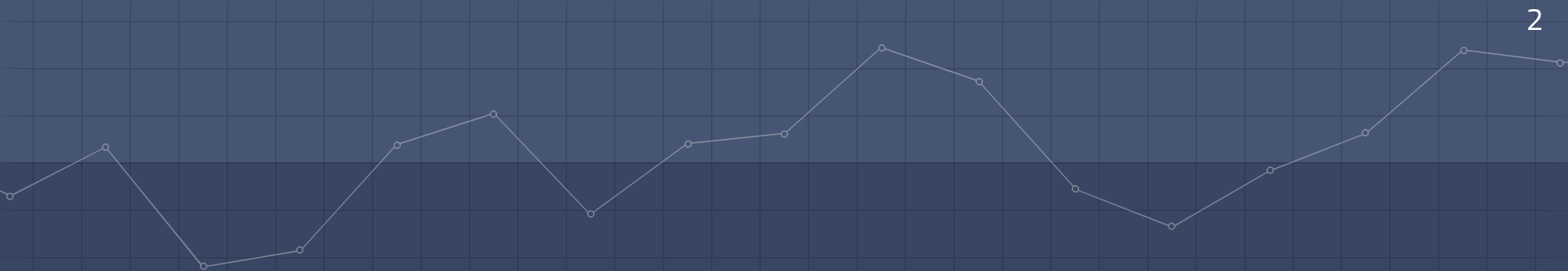


ANTEPROYECTO DEL TFG

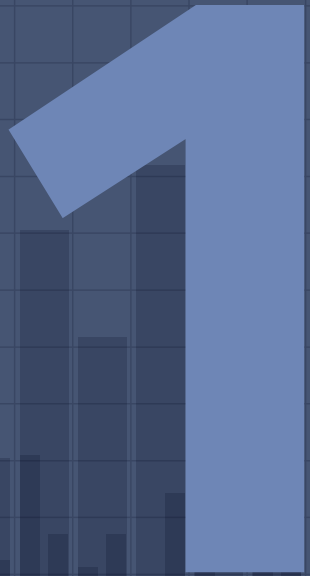
The background features a dark blue grid. Overlaid on the grid is a faint line chart with white circular markers at each data point. The line starts at a low point on the left, rises to a peak, falls to a trough, rises again to a higher peak, falls to another trough, rises to a third peak, falls to a fourth trough, and finally rises to its highest peak on the right before a slight decline.

ROCÍO SOSA RUIZ
Ingeniería Informática



"Without big data, you are blind and deaf and in the middle of a freeway." – Geoffrey Moore

TÍTULO DEL PROYECTO Y MODALIDAD



TÍTULO: 'HOODFINDER'

LOGO APP:



MODALIDAD: Ingeniería

Es un proyecto de ingeniería debido a que es una aplicación donde los clientes interactúan y se obtienen unos resultados tras un estudio de datos que le transmiten una información nueva al cliente de manera visual y agradable.

DESCRIPCIÓN DEL DESARROLLO

2

EL QUÉ

El proyecto a desarrollar consiste en una aplicación web que le sirve de utilidad a los clientes que quieran venir a vivir a Madrid y no tengan información sobre los distintos barrios y distritos de la ciudad.

A través de un formulario inicial donde el cliente responde a una serie de preguntas sobre algunas preferencias de su barrio ideal, se estudia la afinidad de este cliente con los distintos barrios. De manera que cuando haga el envío del formulario, automáticamente le mostrará una página donde se indica el barrio ganador y una serie de información sobre ese barrio.

Además, se mostrará otra sección en la aplicación donde se mostrarán datos curiosos a través de gráficas y datos estadísticos sobre los diferentes barrios y distritos de la ciudad.

DATASET

Para esta aplicación se han necesitado de 4 set de datos iniciales distintos que se han juntado en un único set de datos final.

1. El primer dataset y el principal es extraído del Portal de datos abiertos del Ayuntamiento de Madrid. El link de descarga es:

<https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=71359583a773a510VgnVCM2000001f4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default>

Es una base de datos que ofrece el Ayuntamiento de Madrid que incluye información sociodemográfica de los barrios y distritos de la ciudad de Madrid.

2. El segundo dataset es extraído del Portal de datos abiertos del Ayuntamiento de Madrid. El link de descarga es:

<https://datos.madrid.es/portal/site/egob/menuitem.3efdb29b813ad8241e830cc2a8a409a0/?vgnextoid=55a082257a859610VgnVCM1000001d4a900aRCRD&vgnnextchannel=6db862c549810510VgnVCM1000008a4a900aRCRD&vgnextfmt=default>

Es una base de datos que ofrece el Ayuntamiento de Madrid que refleja la tasa de criminalidad de la ciudad de Madrid por barrios.

DATASET

3. El tercer dataset es extraído del Catálogo de datos abiertos de la Comunidad de Madrid. El link de descarga es:
https://datos.comunidad.madrid/catalogo/dataset/centros_educativos/resource/9578660d-f9de-48f4-a387-7f1a8333338b?inner_span=True
Es una base de datos con información sobre los centros educativos en cada barrio de la Comunidad de Madrid.
4. El cuarto dataset es extraído del Portal de Datos Abiertos de Esri España. El link de descarga es:
https://opendata.esri.es/datasets/7698bfcee79142918ed0f040505adfc0_0
Es una base de datos con información sobre las diferentes Universidades tanto públicas como privadas que existen en España (sólo necesitaremos las de Madrid).

DATASET FINAL (132 barrios)

	A	B	C	D	E	F	G	H	I	J
	Distrito	Barrio	Edad media de la población	Personas con nacionalidad española	Personas con nacionalidad extranjera	Renta neta media anual de los hogares (Urban Audit). 2016	Nº de Bibliotecas	Centros deportivos Municipales	Areas verde m2	Detencion X barrio
1										
2	CENTRO	Palacio	46	82%	18%	90%	1	0	288.625	26
3	CENTRO	Embajadores	43	74%	26%	68%	2	0	45.388	26
4	CENTRO	Cortes	45	77%	23%	94%	0	0	45.187	26
5	CENTRO	Justicia	44	79%	21%	105%	1	2	43.505	26
6	CENTRO	Universidad	44	78%	22%	80%	1	0	16.726	26
7	CENTRO	Sol	45	73%	27%	80%	0	0	6.548	26
8	ARGANZL	Imperial	46	93%	7%	90%	0	1	123.029	3
9	ARGANZL	Las Acacias	46	93%	7%	115%	1	1	281.585	3
10	ARGANZL	La Chopera	46	87%	13%	82%	0	0	135.271	3
11	ARGANZL	Legazpi	38	93%	7%	127%	0	1	578.260	3
12	ARGANZL	Las Delicias	43	89%	11%	102%	1	1	126.762	3
13	ARGANZL	Palos de Moguer	46	85%	15%	87%	0	1	15.730	3
14	ARGANZL	Atocha	39	94%	6%	97%	0	0	42.243	3
15	RETIRO	Pacífico	48	91%	9%	93%	0	2	30.086	2
16	RETIRO	Adelfas	45	93%	7%	116%	1	0	59.058	2
17	RETIRO	La Estrella	49	96%	4%	150%	0	0	193.723	2
18	RETIRO	Ibiza	47	89%	11%	118%	0	0	1.617	2
19	RETIRO	Los Jerónimos	48	90%	10%	176%	1	2	1.240.013	2
20	RETIRO	Niño Jesús	46	94%	6%	172%	0	0	26.715	2
21	SALAMAN	Recoletos	46	83%	17%	201%	0	0	39.142	10
22	SALAMAN	Goya	46	86%	14%	128%	0	1	5.443	10
23	SALAMAN	Fuente del Berro	47	88%	12%	105%	0	1	143.267	10
24	SALAMAN	Guindalera	47	89%	11%	118%	1	1	207.381	10
25	SALAMAN	Lista	46	86%	14%	125%	0	0	10.872	10
26	SALAMAN	Castellana	45	84%	16%	183%	0	0	26.663	10
27	CHAMAR	El Viso	44	88%	11%	222%	0	0	60.429	4

PARA QUÉ

Este proyecto me parece interesante y muy útil debido a que cuando una persona decide mudarse a otra ciudad/país suele estar muy perdido a la hora de elegir el barrio en el que vivir si no conoce a nadie que le pueda ayudar.

Debido a que cada uno tiene sus circunstancias económicas, sociales y culturales es importante tener datos de este tipo sobre cada barrio para elegir según tus gustos y comodidades.

Además de ayudarte a elegir el barrio según tus preferencias, se muestran una serie de datos curiosos en forma de gráficas sobre los barrios y distritos de Madrid, de manera que puedas comparar unos con otros de forma visual y agradable.

Con este proyecto se consigue contribuir al análisis e investigación de los diferentes dataset, descubriendo patrones y tendencias estructurando la información que se ha obtenido de una manera que sea comprensible para su utilización. Además, al implementarlo en una aplicación web es más visual para ver los resultados del estudio realizado.

MOTIVACIÓN

3

POR QUÉ

Desde un principio tenía clara la idea de hacer un análisis de datos a partir de un dataset y usarlos de alguna manera, debido a que es un tema que me interesa mucho y creo que puede aportar valor.

Sólo me faltaba pensar en un tema a estudiar o a analizar. Pensé en preocupaciones mundiales o sucesos importantes donde existiera información y datos suficientes, y por otro lado, que los resultados pudieran contribuir a disponer de más información valiosa sobre este tema. Estuve investigando en diferentes bancos de datos en Internet y encontré el banco de datos del Ayuntamiento de Madrid.

En este portal encontré el set de datos con mucha información sociodemográfica de los distintos barrios y distritos de Madrid. Pensé en que sería útil poder ayudar a las personas extranjeras a elegir qué barrio le conviene vivir según sus preferencias en forma de una aplicación y que te diera datos válidos de cada barrio de forma visual.

Otros temas que me interesaron: factores que influyen en el terrorismo, PIBpc, accidentes de coches, activaciones del SAMUR, nivel de contaminación del aire en Madrid...

Pero finalmente he decidido este tema porque es el que mayor utilidad tiene en el mundo real.

OBJETIVOS A ALCANZAR



4

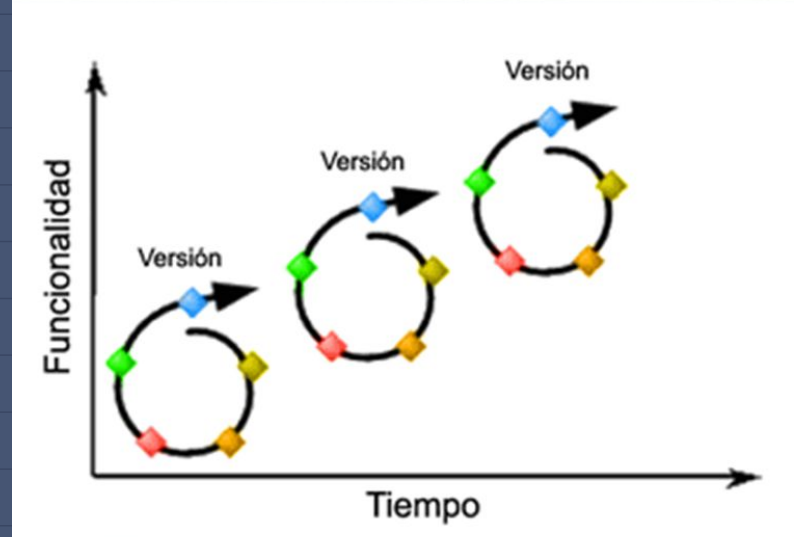
Objetivo principal

El objetivo principal de este proyecto es conseguir realizar una aplicación que ayude e informe al cliente del barrio más afín a sus gustos en la ciudad de Madrid y además aportarle en forma de gráficas y estadísticas nueva información sobre cada barrio y distrito de la ciudad.

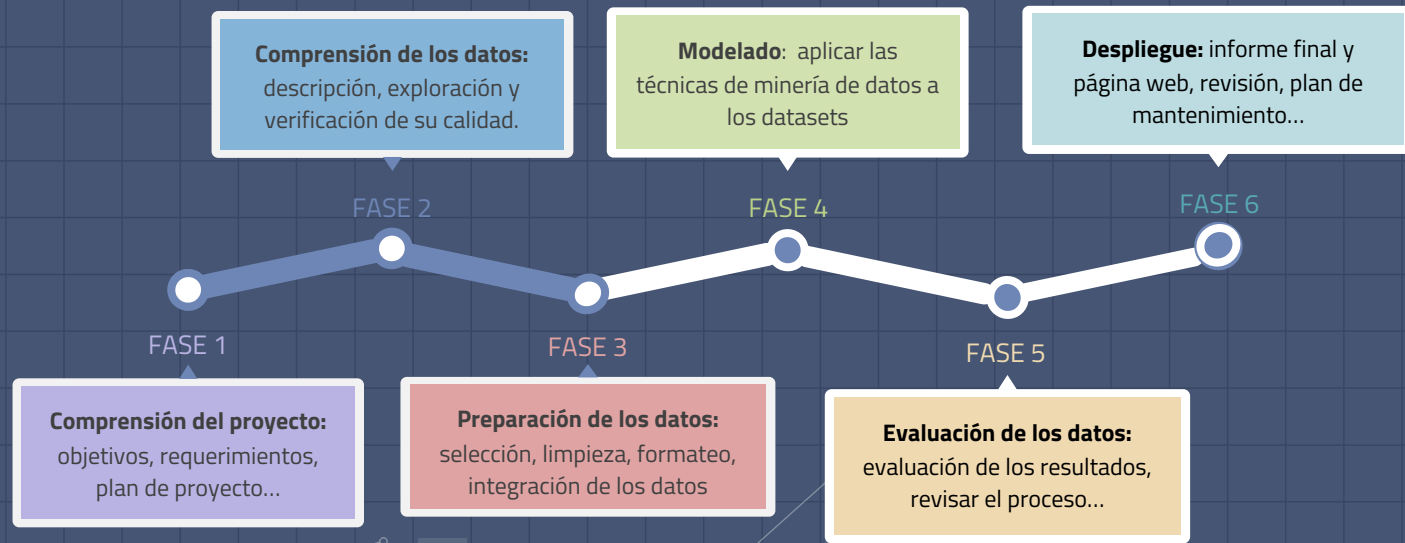
Objetivos específicos

Para poder conseguir el objetivo principal se usará el modelo incremental y la metodología CRISP-DM usadas como guías de referencia para el correcto desarrollo del proyecto, explicadas en las siguientes diapositivas.

MODELO INCREMENTAL



FASES CRISP – DM (datos curiosos)



RESULTADOS ESPERADOS Y VIABILIDAD

5

Resultados: Una aplicación donde haya una interacción cliente-servidor, de manera que el cliente envía unos datos al servidor, sacados del formulario de respuesta múltiple, y el servidor le devuelve una respuesta que sería el barrio elegido según los datos seleccionados y además una información detallada del barrio. Además, un estudio de datos con gráficas de modo que se muestre una interfaz con datos curiosos del dataset.

Plazo estimado del desarrollo técnico: aproximadamente se estima un plazo de 2 meses y medio del desarrollo de la aplicación y el estudio de los datos. (Convocatoria extraordinaria de Enero 2021)

Conocimientos necesarios que ya se tienen: manejo de Excel, uso del IDE eclipse con lenguaje Java, Vaadin, Spring Framework, SQL en MySQL Workbench. Además uso del entorno Python/Anaconda y librerías Machine Learning, análisis de datasets y visualización de datos (preprocesamiento de los datos e interpretación de gráficos). Despliegue de aplicaciones en Heroku y uso de Git durante todo el desarrollo del proyecto.

Conocimientos que es necesario adquirir: irán surgiendo a lo largo del desarrollo del proyecto, pero es posible que nuevas librerías, nuevos conocimientos de Spring y de Vaadin10+, sentencias SQL... Los conocimientos técnicos los obtendré a partir de algunas páginas como kaggle, DataCamp, Stack Overflow, Youtube...

Asignaturas del grado con las que se relaciona: Fundamentos de Ingeniería Informática (1º), Desarrollo Web (2º), Inteligencia Artificial I (2º), Ingeniería del Software I y II (2º y 3º), Estadística (3º), Desarrollo e Integración de Software (3º), Data Mining (4º), Planificación y Gestión de Proyectos Informáticos (4º), Tecnologías Avanzadas de Programación (4º)

BIBLIOGRAFÍA

<https://www.slidescarnival.com/es>

<https://moodleufv.ufv.es/>

PDF de "Presentación de la asignatura" obtenido del aula virtual de la asignatura de Data Mining

PDF de "Introducción al Data Mining" (Tema 1) obtenido del aula virtual de la asignatura de Data Mining

http://www.oldemarrodriguez.com/yahoo_site_admin/assets/docs/Documento_CRISP-DM.2385037.pdf

https://datos.madrid.es/FWProjects/egob/Catalogo/SectorPublico/Ficheros/PANEL_INDICADORE S_2019.pdf

<https://datos.madrid.es/portal/site/egob/menuitem.3efdb29b813ad8241e830cc2a8a409a0/?vgnextoid=55a082257a859610VgnVCM1000001d4a900aRCRD&vgnextchannel=6db862c549810510VgnVCM1000008a4a900aRCRD&vgnextfmt=default>

¡Gracias!

