP...	BASTANT SATISFETA (7 A 8)	INSEGURETAT	TELEGRAM	ESQUERRA	OBLIGATORIS DE 1.501 A 2.000 EUROS
4	2022	SANT MARTÍ	HOME	DE 35 A 44 ANYS	TREBALLA PER COMPTE D'ALTRI (COM ASSALARIAT/ADA)	BASTANT SATISFETA (7 A 8)	NO HO SAP	FACEBOOK	CENTRE	SECUNDARIS GENERALS DE 500 A 1.000 EUROS
...	...	...	...	...	...	...	...	...	...	...
5995	2022	CIUTAT VELLA	DONA	DE 45 A 54 ANYS	TREBALLA PER COMPTE D'ALTRI (COM ASSALARIAT/ADA)	POC/GENS SATISFETA (0 A 4)	CONTAMINACIÓ I MEDI AMBIENT	WHATSAPP	ESQUERRA	SECUNDARIS GENERALS DE 2.501 A 3.000 EUROS
5996	2022	CIUTAT VELLA	DONA	DE 25 A 34 ANYS	TREBALLA PER COMPTE D'ALTRI (COM ASSALARIAT/ADA)	BASTANT SATISFETA (7 A 8)	PROBLEMES ECONÒMICS	INSTAGRAM	ESQUERRA	SECUNDARIS PROFESSIONALS DE 1.001 A 1.500 EUROS
5997	2022	CIUTAT VELLA	DONA	DE 25 A 34 ANYS	TREBALLA PER COMPTE PROPÍ (COM AUTÒNOM/A O AMB...)	BASTANT SATISFETA (7 A 8)	NO HO SAP	INSTAGRAM	NO HO SAP / NO LI INTERESSA LA POLÍTICA	POST UNIVERSITARIS NO HO SAP
5998	2022	CIUTAT VELLA	HOME	DE 45 A 54 ANYS	TREBALLA PER COMPTE PROPÍ (COM AUTÒNOM/A O AMB...)	BASTANT SATISFETA (7 A 8)	TURISME	WHATSAPP	ESQUERRA	UNIVERSITARIS DE 2.001 A 2.500 EUROS
5999	2022	CIUTAT VELLA	HOME	DE 45 A 54 ANYS	ESTÀ A L'ATUR	BASTANT SATISFETA (7 A 8)	NETEJA	WHATSAPP	CENTRE ESQUERRA	UNIVERSITARIS DE 3.001 A 5.000 EUROS

6000 rows x 11 columns

Figura 4: Dataframe de datos de opinión



## 4.2 Datos de inseguridad

Los datos iniciales a partir de los cuales se estudia la inseguridad presente en Barcelona es un csv extraído del portal web de los Mossos D'Esquadra llamado "Fets\_penals\_coneguts\_\_fets\_coneguts\_resolts\_i\_detencions.csv". Este archivo consiste en 10 columnas y 209.645 filas en el cual se recogen todos los delitos conocidos en Barcelona según varias categorías, como el tipo de delito, el año y el distrito en el que suceden. En la siguiente figura se muestra el csv en cuestión leído en Jupyter Notebook:

Mes	Nom mes	Any	Regió Policial (RP)	Àrea Bàsica Policial (ABP)	Títol Codi Penal	Tipus de fet	Coneguts	Resolts	Detencions
0	1	gener	2011	RP Metropolitana Nord	ABP Arenys de Mar	De les falsedats De la usurpació de l'estat civil	2	NaN	NaN
1	1	gener	2011	RP Metropolitana Nord	ABP Arenys de Mar	De les falsedats Falsedats documentals	2	1.0	NaN
2	1	gener	2011	RP Metropolitana Nord	ABP Arenys de Mar	De les lesions Lesions	27	27.0	16.0
3	1	gener	2011	RP Metropolitana Nord	ABP Arenys de Mar	De les tortures i altres delictes contra la in... Tracte degradant / vexatori	6	6.0	2.0
4	1	gener	2011	RP Metropolitana Nord	ABP Arenys de Mar	Delictes contra el patrimoni i contra l'ordre ... Apropiació indeguda	4	4.0	4.0
...	...	...	...	...	...	...	...	...	...
209640	12	desembre	2021	RP Terres de l'Ebre	ABP Terra Alta - Ribera d'Ebre	Delictes contra la llibertat Amenaces	11	10.0	3.0
209641	12	desembre	2021	RP Terres de l'Ebre	ABP Terra Alta - Ribera d'Ebre	Delictes contra la llibertat Detenció il·legal	1	1.0	NaN
209642	12	desembre	2021	RP Terres de l'Ebre	ABP Terra Alta - Ribera d'Ebre	Delictes contra la llibertat i la indemnitat s... Agressions sexuals	1	1.0	NaN
209643	12	desembre	2021	RP Terres de l'Ebre	ABP Terra Alta - Ribera d'Ebre	Delictes contra les relacions familiars Contra els drets i deures familiars	2	2.0	NaN
209644	12	desembre	2021	RP Terres de l'Ebre	ABP Terra Alta - Ribera d'Ebre	Dels delictes contra la seguretat col·lectiva Contra la salut pública	2	2.0	1.0

209645 rows x 10 columns

Figura 5: Fichero csv "Fets\_penals\_coneguts\_\_fets\_coneguts\_resolts\_i\_detencions.csv"

## 4.3 Datos de limpieza

Los datos iniciales a partir de los cuales se analiza el estado de limpieza de Barcelona parten de un csv llamado "estat\_espais\_prioritaris\_2019.csv". Dicho archivo, extraído del portal de datos Open Data BCN, consta de 600 filas y 14 columnas. Estos datos, recogidos en 2019, fueron originados a partir de las observaciones realizadas de elementos individuales considerados basura en distintos espacios distintivos de los diferentes distritos de Barcelona. En las filas encontramos cuantas visitas se han realizado en cada espacio y la media de objetos diferentes considerados basura, así como otros campos como año, espacio y distrito. A continuación se muestra el csv en cuestión leído en Jupyter Notebook:

ID_Espai	Nom_Espai	Nom_Distri	C_Distri	Nom_Barrri	C_Barrri	Fase	Àrea	Mes	N_visites	Exc	B_dom	B_ind	Vol
0	1	Carrer de Josep Anselm Clavé, 2	Ciutat Vella	1	el Barri Gòtic	2	Fase 1   2	809.97	2	1	0.00	1.00	0.00
1	2	Carrer de la Maquinista, 40	Ciutat Vella	1	la Barceloneta	3	Fase 1	1015.35	2	1	0.00	0.00	0.00
2	3	Carrer de les Carretes, 47	Ciutat Vella	1	el Raval	1	Fase 1   2	208.51	2	1	0.00	1.00	0.00
3	4	Carrer de Sant Pere Mitjà, 88	Ciutat Vella	1	Sant Pere, Santa Caterina i la Ribera	4	Fase 1	267.80	2	1	0.00	1.00	0.00
4	5	Carrer de Sant Rafael, 38 (Pista Negra)	Ciutat Vella	1	el Raval	1	Fase 1   2	272.65	2	1	0.00	2.00	0.00
...	...	...	...	...	...	...	...	...	...	...	...	...	...
595	56	Carrer del Maresme	Sant Martí	10	el Besòs i el Maresme	70	Fase 1   2	5457.87	11	3	3.33	14.00	0.67
596	57	Parc de Sant Martí	Sant Martí	10	Sant Martí de Provençals	72	Fase 1   2	15052.52	11	3	2.67	1.67	0.00
597	59	Plaça Camí de la Verneda	Sant Martí	10	la Verneda i la Pau	73	Fase 1   2	1033.94	11	3	6.00	0.00	0.00
598	79	Jardins de Xavier Berenguel	Sant Martí	10	el Poblenou	68	Fase 2	2871.78	11	3	3.00	3.33	0.00
599	80	Gran Via (entre Bac de Roda i Fluvià)	Sant Martí	10	Provençals del Poblenou	71	Fase 2	5187.04	11	3	2.33	2.00	1.00

600 rows x 14 columns

Figura 6: Fichero csv "estat\_espais\_prioritaris\_2019.csv"

## 4.4 Datos de vivienda

Los datos iniciales a partir de los que se realiza el análisis de vivienda son extraídos de dos archivos csv, obtenidos del portal Open Data BCN.

El primero, llamado “2022\_comp\_vend\_preu\_trim.csv” consta de 438 filas y 8 columnas. Contiene la información del precio de las operaciones de compra-venta por distrito y barrio. Ofrece tanto información del precio total como por euros/m2. Se muestra en la siguiente figura:

	Any	Trimestre	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Preu_mitja_habitatge	Valor
0	2022	1	1	Ciutat Vella	1	el Raval	Total. Milers d'euros	226.1
1	2022	1	1	Ciutat Vella	2	el Barri Gòtic	Total. Milers d'euros	453.0
2	2022	1	1	Ciutat Vella	3	la Barceloneta	Total. Milers d'euros	215.4
3	2022	1	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera	Total. Milers d'euros	353.3
4	2022	1	2	Eixample	5	el Fort Pienc	Total. Milers d'euros	340.0
...	...	...	...	...	...	...	...	...
433	2022	1	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou	Usat. Euros/m2 construït	6151.6
434	2022	1	10	Sant Martí	70	el Besòs i el Maresme	Usat. Euros/m2 construït	2593.9
435	2022	1	10	Sant Martí	71	Provençals del Poblenou	Usat. Euros/m2 construït	3806.1
436	2022	1	10	Sant Martí	72	Sant Martí de Provençals	Usat. Euros/m2 construït	2870.6
437	2022	1	10	Sant Martí	73	la Verneda i la Pau	Usat. Euros/m2 construït	2211.0

438 rows x 8 columns

Figura 7: Fichero csv “2022\_comp\_vend\_preu\_trim.csv”

El segundo archivo, llamado “2019\_atles\_renda\_bruta\_persona.csv” muestra la renta bruta media por sección censal de Barcelona, según barrio y distrito. Consta de 7 columnas y 1068 filas. Se muestra a continuación:

	Any	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Seccio_Censal	Import_Renda_Bruta_€
0	2019	1	Ciutat Vella	1	el Raval	1	13022
1	2019	1	Ciutat Vella	1	el Raval	2	10797
2	2019	1	Ciutat Vella	1	el Raval	3	10281
3	2019	1	Ciutat Vella	1	el Raval	4	12875
4	2019	1	Ciutat Vella	1	el Raval	5	10810
...	...	...	...	...	...	...	...
1063	2019	10	Sant Martí	73	la Verneda i la Pau	143	14130
1064	2019	10	Sant Martí	65	el Clot	234	17714
1065	2019	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou	235	15756
1066	2019	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou	236	16748
1067	2019	10	Sant Martí	71	Provençals del Poblenou	237	14993

1068 rows x 7 columns

Figura 8: Fichero csv “2019\_atles\_renda\_bruta\_persona.csv”

## 5. Metodología y resultados

### 5.1 Encuesta de opinión

En este punto se realiza un estudio exploratorio de los datos extraídos de la encuesta de opinión de Barcelona. Se parte del dataframe obtenido en el punto 4.1 y se analiza en orden cada variable disponible, es decir, los datos presentes en cada columna del dataframe.

#### Variable NOM\_DISTRICTE

En esta columna del dataframe encontramos el nombre del distrito donde reside la persona encuestada. A continuación se muestra una cuenta de cada valor presente dentro de la variable y un grafico de barras con dicha información, ambos realizados con Python y visualizados en Jupyter Notebook.

```
NOM_DISTRICTE
EIXAMPLE          924
SANT MARTÍ       770
SANTS - MONTJUÏC 616
NOU BARRIS       616
CIUTAT VELLA     616
HORTA - GUINARDÓ 616
SANT ANDREU      481
GRÀCIA           462
SARRIÀ - SANT GERVASI 462
LES CORTS        437
Name: count, dtype: int64
```

Figura 9: Cuenta de NOM\_DISTRICTE

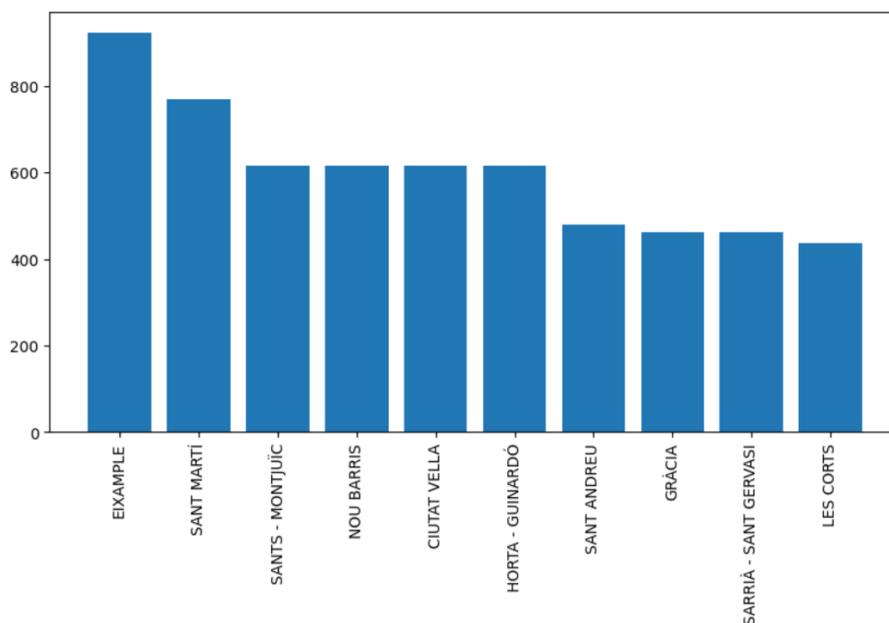


Figura 10: Gráfico de barras de NOM\_DISTRICTE

Se puede observar que no existe la misma cantidad de personas encuestadas de cada distrito. Sin embargo, la cantidad de personas de cada distrito se corresponde mucho con la proporción de personas que vive en cada uno de ellos.

### Variable SEXE

En esta variable se indica el sexo de la persona encuestada. Las siguientes figuras muestran la cuenta de los valores diferentes dentro de la columna y un gráfico de barras comparativo:

```
SEXE
DONA    3187
HOME    2813
Name: count, dtype: int64
```

Figura 11: Cuenta de SEXE

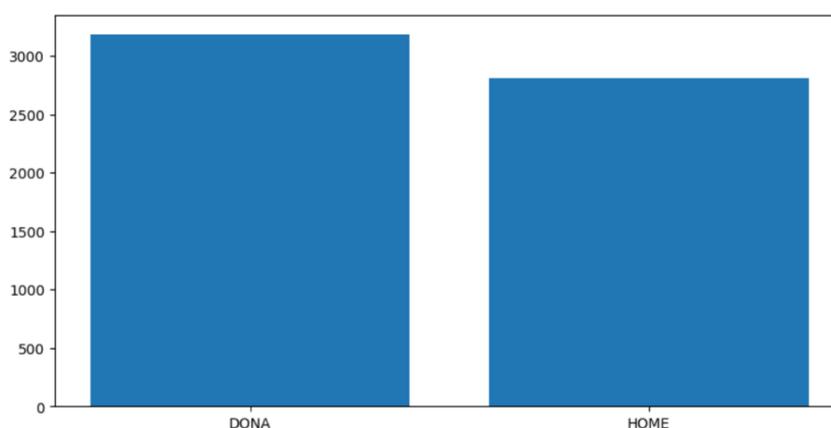


Figura 12: Gráfico de barras de SEXE

Se distingue una pequeña diferencia entre las cantidades, pues hay más encuestadas mujeres que encuestados hombres. Esto también está en proporción de la cantidad de mujeres y hombres que habitan en Barcelona.

### Variable EDAT1899\_1A6

Esta variable indica la edad de la persona encuestada. Las opciones de respuestas eran un gradiente de 5 intervalos de edad, partiendo con el de 18 a 24 años, siguiendo con el de 25 a 34, 35 a 44, 45 a 54, 55 a 64 y terminando con 65 años o más.

A continuación se muestra la cuenta de los diferentes valores dentro de la variable y un gráfico de barras ordenado según el gradiente de edad:

```

EDAT1899_1A6
DE 65 ANYS I MÉS      1473
DE 35 A 44 ANYS      1084
DE 25 A 34 ANYS      1065
DE 45 A 54 ANYS      1026
DE 55 A 64 ANYS       871
DE 18 A 24 ANYS       481
Name: count, dtype: int64

```

Figura 13: Cuenta de EDAT1899\_1A6

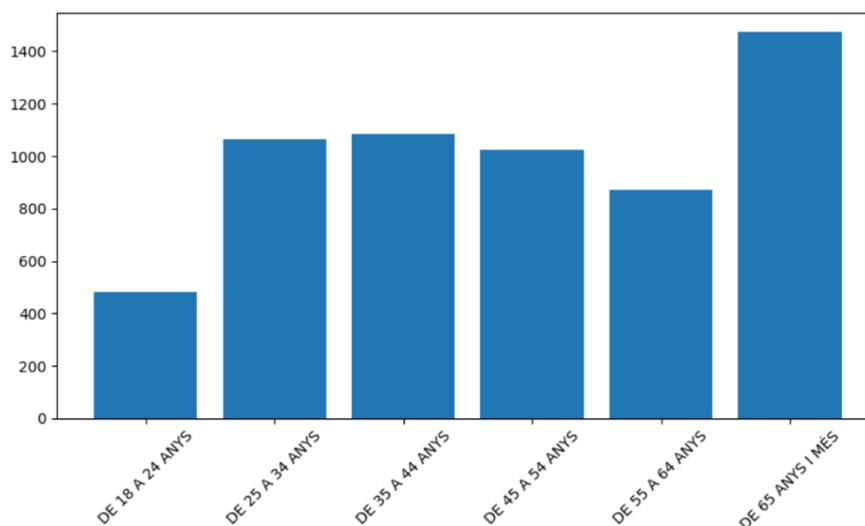


Figura 14: Gráfico de barras de EDAT1899\_1A6

Se observa que en el intervalo de entre 18 y 24 años hay menos encuestados que en el resto, y que en el de mayores de 65 años hay mas. Sin embargo estos son también el intervalo mas escueto en años y el mas extenso en años.

### Variable SITLABORAL\_1A9

Esta variable muestra la situación laboral de las personas encuestadas. Se observan las diferentes opciones:

```

SITLABORAL_1A9
TREBALLA PER COMPTE D'ALTRI (COM ASSALARIAT/ADA)      2666
COBRA JUBILACIÓ, PREJUBILACIÓ O PENSIÓ / INCAPACITACIÓ LABORAL  1676
TREBALLA PER COMPTE PROPI (COM AUTÒNOM/A O AMB CONTRACTACIÓ DE PERSONAL)  559
ESTUDIA                                                401
ESTÀ A L'ATUR                                          380
TASQUES DE LA LLAR (NO REMUNERADES)                  275
TREBALLA PER COMPTE D'ALTRI ACTUALMENT AFECTAT/DA PER UN ERO AMB REDUCCIÓ O ERTO  21
TREBALLA PER COMPTE PROPI PERÒ ACTUALMENT AMB ATURADA D'ACTIVITAT  18
RENDISTA                                              4
Name: count, dtype: int64

```

Figura 15: Cuenta de SITLABORAL\_1A9

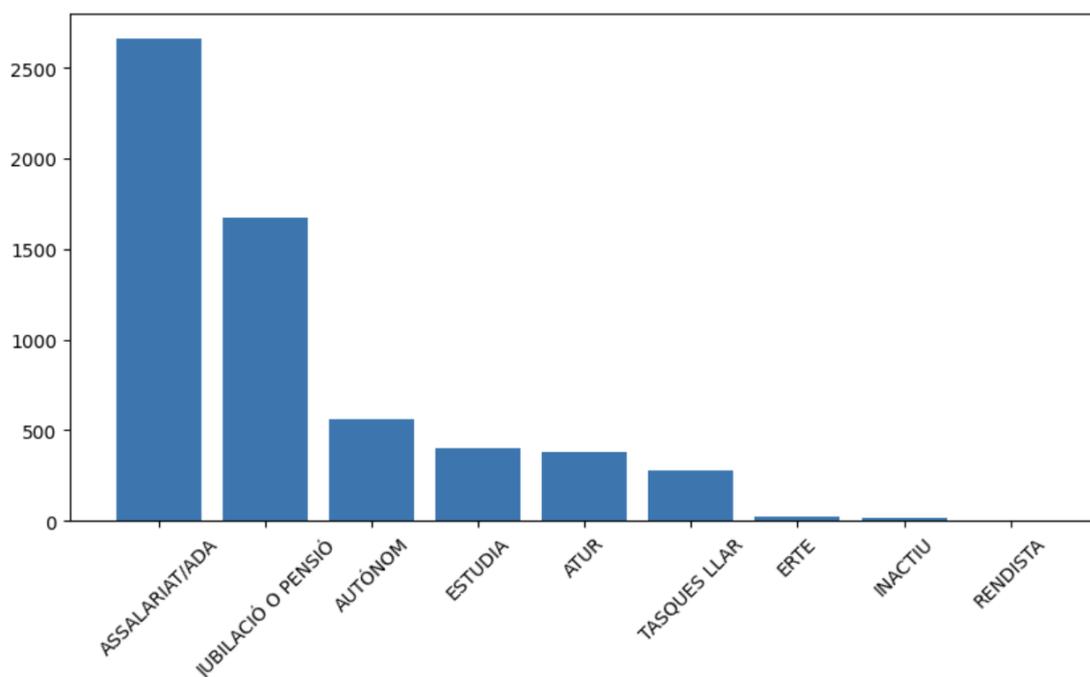


Figura 16: Gráfico de barras de SITLABORAL\_1A9

Se observa una gran cantidad de asalariados, seguido de jubilados/pensionistas.

### Variable SATISF\_RES\_BCN\_1A4

Esta variable indica la satisfacción del encuestado de vivir en Barcelona, puntuado con un valor de 0 a 10 y agrupado mediante un gradiente de 4 opciones, siendo de 0 a 4 la mas baja, 5 a 6 la siguiente, 7 a 8, y 9 a 10. También existe la opción NS/NC. A continuación se muestran los valores, la cuenta y un gráfico de barras:

```

SATISF_RES_BCN_1A4
BASTANT SATISFET/A (7 A 8)      2825
MOLT SATISFET/A (9 A 10)      1581
SATISFET/A (5 A 6)            1124
POC/GENS SATISFET/A (0 A 4)   458
NO HO SAP / NO CONTESTA       12
Name: count, dtype: int64
    
```

Figura 17: Cuenta de SATISF\_RES\_BCN\_1A4

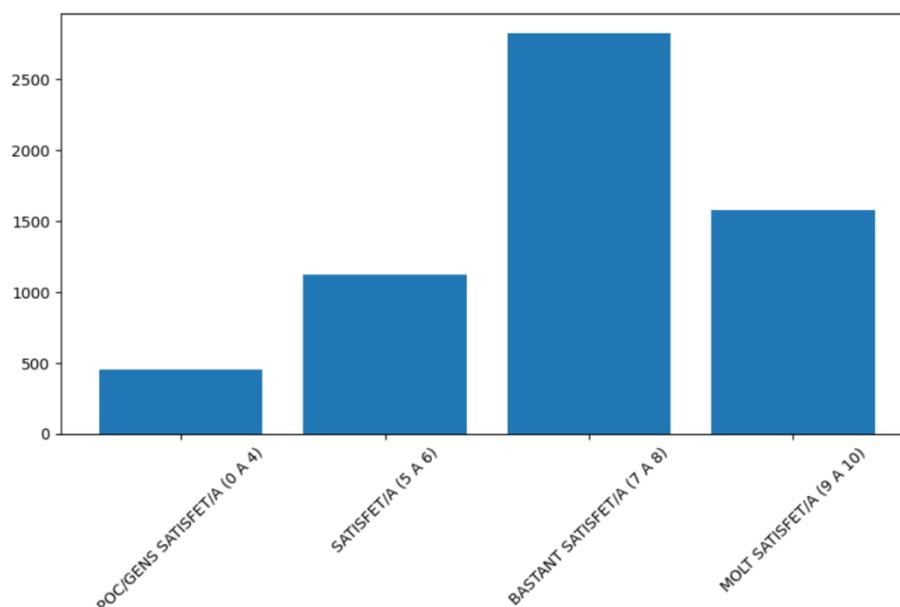


Figura 18: Gráfico de barras de SATSF\_RES\_BCN1A4

Se observa que en general los habitantes encuestados están bastante satisfechos de vivir en Barcelona. Más de la mitad de la población estudiada puntúa su grado de satisfacción con un 7 o superior.

### Variable PROB\_BCN\_AGR

Esta variable corresponde al problema principal de vivir en Barcelona para cada persona encuestada. Existen casi 50 opciones pero solo trabajaremos con las más frecuentes.

A continuación se muestra la cuenta de los valores y un gráfico de barras representativo de las 7 opciones más votadas:

PROB_BCN_AGR	
INSEGURETAT	1166
NETEJA	730
ACCÉS A L'HABITATGE	398
CONTAMINACIÓ I MEDI AMBIENT	335
CONGESTIÓ DE TRÀNSIT	315
NO HO SAP	300
PROBLEMES ECONÒMICS	254
TRANSPORTS, INFRAESTRUCTURES I COMUNICACIONS	213
ATUR / CONDICIONS DE TREBALL	202
VALORS SOCIALS NEGATIUS / MANCA D'EDUCACIÓ, CIVISME	184
GESTIÓ POLÍTICA MUNICIPAL	168
GESTIÓ I ORGANITZACIÓ DEL TRÀNSIT	168
TURISME	158
EXCLUSIÓ SOCIAL (POBRESA, DESIGUALTAT)	126
SOROLLS	109
PROBLEMES ASSOCIATS A LA IMMIGRACIÓ	105

Figura 19: Cuenta de PROB\_BCN\_AGR

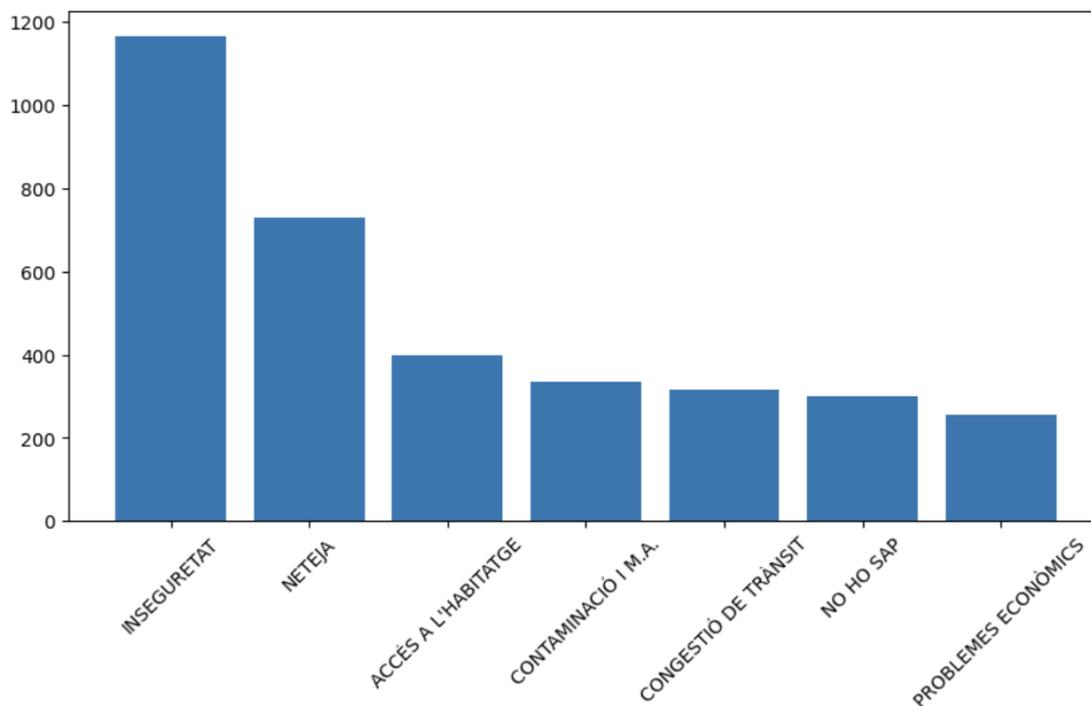


Figura 20: Gráfico de barras de PROB\_BCN\_AGR

Se observa que, con mucha diferencia, la inseguridad es el problema más votado en la encuesta de satisfacción el año 2022. Sigue la limpieza y, en menor medida, el acceso al habitaje.

### Variable US30D\_XARXES\_Q01

Esta variable responde a la red social más utilizada por los habitantes encuestados. A continuación se muestra la cuenta de los valores y el gráfico de barras:

US30D_XARXES_Q01	Count
WHATSAPP	2866
INSTAGRAM	899
FACEBOOK	761
NO N'HA PARTICIPAT / USAT CAP	394
LINKEDIN	367
TWITTER	365
TELEGRAM	276
SIGNAL	49
NO CONTESTA	16
YOUTUBE	4
NO HO SAP	2
PINTEREST	1

Name: count, dtype: int64

Figura 21: Cuenta de US30D\_XARXES\_Q01

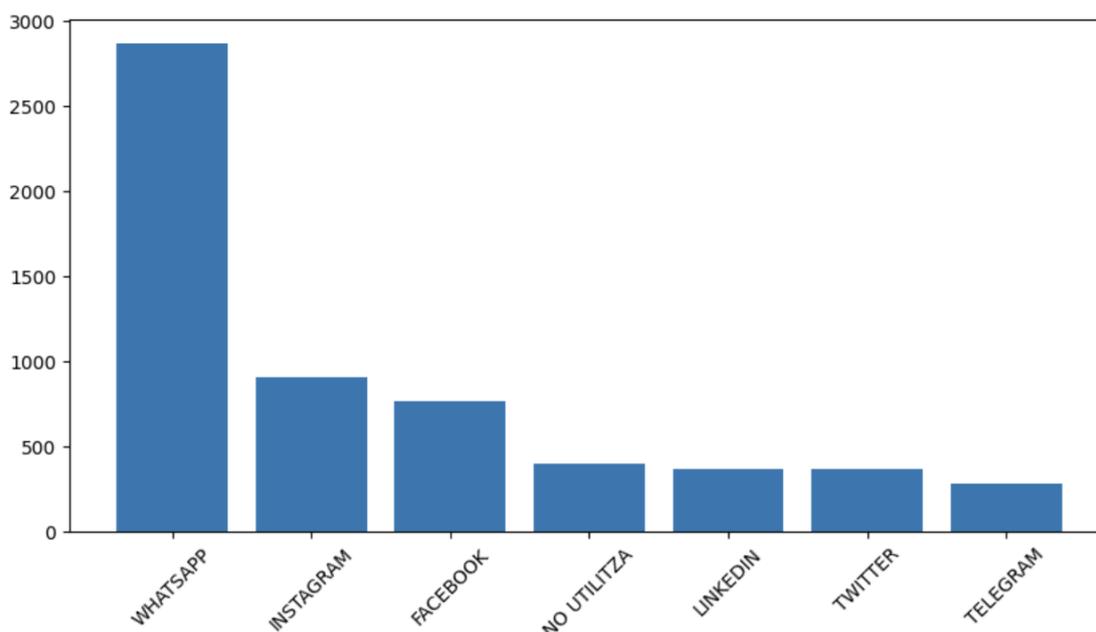


Figura 22: Gráfico de barras de US30D\_XARXES\_Q01

La más utilizada es con mucha diferencia Whatsapp. Después, en menor medida, Instagram y Facebook.

### Variable IDEO\_1A8

Esta variable se corresponde a la ideología política de cada encuestado. Las opciones se corresponden a un gradiente de ideologías desde extrema izquierda a extrema derecha. También se incluyen las opciones Apolítico, No contesta y No sabe/No le interesa la política.

A continuación se muestra la cuenta de los diferentes valores y el gráfico de barras:

IDEO_1A8	
ESQUERRA	2114
NO HO SAP / NO LI INTERESSA LA POLÍTICA	956
CENTRE ESQUERRA	905
CENTRE	859
CENTRE DRETA	304
NO CONTESTA	271
EXTREMA ESQUERRA	199
DRETA	190
APOLÍTIC	180
EXTREMA DRETA	22
Name: count, dtype: int64	

Figura 23: Cuenta de IDEO\_1A8

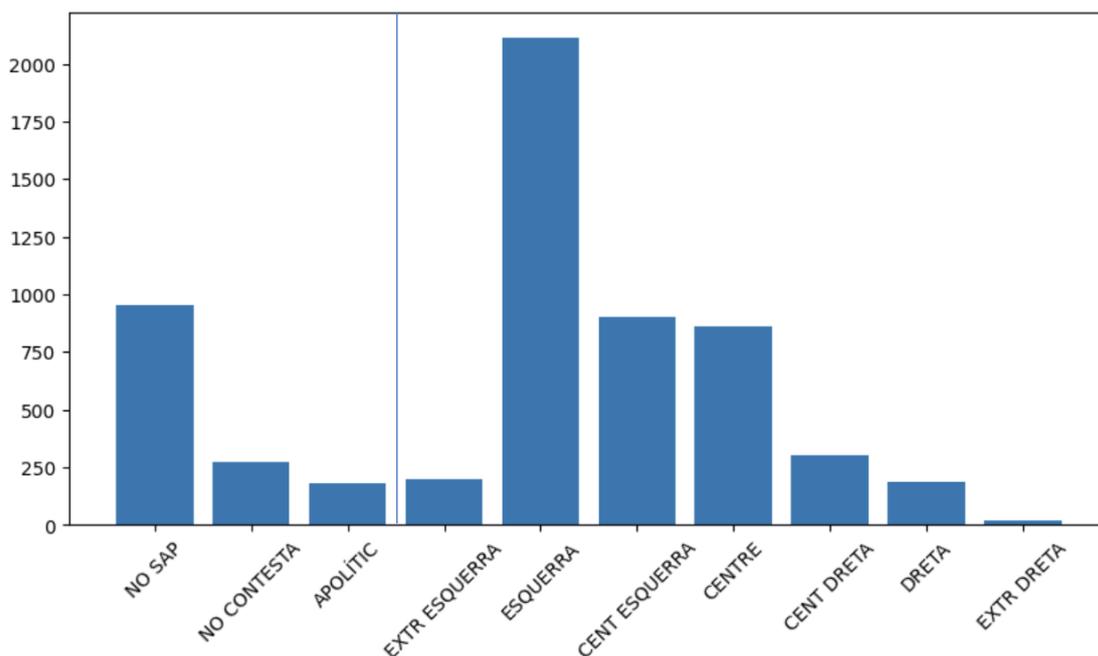


Figura 24: Gráfico de barras de IDEO\_1A8

Se puede observar que, con mucha diferencia, el valor más frecuente es Izquierda. Con menor medida, es seguido por No sabe, Centro Izquierda y Centro.

### Variable ESTUDIS\_1A6

Esta variable indica el grado de estudios que posee cada persona encuestada. A continuación se muestran los diferentes valores presentes en la variable, la cuenta y un gráfico de barras representativo:

ESTUDIS_1A6	
UNIVERSITARIS	1589
POST UNIVERSITARIS	1193
SECUNDARIS GENERALIS	1183
SECUNDARIS PROFESSIONALS	954
OBLIGATORIS	929
NO HA ACABAT ELS ESTUDIS OBLIGATORIS	138
NO CONTESTA	14
Name: count, dtype: int64	

Figura 25: Gráfico de barras de ESTUDIS\_1A6

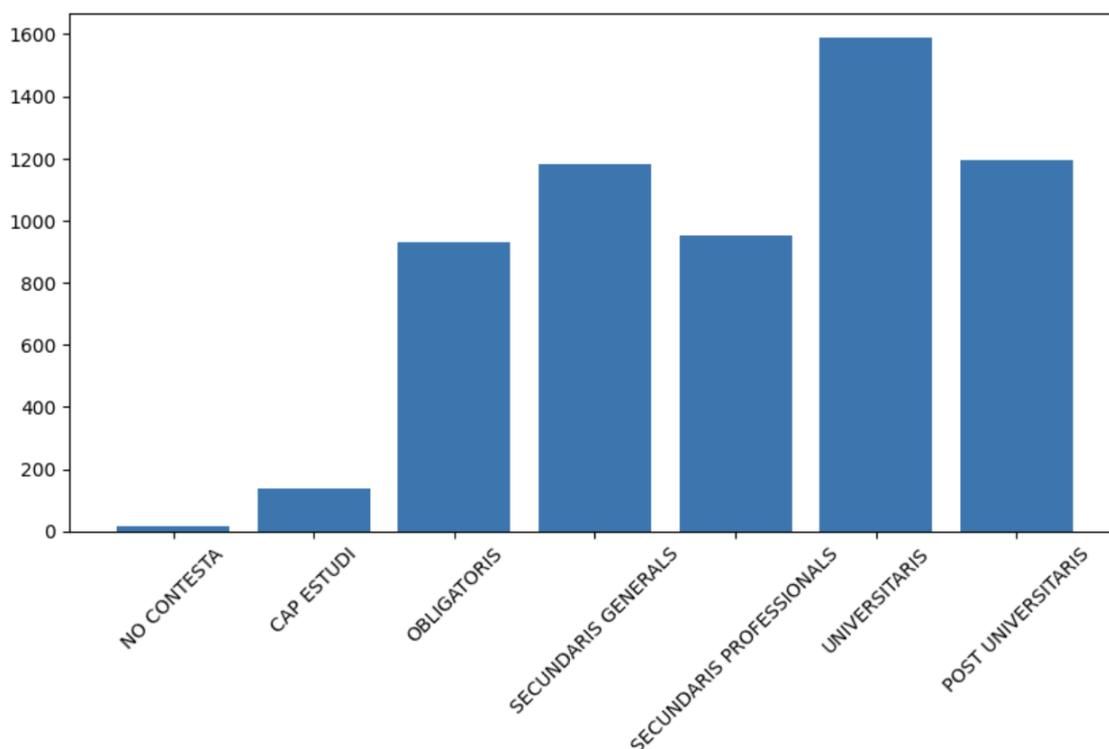


Figura 26: Gráfico de barras de ESTUDIS\_1A6

Se observa que el valor mas frecuente es el de Estudios Universitarios.

### Variable LLAR\_INGRESSOS\_1A10

Esta variable indica la renta mensual de la residencia donde el encuestado vive. Los diferentes valores se corresponden a un gradiente de intervalos de rentas. A continuación se muestra la cuenta de los valores y el gráfico de barras:

```

LLAR_INGRESSOS_1A10
DE 3.001 A 5.000 EUROS    955
DE 1.001 A 1.500 EUROS  774
DE 1.501 A 2.000 EUROS  766
DE 2.501 A 3.000 EUROS  700
DE 500 A 1.000 EUROS    643
DE 2.001 A 2.500 EUROS  637
NO CONTESTA              570
NO HO SAP                 329
DE 5.001 A 7.000 EUROS  304
MENYS DE 500 EUROS      175
DE 7.001 A 9.000 EUROS  75
MÉS DE 9.000 EUROS     72
Name: count, dtype: int64

```

Figura 27: Cuenta de LLAR\_INGRESSOS\_1A10

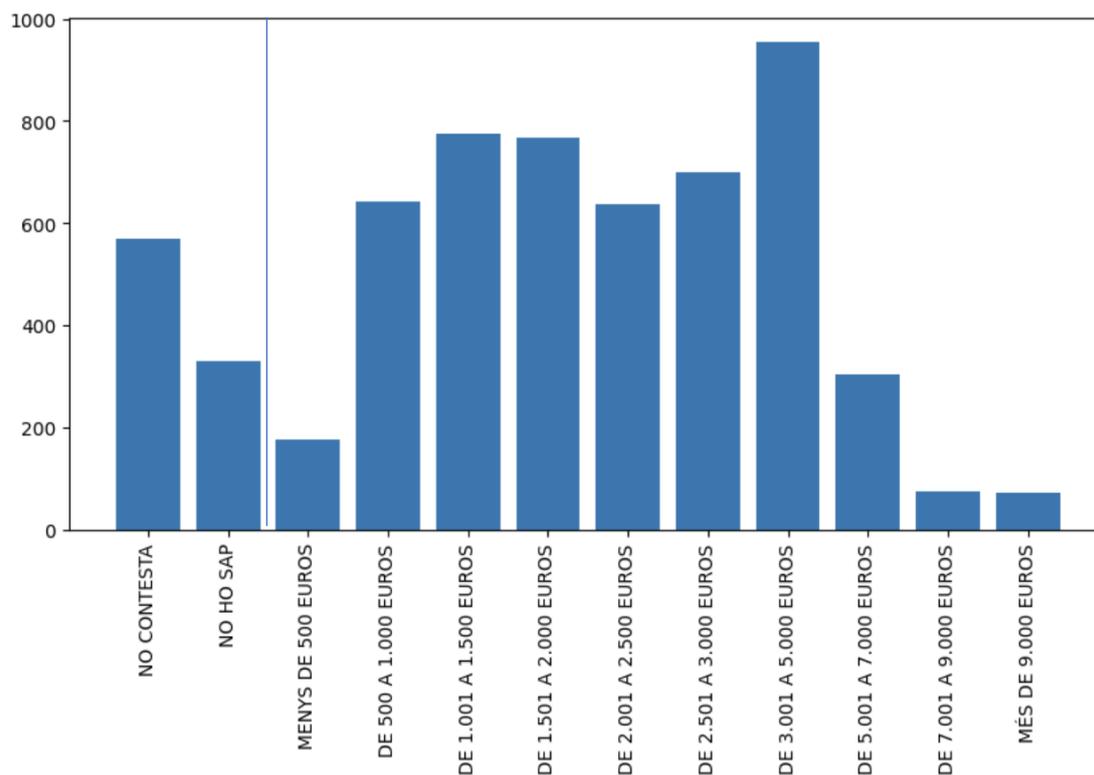


Figura 28: Gráfico de barras de LLAR\_INGRESSOS\_1A10

Se observa que el valor mas frecuente es el correspondiente al intervalo que comprende desde los 3.000 euros a los 5.000 euros.

La encuesta de opinión de Barcelona ofrece unos resultados que parecen caracterizar bien a la población residente. Se observa que el problema más reportado como mas importante es el de inseguridad, seguido de limpieza y de acceso a la vivienda. Se observa también que más de la mitad de los encuestados puntúan su satisfacción de vivir en Barcelona con un 7 o mas.

## 5.2 Estudio de inseguridad

En este estudio se procede a analizar mediante métodos exploratorios univariantes el problema principal de vivir en Barcelona según la encuesta de opinión.

Para llevar a cabo este análisis se utilizan tres indicadores, que son los robos cometidos en cada distrito de Barcelona por cada 10.000 habitantes, las lesiones cometidas en cada distrito por cada 10.000 habitantes y los delitos cometidos en cada distrito por cada 10.000 habitantes.

Primeramente se obtiene el porcentaje de personas de cada distrito de Barcelona que considera que la inseguridad es el principal problema de vivir en Barcelona a partir del archivo detallado en el punto 4.1.

En las dos siguientes figuras se muestra la información obtenida:

```
NOM_DISTRICTE
CIUTAT VELLA           26.461039
HORTA - GUINARDÓ      22.077922
NOU BARRIS            21.103896
SANT MARTÍ            20.389610
SANTS - MONTJUÏC     19.155844
EIXAMPLE              17.748918
LES CORTS             17.620137
SARRIÀ - SANT GERVASI 17.099567
SANT ANDREU          16.839917
GRÀCIA                13.203463
Name: PROB_BCN_AGR, dtype: float64
```

Figura 29: Porcentaje primer problema inseguridad

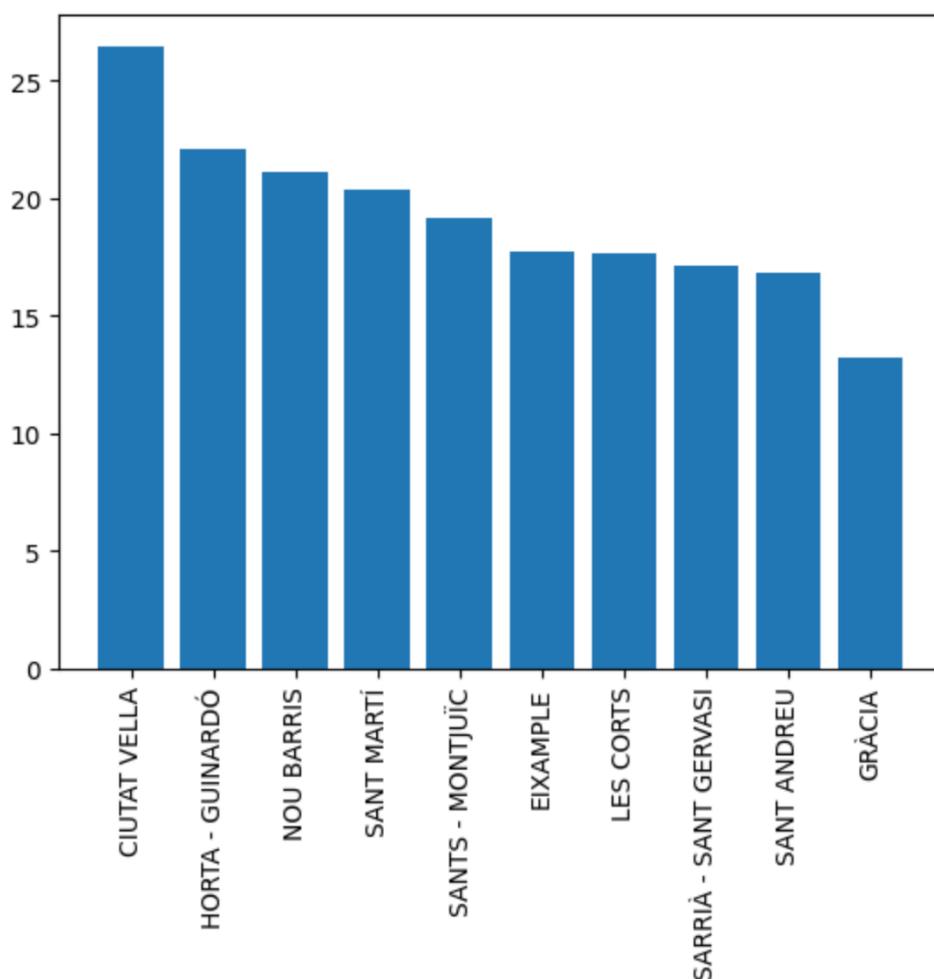


Figura 30: Gráfico de barras porcentaje primer problema inseguridad

Se comienza presentando el indicador de robos. A partir del csv detallado en el punto 4.2 del trabajo se filtran los delitos conocidos en Barcelona por los correspondientes a todos los tipos de robos. Después se agrupan según el distrito en el que han sido cometidos. Se normalizan según la población residente en cada distrito de Barcelona y se calcula el índice de robos por cada 10.000 habitantes en cada distrito. A continuación se muestra la cuenta y un gráfico de barras con dicha información, ordenado según el porcentaje de personas de cada distrito que considera la inseguridad el problemas mas importante:

Àrea Bàsica Policial (ABP)	Robatoris
ABP Ciutat Vella	147.553191
ABP Horta-Guinardó	49.440233
ABP Nou Barris	51.048623
ABP Sant Martí	75.979857
ABP Sants-Montjuïc	81.395980
ABP Eixample	63.926288
ABP Les Corts	76.312576
ABP Sarrià-Sant Gervasi	58.290885
ABP Sant Andreu	62.489987
ABP Gràcia	51.740558

Figura 31: Índice de robos

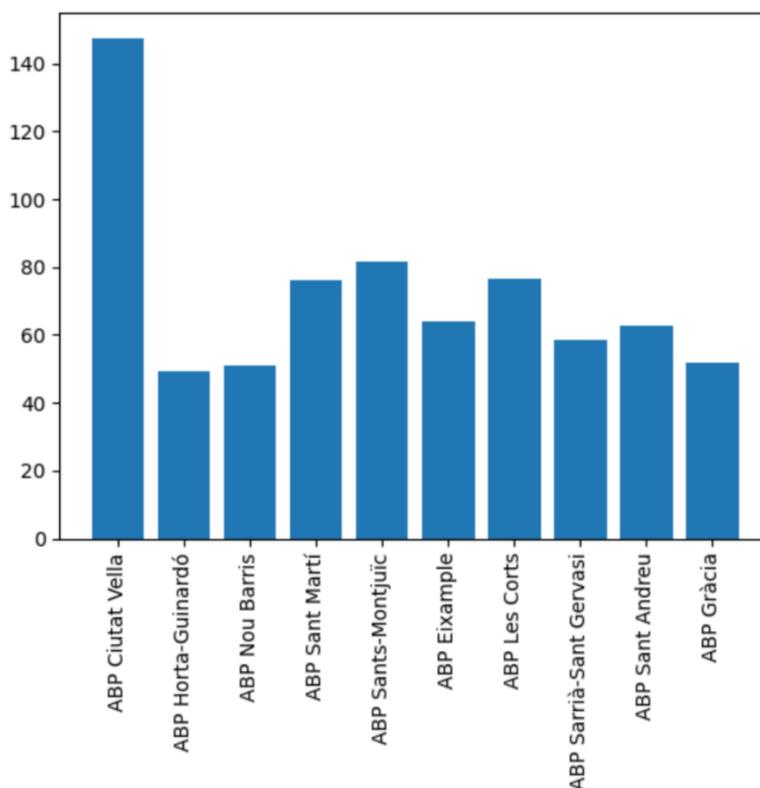


Figura 32: Gráfico de barras índice de robos

El segundo indicador que se calcula es el de lesiones cometidas. De la misma forma que el indicador de robos se obtiene del csv de delitos conocidos en Barcelona. También se normaliza a partir de la población residente en Barcelona y se calcula el índice de lesiones por cada 10.000 habitantes. Se muestra en las siguientes figuras, ordenado según el porcentaje de personas de cada distrito que considera la inseguridad el problemas mas importante:

Àrea Bàsica Policial (ABP)	Lesions
ABP Ciutat Vella	8.520309
ABP Horta-Guinardó	2.250729
ABP Nou Barris	3.063855
ABP Sant Martí	3.096937
ABP Sants-Montjuïc	4.193373
ABP Eixample	3.350884
ABP Les Corts	1.843712
ABP Sarrià-Sant Gervasi	1.548257
ABP Sant Andreu	2.743658
ABP Gràcia	1.732348

Figura 33: Índice de lesiones

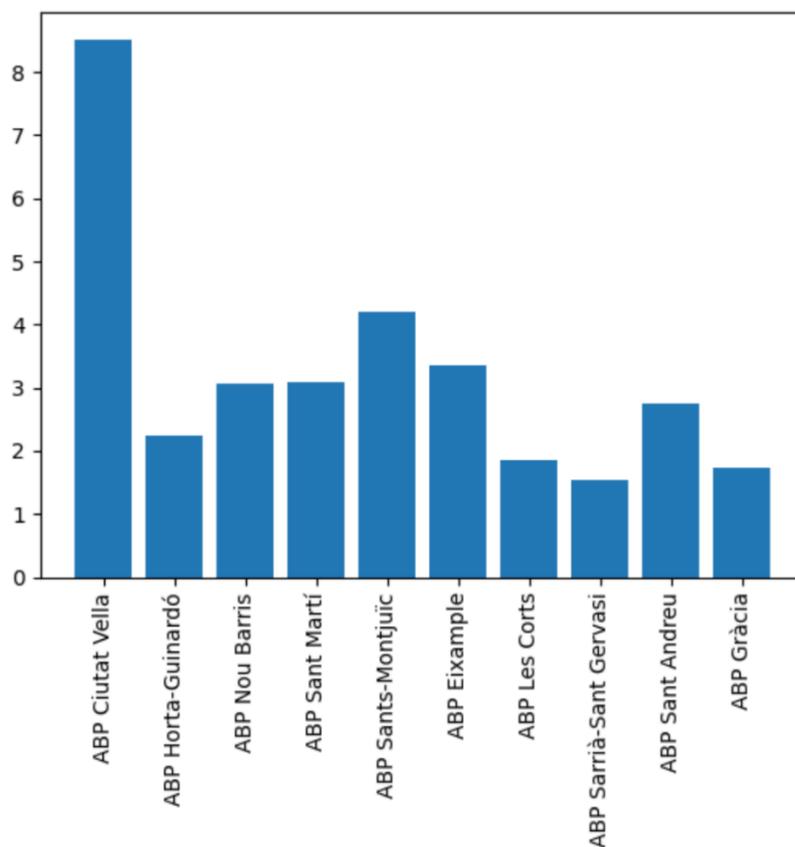


Figura 34: Gráfico de barras de índice de lesiones

El tercer indicador, que corresponde a todos los delitos conocidos, se extrae del mismo archivo csv. En este caso no se filtran las filas según ningún tipo de delito. También se normaliza según los residentes de cada distrito de Barcelona obteniendo el índice de delitos por cada 10.000 habitantes. A continuación se muestra la cuenta y un gráfico de barras, ordenado según el porcentaje de personas de cada distrito que considera la inseguridad el problemas mas importante:

Àrea Bàsica Policial (ABP)	Delictes
ABP Ciutat Vella	156.073501
ABP Horta-Guinardó	51.690962
ABP Nou Barris	54.112478
ABP Sant Martí	79.076794
ABP Sants-Montjuïc	85.589354
ABP Eixample	67.277172
ABP Les Corts	78.156288
ABP Sarrià-Sant Gervasi	59.839142
ABP Sant Andreu	65.233645
ABP Gràcia	53.472906

Figura 35: Índice de delitos

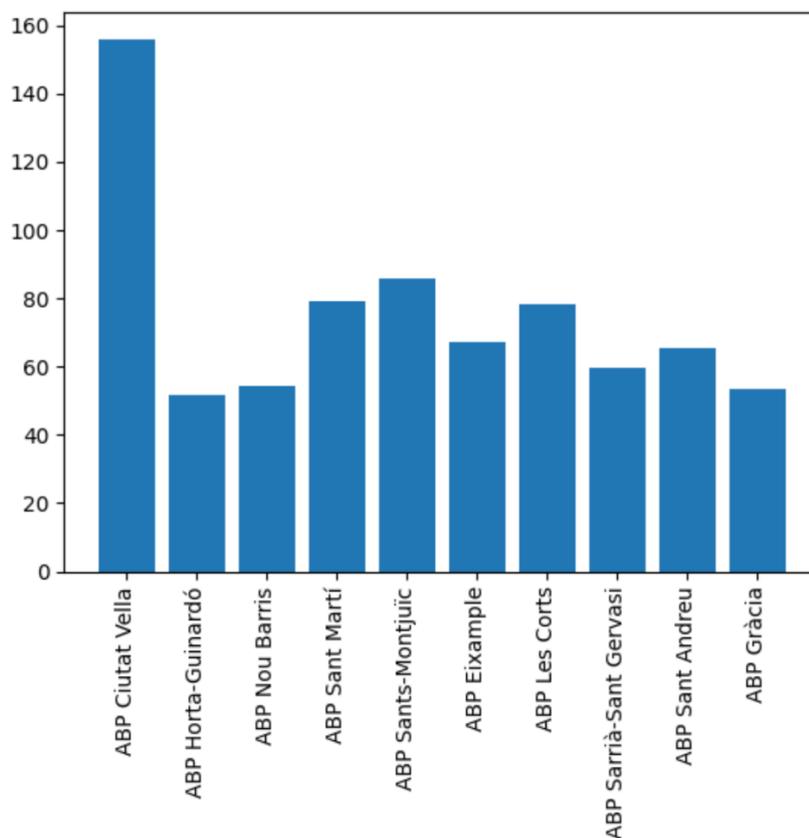


Figura 36: Gráfica de barras del índice de delitos

Se observa que los tres indicadores son muy parecidos entre sí, añadiendo fiabilidad al uso de cualquiera de los tres indicadores.

En los tres indicadores se observa que el distrito donde mas delitos se cometen es Ciutat Vella con muchas diferencia. Con mucha menos diferencia respecto al resto de distritos, le siguen Sants-Montjuïc, Sant Martí, Les Corts i L'Eixample.

Para que los indicadores se correspondan con la proporción de personas de cada distrito que considera que la inseguridad es el problemas mas importante de vivir en Barcelona, se debería ver un orden decreciente en los valores del eje y en cada uno de los tres indicadores. Se observa que no ocurre así. No se obtiene ninguna relación entre el orden de los distritos donde mas se considera la inseguridad el mayor problema y la cantidad de delitos cometidos en estos, a excepción de Ciutat Vella.

Respecto a Ciutat Vella se observa que es el distrito donde más encuestados han declarado que su principal preocupación es la inseguridad y claramente es el distrito con más delitos cometidos con mucha diferencia del segundo en cualquiera de los tres indicadores. Esto añade fiabilidad al sistema utilizado.

Se concluye que, a excepción de Ciutat Vella, la cantidad de encuestados de cada distrito que declara como principal problema la inseguridad no es proporcional a los tres indicadores de delitos de cada distrito, es decir, que la percepción ciudadana de seguridad no se corresponde con los indicadores estudiados. Sin embargo, sí se puede confirmar que ocurre un problema de inseguridad en Ciutat Vella.

### **5.3 Estudio de limpieza**

En este estudio se procede a analizar el segundo problema principal de vivir en Barcelona según la encuesta de opinión mediante técnicas exploratorias univariantes.

Para llevar a cabo este análisis se utiliza un indicador de basura encontrada, obtenido a partir del archivo descrito en el apartado 4.3.

Primeramente se obtiene el porcentaje de personas de cada distrito de Barcelona que considera que la limpieza es el principal problema de vivir en Barcelona.

En las dos siguientes figuras se muestra la información obtenida:

NOM_DISTRICTE	
SANT ANDREU	13.929314
LES CORTS	13.729977
EIXAMPLE	13.636364
SANT MARTÍ	13.376623
SARRIÀ - SANT GERVASI	12.987013
SANTS - MONTJUÏC	12.824675
GRÀCIA	12.337662
HORTA - GUINARDÓ	10.714286
CIUTAT VELLA	10.551948
NOU BARRIS	7.629870

Name: PROB\_BCN\_AGR, dtype: float64

Figura 37: Porcentaje primer problema limpieza

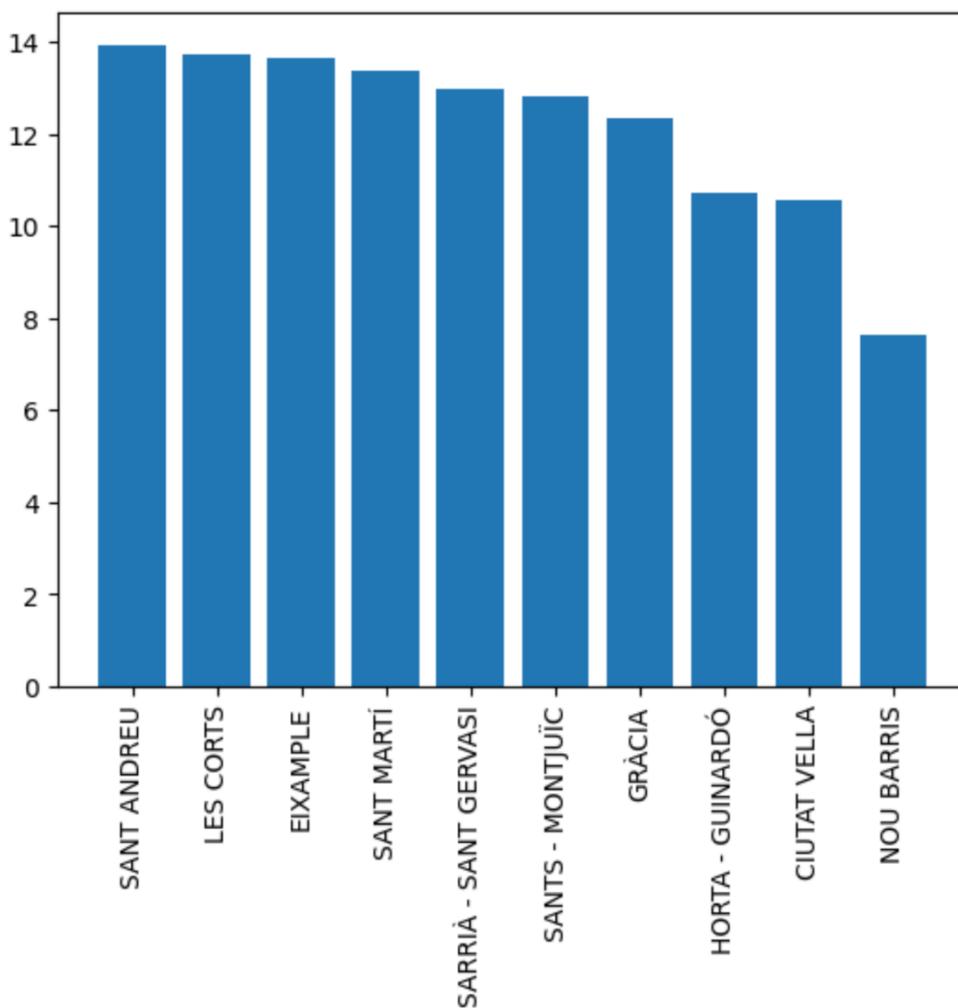


Figura 38: Gráfico de barras porcentaje primer problema limpieza

A continuación se calcula el indicador de basura. Se suman las columnas de tipos de basura, ya que se considera que no importa el tipo de objeto en el suelo público(excrementos, bolsas, otros), si no el hecho de que esté presente. Se suma la media de basura encontrada en cada calle. Se agrupa la basura encontrada en las calles en cada distrito.

En las dos siguientes figuras se muestra la cantidad de basura encontrada en cada distrito de Barcelona mediante una cuenta y un gráfico de barras, ordenados según el porcentaje de personas de cada distrito que considera la limpieza el problemas mas importante:

Basura	
Nom_Distri	
Sant Andreu	277.47
Les Corts	344.31
Eixample	379.40
Sant Martí	481.09
Sarrià-Sant Gervasi	335.40
Sants-Montjuïc	313.72
Gràcia	254.32
Horta-Guinardó	307.43
Ciutat Vella	293.36
Nou Barris	408.12

Figura 39: Cantidad de basura

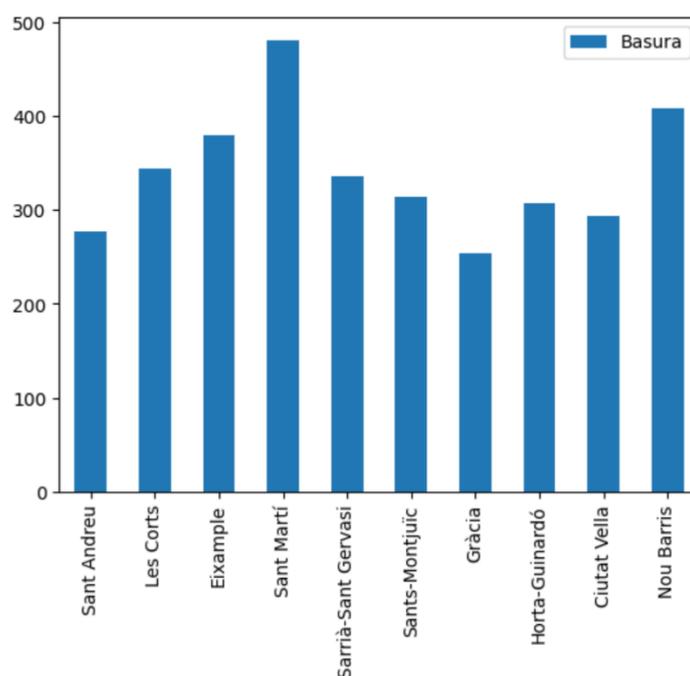


Figura 40: Gráfico de barras de cantidad de basura

Se observa que el distrito donde mas basura se ha encontrado es Sant Martí, seguido de Nou Barris y l'Eixample con menos diferencia ya del resto de distritos.

Para que el indicador de basura se corresponda con la proporción de personas de cada distrito que considera que la inseguridad es el problema mas importante de vivir en Barcelona, se debería ver un orden decreciente en los valores del eje y en cada uno de los tres indicadores. Se observa que no ocurre así. No se obtiene ninguna relación entre el orden de los distritos donde mas se considera la limpieza el mayor problema y la cantidad de basura encontrada en estos. Se observa que el distrito donde más porcentaje de encuestados ha declarado la limpieza como mayor preocupación es uno de los distritos donde menos basura se ha encontrado.

Se concluye que la cantidad de encuestados de cada distrito que declara como principal problema la limpieza no es proporcional a la basura encontrada en cada distrito. La percepción de los ciudadanos respecto a la limpieza no se corresponde con los indicadores estudiados. No se puede afirmar que exista un problema de limpieza.

## 5.4 Estudio de vivienda

En este estudio se procede a analizar el tercer problema principal de vivir en Barcelona según la encuesta de opinión.

Para llevar a cabo este análisis se utilizan dos indicadores obtenidos a partir de los archivos descritos en el apartado 4.4 del presente trabajo. Estos indicadores son el precio de las operaciones de compraventa realizadas en los diferentes distritos de Barcelona en euros/m<sup>2</sup> y la renta por capita de cada distrito de Barcelona.

Primeramente se obtiene el porcentaje de personas de cada distrito de barcelona que considera que el acceso al habitaje es el principal problema de vivir en Barcelona.

En las dos siguientes figuras se muestra la información obtenida:

NOM_DISTRICTE	
GRÀCIA	10.822511
SANTS – MONTJUÏC	8.279221
HORTA – GUINARDÓ	7.954545
SANT ANDREU	7.276507
CIUTAT VELLA	6.818182
EIXAMPLE	6.601732
SANT MARTÍ	5.714286
NOU BARRIS	5.032468
SARRIÀ – SANT GERVASI	3.896104
LES CORTS	3.890160
Name: PR0B_BCN_AGR, dtype: float64	

Figura 41: Porcentaje primer problema vivienda

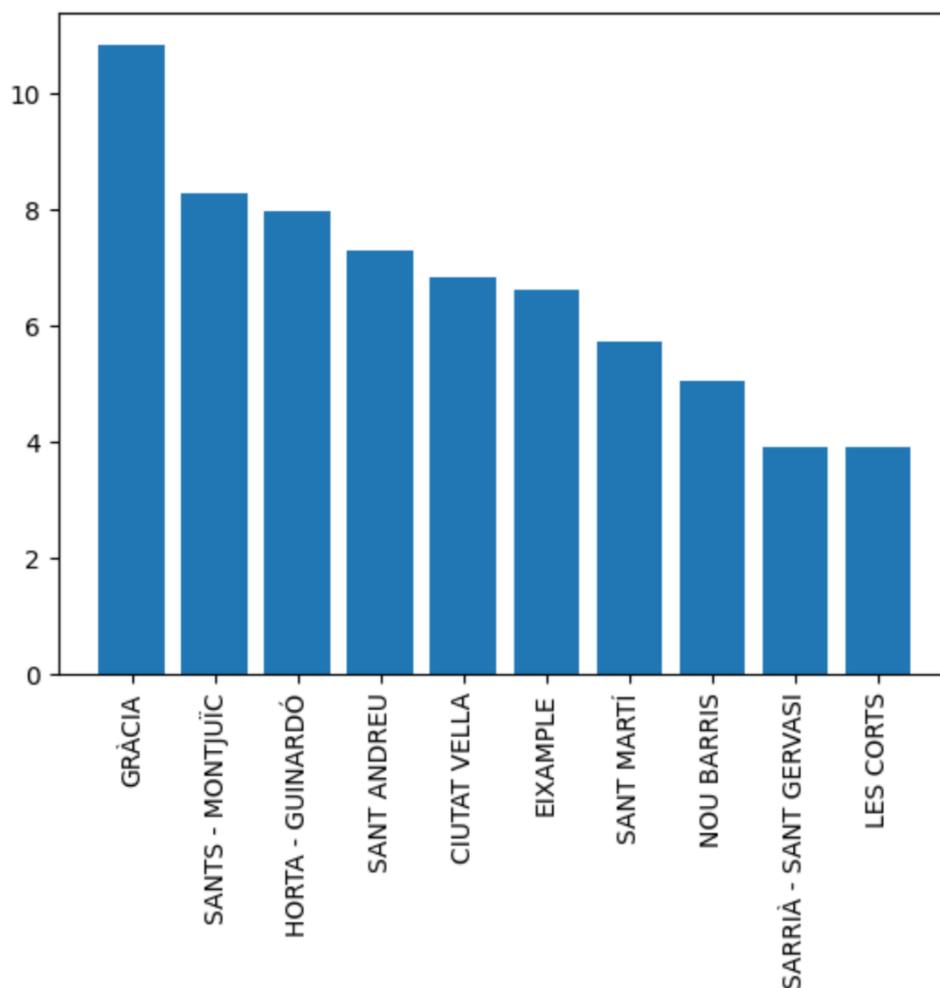


Figura 42: Gráfico de barras primer problema vivienda

A continuación se calcula el indicador de precio de la vivienda. Se filtran las filas del archivo para conservar solamente aquellas donde el valor de precio corresponde a euros/m<sup>2</sup>. Después se calcula el precio medio de cada distrito.

En las dos siguientes figuras se muestra la media del precio en euros de cada metro cuadrado de las operaciones de compra-venta realizadas en Barcelona en 2022 mediante una cuenta y un gráfico de barras ordenado según el porcentaje de personas de cada distrito que considera el acceso a la vivienda el problemas mas importante:

Nom_Districte	
Gràcia	3807.000000
Sants-Montjuïc	3621.114286
Horta-Guinardó	3351.690000
Sant Andreu	3109.150000
Ciutat Vella	4296.875000
Eixample	4678.450000
Sant Martí	4192.750000
Nou Barris	2460.540000
Sarrià-Sant Gervasi	5462.416667
Les Corts	4725.466667

Name: Valor, dtype: float64

Figura 43: Precio operaciones compra-venta

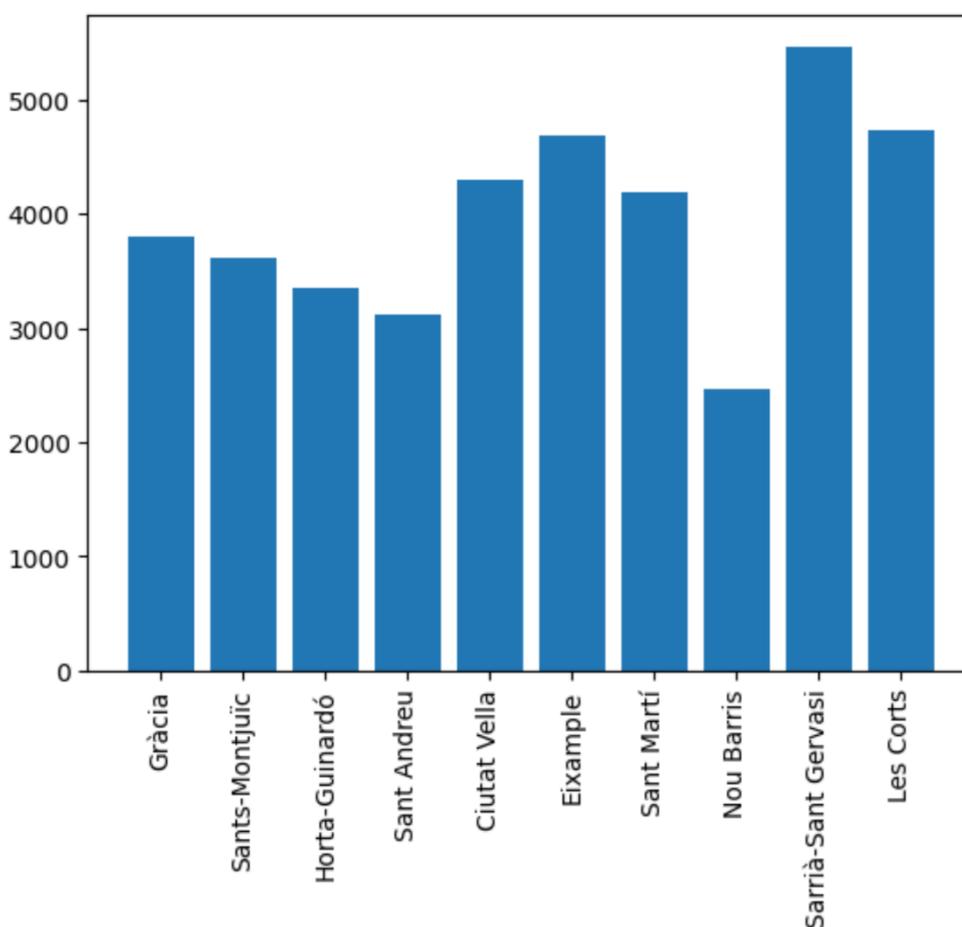


Figura 44: Gráfico de barras de precio operaciones compra-venta

Por último se calcula el indicador de renta por cápita. Se calcula la media de renta por cápita de cada distrito a partir del archivo csv descrito en el apartado 4.4 del trabajo.

En las dos siguientes figuras se muestra la media de la renta por persona de cada distrito en miles de euros mediante una cuenta y un gráfico de barras, ordenados según el porcentaje de personas de cada distrito que considera el acceso a la vivienda el problema mas importante:

```
Nom_Districte
Gràcia                23457.839080
Sants-Montjuïc       17835.547009
Horta-Guinardó       18450.715447
Sant Andreu          18285.726316
Ciutat Vella         14559.555556
Eixample             24546.150289
Sant Martí           18840.503401
Nou Barris           14577.572650
Sarrià-Sant Gervasi  37289.846939
Les Corts            29403.245614
Name: Import_Renda_Bruta_€, dtype: float64
```

Figura 45: Renta por cápita

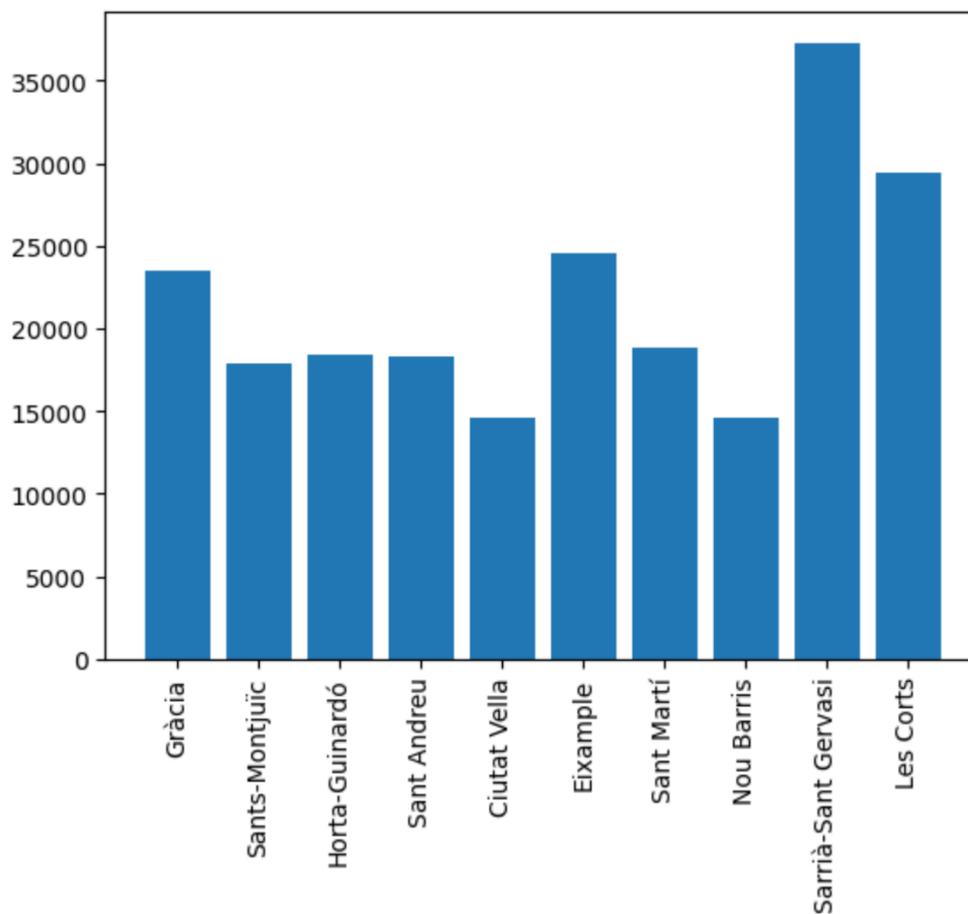


Figura 46: Gráfico de barras de renta por cápita

Se observa que los distritos donde menos porcentaje de la población indica que el acceso a la vivienda es el principal problema de vivir en Barcelona son Sarrià y Les Corts, justamente los dos distritos con más renta por cápita y precio por metro cuadrado. Esto añade fiabilidad a nuestro estudio.

Se observa que la mayor brecha entre renta por cápita y precio por metro cuadrado es en Ciutat Vella, Sant Martí i l'Eixample.

Se observa una brecha comparativa pequeña entre renta por cápita y coste de la vivienda en Gràcia, que es el distrito donde más porcentaje de la población ha indicado que el acceso a la vivienda es su mayor preocupación. No se distingue ninguna relación entre el orden de distritos según cantidad de gente que considera el acceso a la vivienda su mayor preocupación y la renta o el precio de la vivienda, tampoco en la diferencia entre los dos indicadores, a excepción de los dos últimos distritos, los mas ricos de Barcelona.

Sin embargo, observamos que el distrito con una diferencia mayor entre coste y renta por cápita es Ciutat Vella, que destacaba especialmente por la problemática de inseguridad, por lo que los resultados obtenidos hay que tomarlos con prudencia, pues aunque no sea considerado como problema principal puede ser considerado como un problema secundario y faltaría más información para obtener mejores conclusiones.

## 5.5 Clustering

En este punto se realiza el estudio de Clustering sobre los datos de la encuesta de opinión.

Se calculan los clústers a partir de los datos presentados en el apartado 4.1 del trabajo.

En los cálculos iterativos de las técnicas de clustering no se ha tomado en cuenta los distritos porque no interesan las diferencias que aportan para este estudio. Tampoco se ha tomado en cuenta el uso de redes y la situación laboral de los encuestados, ya que tras los estudios exploratorios que se considera que no aportan información relevante si no que influirían negativamente en la obtención de resultados. Esto es debido a que en el uso de redes encontramos una enorme frecuencia en uno de los valores, y respecto a la situación laboral de los encuestados es debido a que el valor más repetido es demasiado frecuente y a que el segundo valor mas frecuente se relaciona claramente con la cantidad de gente de 65 años o mayor presente en la variable de edad. Se mostrará el valor de la variable de distritos en cada clúster aunque no se tome en cuenta en los cálculos.

Primeramente se realizan las iteraciones necesarias con el método de clustering K-Modes para decidir el número de clusters en los que se pueden dividir los datos analizados mediante la técnica de los codos.

En la siguiente figura se muestra la representación gráfica de los resultados obtenidos:



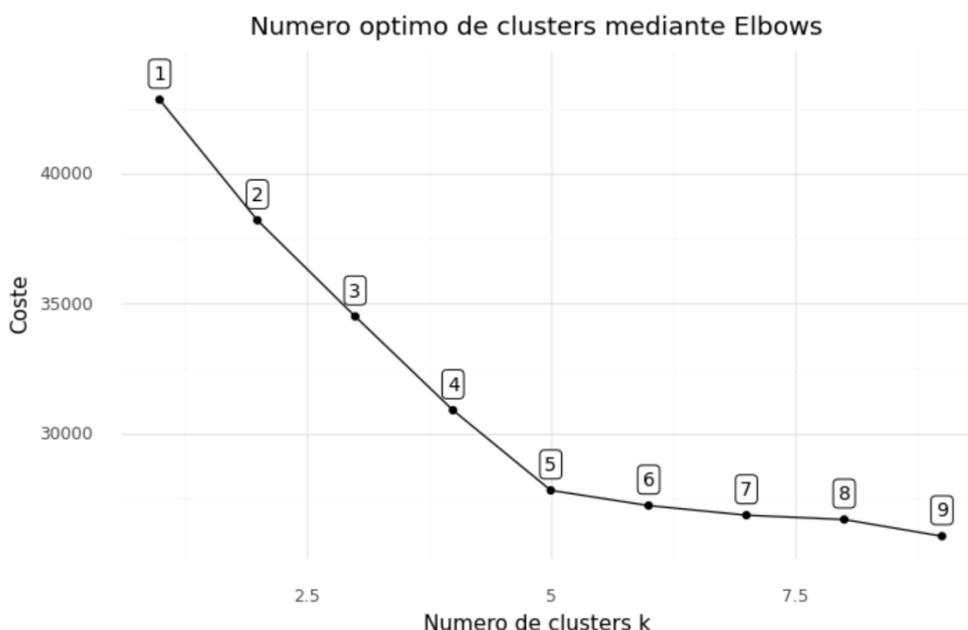


Figura 47: Gráfico número óptimo de clusters

Se puede observar que con 5 clústers hay un claro codo así que se procede a obtenerlos ejecutando la técnica de clustering K-Modes en Jupyter Notebook con Python.

Los clústers obtenidos son los siguientes:

Segment	Total	NOM_DISTRICTE	SEXE	EDAT1899_1A6	SATISF_RES_BCN_1A4	PROB_BCN_AGR	IDEO_1A8	ESTUDIS_1A6	LLAR_INGRESSOS_1A10
0	First	1311	EIXAMPLE	HOME	DE 65 ANYS I MÉS	SATISFET/A (5 A 6)	INSEGURETAT	CENTRE	OBLIGATORIS DE 500 A 1.000 EUROS
1	Second	1455	SANT MARTÍ	HOME	DE 55 A 64 ANYS	BASTANT SATISFET/A (7 A 8)	NETEJA	ESQUERRA SECUNDARIS PROFESSIONALS	DE 1.001 A 1.500 EUROS
2	Third	1191	EIXAMPLE	HOME	DE 25 A 34 ANYS	BASTANT SATISFET/A (7 A 8)	INSEGURETAT	ESQUERRA POST UNIVERSITARIS	DE 2.501 A 3.000 EUROS
3	Fourth	1134	EIXAMPLE	DONA	DE 45 A 54 ANYS	BASTANT SATISFET/A (7 A 8)	INSEGURETAT	CENTRE ESQUERRA UNIVERSITARIS	DE 2.001 A 2.500 EUROS
4	Fifth	909	NOU BARRIS	DONA	DE 65 ANYS I MÉS	MOLT SATISFET/A (9 A 10)	NO HO SAP NO HO SAP / NO LI INTERESSA LA POLÍTICA	SECUNDARIS GENERALS	DE 500 A 1.000 EUROS

Figura 48: Clusters

Esto significa que en los datos obtenidos de la encuesta de opinión de Barcelona se distinguen 5 grupos de personas con características similares.

A continuación se muestra con detalle cada uno de los clústers.

## CLÚSTER 1

La siguiente figura muestra la cuenta de los diferentes valores de la variable de distrito:

```
NOM_DISTRICTE
EIXAMPLE          177
NOU BARRIS        169
HORTA – GUINARDÓ  164
SANT MARTÍ        153
CIUTAT VELLA     133
SANTS – MONTJUÏC 132
SARRIÀ – SANT GERVASI 99
SANT ANDREU       99
LES CORTS         93
GRÀCIA           92
Name: count, dtype: int64
```

Figura 49: Cuenta de NOM\_DISTRICTE de Cluster 1

La siguiente figura muestra la cuenta de los diferentes valores de la variable de sexo:

```
SEXE
HOME    949
DONA    362
Name: count, dtype: int64
```

Figura 50: Cuenta de SEXE de Cluster 1

Se observa que en este clúster hay una gran mayoría de hombres.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de edad:

```
EDAT1899_1A6
DE 65 ANYS I MÉS  579
DE 35 A 44 ANYS  213
DE 45 A 54 ANYS  169
DE 55 A 64 ANYS  133
DE 25 A 34 ANYS  109
DE 18 A 24 ANYS  108
Name: count, dtype: int64
```

Figura 51: Cuenta de EDAT1899\_1A6 de Cluster 1

Este clúster esta formado mayormente por personas con 65 años o más.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de satisfacción de vivir en Barcelona:

```
SATISF_RES_BCN_1A4
SATISFET/A (5 A 6)  590
MOLT SATISFET/A (9 A 10) 297
BASTANT SATISFET/A (7 A 8) 281
POC/GENS SATISFET/A (0 A 4) 141
NO HO SAP / NO CONTESTA 2
Name: count, dtype: int64
```

Figura 52: Cuenta de SATISF\_RES\_BCN de Cluster 1

Este clúster esta formado mayormente por personas que puntúan su satisfacción entre 5 y 6. La cantidad de personas que realiza esta valoración relativa a la satisfacción representa la mitad de personas de todos los encuestados con esta opinión

La siguiente figura muestra la cuenta de los diferentes valores de la variable de mayor problema de vivir en Barcelona:

PROB_BCN_AGR	
INSEGURETAT	409
NETEJA	109
GESTIÓ POLÍTICA MUNICIPAL	59
CONGESTIÓ DE TRÀNSIT	52
PROBLEMES ECONÒMICS	52
NO HO SAP	52
ACCÉS A L'HABITATGE	50
ATUR / CONDICIONS DE TREBALL	48

Figura 53: Cuenta de PROB\_BCN\_AGR de Cluster 1

Se aprecia que el problema de inseguridad es claramente el mas indicado.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ideología política:

IDEO_1A8	
CENTRE	437
ESQUERRA	265
NO HO SAP / NO LI INTERESSA LA POLÍTICA	171
CENTRE ESQUERRA	122
NO CONTESTA	86
CENTRE DRETA	77
APOLÍTIC	65
DRETA	55
EXTREMA ESQUERRA	26
EXTREMA DRETA	7
Name: count, dtype: int64	

Figura 54: Cuenta de IDEO\_1A8 de Cluster 1

La opción ideológica de centro es la mas escogida.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de estudios:

ESTUDIS_1A6	
OBLIGATORIS	526
UNIVERSITARIS	250
SECUNDARIS GENERALS	217
SECUNDARIS PROFESSIONALS	162
POST UNIVERSITARIS	116
NO HA ACABAT ELS ESTUDIS OBLIGATORIS	34
NO CONTESTA	6
Name: count, dtype: int64	

Figura 55: Cuenta de ESTUDIS\_1A6 de Cluster 1

Gran parte de los integrantes de este grupo poseen solamente estudios obligatorios.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ingresos por hogar:

```

LLAR_INGRESSOS_1A10
DE 500 A 1.000 EUROS      247
DE 1.501 A 2.000 EUROS   189
NO CONTESTA              161
DE 3.001 A 5.000 EUROS   160
DE 1.001 A 1.500 EUROS   153
DE 2.001 A 2.500 EUROS   102
NO HO SAP                 90
DE 2.501 A 3.000 EUROS   90
MENYS DE 500 EUROS       50
DE 5.001 A 7.000 EUROS   45
DE 7.001 A 9.000 EUROS   12
MÉS DE 9.000 EUROS       12
Name: count, dtype: int64

```

Figura 55: Cuenta de LLAR\_INGRESSOS\_1A10 de Cluster 1

El tramo de renta más recurrente es el de 500 a 1.000 euros.

Así pues, se observa que el primer clúster esta formado por un perfil de encuestados muy mayoritariamente hombres de avanzada edad, que están poco satisfechos de vivir en Barcelona, de ideología principalmente de centro, estudios obligatorios y pocos ingresos.

Se recuerda que mas de la mayoría de los encuestados estaban muy satisfechos de vivir en Barcelona, este clúster reúne a la mitad de las personas que no lo están.

## CLÚSTER 2

La siguiente figura muestra la cuenta de los diferentes valores de la variable de distrito:

```

NOM_DISTRICTE
SANT MARTÍ          216
EIXAMPLE            204
SANTS - MONTJUÏC   171
HORTA - GUINARDÓ   162
NOU BARRIS         152
SANT ANDREU        141
GRÀCIA             117
CIUTAT VELLA       110
LES CORTS           97
SARRIÀ - SANT GERVASI 85
Name: count, dtype: int64

```

Figura 57: Cuenta de NOM\_DISTRICTE de Cluster 2

La siguiente figura muestra la cuenta de los diferentes valores de la variable de sexo:

```

SEXE
HOME      887
DONA      568
Name: count, dtype: int64

```

Figura 58: Cuenta de SEXE de Cluster 2

Este clúster está formado mayoritariamente por hombres.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de edad:

```

EDAT1899_1A6
DE 55 A 64 ANYS      526
DE 35 A 44 ANYS      242
DE 65 ANYS I MÉS     208
DE 45 A 54 ANYS      200
DE 18 A 24 ANYS      167
DE 25 A 34 ANYS      112
Name: count, dtype: int64

```

Figura 59: Cuenta de EDAT1899\_1A6 de Cluster 2

Este clúster esta formado mayormente por personas de edad media-avanzada.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de satisfacción de vivir en Barcelona:

```

SATISF_RES_BCN_1A4
BASTANT SATISFET/A (7 A 8)      898
MOLT SATISFET/A (9 A 10)        243
SATISFET/A (5 A 6)              177
POC/GENS SATISFET/A (0 A 4)     135
NO HO SAP / NO CONTESTA         2
Name: count, dtype: int64

```

Figura 60: Cuenta de SATISF\_RES\_BCN de Cluster 2

La mayoría de las personas de este clúster opina que esta muy satisfecha de vivir en Barcelona.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de mayor problema de vivir en Barcelona:

```

PROB_BCN_AGR
NETEJA                      361
ACCÉS A L'HABITATGE        105
CONGESTIÓ DE TRÀNSIT        97
INSEGURETAT                 94
CONTAMINACIÓ I MEDI AMBIENT 91
TRANSPORTS, INFRAESTRUCTURES I COMUNICACIONS 73
PROBLEMES ECONÒMICS         57

```

Figura 61: Cuenta de PROB\_BCN\_AGR de Cluster 2

Se observa que la limpieza es la opción más indicada.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ideología política:

IDEO_1A8	
ESQUERRA	860
CENTRE ESQUERRA	151
NO HO SAP / NO LI INTERESSA LA POLÍTICA	110
CENTRE	108
CENTRE DRETA	53
NO CONTESTA	52
EXTREMA ESQUERRA	49
APOLÍTIC	40
DRETA	30
EXTREMA DRETA	2
Name: count, dtype: int64	

Figura 62: Cuenta de IDEO\_1A8 de Cluster 2

La mayoría de las personas que forman este clúster se define de ideología de izquierdas.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de estudios:

ESTUDIS_1A6	
SECUNDARIS PROFESSIONALS	555
UNIVERSITARIS	333
SECUNDARIS GENERALS	314
POST UNIVERSITARIS	127
OBLIGATORIS	107
NO HA ACABAT ELS ESTUDIS OBLIGATORIS	15
NO CONTESTA	4
Name: count, dtype: int64	

Figura 63: Cuenta de ESTUDIS\_1A6 de Cluster 2

El grado de estudios más frecuente en este clúster es el de secundarios profesionales.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ingresos por hogar:

LLAR_INGRESSOS_1A10	
DE 1.001 A 1.500 EUROS	381
DE 3.001 A 5.000 EUROS	248
DE 1.501 A 2.000 EUROS	185
DE 2.501 A 3.000 EUROS	135
DE 2.001 A 2.500 EUROS	132
NO CONTESTA	113
DE 5.001 A 7.000 EUROS	86
NO HO SAP	56
DE 500 A 1.000 EUROS	55
MENYS DE 500 EUROS	28
MÉS DE 9.000 EUROS	20
DE 7.001 A 9.000 EUROS	16
Name: count, dtype: int64	

Figura 64: Cuenta de LLAR\_INGRESSOS\_1ª10 de Cluster 2

Se observa que el segundo clúster es conformado por un perfil de encuestados mayoritariamente hombres, de edad mas bien avanzada, muy satisfechos con vivir en Barcelona, cuyo problema principal es la limpieza, de ideología política de izquierdas, de estudios mayoritariamente secundarios profesionales y renta moderada.

### CLÚSTER 3

La siguiente figura muestra la cuenta de los diferentes valores de la variable de distrito:

```
NOM_DISTRICTE
EIXAMPLE          237
CIUTAT VELLA     168
SANT MARTÍ       135
SANTS - MONTJUÏC 109
GRÀCIA           108
SARRIÀ - SANT GERVASI 108
HORTA - GUINARDÓ 104
LES CORTS        85
SANT ANDREU      77
NOU BARRIS       60
Name: count, dtype: int64
```

Figura 65: Cuenta de NOM\_DISTRICTE de Cluster 3

La siguiente figura muestra la cuenta de los diferentes valores de la variable de sexo:

```
SEXE
HOME    712
DONA    479
Name: count, dtype: int64
```

Figura 66: Cuenta de SEXE de Cluster 3

La mayoría de este clúster son hombres.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de edad:

```
EDAT1899_1A6
DE 25 A 34 ANYS    614
DE 35 A 44 ANYS    266
DE 45 A 54 ANYS    161
DE 65 ANYS I MÉS    77
DE 18 A 24 ANYS     43
DE 55 A 64 ANYS     30
Name: count, dtype: int64
```

Figura 67: Cuenta de EDAT1899\_1a6 de Cluster 3

Se observa que este clúster está formado mayoritariamente por personas jóvenes.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de satisfacción de vivir en Barcelona:

```
SATISF_RES_BCN_1A4
BASTANT SATISFET/A (7 A 8)      829
MOLT SATISFET/A (9 A 10)      203
SATISFET/A (5 A 6)            107
POC/GENS SATISFET/A (0 A 4)    51
NO HO SAP / NO CONTESTA        1
Name: count, dtype: int64
```

Figura 68: Cuenta de SATISF\_RES\_BCN de Cluster 3

La mayoría de personas de este clúster está muy satisfecha de vivir en Barcelona.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de mayor problema de vivir en Barcelona:

```
PROB_BCN_AGR
INSEGURETAT                      285
ACCÉS A L'HABITATGE              140
CONTAMINACIÓ I MEDI AMBIENT      108
CONGESTIÓ DE TRÀNSIT              74
TURISME                           58
PROBLEMES ECONÒMICS               48
TRANSPORTS, INFRASTRUCTURES I    48
COMUNICACIONS
NETEJA                            44
```

Figura 69: Cuenta de PROB\_BCN\_AGR de Cluster 3

La opción más frecuente de este clúster es la inseguridad.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ideología política:

```
IDEO_1A8
ESQUERRA                          542
CENTRE ESQUERRA                    196
CENTRE                              120
NO HO SAP / NO LI INTERESSA LA    116
POLÍTICA
CENTRE DRETA                        83
EXTREMA ESQUERRA                    54
DRETA                                31
NO CONTESTA                          29
APOLÍTIC                             19
EXTREMA DRETA                        1
Name: count, dtype: int64
```

Figura 70: Cuenta de IDEO\_1A8 de Cluster 3

La opción más frecuente de este clúster es la ideología de izquierdas.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de estudios:

```

ESTUDIS_1A6
POST UNIVERSITARIS           751
UNIVERSITARIS                 238
SECUNDARIS GENERALS          123
SECUNDARIS PROFESSIONALS      43
OBLIGATORIS                   35
NO HA ACABAT ELS ESTUDIS OBLIGATORIS  1
Name: count, dtype: int64

```

Figura 71: Cuenta de ESTUDIS\_1A6 de Cluster 3

La gran mayoría de este clúster tiene estudios universitarios o superiores.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ingresos por hogar:

```

LLAR_INGRESSOS_1A10
DE 2.501 A 3.000 EUROS    359
DE 3.001 A 5.000 EUROS    289
DE 1.501 A 2.000 EUROS    118
DE 2.001 A 2.500 EUROS    104
DE 5.001 A 7.000 EUROS     90
NO CONTESTA                81
DE 1.001 A 1.500 EUROS     38
NO HO SAP                   28
DE 500 A 1.000 EUROS       28
DE 7.001 A 9.000 EUROS     21
MENYS DE 500 EUROS         19
MÉS DE 9.000 EUROS         16
Name: count, dtype: int64

```

Figura 72: Cuenta de LLAR\_INGRESSOS\_1A10 de Cluster 3

Se observa que el tercer cluster está conformado por encuestados mayoritariamente hombres, de edad bastante joven, muy satisfechos de vivir en Barcelona, cuyo problema principal es la inseguridad, de ideología política de izquierdas, con estudios universitarios y post-universitarios y con una renta del hogar bastante alta.

#### CLÚSTER 4

La siguiente figura muestra la cuenta de los diferentes valores de la variable de distrito:

```

NOM_DISTRICTE
EIXAMPLE          187
SANT MARTÍ       157
SARRIÀ - SANT GERVASI 126
CIUTAT VELLA    113
LES CORTS        103
HORTA - GUINARDÓ  98
SANTS - MONTJUÏC 95
NOU BARRIS       94
GRÀCIA           88
SANT ANDREU      73
Name: count, dtype: int64

```

Figura 73: Cuenta de NOM\_DISTRICTE de Cluster 4

La siguiente figura muestra la cuenta de los diferentes valores de la variable de sexo:

```

SEXE
DONA    987
HOME    147
Name: count, dtype: int64

```

Figura 74: Cuenta de SEXE de Cluster 4

Se observa una gran mayoría de mujeres.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de edad:

```

EDAT1899_1A6
DE 45 A 54 ANYS    438
DE 35 A 44 ANYS   230
DE 65 ANYS I MÉS  155
DE 25 A 34 ANYS   125
DE 55 A 64 ANYS   108
DE 18 A 24 ANYS    78
Name: count, dtype: int64

```

Figura 75: Cuenta de EDAT1899\_1<sup>a</sup>6 de Cluster 4

La siguiente figura muestra la cuenta de los diferentes valores de la variable de satisfacción de vivir en Barcelona:

```

SATISF_RES_BCN_1A4
BASTANT SATISFET/A (7 A 8)  674
SATISFET/A (5 A 6)         188
MOLT SATISFET/A (9 A 10)   168
POC/GENS SATISFET/A (0 A 4) 103
NO HO SAP / NO CONTESTA     1
Name: count, dtype: int64

```

Figura 76: Cuenta de SATISF\_RES\_BCN de Cluster 4

La mayoría de personas del clúster están bastante satisfechas de vivir en Barcelona.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de mayor problema de vivir en Barcelona:

PROB_BCN_AGR	
INSEGURETAT	302
NETEJA	130
ACCÉS A L'HABITATGE	65
CONTAMINACIÓ I MEDI AMBIENT	64
CONGESTIÓ DE TRÀNSIT	61
PROBLEMES ECONÒMICS	46
TRANSPORTS, INFRAESTRUCTURES I COMUNICACIONS	41

Figura 77: Cuenta de PROB\_BCN\_AGR de Cluster 4

Inseguridad es la opción más frecuente.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ideología política:

IDEO_1A8	
CENTRE ESQUERRA	393
ESQUERRA	245
CENTRE	141
NO HO SAP / NO LI INTERESSA LA POLÍTICA	113
CENTRE DRETA	65
EXTREMA ESQUERRA	49
NO CONTESTA	48
DRETA	46
APOLÍTIC	27
EXTREMA DRETA	7
Name: count, dtype: int64	

Figura 78: Cuenta de IDEO\_1<sup>a</sup>8 de Cluster 4

La ideología más frecuente en este clúster es la de centro-izquierda.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de estudios:

ESTUDIS_1A6	
UNIVERSITARIS	704
POST UNIVERSITARIS	155
SECUNDARIS GENERALS	124
SECUNDARIS PROFESSIONALS	91
OBLIGATORIS	48
NO HA ACABAT ELS ESTUDIS OBLIGATORIS	9
NO CONTESTA	3
Name: count, dtype: int64	

Figura 79: Cuenta de ESTUDIS\_1<sup>a</sup>6 de Cluster 4

La mayoría de personas de este clúster tienen estudios universitarios o superiores.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ingresos por hogar:

```

LLAR_INGRESSOS_1A10
DE 2.001 A 2.500 EUROS    256
DE 3.001 A 5.000 EUROS    219
DE 1.501 A 2.000 EUROS    152
NO CONTESTA              112
DE 2.501 A 3.000 EUROS    84
DE 1.001 A 1.500 EUROS    79
DE 5.001 A 7.000 EUROS    70
NO HO SAP                 47
DE 500 A 1.000 EUROS      46
MENYS DE 500 EUROS        31
MÉS DE 9.000 EUROS        19
DE 7.001 A 9.000 EUROS    19
Name: count, dtype: int64

```

Figura 80: Cuenta de LLAR\_INGRESSOS\_1ª10 de Cluster 4

Se observa que el cuarto clúster esta formado por un perfil de encuestados mayoritariamente mujeres, de edad media, bastante satisfechas de vivir en Barcelona, cuyo principal problema es la inseguridad, de ideología política principalmente de centro-izquierda, con estudios universitarios y post-universitarios y rentas del hogar bastantes altas.

## CLUSTER 5

La siguiente figura muestra la cuenta de los diferentes valores de la variable de distrito:

```

NOM_DISTRICTE
NOU BARRIS          141
EIXAMPLE            119
SANT MARTÍ          109
SANTS - MONTJUÏC   109
CIUTAT VELLA        92
SANT ANDREU         91
HORTA - GUINARDÓ    88
LES CORTS           59
GRÀCIA              57
SARRIÀ - SANT GERVA SI 44
Name: count, dtype: int64

```

Figura 81: Cuenta de NOM\_DISTRICTE de Cluster 5

La siguiente figura muestra la cuenta de los diferentes valores de la variable de sexo:

```

SEXE
DONA    791
HOME    118
Name: count, dtype: int64

```

Figura 82: Cuenta de SEXE de Cluster 5

Se observa una gran mayoría de mujeres.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de edad:

```
EDAT1899_1A6
DE 65 ANYS I MÉS      454
DE 35 A 44 ANYS      133
DE 25 A 34 ANYS      105
DE 18 A 24 ANYS       85
DE 55 A 64 ANYS       74
DE 45 A 54 ANYS       58
Name: count, dtype: int64
```

Figura 83: Cuenta de EDAT1899\_1<sup>a</sup>6 de Cluster 5

La franja de edad más frecuente es claramente de 65 años y más.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de satisfacción de vivir en Barcelona:

```
SATISF_RES_BCN_1A4
MOLT SATISFET/A (9 A 10)    670
BASTANT SATISFET/A (7 A 8)  143
SATISFET/A (5 A 6)          62
POC/GENS SATISFET/A (0 A 4)  28
NO HO SAP / NO CONTESTA     6
Name: count, dtype: int64
```

Figura 84: Cuenta de SATISF\_RES\_BCN de Cluster 5

La mayoría de personas en este clúster están muy satisfechas de vivir en Barcelona.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de mayor problema de vivir en Barcelona:

```
PROB_BCN_AGR
NO HO SAP                161
NETEJA                   86
INSEGURETAT              76
ATUR / CONDICIONS DE TREBALL 56
PROBLEMES ECONÒMICS      51
VALORS SOCIALS NEGATIUS / MANCA D'EDUCACIÓ, CIVISME 49
ACCÉS A L'HABITATGE     38
```

Figura 85: Cuenta de PROB\_BCN\_AGR de Cluster 5

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ideología política:

IDEO_1A8	
NO HO SAP / NO LI INTERESSA LA POLÍTICA	446
ESQUERRA	202
NO CONTESTA	56
CENTRE	53
CENTRE ESQUERRA	43
APOLÍTIC	29
DRETA	28
CENTRE DRETA	26
EXTREMA ESQUERRA	21
EXTREMA DRETA	5
Name: count, dtype: int64	

Figura 86: Cuenta de IDEO\_1A8 de Cluster 5

La siguiente figura muestra la cuenta de los diferentes valores de la variable de estudios:

ESTUDIS_1A6	
SECUNDARIS GENERALS	405
OBLIGATORIS	213
SECUNDARIS PROFESSIONALS	103
NO HA ACABAT ELS ESTUDIS OBLIGATORIS	79
UNIVERSITARIS	64
POST UNIVERSITARIS	44
NO CONTESTA	1
Name: count, dtype: int64	

Figura 87: Cuenta de ESTUDIS\_1A6 de Cluster 5

La mayoría de personas de este clúster tienen estudios secundarios y obligatorios.

La siguiente figura muestra la cuenta de los diferentes valores de la variable de ingresos por hogar:

LLAR_INGRESSOS_1A10	
DE 500 A 1.000 EUROS	267
DE 1.001 A 1.500 EUROS	123
DE 1.501 A 2.000 EUROS	122
NO HO SAP	108
NO CONTESTA	103
MENYS DE 500 EUROS	47
DE 2.001 A 2.500 EUROS	43
DE 3.001 A 5.000 EUROS	39
DE 2.501 A 3.000 EUROS	32
DE 5.001 A 7.000 EUROS	13
DE 7.001 A 9.000 EUROS	7
MÉS DE 9.000 EUROS	5
Name: count, dtype: int64	

Figura 88: Cuenta de LLAR\_INGRESSOS\_1A10 de Cluster 5

El quinto y último clúster esta formado principalmente por mujeres de edad avanzada muy satisfechas de vivir en Barcelona, que no saben cual consideran su principal problema de vivir en Barcelona ni su ideología política, con estudios secundarios y obligatorios y rentas del hogar bajas.

Tras explorar los 5 clústers se identifican varias relaciones entre algunas variables y las respuestas que ofrecen. Cuanto más predomina la ideología política de centro más insatisfecho se declara el encuestado de vivir en Barcelona y su principal problema de vivir en ella es la inseguridad, como se puede observar en el primer y cuarto clúster. El segundo y tercer clúster están muy satisfechos de vivir en Barcelona, siendo de ideología política de izquierdas, y sus principales problemas son la inseguridad y la limpieza. Sin embargo, al estar estos dos clústers tan satisfechos de vivir en Barcelona se puede desestimar que consideren estos dos problemas como críticos. Esto refuerza la poca relación entre los indicadores de los tres análisis de los problemas principales y las opiniones segregadas por distritos.

## 6. Planificación

Debido a que el trabajo se presenta en el terminio de prorroga, se muestra la planificación llevada a cabo finalmente, no la pretendida inicialmente.

A continuación se adjunta un diagrama de Gant simple que muestra la realización del proyecto según el marco temporal:



Figura 89: Diagrama de Gant

## 7. Estudio económico

El estudio económico se basa en un presupuesto a partir del coste de personal, licencias de los programas utilizados y el coste del material empleado.

### Coste de personal

Para realizar este trabajo se han empleado 290h a un coste de 15€. Por lo tanto, el coste de personal total es el siguiente:

$$290 \text{ h} \times \frac{15\text{€}}{\text{h}} = 4350\text{€}$$

### Coste de licencias

La mayoría de programas utilizados en el desarrollo del trabajo son gratuitos a excepción de Microsoft Office, que cuesta 69€/año.

Por lo tanto, el coste de licencias es el siguiente:

$$\frac{69\text{€}}{\text{año}} \times \frac{1 \text{ año}}{12 \text{ meses}} \times 6 \text{ meses} = 34,5\text{€}$$

### Coste de material

Para realizar este trabajo se ha utilizado un ordenador iMac con un coste de adquisición de 1200€. Teniendo en cuenta que un ordenador se amortiza linealmente en 5 años, el coste de material es el siguiente:

$$\frac{1200\text{€}}{5 \text{ años}} \times \frac{1 \text{ año}}{12 \text{ meses}} \times 6 \text{ meses} = 120\text{€}$$

### Coste total

Realizando la suma de los costes anteriores se obtiene el coste total siguiente:

$$4350\text{€} + 34,5\text{€} + 120\text{€} = 4504,5\text{€}$$

Añadiendo el IVA del 21% el coste total definitivo asciende a la cuantía siguiente:

$$4504,5\text{€} \times 1,21 = 5450,45\text{€}$$

Se considera un coste reducido, pues el coste total consiste principalmente en el coste de personal.

## 8. Estudio ambiental

El impacto ambiental ocasionada por este proyecto es prácticamente inexistente pues se ha realizado íntegramente utilizando las capacidades computacionales de un ordenador.

Sin embargo, es posible calcular el gasto energético y las emisiones producidas por el uso del ordenador, considerando que la potencia consumida por un ordenador es de 0,1 kW y que cada 1kWh de energía eléctrica comporta una emisión de 0,45 kg de CO<sub>2</sub> en su generación.

Por lo tanto, el gasto energético es el siguiente:

$$0,1 \text{ kW} \times 290 \text{ h} = 29 \text{ kWh}$$

Y la emisión la siguiente:

$$29 \text{ kWh} \times \frac{0,45 \text{ kg CO}_2}{1 \text{ kWh}} = 13,05 \text{ kg CO}_2$$

Comparando el gasto energético y la emisión con cualquier otro proyecto de ingeniería o proceso industrial resulta prácticamente nulo.

## 9. Estudio social y de igualdad de género

A continuación se presentan varios comentarios respectivos al estudio social y de igualdad de género del trabajo:

- El acceso a la tecnología trabajada no discrimina a ningún colectivo en especial, es tecnología de código libre y con acceso abierto. Sin embargo, las personas en situación de pobreza económica no pueden permitirse el equipo electrónico necesario para acceder a él.
- El trabajo ha sido realizado por un hombre y el director del trabajo ha sido otro hombre.
- Durante el desarrollo del trabajo se ha tenido en cuenta constantemente la opinión de mujeres y hombres y de habitantes con todo tipo de rentas y estudios, así como del impacto que sufren dichos colectivos por los asuntos analizados.
- Se ha usado un lenguaje no sexista.
- Siempre se ha dado la misma importancia a la opinión y impacto de los asuntos estudiados a todos los individuos, independientemente de género, renta, estudios...
- Los resultados del trabajo favorecen a identificar problemas de toda la población. Se identifica durante la realización del proyecto problemas que afectan a colectivos vulnerables.
- El trabajo realizado se ajusta al objetivo de desarrollo sostenible (ODS) de Ciudades y Comunidades Sostenibles.

## 10. CONCLUSIONES

Se concluye que la encuesta de opinión de Barcelona ofrece unos resultados que parecen caracterizar bien a la población residente. Se observa que el problema más reportado como más importante es el de inseguridad, seguido de limpieza y de acceso a la vivienda. Se observa también que más de la mitad de los encuestados puntúan su satisfacción de vivir en Barcelona con un 7 o más.

La cantidad de encuestados de cada distrito que declara como principal problema la inseguridad no es proporcional a los tres indicadores de delitos de cada distrito, es decir, que la percepción ciudadana de seguridad no se corresponde con los indicadores estudiados. Sin embargo, en Ciutat Vella sí se corresponde, se puede confirmar que ocurre un problema de inseguridad en Ciutat Vella.

La cantidad de encuestados de cada distrito que declara como principal problema la limpieza no es proporcional a la basura encontrada en cada distrito. La percepción de los ciudadanos respecto a la limpieza no se corresponde con los indicadores estudiados. No se puede afirmar que exista un problema de limpieza.

La cantidad de encuestados de cada distrito que declara como principal problema el acceso a la vivienda tampoco es proporcional a los indicadores de cada distrito.

Tras el análisis de Clustering se descubre que los encuestados de Barcelona se pueden dividir en 5 grupos:

El primero consta de una gran mayoría de hombres de edad bastante avanzada con ideología política de centro, poco satisfechos de vivir en Barcelona y cuyo problema principal es la inseguridad, con estudios mayoritariamente secundarios y rentas del hogar más bien bajas.

En el segundo se observa un perfil de encuestados de mayoría hombres, de edad más bien avanzada, con ideología de izquierdas, muy satisfechos de vivir en Barcelona, cuyo problema principal es la limpieza, con estudios mayoritariamente secundarios profesionales y rentas de hogar media-baja.

El tercer cluster está formado por una mayoría de hombres, con edad más bien joven, muy satisfechos de vivir en Barcelona, con una ideología de izquierdas, cuyo problema principal es la inseguridad, estudios post universitarios principalmente, y rentas medias-altas.

El cuarto grupo consta de una gran mayoría de mujeres, de edad media, bastante satisfechas de vivir en Barcelona, cuyo problema principal es la inseguridad, ideología principalmente de centro-izquierda, con estudios universitarios y rentas medias-altas.

El quinto grupo esta formado por una gran mayoría de mujeres, de edad avanzada, muy satisfechas de vivir en Barcelona, que no saben cual es su principal problema ni su ideología política, con estudios secundarios y obligatorios y rentas bajas.

Se aprecia que los clústers con ideología principalmente de centro son los que menos satisfechos están de vivir en Barcelona y indican inseguridad como problema principal.

Los clústers con ideología principalmente de izquierdas reparten su opinión entre inseguridad y limpieza.

## 11. Bibliografia

[1] <https://www.python.org/about/>

[2] <http://pandas.pydata.org/>

[3] <https://matplotlib.org/>

[4] Wes McKinney. Python for Data Analysis. O'Reilly, first edition 2012 october.

[5] <https://novagenio.medium.com/aprendizaje-no-supervisado-clustering-para-todos-con-python-2cc2bcaafae9>

[6] <https://www.unir.net/ingenieria/revista/clustering/>

[7] <https://www.analyticslane.com/2022/10/14/como-funciona-k-modes-e-implementacion-en-python/>

[8] <https://medium.com/@jonathanrmzg/k-modes-algorithm-d4b54aca1e01>

[9] <https://jarroba.com/seleccion-del-numero-optimo-clusters/>