

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA INFORMÁTICA

# **ANÁLISIS DE SENTIMIENTOS Y CLÚSTERING DE ERRORES EN USUARIOS DE APP INVENTOR**

AUTORA: CLAUDIA SORIANO ROLDÁN

Puerto Real, Septiembre 2023



**UCA**

Universidad  
de Cádiz

Escuela Superior  
de Ingeniería

## TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA INFORMÁTICA

# ANÁLISIS DE SENTIMIENTOS Y CLÚSTERING DE ERRORES EN USUARIOS DE APP INVENTOR

DIRECTOR: IVÁN RUIZ RUBE  
CO-DIRECTOR: RUBÉN BAENA PÉREZ  
AUTORA: CLAUDIA SORIANO ROLDÁN

Puerto Real, septiembre de 2023



# **Declaración personal de autoría**

---

Claudia Soriano Roldán con DNI 45339131J, estudiante del Grado de Ingeniería Informática en la Escuela Superior de Ingeniería de la Universidad de Cádiz, como autor de este documento académico titulado Análisis de Sentimientos y Clustering de errores en usuarios de App Inventor y presentado como Trabajo Final de Grado.

## **DECLARO QUE**

Es un trabajo original, que no copio ni utilice parte de obra alguna sin mencionar de forma clara y precisa su origen tanto en el cuerpo del texto como en su bibliografía y que no empleo datos de terceros sin la debida autorización, de acuerdo con la legislación vigente. Asimismo, declaro que soy plenamente consciente de que no respetar esta obligación podrá implicar la aplicación de sanciones académicas, sin perjuicio de otras actuaciones que pudieran iniciarse. En Puerto Real, a 4 de Septiembre de 2023.

Fdo: Claudia Soriano Roldán

**Copyright © 2023 CLAUDIA SORIANO ROLDÁN.**

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled “GNU Free Documentation License”.



# Agradecimientos

---

*En primer lugar, quiero agradecer a mi familia y en especial a mi madre por estar siempre junto a mi y apoyarme en este proceso de aprendizaje. Vuestros ánimos y vuestra confianza inquebrantable en mi me han dado las fuerzas y la motivación necesaria para continuar.*

*A mis tutores, quiero expresar mi gratitud por vuestra orientación, consejos y por vuestro compromiso y tiempo dedicado en ayudarme.*

*Finalmente a mis amigos, quiero agradecerles por estar siempre ahí, incluso en la distancia habéis sabido apoyarme y animarme en los momentos difíciles.*

*Gracias a todos de nuevo por ser parte de esta etapa, este logro no es solo mío sino que también es vuestro.*

*Mi gratitud más sincera,  
Claudia Soriano Roldán*



# Resumen

---

En este Trabajo de Fin de Grado se aborda la creación de un software que realiza un análisis de las preguntas y respuestas de usuarios sobre MIT AppInventor, consultando como fuentes el foro oficial de la herramienta (MIT App Inventor Community), y un foro externo (Stack Overflow). Para ello, se han usado técnicas de web scraping, de procesamiento de lenguaje natural (Clustering y Clasificación de texto) y de minado de datos.

Este sistema tiene como principal objetivo clasificar y exponer los problemas o dudas que encuentran los usuarios de AppInventor y presentar los resultados en forma de infografías en un cuadro de mandos web que el usuario podrá explorar con detalle.

**Palabras clave:** App Inventor, clustering, clasificación, web scraping, procesamiento de lenguaje natural, minado de datos.



# Índice general

---

Índice de figuras	xii
Índice de tablas	xiv
<b>I Prolegómeno</b>	<b>1</b>
<b>1. Introducción</b>	<b>3</b>
1.1. Motivación y contexto . . . . .	3
1.2. Alcance y Objetivos . . . . .	3
1.3. Glosario de términos . . . . .	4
1.4. Organización del documento . . . . .	5
<b>2. Estado del arte</b>	<b>9</b>
2.1. Web Scraping . . . . .	9
2.2. Procesamiento del Lenguaje Natural . . . . .	9
2.3. Clustering . . . . .	11
<b>3. Plan de gestión de proyecto</b>	<b>15</b>
3.1. Metodología de desarrollo . . . . .	15
3.2. Planificación del proyecto . . . . .	15
3.3. Costes . . . . .	17
3.4. Riesgos . . . . .	17
<b>II Desarrollo</b>	<b>21</b>
<b>4. Requisitos del Sistema</b>	<b>22</b>
4.1. Objetivos del Sistema . . . . .	22
4.2. Requisitos funcionales . . . . .	22
4.2.1. Subproyecto I: Web Scraper . . . . .	22
4.2.2. Subproyecto II: Procesado de datos . . . . .	22
4.2.3. Subproyecto III: Web . . . . .	23
4.3. Requisitos no funcionales . . . . .	25
4.4. Requisitos de información . . . . .	25
<b>5. Análisis del Sistema</b>	<b>29</b>
5.1. Modelo Conceptual . . . . .	29
5.2. Modelo de Casos de Uso . . . . .	30
5.3. Modelo de Interfaz de Usuario . . . . .	35
<b>6. Diseño del Sistema</b>	<b>41</b>
6.1. Diseño de la Arquitectura del Sistema . . . . .	41

6.2.	Diseño Físico de Datos . . . . .	43
6.2.1.	Tabla questions . . . . .	43
6.2.2.	Tabla answers . . . . .	43
6.2.3.	Tabla users . . . . .	44
6.2.4.	Tabla categories . . . . .	44
6.2.5.	Tabla tags . . . . .	45
6.2.6.	Tabla AI tags . . . . .	45
6.2.7.	Tabla AI answers . . . . .	45
6.3.	Diseño de la Interfaz de Usuario . . . . .	46
<b>7.</b>	<b>Construcción del Sistema</b>	<b>49</b>
7.1.	Entorno de Construcción . . . . .	49
7.1.1.	Python . . . . .	49
7.1.2.	PyCharm . . . . .	50
7.1.3.	OpenAI API . . . . .	50
7.1.4.	Django . . . . .	50
7.1.5.	MySQL . . . . .	50
7.1.6.	GitHub . . . . .	51
7.1.7.	StartBootstrap . . . . .	51
7.1.8.	LateX . . . . .	51
7.2.	Código Fuente . . . . .	52
7.2.1.	Subproyecto I: Web Scraper . . . . .	52
7.2.2.	Subproyecto II: Procesado de datos . . . . .	53
7.2.3.	Subproyecto III: AppInventor Dashboard . . . . .	54
<b>8.</b>	<b>Pruebas del Sistema</b>	<b>57</b>
8.1.	Estrategia . . . . .	57
8.2.	Pruebas de Sistema . . . . .	57
8.2.1.	Pruebas Funcionales . . . . .	57
8.2.2.	Pruebas No Funcionales . . . . .	59
<b>9.</b>	<b>Despliegue del Sistema</b>	<b>63</b>
9.1.	Arquitectura Física . . . . .	63
9.2.	Instrucciones de despliegue . . . . .	63
9.2.1.	Requisitos previos . . . . .	64
9.2.2.	Inventario de componentes . . . . .	65
9.2.3.	Procedimientos de instalación . . . . .	65
9.2.4.	Pruebas de implantación . . . . .	65
<b>III</b>	<b>Epílogo</b>	<b>66</b>
<b>10.</b>	<b>Conclusiones</b>	<b>67</b>
10.1.	Objetivos alcanzados . . . . .	67
10.2.	Lecciones aprendidas . . . . .	69
10.3.	Trabajo futuro . . . . .	69
<b>11.</b>	<b>Información sobre Licencia</b>	<b>71</b>

<b>Información sobre Licencia</b>	<b>71</b>
<b>GNU Free Documentation License</b>	<b>72</b>
1. APPLICABILITY AND DEFINITIONS . . . . .	72
2. VERBATIM COPYING . . . . .	74
3. COPYING IN QUANTITY . . . . .	74
4. MODIFICATIONS . . . . .	75
5. COMBINING DOCUMENTS . . . . .	76
6. COLLECTIONS OF DOCUMENTS . . . . .	77
7. AGGREGATION WITH INDEPENDENT WORKS . . . . .	77
8. TRANSLATION . . . . .	77
9. TERMINATION . . . . .	78
10. FUTURE REVISIONS OF THIS LICENSE . . . . .	78
11. RELICENSING . . . . .	78
<b>Bibliografía</b>	<b>15</b>
<b>IV Anexos</b>	<b>19</b>
<b>A. Manual del desarrollador</b>	<b>20</b>
A.1. Introducción . . . . .	20
A.2. Preparación del entorno de trabajo . . . . .	20
A.3. Consideraciones generales sobre el desarrollo . . . . .	21
A.4. Instrucciones para construcción . . . . .	22
A.5. Ejecución de los scripts de web scraping y de análisis de datos . . . . .	22
<b>B. Manual de usuario</b>	<b>25</b>
B.1. Introducción . . . . .	25
B.2. Ejecutar la recogida de los datos . . . . .	25
B.3. Ejecutar el análisis de los datos . . . . .	25
B.4. Uso de la aplicación web . . . . .	25
B.4.1. Menú de navegación . . . . .	25
B.4.2. Home . . . . .	25
B.4.3. Questions . . . . .	26
B.4.4. Answers . . . . .	26
B.4.5. Categories . . . . .	27
B.4.6. Tags . . . . .	27
B.4.7. Users . . . . .	29
B.4.8. Uso de tablas y gráficas . . . . .	29
<b>C. Extractos de código fuente</b>	<b>31</b>
C.1. Introducción . . . . .	31
C.2. Extractos de código . . . . .	31
<b>D. Test de usabilidad</b>	<b>35</b>
D.1. Introducción . . . . .	35
D.2. Resultados . . . . .	35

<b>E. Test de portabilidad</b>	<b>39</b>
E.1. Introducción . . . . .	39
E.2. Resultados . . . . .	39

# Índice de figuras

---

2.1. Algoritmos de clústering [4] . . . . .	12
3.1. Diagrama de Gantt . . . . .	20
5.1. Diagrama conceptual UML . . . . .	29
5.2. Diagrama de Casos de Uso . . . . .	30
5.3. Pantalla principal / Home . . . . .	35
5.4. Pantalla Questions . . . . .	36
5.5. Pantalla Answers . . . . .	37
5.6. Pantalla Categories . . . . .	38
5.7. Pantalla Tags . . . . .	39
5.8. Pantalla Users . . . . .	40
6.1. Diagrama de Contexto del sistema . . . . .	41
6.2. Diagrama de Contenedores . . . . .	42
6.3. Diagrama de Componentes . . . . .	42
6.4. Diagrama de la base de datos . . . . .	46
6.5. Muestra de colores . . . . .	47
6.6. Diseño pantalla Home . . . . .	48
6.7. Diseño pantalla Questions . . . . .	48
7.1. Logo de Python . . . . .	49
7.2. Logo de OpenAI . . . . .	50
7.3. Logo de Django . . . . .	50
7.4. Logo de GitHub . . . . .	51
7.5. Logo de StartBootstrap . . . . .	51
7.6. Logo de LaTeX . . . . .	51
7.7. Ejecución de get_k() . . . . .	54
8.1. Resumen de resultados del análisis SonarCloud . . . . .	60
8.2. Resumen de resultados del análisis Mantenibilidad . . . . .	61
8.3. Resumen de resultados del análisis de vulnerabilidades . . . . .	61
8.4. Resumen de resultados del análisis de seguridad . . . . .	61
9.1. Arquitectura Física . . . . .	63
10.1. El clúster descubre una posible nueva categoría para el foro: 'Localización y GPS' . . . . .	67
10.2. El clúster descubre una posible nueva categoría para el foro: 'Uso de Componente WebViewer.' . . . . .	68
10.3. El clúster descubre una posible nueva categoría para el foro: 'Problemas con Emuladores' . . . . .	68
B.1. Pantalla Questions . . . . .	26
B.2. Pantalla Home . . . . .	26

B.3. Pantalla Answers . . . . .	27
B.4. Pantalla Categories . . . . .	27
B.5. Pantalla Tags . . . . .	28
B.6. Pantalla Users . . . . .	29
C.1. Código de get_k() . . . . .	31
C.3. Código de get_embedding() . . . . .	31
C.2. Código de apply_kmeans() . . . . .	32
C.4. Código de answers_diagnosis() . . . . .	32
C.5. Código de get_tags_description() . . . . .	33
D.1. Edades de los usuarios encuestados . . . . .	35
D.2. Conocimientos informáticos de los encuestados . . . . .	35
D.3. Nivel de inglés de los encuestados . . . . .	36
D.4. Pregunta 1 . . . . .	36
D.5. Pregunta 2 . . . . .	37
D.6. Pregunta 3 . . . . .	37
D.7. Sugerencia de los usuarios . . . . .	37
E.1. Resultados portabilidad web (Escritorio) . . . . .	39
E.2. Resultados portabilidad web (Móvil) . . . . .	40

# Índice de tablas

---

3.1. Costes humanos . . . . .	17
3.2. Costes equipo hardware . . . . .	17
3.3. Costes API OpenAI . . . . .	17
3.4. Riesgo 01 - Expectativas poco realistas . . . . .	18
3.5. Riesgo 02 - Desconocimiento de las tecnologías . . . . .	18
3.6. Riesgo 03 - Cambio en el HTML de las webs de terciers . . . . .	18
3.7. Riesgo 04 - La API de la que se depende deja de funcionar . . . . .	19
3.8. Riesgo 05 - Sobrecoste presupuestario . . . . .	19
3.9. Riesgo 06 - Baja por enfermedad del personal . . . . .	19
5.1. CU-01 Consultar preguntas . . . . .	30
5.2. CU-02 Consultar información adicional de una pregunta . . . . .	31
5.3. CU-03 Consultar respuestas . . . . .	31
5.4. CU-04 Consultar información adicional de una respuesta . . . . .	31
5.5. CU-05 Consultar categorías . . . . .	32
5.6. CU-06 Consultar información adicional de una categoría . . . . .	32
5.7. CU-07 Consultar tags . . . . .	32
5.8. CU-08 Consultar información adicional de un tag . . . . .	33
5.9. CU-09 Consultar usuarios . . . . .	33
5.10. CU-10 Consultar información adicional de un usuario . . . . .	33
5.11. CU-11 Ejecutar recogida de datos . . . . .	34
5.12. CU-12 Ejecutar procesado y análisis de datos . . . . .	34





# **Parte I**

## **Prolegómeno**



# 1. Introducción

---

A continuación, se describe la motivación del presente proyecto y su alcance. También se incluye un glosario de términos y la organización del resto de la presente documentación.

## 1.1. Motivación y contexto

En las últimas décadas, debido a la fuerte digitalización de la sociedad y del mercado, muchas personas deciden acercarse al mundo de la programación, gran parte de ellas por primera vez, algunos incluso de forma autodidacta.

Por ello es ahora cuando el uso de foros dedicados a la programación se encuentra en uno de sus momentos más álgidos. La actividad en estos foros crece cada vez más y se han convertido en una fuente valiosísima de consulta y comunicación para usuarios de todos los niveles que quieren resolver sus dudas o exponer sus ideas.

Sin embargo, la información que albergan estos sitios es mucha más de la que podemos ver a simple vista. A partir de ciertos datos de los foros, como las preguntas de los usuarios y sus respuestas, podemos obtener información subyacente que arroje una nueva luz y aporten una nueva perspectiva desde la que mirar los datos iniciales.

Debido al gran volumen de datos, resulta inviable analizar cada consulta de los usuarios de manera manual e individual. Es por esto que en este trabajo se aborda el diseño de una herramienta que obtiene y analiza de manera automática todas las preguntas y respuestas de los usuarios mediante técnicas de web scraping y Procesamiento de lenguaje natural y que genera un feedback ordenado y detallado en forma de infografías en un cuadro de mandos web.

Para este proyecto se han elegido analizar las interacciones en los foros de usuarios de IDE App Inventor, ya que es usado tanto por usuarios no experimentados en programación, al tratarse de un sencillo IDE visual centrado en bloques, como por usuarios más experimentados, pues es capaz de crear aplicaciones potentes con un impacto real.

## 1.2. Alcance y Objetivos

Diseñar e implementar una herramienta software que realiza un análisis de las preguntas y respuestas de usuarios sobre MIT AppInventor, consultado como fuentes el foro oficial de la herramienta (MIT App Inventor Community), y un foro externo (Stack Overflow). Para ello, se han usado técnicas de web scraping, de procesamiento de lenguaje natural (Clustering y Clasificación de texto) y de minado de datos.

Este sistema tiene como principal objetivo clasificar y exponer los problemas o dudas que encuentran los usuarios de AppInventor y presentar los resultados en forma de infografías en un cuadro de mandos web que el usuario podrá explorar con detalle.

### 1.3. Glosario de términos

- App Inventor: App Inventor es un entorno de desarrollo de software actualmente mantenido por el Instituto de Tecnología de Massachusetts (MIT).
- Base de datos (BD): Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático.
- Clasificación: La clasificación es una técnica en el campo del aprendizaje automático y la minería de datos que se utiliza para asignar objetos o datos a categorías o etiquetas predefinidas. En esencia, se trata de construir un modelo que pueda aprender a distinguir entre diferentes clases o categorías en función de las características de los datos de entrada.
- Clustering: Es una técnica en el campo de la minería de datos y el aprendizaje automático que se utiliza para organizar un conjunto de datos en grupos o clústeres basados en la similitud entre sus elementos. El objetivo principal del clustering es identificar patrones intrínsecos en los datos y agrupar objetos que sean más similares entre sí que con los objetos en otros grupos.
- Dashboard: Herramienta de gestión de la información que monitoriza, analiza y muestra de manera visual las métricas y datos fundamentales para hacer un seguimiento del estado de una empresa, un departamento, una campaña o un proceso específico.
- Word Embedding: Nombre de un conjunto de modelos de lenguaje y técnicas de aprendizaje en procesamiento del lenguaje natural (PLN) en donde las palabras o frases del lenguaje natural son representadas como vectores de números reales.
- Minado de datos: Conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos que expliquen el comportamiento de estos datos.
- Procesamiento de lenguaje natural (NLP): Campo de las ciencias de la computación, de la inteligencia artificial y de la lingüística que estudia las interacciones entre las computadoras y el lenguaje humano.
- Web scraping: Proceso de extracción de contenidos y datos de sitios web mediante software.

## 1.4. Organización del documento

A continuación se detalla la estructura de este documento y el contenido de cada uno de sus capítulos.

- **Capítulo 1: Introducción.**

En este capítulo se describe la motivación y el contexto del presente proyecto y sus objetivos. También incluye un glosario de términos y la organización del resto de la documentación.

- **Capítulo 2: Estado del arte.**

En este capítulo se describen las técnicas y alternativas tecnológicas existentes que ayudan a resolver la problemática de este proyecto.

- **Capítulo 3: Plan de gestión de proyecto.**

En este capítulo se describen los aspectos relativos a la gestión del proyecto: metodología, organización, costes y planificación.

- **Capítulo 4: Requisitos del Sistema.**

En este capítulo se presentan los objetivos y el catálogo de requisitos del sistema.

- **Capítulo 5: Análisis del Sistema.**

Este capítulo cubre el análisis del sistema a desarrollar, presentando los casos de uso y haciendo uso del lenguaje de modelado UML.

- **Capítulo 6: Diseño del Sistema.**

En este capítulo se recoge el diseño de la arquitectura del sistema informático, el diseño físico de datos y un diseño detallado de la interfaz de usuario

- **Capítulo 7: Construcción del Sistema.**

Este capítulo trata sobre todos los aspectos relacionados con la implementación del sistema en código, haciendo uso de un determinado entorno tecnológico.

- **Capítulo 8: Pruebas del Sistema.**

En este capítulo se describen las pruebas de sistema que el mismo cumple con todos los requisitos establecidos.

- **Capítulo 9: Despliegue del Sistema.**

Este capítulo recoge la arquitectura física planteada para el sistema, las instrucciones para su despliegue y las instrucciones para la operación y mantenimiento del nivel de servicio.

- **Capítulo 10: Conclusiones.**

En este capítulo se detallan las lecciones aprendidas tras el desarrollo del proyecto y se identifican las posibles oportunidades de mejora sobre el software desarrollado.

- **Información sobre licencia.**

Información relativa a la licencia seleccionada para la documentación y software del proyecto.

- **Bibliografía.**

- **Manual de desarrollador.**

En él se recogen las instrucciones necesarias para evolucionar el software.

- **Manual de usuario.**

En él se recogen las instrucciones de uso del software. Este manual se dirige al usuario final del software objetivo de este proyecto.





## 2. Estado del arte

---

A continuación se describen las técnicas y alternativas tecnológicas existentes que ayudan a resolver la problemática de este proyecto.

### 2.1. Web Scraping

El web scraping es el proceso de extracción de los datos de una página web. Consiste en hacer una llamada HTTP al servidor del sitio web para descargar su HTML y posteriormente parsearlo y extraer los datos deseados [16].

Desde la creación de la Red Informática Mundial, los desarrolladores comenzaron a escribir scripts capaces de automatizar la extracción de información. Con el paso de los años, la importancia de internet y de las webs no ha dejado de crecer, y con ello también ha crecido el interés en el uso del web scraping.

En la última década, muchas webs comenzaron a ofrecer algunos de sus datos mediante APIs. Sin embargo, no todas las webs tienen una, no todas son gratuitas y además la cantidad de datos ofrecidos es limitada, por lo tanto no podemos depender de ellas para obtener los datos que necesitamos [8]. No podemos olvidar que vivimos en la era de la información, y que los datos que se pueden obtener mediante web scraping pueden servir para:

- Investigaciones académicas.
- Análisis del mercado y de los precios.
- Caza de tendencias.
- Marketing.
- Monitorización propia y de la competencia.

Por eso en los últimos años han surgido numerosas herramientas y librerías dedicadas al web scraping y que facilitan su uso, como por ejemplo Scrapy<sup>1</sup>, Octoparse<sup>2</sup> o BeautifulSoup<sup>3</sup>.

### 2.2. Procesamiento del Lenguaje Natural

El estudio del campo del procesamiento de lenguaje natural comenzó alrededor de 1940, tras la Segunda Guerra Mundial. En este contexto, se reconoció la importancia de la traducción e interpretación automática del lenguaje, y se deseaba crear una máquina capaz de llevar a cabo esta tarea.

---

<sup>1</sup>Scrapy - <https://scrapy.org>

<sup>2</sup>Octoparse - <https://www.octoparse.com>

<sup>3</sup>Beautiful Soup - <https://www.crummy.com/software/BeautifulSoup/>

Sin embargo, no fue un proceso fácil, ya que entre los investigadores existía gran división de ideas sobre como afrontar el problema. No fué hasta décadas más tarde cuando sus esfuerzos se unieron bajo un mismo punto de vista: el uso del empirismo y de los modelos probabilísticos [17].

Gracias a esto, en las últimas décadas, los estudios, la información y las herramientas dedicadas a este campo han aumentado en gran número y han surgido técnicas y modelos revolucionarios como:

- **RNN<sup>4</sup>[2]**: las Redes Neuronales Recurrentes son una variante de las redes neuronales prealimentadas. Las RNNs son de naturaleza recurrente ya que aplican la misma función para cada punto de entrada y su resultado depende del dato procesado de su célula anterior. Es decir, para realizar las decisiones, considera los datos de entrada actuales y los datos resultantes que ha aprendido de la entrada anterior.
- **LSTM<sup>5</sup>[2]** : son una versión modificada de las RNN con mayor facilidad para recordar datos previos. Sin embargo, hay un límite de cuánta información pueden guardar, lo que reduce la longitud de las secuencias que puede recordar a tan solo unos pocos centenares de palabras y son muy difíciles de entrenar debido a sus elevados requisitos computacionales.
- **Mecanismo de Atención[2]**: es un mecanismo que permite más paralelismo que las RNN durante el entrenamiento, lo que mejora su eficiencia y rapidez de computación.  
El transformador usa este mecanismo para codificar información acerca del contexto de una palabra determinada en un vector de palabras. El mecanismo de atención permite a la red centrarse en palabras relevantes contextualmente tanto en la entrada como en sus salidas anteriores para determinar el próximo resultado.
- **Modelo BERT<sup>6</sup>[2]**: introducido por Google, muestra el máximo potencial del Mecanismo de Atención, batiendo récords y posicionándose en lo más alto de los Benchmarks.  
El modelo produce múltiples conjuntos de representaciones de la secuencia de entrada, cada una codificando una característica diferente de la entrada. BERT se entrena usando un modelo de lenguaje profundamente bidireccional (predecir una palabra oculta en una frase, dadas el resto de palabras).  
Con el éxito de BERT, surgieron más versiones con mejoras como ALBERT, RoBERTA o DistilBERT.
- **GPT<sup>7</sup> [3]** : modelo de lenguaje desarrollado por OpenAI. Estos modelos se basan en la arquitectura de transformadores y se entrenan en un proceso de dos pasos:
  1. Pre-entrenamiento: el modelo se entrena con gran cantidad de datos de texto. El modelo aprende a predecir la siguiente palabra en una oración dadas las anteriores.

---

<sup>4</sup>RNN - Recurrent Neural Network.

<sup>5</sup>LSTM - Long short-term memory.

<sup>6</sup>BERT - Bidirectional Encoder Representations from Transformers.

<sup>7</sup>GPT - Generative Pre-trained Transformer.

2. Afinado (fine tuning): después del pre-entrenamiento, el modelo se puede ajustar a tareas más específicas, alimentándolo con datos etiquetados específicos de la tarea deseada.

Los modelos GPT han logrado colocarse a la vanguardia en tareas de diversa índole como clasificación de texto, traducción de idiomas, respuesta de preguntas, agentes conversacionales, chat bots etc.

Algunos ejemplos de la aplicación del procesamiento de lenguaje natural muy usados en la actualidad son:

- Motores de búsqueda, por ejemplo Google.
- Asistentes inteligentes, como Siri o Alexa.
- Análisis de textos, por ejemplo para el análisis de sentimientos o la clasificación.
- Chatbots altamente capaces, como el de ChatGPT.

### 2.3. Clustering

Es una técnica de análisis de datos usada para agrupar entidades basándose en sus características y similitudes.

El objetivo general del clustering es descubrir grupos o patrones ocultos en el conjunto de datos.

Hoy en día existen diferentes aproximaciones y algoritmos para resolver esta tarea, como por ejemplo:

- **Algoritmo K-Means[9]:** agrupa los objetos en K grupos. El agrupamiento se realiza minimizando la suma de las distancias entre cada objeto y el centroide de su grupo. Los centroides se ajustan y se reasignan hasta que no se mueven más. Este es uno de los algoritmos de clustering mas utilizados debido a que es fácil de implementar, es computacionalmente eficiente incluso en grandes datasets y puede adaptarse a diferentes tipos de datos de entrada y de funciones de distancia.
- **DBSCAN[9]:** es un algoritmo basado en la densidad. Los clústeres son regiones densas en el espacio de datos, separadas por regiones de menor densidad de puntos. A diferencia de Kmeans, no requiere que el usuario especifique un número de clústeres, sin embargo su función de agrupamiento depende del parámetro épsilon y los puntos mínimos, que no se pueden seleccionar por separado para cada clúster y pueden ser difíciles de determinar.
- **Fuzzy C-Means[9]:** este algoritmo funciona asignando a cada objeto una puntuación de pertenencia a cada cluster. Cuanto mas cerca esté del centro de un cluster, más puntuación tendrá. Por lo tanto nos permite realizar una partición difusa de los datos. Al igual que K-means requiere predeterminar el número de clústeres K y su ejecución es más lenta.
- **STING[9]:** se basa en el uso de cuadrículas. El espacio se divide en celdas rectangulares y estas se vuelven a dividir en más celdas resultando en una cuadrícula con diferentes niveles de resolución. La estructura en cuadrícula lo convierte en

altamente paralelizable, sin embargo los lazos de pertenencia a sus clústeres son únicamente horizontales o verticales, por lo que los lazos diagonales no se detectan.

Como se puede observar, todas las aproximaciones poseen sus ventajas y desventajas y los resultados pueden depender de la forma de los datos en el espacio, de su cantidad y de la elección de parámetros por parte del usuario.

En la figura 2.1 se muestran algunos de los algoritmos anteriormente explicados junto con sus características.

<b>Algorithm Name</b>	<b>Algorithm Type</b>	<b>Dataset Type</b>	<b>Cluster Shape</b>	<b>Time Complexity</b>
K-Means	Partitional	Numerical	Non-Convex	$O(n k d)$
Fuzzy C-Means	Partitional	Numerical	Spherical	$O(n)$
ROCK	Partitional	Numerical & Categorical	Arbitrary	$O(n^2+n m m a + n^2 \log n)$
CACTUS	Hierarchical	Categorical	Hyper Rectangular	$O(cN)$
DBSCAN	Hierarchical	Numerical	Arbitrary	$O(n \log n)$
OPTICS	Hierarchical	Numerical	Arbitrary	$O(n \log n)$
STING	Density Based	Spatial	Arbitrary	$O(k)$

Figura 2.1: Algoritmos de clústering [4]





# **3. Plan de gestión de proyecto**

---

## **3.1. Metodología de desarrollo**

Para este proyecto se ha usado una metodología de desarrollo en cascada. Las diferentes etapas han sido desarrolladas de forma diferenciada y siguiendo un orden riguroso, puesto que debido a las características y naturaleza del proyecto, una etapa posterior no puede comenzar a desarrollarse sin que la etapa anterior esté correctamente finalizada.

## **3.2. Planificación del proyecto**

A continuación se presentan las diferentes etapas en las que se ha dividido el desarrollo de este proyecto. Adicionalmente se incluye un diagrama de Gantt con la planificación de los tiempos (Figura 3.1).

1. Investigación de técnicas y herramientas para el Minado de Datos y el Procesamiento de Lenguaje Natural (NLP):
  - 1.1. Estudio de la evolución del campo NLP a través del tiempo.
  - 1.2. Estudio del estado del arte de NLP.
  - 1.3. Estudio del estado del arte del Minado de Datos.
  - 1.4. Estudio de distintas técnicas y herramientas de minado de datos.
2. Desarrollo de procesos para la extracción, transformación y carga de datos (ETL):
  - 2.1. Investigación de diferentes técnicas y herramientas de web scraping.
  - 2.2. Configuración y despliegue de entorno para el desarrollo y ejecución del proceso ETL.
  - 2.3. Desarrollo e implementación de procesos para la obtención de datos en portales web.
  - 2.4. Tratamiento y transformación de los datos obtenidos.
  - 2.5. Carga de datos en la herramienta de procesamiento de lenguaje natural y datamining.
3. Implementación de técnicas y herramientas de Datamining y NLP:
  - 3.1. Configuración y despliegue de entorno para el desarrollo y ejecución de algoritmos de NLP.
  - 3.2. Implementación y aplicación de distintas técnicas y algoritmos de NLP, y realización de benchmark.

- 3.3. Comparativa de resultados y elección de algoritmos que mejor se ajusten al problema.
  - 3.4. Configuración de herramientas de minado de datos.
  - 3.5. Aplicación de distintas técnicas y herramientas de minado de datos.
4. Análisis de resultados:
    - 4.1. Desarrollo de aplicación web para el control de la herramienta y la visualización de los resultados.
    - 4.2. Representación visual y análisis de resultados.
    - 4.3. Obtención y presentación de conclusiones.
  5. Elaboración de la documentación y entregables del proyecto.

### 3.3. Costes

El coste humano (Tabla 3.1) de este proyecto se ha calculado mediante el sueldo promedio de un desarrollador software full-stack en España. Las horas totales realizadas son una estimación.

Para el coste de los recursos materiales se ha tenido en cuenta el precio de amortización mensual del equipo en el que se ha desarrollado el proyecto (Tabla 3.2) y el precio por cada 1000 tokens de la API de OpenAI (Tabla 3.3). El resto de herramientas (IDE, librerías, etc.) se han usado en su versión gratuita por lo tanto no aplican para el cálculo.

Tabla 3.1

*Costes humanos*

Costes humanos		
Sueldo/hora	Horas	Total
20€	400	8000€

Tabla 3.2

*Costes equipo hardware*

Costes equipo hardware		
Precio mensual	Meses	Total
35.40€	5	177€

Tabla 3.3

*Costes API OpenAI*

Costes API OpenAI		
Precio 1000 tokens	Miles de tokens procesados	Total
0,024 €	100	24€

Finalmente, sumando costes humanos y materiales, el proyecto tiene un **coste total de 8201€**.

### 3.4. Riesgos

A continuación se detallan los posibles riesgos del proyecto, indicando su impacto y su probabilidad de ocurrencia, además se detalla la estrategia a seguir para reducir su impacto o evitar su ocurrencia.

Tabla 3.4

*Riesgo 01 - Expectativas poco realistas*

<b>Riesgo 01</b>	<b>Expectativas poco realistas</b>
Descripción	Las partes interesadas del equipo pueden tener expectativas poco realistas sobre lo que puede lograr el producto final.
Magnitud	Crítica
Sistema impactado	Totalidad del proyecto
Probabilidad de ocurrencia	-
Estrategia	Realizar primero una planificación cuidadosa del producto. Detallar desde un principio cuáles van a ser sus funcionalidades y objetivos necesarios de forma clara.

Tabla 3.5

*Riesgo 02 - Desconocimiento de las tecnologías*

<b>Riesgo 02</b>	<b>Desconocimiento de las tecnologías necesarias para desarrollar el proyecto</b>
Descripción	Se desconocen las tecnologías que hay que aplicar para desarrollar el proyecto. Se desconoce si existen herramientas que puedan servir de ayuda.
Magnitud	Crítica
Sistema impactado	Totalidad del proyecto
Probabilidad de ocurrencia	-
Estrategia	El desarrollador deberá tener las base de conocimiento necesaria para realizar la tarea. Se debe dedicar un tiempo al estudio y al descubrimiento de las tecnologías y tenerlo en cuenta en la planificación.

Tabla 3.6

*Riesgo 03 - Cambio en el HTML de las webs de terceros*

<b>Riesgo 03</b>	<b>Cambio en el HTML de las web de terceros</b>
Descripción	El funcionamiento de partes del producto final dependen directamente del código HTML de páginas de terceros. Si por su parte se modifica, el funcionamiento del producto se verá afectado.
Magnitud	Crítica
Sistema impactado	Recogida de datos
Probabilidad de ocurrencia	50 %
Estrategia	En este caso se deberá de revisar y modificar cuanto antes el código fuente de la recogida de datos para adaptarlo a la nueva estructura de la web.

Tabla 3.7

*Riesgo 04 - La API de la que se depende deja de funcionar*

<b>Riesgo 04</b>	<b>La API de la que se depende deja de funcionar.</b>
Descripción	El funcionamiento de partes del producto final dependen directamente del funcionamiento de una API externa. Si esta deja de funcionar, parte del procesado de los datos deja de funcionar.
Magnitud	Grave
Sistema impactado	Procesado de datos
Probabilidad de ocurrencia	60 %
Estrategia	En este caso se deberá esperar un tiempo prudencial, revisar en la documentación si la API ha sido actualizada y mandar un mensaje al soporte si el problema persiste.

Tabla 3.8

*Riesgo 05 - Sobrecoste presupuestario*

<b>Riesgo 05</b>	<b>Sobrecoste presupuestario</b>
Descripción	Debido al coste del equipamiento y el uso de herramientas de tipo pago por uso (la API de OpenAI), retrasos o imprevistos en el desarrollo pueden producir sobrecostes no considerados inicialmente.
Magnitud	Grave
Sistema impactado	Procesamiento de datos
Probabilidad de ocurrencia	75 %
Estrategia	Durante la elaboración inicial del presupuesto hay que tener en cuenta estos imprevistos y dejar un margen de precio suficiente.

Tabla 3.9

*Riesgo 06 - Baja por enfermedad del personal*

<b>Riesgo 06</b>	<b>Baja por enfermedad del personal</b>
Descripción	El desarrollador se pone enfermo y debe darse de baja unos días.
Magnitud	Medio
Sistema impactado	-
Probabilidad de ocurrencia	50 %
Estrategia	Durante la planificación del proyecto y la estimación del tiempo de entrega hay que tener en cuenta las bajas imprevistas del personal, y más aún cuando hay un único desarrollador.

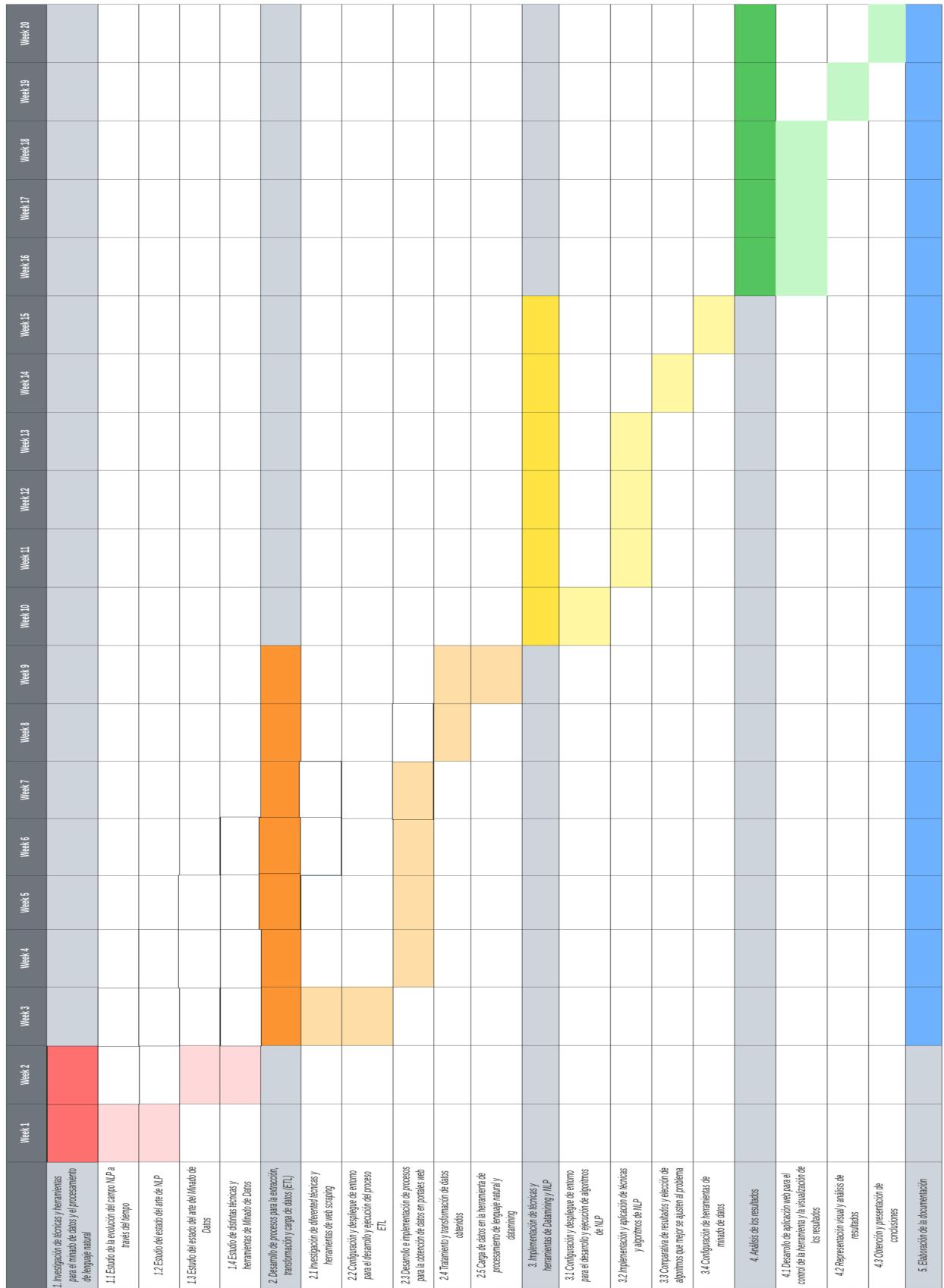


Figura 3.1: Diagrama de Gantt

# **Parte II**

## **Desarrollo**

# 4. Requisitos del Sistema

---

En este capítulo se presentan los objetivos y el catálogo de requisitos del nuevo sistema informático.

## 4.1. Objetivos del Sistema

El objetivo principal de este sistema es recoger, clasificar y exponer los problemas y dudas que encuentran los usuarios de App Inventor y presentar los resultados en un cuadro de mandos web que el usuario podrá explorar con detalle para obtener información adicional, ayudarle a realizar comparativas y a obtener recomendaciones.

## 4.2. Requisitos funcionales

A continuación se definen los requisitos funcionales de cada uno de los tres subproyectos que componen el producto final.

### 4.2.1. Subprojeto I: Web Scraper

- **RF01: Recoger los datos necesarios del foro de StackOverflow.**  
La herramienta será capaz de obtener los datos necesarios de la página web de StackOverflow y posteriormente almacenarlos en la base de datos.
- **RF02: Recoger los datos necesarios del foro MIT AppInventor Community.**  
La herramienta será capaz de obtener los datos necesarios de la página web de MIT AppInventor Community y posteriormente almacenarlos en la base de datos.

### 4.2.2. Subprojeto II: Procesado de datos

- **RF03: Clasificar las preguntas según las categorías del foro oficial de App Inventor.**  
El sistema deberá asignar una de las categorías del foro oficial a aquellas preguntas que no tengan una asignada inicialmente.
- **RF04: Asignar tags recomendados a las preguntas.**  
El sistema deberá asignar una serie de tags a aquellas preguntas que no tengan algún tag asignado inicialmente.

- **RF05: Obtener el análisis de los sentimientos de las preguntas.**  
El sistema será capaz de obtener un análisis detallado de los sentimientos de las preguntas.
- **RF06: Obtener el análisis de los sentimientos de las respuestas.**  
El sistema será capaz de obtener un análisis detallado de los sentimientos de las respuestas.
- **RF07: Recomendar respuestas.**  
El sistema será capaz de recomendar la mejor respuesta de entre todas las existentes para una pregunta.
- **RF08: Recomendar preguntas similares.**  
El sistema será capaz de recomendar una serie de preguntas similares a cada pregunta dada.

#### 4.2.3. Subproyecto III: Web

- **RF09: Consultar y filtrar listado de preguntas.**  
La herramienta será capaz de generar y mostrar un listado de todas las preguntas existentes. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario.
- **RF10: Mostrar todas las respuestas de una pregunta concreta.**  
Elegida una pregunta, la herramienta mostrará un listado de todas su respuestas.
- **RF11: Mostrar el análisis de sentimientos de una pregunta concreta.**  
Elegida una pregunta, la herramienta mostrará en una gráfica el resultado del análisis de sentimientos.
- **RF12: Mostrar la respuesta aceptada de una pregunta concreta.**  
Elegida una pregunta, la herramienta mostrará (si existe) la respuesta aceptada.
- **RF13: Mostrar la respuesta aceptada recomendada por la IA de una pregunta concreta.**  
Elegida una pregunta, si no existe respuesta aceptada, la herramienta será capaz de mostrar una respuesta recomendada.
- **RF14: Mostrar un listado de tags recomendados para una pregunta concreta.**  
Elegida una pregunta, si no tiene tags, el sistema será capaz de mostrar una serie de tags recomendados.
- **RF015: Consultar y filtrar listado de respuestas.**  
La herramienta será capaz de generar y mostrar un listado de todas las respuestas existentes. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario.
- **RF16: Mostrar la pregunta a la que contesta una respuesta concreta.**  
Elegida una respuesta, se mostrará la pregunta a la que contesta.
- **RF17: Mostrar una comparativa del análisis de sentimientos de una pregunta y de su respuesta.**

Elegida una respuesta, se mostrará una gráfica comparativa del análisis de los sentimientos de dicha respuesta y de su pregunta.

- **RF18: Consultar y filtrar listado de categorías.**

La herramienta será capaz de generar y mostrar un listado de todas las categorías existentes. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario.

- **RF19: Mostrar una comparativa del análisis de sentimientos entre todas las categorías.**

La herramienta mostrará una gráfica comparativa del análisis de sentimientos de todas las categorías existentes. Esta gráfica puede mostrar la información de todas las categorías o solo de las que el usuario seleccione.

- **RF20: Mostrar una linea temporal comparativa de todas las categorías.**

La herramienta mostrará una gráfica de líneas temporales del uso de todas las categorías existentes. Esta gráfica puede mostrar la información de todas las categorías o solo de las que el usuario seleccione. Además el rango de tiempo mostrado se puede acotar a voluntad del usuario.

- **RF21 Mostrar las preguntas más relevantes de una categoría concreta.**

Elegida una categoría, la herramienta mostrará un listado de las preguntas más relevantes de dicha categoría.

- **RF22: Mostrar los tags más usados en una categoría concreta.**

Elegida una categoría, la herramienta mostrará una gráfica con los tags más relevantes en dicha categoría.

- **RF23: Mostrar los usuarios más activos en una categoría concreta.**

Elegida una categoría, la herramienta mostrará un listado con los usuarios que más participan en dicha categoría.

- **RF24: Mostrar un mapa de calor que señala las localizaciones dónde se realizan consultas sobre una categoría concreta.**

Elegida una categoría, la herramienta mostrará un mapa de calor con las localizaciones donde se realizan las consultas sobre la categoría dada.

- **RF25: Consultar y filtrar listado de tags.**

La herramienta será capaz de generar y mostrar un listado de todos los tags existentes. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario.

- **RF26: Mostrar una comparativa del número de apariciones de un tag concreto en todas las categorías.**

Dado un tag, la herramienta será capaz de realizar una gráfica comparativa de las apariciones de dicho tag en cada una de las categorías.

- **RF27: Consultar y filtrar listado de usuarios.**

La herramienta será capaz de generar y mostrar un listado de todos los usuarios existentes. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario.

- **RF28: Mostrar una comparativa del número de preguntas de un usuario concreto en cada categoría.**

Dado un usuario, la herramienta será capaz de generar un gráfico comparativo del número de preguntas realizadas por el usuario en cada categoría.

- **RF29: Mostrar una comparativa del análisis de sentimientos de todas las preguntas y respuestas de un usuario concreto.**  
Dado un usuario, la herramienta generará un gráfico de barras comparativo del análisis de sentimientos de todas sus preguntas y de todas sus respuestas.
- **RF30: Mostrar un listado de todas las preguntas hechas por un usuario concreto.**  
Dado un usuario, la herramienta será capaz de generar y mostrar un listado de todas sus preguntas. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario.
- **RF31: Mostrar un listado de todas las respuestas hechas por un usuario concreto.**  
Dado un usuario, la herramienta será capaz de generar y mostrar un listado de todas sus respuestas. A esta lista se le podrán aplicar filtros múltiples a voluntad del usuario. .

### 4.3. Requisitos no funcionales

- **RNF01 Usabilidad.**  
La herramienta será sencilla de usar, los datos se presentaran de manera clara y ordenada. Se incluye adicionalmente un manual de usuario.
- **RNF02 Portabilidad.**  
Se podrá acceder a la herramienta de visualización a través de cualquier navegador convencional (Opera,Firefox,Safari etc.)
- **RNF04 Mantenibilidad.**  
El código de la herramienta debe de ser legible y facil de mantener para los desarrolladores encargados.
- **RNF05 Seguridad.**  
La herramienta debe asegurar que los datos introducidos y recogidos, así como los datos obtenidos después del procesado, estén protegidos y no sean alterados de manera no autorizada.

### 4.4. Requisitos de información

- **RI01 Almacenar y gestionar los datos de las preguntas.**  
El sistema debe ser capaz de obtener y almacenar los datos necesarios sobre las preguntas como pueden ser: su autor, su texto, la fecha de publicación, su número de visualizaciones o su puntuación en el foro.
- **RI02 Almacenar y gestionar los datos de las respuestas.**  
El sistema debe ser capaz de obtener y almacenar los datos necesarios sobre las respuestas. Por ejemplo: el texto, su fecha de publicación o el autor.

- **RI03 Almacenar y gestionar los datos de las categorías.**

El sistema debe ser capaz de obtener y almacenar los datos necesarios sobre las respuestas. Todas las categorías poseen un nombre y una descripción.

- **RI04 Almacenar y gestionar los datos de los usuarios.**

El sistema debe ser capaz de obtener y almacenar los datos necesarios sobre los usuarios, como por ejemplo su nombre de usuario, su perfil y su localización aproximada.

- **RI05 Almacenar y gestionar los datos de los tags.**

El sistema debe ser capaz de obtener y almacenar los datos necesarios sobre los tags. Todos los tags poseen un nombre y una descripción.





# 5. Análisis del Sistema

Este capítulo cubre el análisis del sistema de información a desarrollar, haciendo uso del lenguaje de modelado UML.

## 5.1. Modelo Conceptual

A partir de los requisitos de información, se ha desarrollado un diagrama conceptual de clases UML (5.1), que identifica las clases, atributos, relaciones y reglas de derivación necesarias.

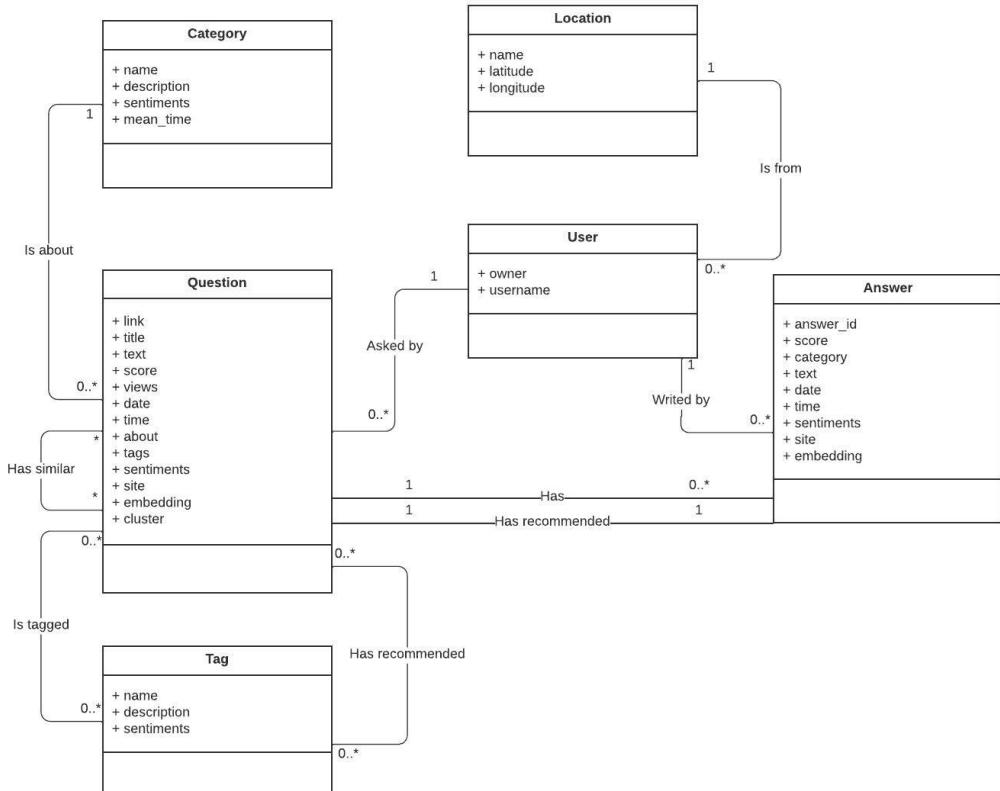


Figura 5.1: Diagrama conceptual UML

## 5.2. Modelo de Casos de Uso

A continuación se detallan los casos de uso del sistema mediante un diagrama (5.2) y su descripción.

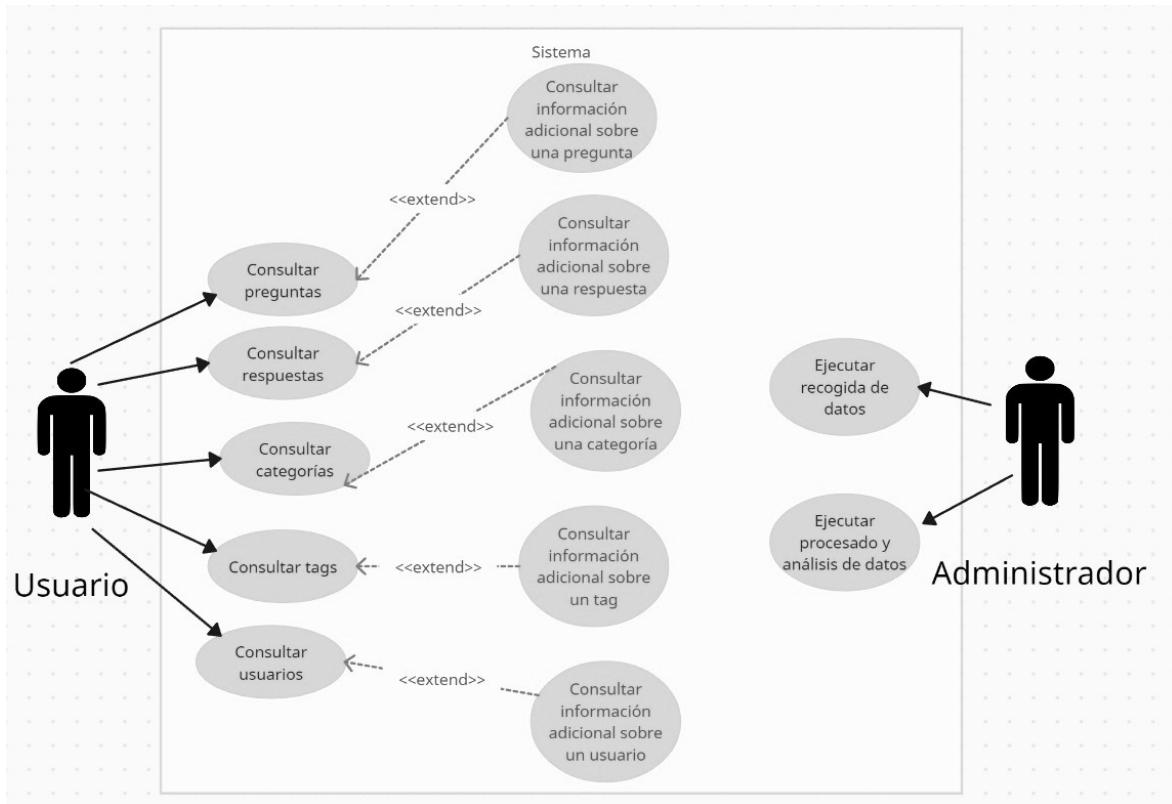


Figura 5.2: Diagrama de Casos de Uso

Tabla 5.1

*CU-01 Consultar preguntas*

CU-01	Consultar preguntas
Actor	Usuario
Precondición	El usuario ha accedido a la web
Escenario	1. El usuario pulsa "Questions & Answers." en el menú. 2. El sistema despliega un submenú. 3. El usuario pulsa "Questions." en el submenú.
Postcondición	El sistema se dirige a la sección "Questions" y muestra el listado de preguntas.
Escenario alternativo	-

Tabla 5.2

*CU-02 Consultar información adicional de una pregunta*

<b>CU-02</b>	<b>Consultar información adicional sobre una pregunta concreta</b>
Actor	Usuario
Precondición	El usuario ha accedido a la sección "Questions" de la web.
Escenario	1. El usuario selecciona una de las preguntas de la tabla.
Postcondición	El sistema muestra toda la información adicional de la pregunta seleccionada en forma de tablas y gráficas.
Escenario alternativo	-

Tabla 5.3

*CU-03 Consultar respuestas*

<b>CU-03</b>	<b>Consultar respuestas</b>
Actor	Usuario
Precondición	El usuario ha accedido a la web
Escenario	1. El usuario pulsa "Questions & Answers." en el menú. 2. El sistema despliega un submenú. 3. El usuario pulsa ".Answers." en el submenú.
Postcondición	El sistema se dirige a la sección ".Answers" muestra el listado de respuestas.
Escenario alternativo	-

Tabla 5.4

*CU-04 Consultar información adicional de una respuesta*

<b>CU-04</b>	<b>Consultar información adicional sobre una respuesta concreta</b>
Actor	Usuario
Precondición	El usuario ha accedido a la sección ".Answers" de la web.
Escenario	1. El usuario selecciona una de las respuestas de la tabla.
Postcondición	El sistema muestra toda la información adicional de la respuesta seleccionada en forma de tablas y gráficas.
Escenario alternativo	-

Tabla 5.5

*CU-05 Consultar categorías*

<b>CU-05</b>	<b>Consultar categorías</b>
Actor	Usuario
Precondición	El usuario ha accedido a la web
Escenario	1. El usuario pulsa "Categories" en el menú.
Postcondición	El sistema se dirige a la sección "Categories" muestra el listado de categorías, su análisis de sentimientos y su línea temporal.
Escenario alternativo	-

Tabla 5.6

*CU-06 Consultar información adicional de una categoría*

<b>CU-06</b>	<b>Consultar información adicional sobre una categoría</b>
Actor	Usuario
Precondición	El usuario ha accedido a la sección "Categories" de la web.
Escenario	1. El usuario selecciona una de las categorías en el desplegable.
Postcondición	El sistema muestra toda la información adicional de la categoría seleccionada en forma de tablas y gráficas.
Escenario alternativo	-

Tabla 5.7

*CU-07 Consultar tags*

<b>CU-07</b>	<b>Consultar tags</b>
Actor	Usuario
Precondición	El usuario ha accedido a la web
Escenario	1. El usuario pulsa "Tags" en el menú.
Postcondición	El sistema se dirige a la sección "Tags" muestra el listado de tags.
Escenario alternativo	-

Tabla 5.8

*CU-08 Consultar información adicional de un tag*

<b>CU-06</b>	<b>Consultar información adicional sobre un tag</b>
Actor	Usuario
Precondición	El usuario ha accedido a la sección "Tag" de la web.
Escenario	1. El usuario selecciona uno de los tags de la lista.
Postcondición	El sistema muestra toda la información adicional del tag seleccionado en forma de tablas y gráficas.
Escenario alternativo	-

Tabla 5.9

*CU-09 Consultar usuarios*

<b>CU-09</b>	<b>Consultar usuarios</b>
Actor	Usuario
Precondición	El usuario ha accedido a la web
Escenario	1. El usuario pulsa "Users" en el menú.
Postcondición	El sistema se dirige a la sección "Users" y muestra el listado de usuarios.
Escenario alternativo	-

Tabla 5.10

*CU-10 Consultar información adicional de un usuario*

<b>CU-10</b>	<b>Consultar información adicional sobre un usuario</b>
Actor	Usuario
Precondición	El usuario ha accedido a la sección "Users" de la web.
Escenario	1. El usuario selecciona uno de los usuarios de la lista.
Postcondición	El sistema muestra toda la información adicional del usuario seleccionado en forma de tablas y gráficas.
Escenario alternativo	-

Tabla 5.11

*CU-11 Ejecutar recogida de datos*

<b>CU-11</b>	<b>Ejecutar la recogida de datos</b>
Actor	Administrador
Precondición	-
Escenario	1. El administrador ejecuta el script de recogida de datos. 2. El sistema realiza el proceso de web scraping para recoger los datos.
Postcondición	Los datos recogidos quedan almacenados en la base de datos.
Escenario alternativo	-

Tabla 5.12

*CU-12 Ejecutar procesado y análisis de datos*

<b>CU-11</b>	<b>Ejecutar procesado y análisis de datos</b>
Actor	Administrador
Precondición	-
Escenario	1. El administrador ejecuta el script de procesado y análisis de datos. 2. El sistema ejecuta el procesado de los datos.
Postcondición	Los nuevos datos quedan almacenados en la base de datos.
Escenario alternativo	-

### 5.3. Modelo de Interfaz de Usuario

A continuación se muestran imágenes del prototipo de la interfaz de usuario en las figuras 5.3, 5.4, 5.5, 5.6, 5.7 y 5.8

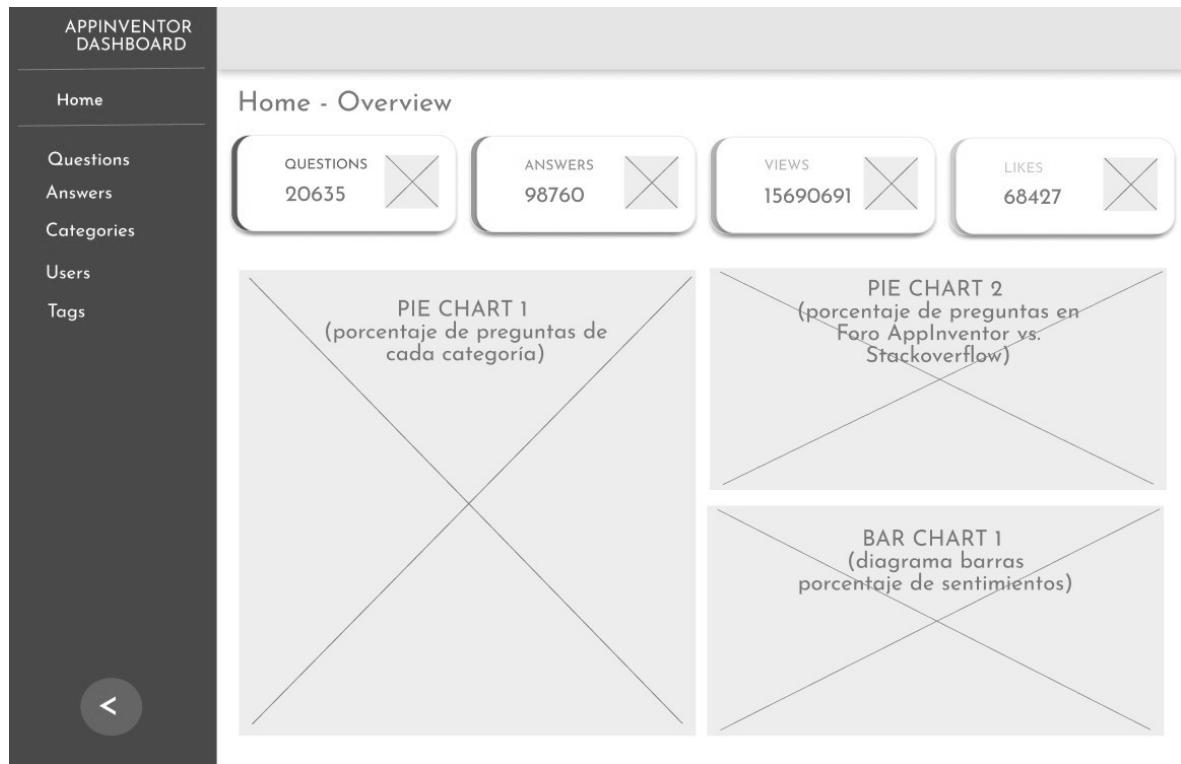


Figura 5.3: Pantalla principal / Home

**Questions**

Buscador por ID      x Filtro multiple 1      x Filtro multiple 2      x Filtro multiple 3

**DATA TABLE 1**  
(tabla preguntas)

**DATA TABLE 2**  
(tabla de respuestas de la pregunta seleccionada)

**BAR CHART 1**  
(diagrama barras porcentaje de sentimientos de la pregunta seleccionada)

**DATA TABLE 4**  
(Tabla de tags recomendados por la IA)

**ACCEPTED ANSWER**  
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed velit nunc, tincidunt id velit et, efficitur gravida ipsum. Quisque eu sodales augue, facilisis viverra nibh. Etiam euismod leo eleifend dolor blandit lobortis. Etiam ut aliquam justo, ac gravida nisl. Aliquam vitae eros a diam imperdiet vulputate. Phasellus consectetur mauris nibh.

**IA SUGESTED ANSWER**  
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed velit nunc, tincidunt id velit et, efficitur gravida ipsum. Quisque eu sodales augue, facilisis viverra nibh. Etiam euismod leo eleifend dolor blandit lobortis. Etiam ut aliquam justo, ac gravida nisl. Aliquam vitae eros a diam imperdiet vulputate. Phasellus consectetur mauris nibh.

Figura 5.4: Pantalla Questions

## Answers

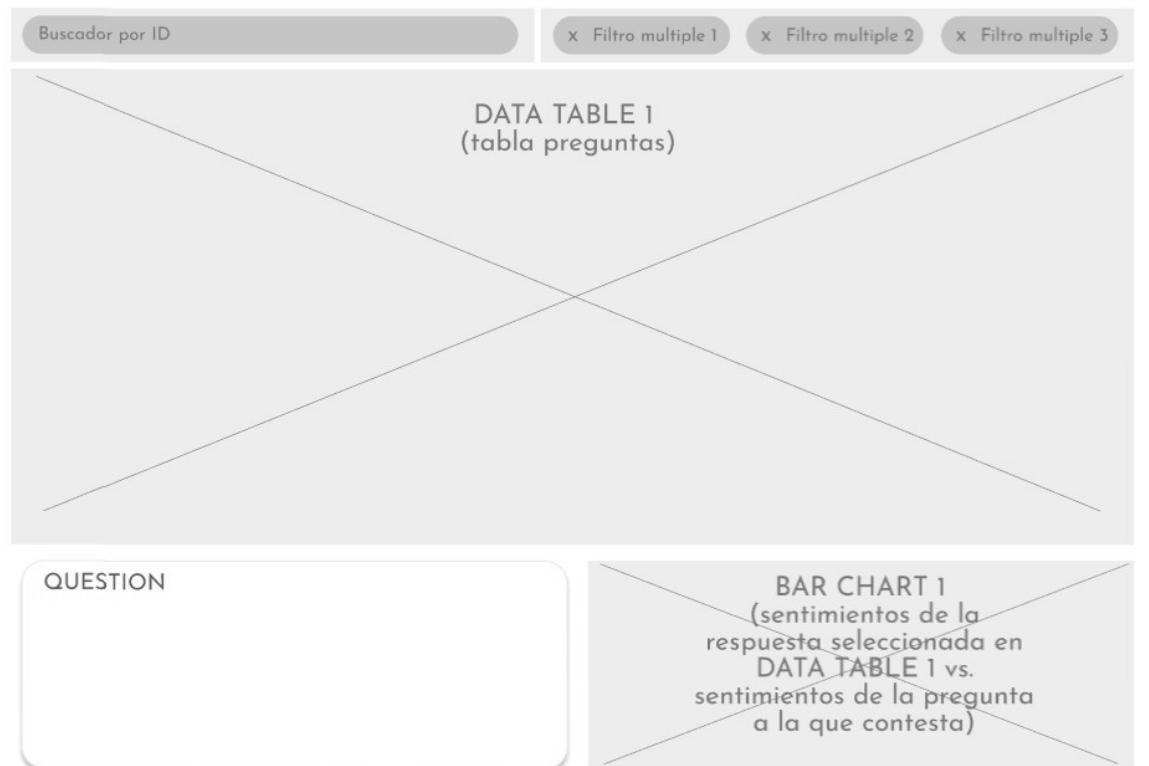


Figura 5.5: Pantalla Answers

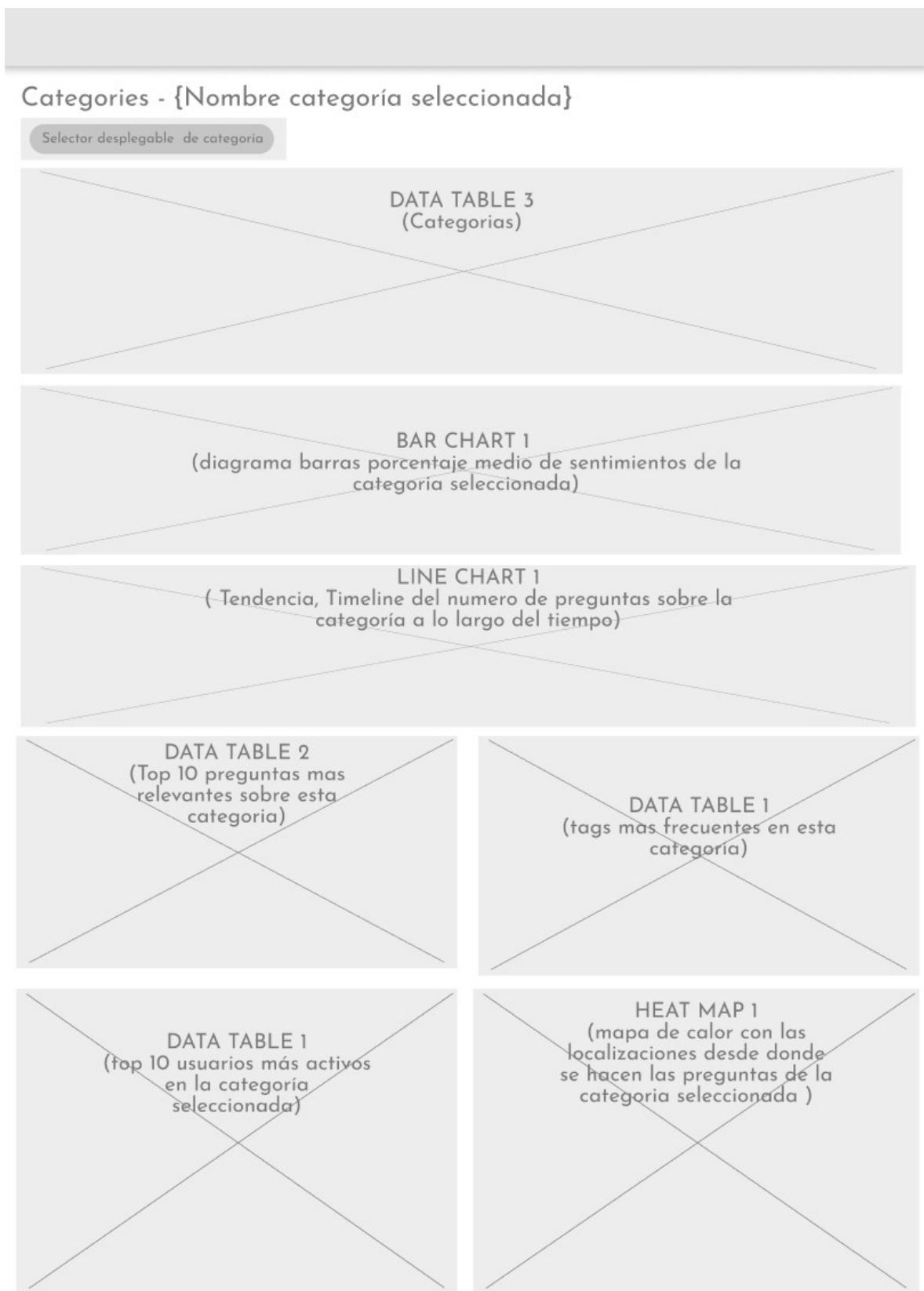


Figura 5.6: Pantalla Categories

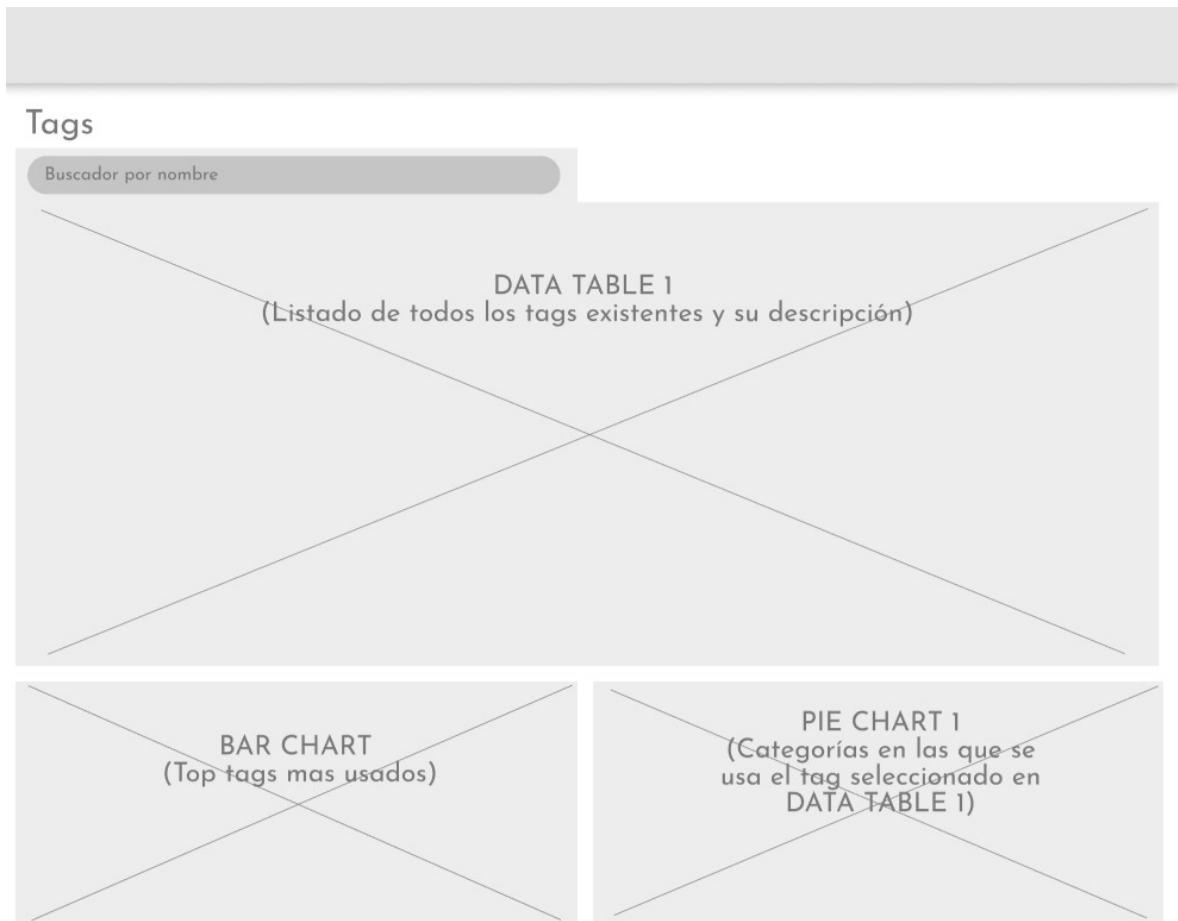
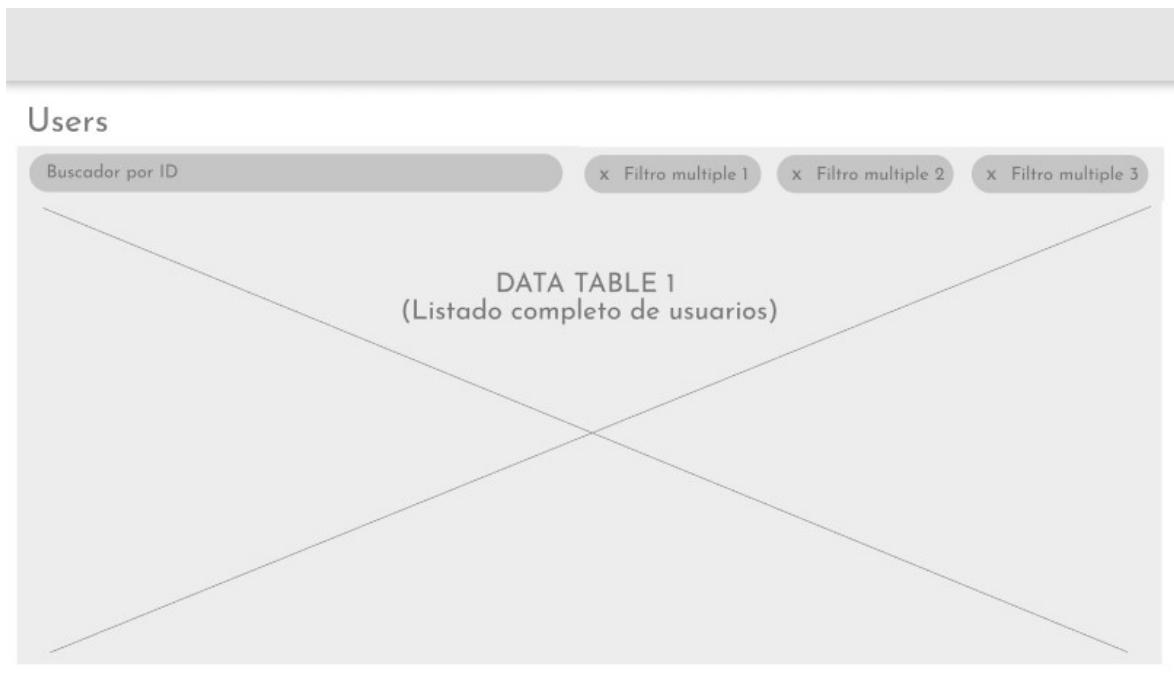


Figura 5.7: Pantalla Tags



User - {Usuario seleccionado en DATA TABLE 1}

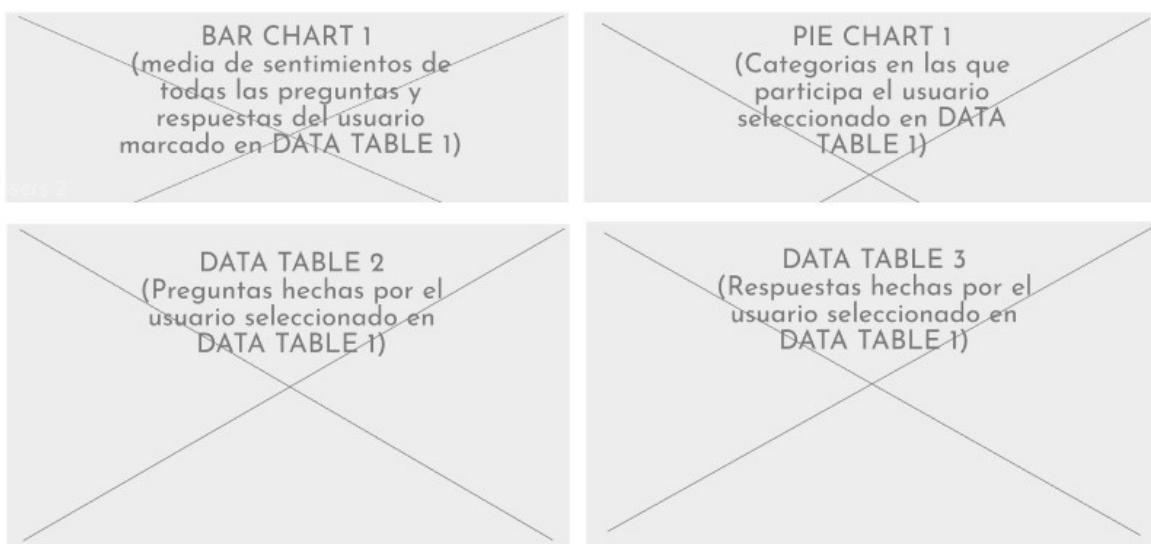


Figura 5.8: Pantalla Users

# 6. Diseño del Sistema

---

En este capítulo se recoge el diseño de la arquitectura del sistema informático, el diseño físico de datos y el diseño detallado de la interfaz de usuario.

## 6.1. Diseño de la Arquitectura del Sistema

En esta sección se define la arquitectura general del sistema, especificando la infraestructura tecnológica necesaria para dar soporte al software y la estructura de los componentes que lo forman.

Nos centraremos, en este caso, en la arquitectura lógica o de desarrollo del software. Siguiendo el modelo C4 [5] se especifican a continuación los diagramas de Contexto del sistema [6.1](#), de Contenedores [6.2](#) y de Componentes [6.3](#).

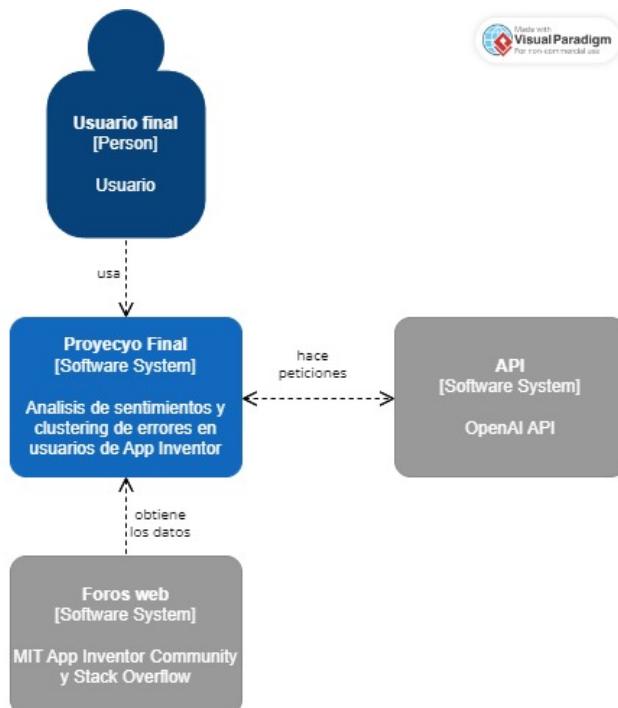


Figura 6.1: Diagrama de Contexto del sistema

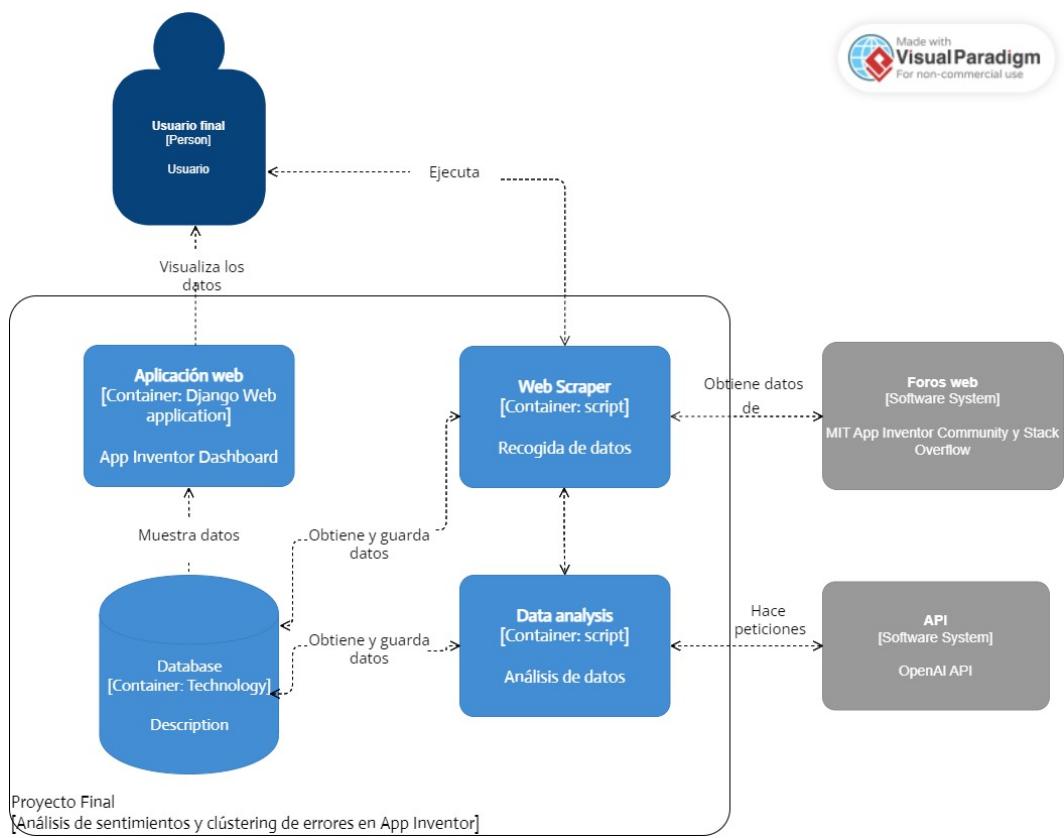


Figura 6.2: Diagrama de Contenedores

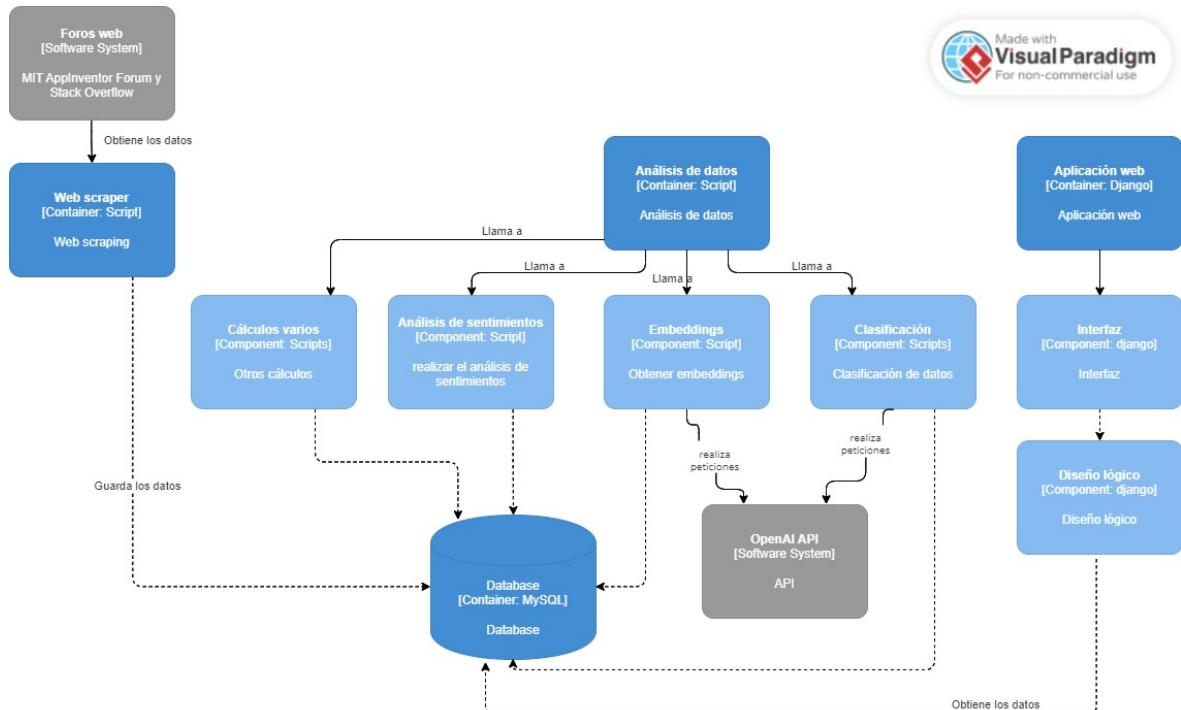


Figura 6.3: Diagrama de Componentes

## 6.2. Diseño Físico de Datos

En esta sección se define la estructura física de datos que utilizará el sistema. Se incluye a continuación una descripción de las tablas de la base de datos y más abajo un diagrama (Figura 6.4) que detalla su estructura y sus relaciones.

### 6.2.1. Tabla questions

Tabla que contiene la información referente a cada pregunta de los foros.

- *link*: clave numérica única que identifica a la pregunta y con la que se puede reconstruir el link de la pregunta.
- *title*: título de la pregunta.
- *owner*: clave numérica única que identifica al usuario creador de la pregunta y con la que se puede reconstruir el link hacia su perfil.
- *username*: nombre del usuario que crea la pregunta.
- *text*: texto de la pregunta.
- *score*: puntuación de la pregunta.
- *views*: numero de visitas que ha recibido la pregunta.
- *date*: fecha de creación de la pregunta en el foro.
- *time*: hora de creación de la pregunta en el foro.
- *about*: tema o categoría de la pregunta.
- *sentiments*: valor de cada sentimiento de la pregunta.
- *site*: valor numérico que controla si la pregunta es del foro de App Inventor o de StackOverflow.
- *embeddings*: embedding del texto de la pregunta.
- *similar*: preguntas similares.

### 6.2.2. Tabla answers

Tabla que contiene la información referente a cada respuesta de los foros.

- *question id*: clave numérica que hace referencia al link de la pregunta a la que contesta la respuesta.
- *answer id*: clave numérica única que identifica a la respuesta.
- *owner*: clave numérica que identifica al usuario que ha creado la respuesta.
- *score*: puntuación de la pregunta.

- *category*: identifica si la respuesta es una respuesta aceptada.
- *text*: texto de la pregunta.
- *date*: fecha de creación de la respuesta.
- *time*: hora de creación de la respuesta.
- *sentiments*: valor de cada sentimiento de la pregunta.
- *site*: valor numérico que controla si la respuesta es del foro de App Inventor o de StackOverflow.
- *embeddings*: embedding del texto de la pregunta.

### 6.2.3. Tabla users

Tabla que contiene la información de los usuarios.

- *owner*: clave numérica que identifica al usuario.
- *username*: nombre del usuario en el foro.
- *location*: localización del usuario.
- *latitude*: latitud de la localización.
- *longitude*: longitud de la localización.

### 6.2.4. Tabla categories

Tabla que contiene la información referente a cada categoría en la que el foro de App Inventor divide sus preguntas.

- *id*: clave numérica que identifica a la categoría.
- *name*: nombre de la categoría.
- *description*: descripción de la categoría.
- *anger*: valor del sentimiento en la categoría.
- *disgust*: valor del sentimiento en la categoría.
- *fear*: valor del sentimiento en la categoría.
- *joy*: valor del sentimiento en la categoría.
- *neutral*: valor del sentimiento en la categoría.
- *sadness*: valor del sentimiento en la categoría.
- *surprise*: valor del sentimiento en la categoría.
- *tags*: tags usados en la categoría junto a su numero de apariciones.

- *top users*: usuarios que participan en la categoría junto a su número de participaciones.
- *ans mean time*: tiempo medio de respuesta aceptada en la categoría.
- *questions*: número de preguntas en la categoría.
- *answers*: número de respuestas en la categoría.
- *likes*: número de likes de la categoría.
- *views*: número de visitas de la categoría.

### 6.2.5. Tabla tags

Tabla que contiene la información referente a cada tag existente.

- *name*: nombre del tag.
- *description*: descripción del tag.
- *anger*: valor del sentimiento en el tag.
- *disgust*: valor del sentimiento en el tag.
- *fear*: valor del sentimiento en el tag.
- *joy*: valor del sentimiento en el tag.
- *neutral*: valor del sentimiento en el tag.
- *sadness*: valor del sentimiento en el tag.
- *surprise*: valor del sentimiento en el tag.

### 6.2.6. Tabla AI tags

Tabla que contiene los tags sugeridos por la IA para cada pregunta.

- *question id*: id de la pregunta.
- *suggested tags*: tags recomendados por la IA.

### 6.2.7. Tabla AI answers

Tabla que contiene las respuestas recomendadas por la IA para cada pregunta.

- *question id*: id de la pregunta.
- *ai answer*: respuesta creada por la IA.
- *ai suggested answer*: respuesta existente recomendada por la IA.

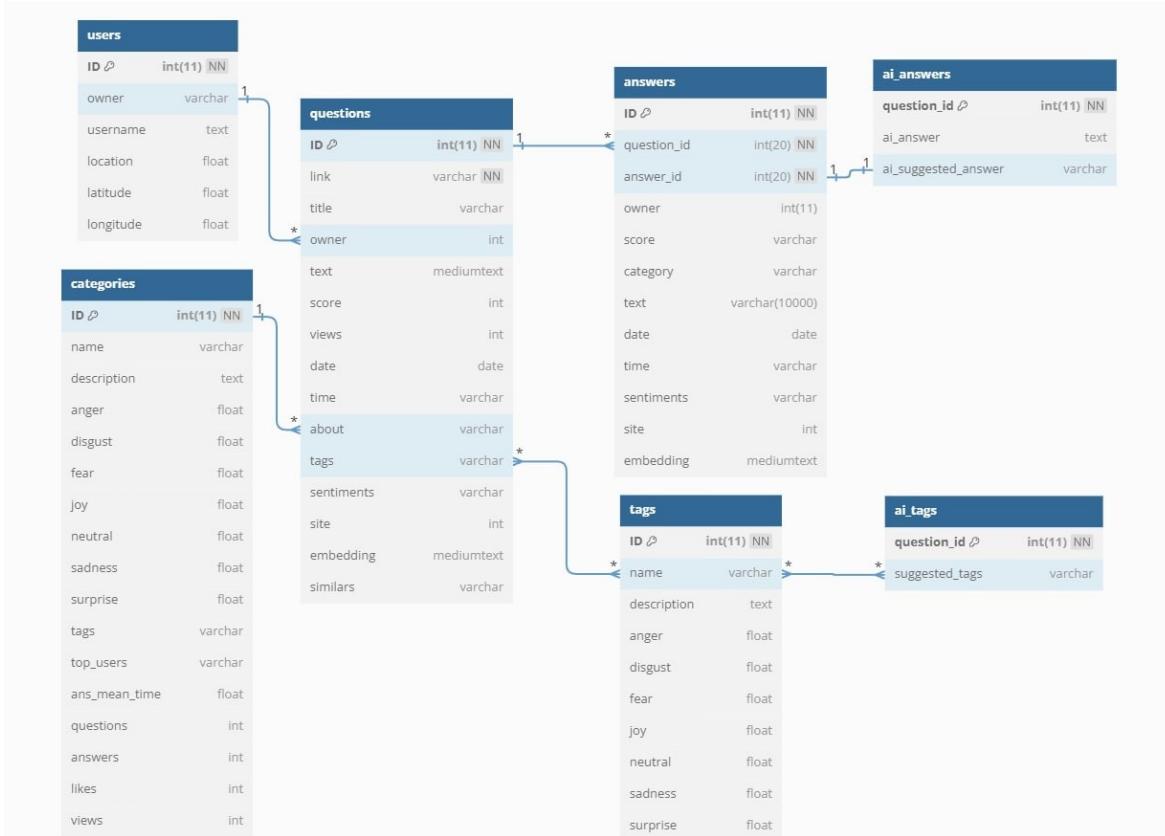


Figura 6.4: Diagrama de la base de datos

### 6.3. Diseño de la Interfaz de Usuario

A la hora de desarrollar la interfaz de usuario se han tenido en cuenta las siguientes pautas clave de diseño:

- **Sencillez, claridad y concisión:** debemos tener siempre como objetivo crear una interfaz sencilla que pueda ser utilizada de manera intuitiva. Por lo tanto la interfaz de usuario está dividida en secciones con un nombre y una función muy claros (Ejemplos: Questions, Answers, Categories, Tags) que evitan cualquier tipo de ambigüedad.

A pesar de la gran cantidad de datos mostrados en cada sección, se ha asegurado que la cantidad de gráficas y cuadros mostrados en cada sección no resulten excesivos y que haya un espaciado suficiente entre ellas para no provocar cansancio visual al usuario.

- **Familiaridad y contextualidad:** todas las funciones referentes a cada objeto están situadas en la misma sección y todas las herramientas adicionales de ayuda (filtros de tablas, botones de descarga de las gráficas etc) están integrados en los elementos a los que afectan de modo que sean fáciles de encontrar. Además el diseño de los iconos representa objetos conocidos de la vida real que el usuario reconoce fácilmente, por ejemplo la función de descargar una gráfica como imagen se representa con un ícono de una cámara de fotos, los likes recibidos se representan mediante un ícono de un corazón etc.

- **Interactividad o “feedback”:** la interfaz del usuario es reactiva de modo que simplemente pasando el ratón sobre las tablas o gráficas el sistema reacciona mostrando información mas detallada, consiguiendo que la página se sienta ”viva” y dinámica.
- **Posición y colores:** se ha tenido muy en cuenta el lugar clave para situar los botones y los enlaces. La web cuenta con un menú de navegación lateral colapsable que siempre está visible para el usuario, desde el se pueden acceder a todas las secciones de la web. La página web también cuenta con un botón de ”volver hacia arriba” que aparece cuando el usuario scrolla verticalmente hacia abajo para que no tenga que volver manualmente a la parte superior de la página.  
El color de acento principal elegido para la aplicación ha sido el azul #636efa, para el fondo se ha usado el blanco #ffffff y para los textos se ha elegido el gris oscuro #5c5e65.

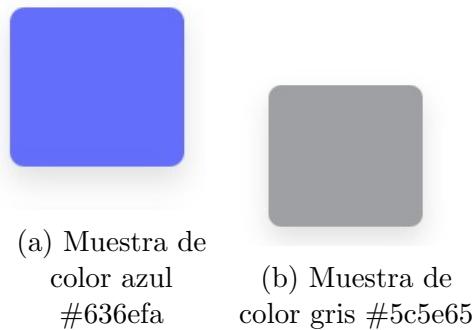


Figura 6.5: Muestra de colores

Las secciones en las que se divide la web son:

- Home: Página inicial. Muestra la información más general.
- Questions: Página que muestra toda la información relacionada con las preguntas.
- Answers: Página que muestra toda la información relacionada con las respuestas.
- Categories: Página que muestra toda la información relacionada con las categorías.
- Users: Página que muestra toda la información relacionada con los usuarios.
- Tags: Página que muestra toda la información relacionada con los tags.

A continuación se muestran el diseño del menú de navegación y algunas de las pantallas de la web. (Figuras 6.6, 6.7)



Figura 6.6: Diseño pantalla Home

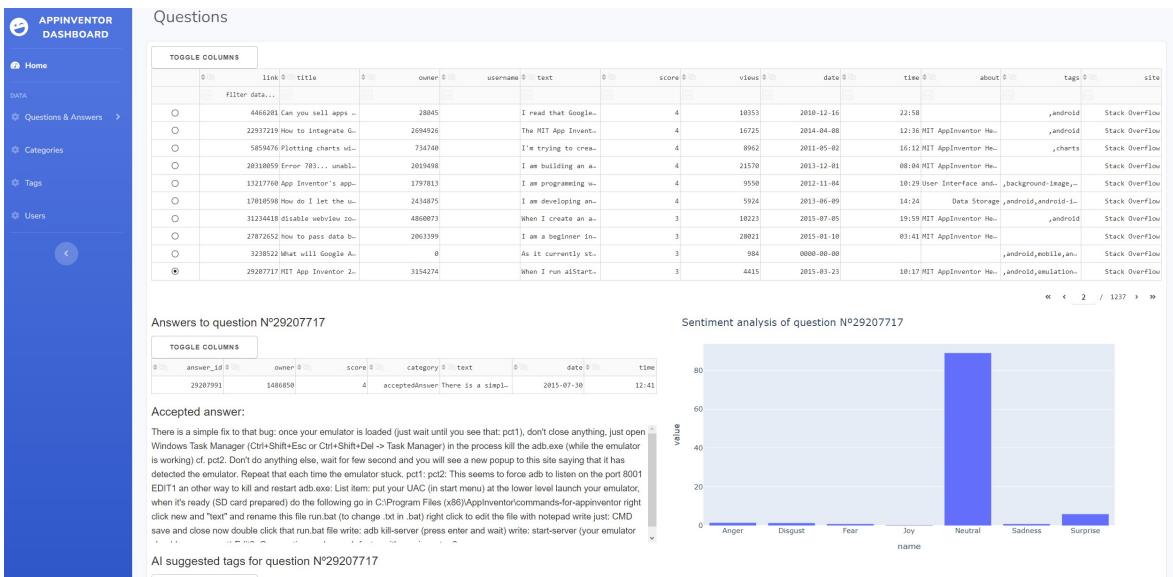


Figura 6.7: Diseño pantalla Questions

# 7. Construcción del Sistema

---

Este capítulo trata todos los aspectos relacionados con la implementación del sistema en código, haciendo uso de un determinado entorno tecnológico.

## 7.1. Entorno de Construcción

En esta sección se indica el marco tecnológico utilizado para la construcción del sistema: entorno de desarrollo, lenguaje de programación, herramientas de ayuda a la construcción y control de versiones.

### 7.1.1. Python

Python<sup>1</sup> es un lenguaje de programación interpretado muy utilizado en el desarrollo de aplicaciones web, software y en el análisis de datos.

Además debido a la gran popularidad de este lenguaje, existen gran cantidad de frameworks y librerías adicionales. Para la realización del web scraping se ha utilizado la librería Beautiful Soup para parsear el código HTML de los foros web.

Adicionalmente debido a las características de uno de los foros web y de su código HTML se ha utilizado Selenium Web Driver<sup>2</sup> para automatizar el scrolleado vertical y obligar a cargar el código. Para el procesado y el análisis de los datos se han usado las librerías TensorFlow<sup>3</sup>, Pandas<sup>4</sup> y NumPy<sup>5</sup>.

Para la representación de los datos y la creación de gráficas se ha usado la librería Plotly<sup>6</sup> y NumPy<sup>7</sup>.



Figura 7.1: Logo de Python

---

<sup>1</sup>Python - <https://www.python.org/downloads/>

<sup>2</sup>Selenium Web Driver - <https://www.selenium.dev/documentation/webdriver/>

<sup>3</sup>TensorFlow - <https://www.tensorflow.org/guide?hl=es-419>

<sup>4</sup>Pandas - <https://pandas.pydata.org>

<sup>5</sup>NumPy - <https://numpy.org>

<sup>6</sup>Plotly - <https://dash.plotly.com/>

<sup>7</sup>NumPy - <https://numpy.org>

### **7.1.2. PyCharm**

Este software ha sido programado en el entorno de desarrollo PyCharm<sup>8</sup>. Es un entorno de desarrollo integrado (IDE) utilizado en programación informática, concretamente para el lenguaje de programación Python. Está desarrollado por la empresa JetBrains. Se ha elegido debido a sus características y a su facilidad de uso. Además, gracias a la Universidad de Cádiz esta herramienta es gratuita para sus estudiantes.

### **7.1.3. OpenAI API**

OpenAI es la compañía desarrolladora de ChatGPT, y su API<sup>9</sup> proporciona una forma de integrar las capacidades de procesamiento del lenguaje natural de ChatGPT en sus aplicaciones. En este proyecto se ha utilizado para la obtención de los embeddings y para la clasificación del texto.



Figura 7.2: Logo de OpenAI

### **7.1.4. Django**

Para la creación del cuadro de mandos web de este proyecto se ha usado Django.<sup>10</sup> Es un framework para el desarrollo de webs en Python. Se ha elegido por ser uno de los frameworks más extendidos, rápidos, seguros y escalables.



Figura 7.3: Logo de Django

### **7.1.5. MySQL**

Para el almacenaje de los datos de este proyecto se ha utilizado MySQL.<sup>11</sup> Es un sistema de gestión de bases de datos relacional y es uno de los más utilizados debido a que es sencillo, potente y de código abierto.

---

<sup>8</sup>PyCharm - <https://www.jetbrains.com/es-es/pycharm/>

<sup>9</sup>OpenAI API - <https://openai.com/blog/openai-api>

<sup>10</sup>Django - <https://www.djangoproject.com>

<sup>11</sup>MySQL - <https://www.mysql.com>

### **7.1.6. GitHub**

Para guardar los ficheros de código del proyecto y para su control de versiones se ha usado GitHub.



Figura 7.4: Logo de GitHub

### **7.1.7. StartBootstrap**

Conjunto de plantillas y recursos gratuitos de diseño web que facilitan la creación de sitios web modernos y atractivos. Está diseñado para ayudar a los desarrolladores y diseñadores a comenzar rápidamente con el desarrollo de sitios web al proporcionarles una base sólida de diseño y funcionalidad.



Figura 7.5: Logo de StartBootstrap

### **7.1.8. LateX**

La documentación y la memoria de este proyecto han sido generados en LateX al resultar una herramienta idónea para crear documentos técnicos.



Figura 7.6: Logo de LaTeX

## 7.2. Código Fuente

Este proyecto está dividido en tres subproyectos bien diferenciados en contenido y utilidad.

El primer subproyecto se encarga de la recogida de datos, el segundo se encarga de analizar y transformar estos datos para obtener información adicional, una vez guardados en la base de datos, el tercer subproyecto (aplicación web) se encarga de mostrar la información en forma de infográficas y tablas.

La totalidad del código fuente de este proyecto está disponible para su descarga y visualización en el siguiente repositorio: <https://github.com/clausoriano/Clustering-y-clasificacion-de-errores-en-usuarios-de-AppInventor>

### 7.2.1. Subproyecto I: Web Scraper

Se encarga de realizar el web scraping de los foros de StackOverflow y de MIT App Inventor Community, de esta manera realiza la recogida y la carga inicial de los datos en la base de datos.

La recogida de datos de Stack Overflow se realiza mediante el uso de BeautifulSoup. Para la recogida de datos del foro MIT App Inventor Community se utiliza adicionalmente Selenium Web Driver para cargar el HTML de la página. A continuación se explican cada una de sus funciones:

- *obtain\_questions\_so()*: obtener todos los links de las preguntas de Stack Overflow. Se ha utilizado la librería BeautifulSoup para parsear el código HTML de la web.
- *obtain\_questions\_forum()*: obtener los links de las preguntas de MIT AppInventor Community, adicionalmente también se obtienen otros datos útiles en formato JSON.
- *visit\_links\_so()*: visitar todos los links y obtener los datos de las preguntas y respuestas de Stack Overflow
- *visit\_links\_forum()*: visitar todos los links y obtener los datos de las preguntas y respuestas de MIT AppInventor Community. Debido a las características de la web y de su código HTML se ha utilizado Selenium Web Driver para automatizar el scrollado vertical y obligar a cargar el código HTML.
- *obtain\_owners()*: función auxiliar encargada de obtener los datos de los usuarios.
- *obtain\_text()*: función auxiliar encargada de obtener el texto de las preguntas.
- *obtain\_answers()*: función auxiliar encargada de obtener las respuestas.
- *obtain\_tags()*: función auxiliar encargada de obtener los tags de las preguntas.

### 7.2.2. Subproyecto II: Procesado de datos

Se encarga del procesado de los datos anteriormente recogidos por el web scraper. Al procesar estos datos, se obtienen datos adicionales que también serán guardados en la Base de datos.

A continuación se explica el contenido y la función de cada uno de sus ficheros:

- **cat\_classification\_gpt\_api.py:** fichero encargado de realizar la clasificación por categorías de aquellas preguntas que no tienen una asignada. Para la clasificación se utiliza el modelo gpt-3.5-turbo de OpenAI [13].
- **sentiment\_analysis.py:** fichero encargado de realizar el análisis de sentimientos de las preguntas, respuestas, usuarios y categorías. El modelo de clasificación de sentimientos seleccionado para este proyecto ha sido Emotion English DistilRoBERTa-base [7], una versión afinada del modelo DistilBERT que predice las 6 emociones básicas de Ekman [18] además de la emoción neutra.  
Puede visualizar un extracto de código de la función de análisis de sentimientos en el anexo C.4.
- **tags.py:** fichero dedicado a gestionar los tags y a realizar la clasificación y la recomendación automática de tags para las preguntas. Debido a que los tags no tienen una descripción en el foro, debemos generar una para cada tag para posteriormente poder realizar la clasificación. Para ambas tareas se ha utilizado también el modelo gpt-3.5-turbo de OpenAI [13].  
Puede ver un ejemplo de llamada al modelo en el anexo C.5.
- **location\_gpt\_api.py:** fichero encargado de proporcionar coordenadas para su posterior representación en un mapa.
- **embeddings\_gpt\_api.py:** fichero dedicado a proporcionar el embedding del texto de las preguntas y respuestas. Para realizar los embeddings se ha usado el modelo text-embedding-ada-002 [12] de la API de OpenAI. Estos embeddings son necesarios para después poder realizar el clustering.  
Puede ver un extracto de código con la llamada al modelo de embedding en el anexo C.3
- **clustering.py:** fichero que se encarga de realizar el clustering de las preguntas. Una vez obtenidos los resultados, estos se usan para crear un algoritmo recomendador de preguntas similares.

Cabe destacar que esta fichero contiene una función auxiliar que ayuda a obtener la K para la aplicación del algoritmo KMeans. Puede verla en C.1.

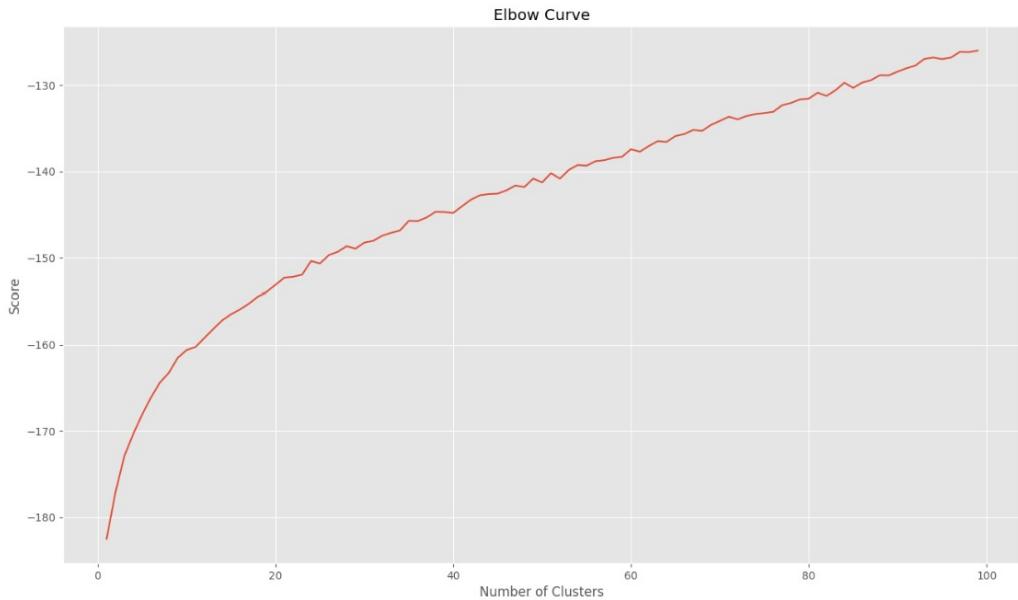


Figura 7.7: Ejecución de get\_k()

Cómo se puede observar en la gráfica resultante 7.7, a partir de  $K = 20$  la pendiente comienza a suavizarse. Así que se ejecuta K-means con valores  $K$  cercanos o iguales a 20 y se comparan resultados para finalmente elegir la  $K$  que se considere óptima. Puede ver la función de aplicación de k-medios en el anexo C.2.

El algoritmo recomendador de preguntas recoge las preguntas del mismo clúster y calcula su similitud mediante la similitud del coseno (7.1). Se ha elegido esta forma de calcular la distancia ya que es la fórmula recomendada para los embeddings generados por ada.

$$\cos(\theta) = \frac{A \cdot B}{\|A\|_2 \|B\|_2} \quad (7.1)$$

### 7.2.3. Subproyecto III: AppInventor Dashboard

Se encarga de mostrar los datos en una web en forma de tablas e infografías legibles para el usuario. Debido a que se ha usado Django para la creación de la web, la organización de los directorios y ficheros de este subproyecto atiende a las directivas del framework.

De esta forma, la web está compuesta por 6 'apps' que corresponden con cada parte de la web:

- *home*: página principal .
- *questions*: página de preguntas.
- *answers*: página de respuestas.
- *categories*: página de categorías.
- *users*: página de usuarios.

- *tags*: página de tags.

A su vez, cada una de estos directorios o 'apps' contienen los siguientes ficheros siguiendo el modelo del framework:

- *dash\_apps*: fichero adicional con el código para generar las gráficas.
- *templates*: fichero con archivos html de la app.
- *admin.py*: archivo que registra los modelos.
- *apps.py*: archivo que contiene configuración adicional de la aplicación.
- *models.py*: archivo de modelos.
- *urls.py*: archivo que gestiona las urls de la app.
- *views.py*: archivo con la lógica de la aplicación.



# 8. Pruebas del Sistema

---

En este capítulo se presenta el plan de pruebas del sistema, ya sean manuales (mediante listas de comprobación) o automatizadas mediante algún software específico de pruebas.

## 8.1. Estrategia

El entorno en el que se han realizado las pruebas tiene las mismas características que el entorno de ejecución final. Para asegurar el correcto funcionamiento de la obtención y el análisis de los datos se han hecho comprobaciones sobre la integridad tanto de los datos antes y después de procesarlos.

Para las pruebas de la aplicación web se ha elaborado un listado o guión de comprobaciones a realizar para asegurar que la navegación y los resultados obtenidos son los esperados.

## 8.2. Pruebas de Sistema

Aquí se describen las pruebas de sistema realizadas, de modo que se asegure que el sistema cumple con todos los requisitos establecidos: funcionales, de almacenamiento y no funcionales.

### 8.2.1. Pruebas Funcionales

Para el proceso de recogida de datos:

1. Se preselecciona una serie de links de los foros con valores 'normales' y valores atípicos (ej: pregunta eliminada, usuario borrado, link inexistente etc).
2. Ejecución del script de web scraping sobre unos links de Stack Overflow preseleccionados y posterior comprobación de que los datos obtenidos coinciden con los contenidos en los links.
3. Ejecución del script de web scraping sobre unos links de MIT App Inventor Community preseleccionados y posterior comprobación de que los datos obtenidos coinciden con los contenidos en los links.
4. Comprobación de que estos datos se han guardado correctamente en la BD mediante una query y una visualización de los datos.

Para el procesado de los datos se ha elegido un conjunto de datos representativos que se han utilizado para las siguientes pruebas:

1. Embedding: se comprueba que todos los vectores resultantes de aplicar el embedding tienen el mismo formato y las mismas dimensiones.
2. Clasificación de los datos en categorías: se ejecuta la clasificación por categorías sobre el conjunto de datos, una vez obtenidos, y teniendo en cuenta la descripción oficial del foro de cada categoría se comprueba uno a uno que la categoría asignada por el algoritmo tiene sentido.
3. Recomendación de preguntas similares (clustering): se ejecuta sobre el conjunto de pruebas el algoritmo de K-Means con diferentes valores de K de forma experimental, para dar con aquella que consideremos más óptima.
4. Sentiment Analysis: se ejecuta el análisis de sentimientos sobre el conjunto de pruebas, se comprueba visualmente que los valores devueltos tienen sentido y se comprueba que se guardan en el formato elegido en la base de datos.
5. Respuesta recomendada por la IA: se ejecuta el algoritmo de recomendación de respuesta sobre los datos seleccionados y se comprueba individualmente si cada resultado generado tiene sentido.
6. Tags recomendados por la IA: se ejecuta el algoritmo de recomendación de tags y se comprueba que los resultados generados tienen sentido.

En la aplicación web se han realizado estas pruebas:

1. Pruebas de menú de navegación: se hace click en cada sección del menú y se comprueba que lleva a la sección indicada por el botón.
2. Consultar listado de preguntas: se hace click en preguntas y se comprueba que la longitud de el listado de preguntas mostrado en la web coincide con la de la tabla 'questions' de la BD. Adicionalmente se comprueba el funcionamiento de los filtros introduciendo valores en cada uno de ellos.
3. Consultar información adicional sobre una pregunta concreta: se elige una pregunta del listado de preguntas y se comprueba que se muestra el listado de sus respuestas, una gráfica del análisis de sentimientos, la respuesta aceptada o recomendada por la IA según proceda y el listado de sus tags recomendados. Se contrasta la información obtenida en las gráficas con la información obtenida de la base de datos para asegurar que las gráficas están representando los datos correctamente.
4. Consultar listado de respuestas: se hace click en respuestas y se comprueba que la longitud de el listado de respuestas mostrado en la web coincide con la de la tabla 'answers' de la BD. Adicionalmente se comprueba el funcionamiento de los filtros introduciendo valores en cada uno de ellos.
5. Consultar información adicional sobre una respuesta concreta: se elige una respuesta del listado de respuestas y se comprueba que se muestra también la pregunta a la que contesta, y una gráfica comparativa de los sentimientos de la respuesta con los de su pregunta. Se contrasta la información obtenida en la gráfica con la información obtenida de la base de datos para asegurar que se están representando los datos correctamente.

6. Consultar listado de categorías: se hace click en 'categories' y se comprueba que la longitud de el listado de categorías mostrado en la web coincide con la de la tabla 'answers' de la BD y con las existentes en el foro de MIT App Inventor Community. Adicionalmente se comprueba el funcionamiento de los filtros introduciendo valores en cada uno de ellos.
7. Consultar información adicional sobre una categoría: se selecciona una categoría en el desplegable y se comprueba que la web muestra un listado de las preguntas más relevantes en esa categoría y otro de los usuarios más activos, una gráfica con los tags más utilizados en ella y un mapa de calor. Se contrasta la información obtenida en los gráficos con la información obtenida de la base de datos para asegurar que se están representando los datos correctamente.
8. Consultar listado de tags: se hace click en tags y se comprueba que la longitud de el listado de tags mostrado en la web coincide con la de la tabla 'tags' de la BD. Adicionalmente se comprueba el funcionamiento de los filtros introduciendo valores en cada uno de ellos.
9. Consultar información adicional sobre un tag: se selecciona un tag en el listado y se comprueba que el sistema muestra una comparativa del número de apariciones del tag en cada categoría.
10. Consultar listado de usuarios: se hace click en 'users' y se comprueba que la longitud de el listado de usuarios mostrado en la web coincide con la de la tabla 'users' de la BD. Adicionalmente se comprueba el funcionamiento de los filtros introduciendo valores en cada uno de ellos.
11. Consultar información adicional sobre un usuario: se selecciona un usuario del listado mostrado y se comprueba que se obtiene una gráfica con el análisis de sus sentimientos, un listado de todas las preguntas y otro de sus respuestas. Se realizan querys equivalentes a la base de datos para comprobar que los datos representados son los correctos

### **8.2.2. Pruebas No Funcionales**

#### **Usabilidad**

La aplicación web ha sido probada por usuarios de edades diferentes y con distintos niveles de conocimientos tecnológicos a los que se les ha pedido que realicen una serie de tareas:

1. Consultar listado de preguntas.
2. Obtener más información sobre una pregunta cualquiera.
3. Dado un nombre de usuario, encontrar todas las preguntas que ha realizado este usuario.
4. Consultar listado de respuestas.
5. Obtener más información sobre una respuesta cualquiera.

6. Dado un nombre de usuario, encontrar todas las respuestas que ha realizado este usuario.
7. Consultar listado de categorías.
8. Filtrar la gráfica de sentimientos para una única categoría y descargarla.
9. Dado el nombre de un tag, obtener su descripción y descargar una gráfica sobre su uso en cada categoría.

Después de usar la aplicación, los usuarios contestaron un test anónimo para calificar la usabilidad del sitio. Los resultados del test de usabilidad pueden verse en el anexo [D.1](#).

## Portabilidad

Se han abierto pestañas con la herramienta web en los navegadores Opera, Firefox y Safari para comprobar que funciona correctamente.

Adicionalmente en el anexo [E.1](#) se incluye el informe resultante realizado por la herramienta de evaluación de webs LightHouse<sup>1</sup>.

## Mantenibilidad

Para hacer una valoración de la mantenibilidad del código se ha utilizado la herramienta SonarCloud<sup>2</sup>. Con esta herramienta se han ejecutado análisis directamente sobre el repositorio de GitHub.

Tras analizar el código se genera un informe (Figura 8.1) con problemas de mantenibilidad a solucionar (violación de convenciones en el nombrado de variables, código redundante, comentarios innecesarios etc.) y una calificación en base a la deuda técnica <sup>3</sup>. En el caso de este proyecto el informe final ha otorgado la calificación más alta (A) (Figura 8.2) con una valoración de deuda técnica de 3h, que representa un ratio de deuda técnica del 0.05 %.



Figura 8.1: Resumen de resultados del análisis SonarCloud

<sup>1</sup>LighthHouse - <https://chrome.google.com/webstore/detail/lighthouse>

<sup>2</sup>SonarCloud - <https://sonarcloud.io>

<sup>3</sup>Deuda técnica - Costo del trabajo adicional causado por la elección de la solución más rápida en lugar de la más efectiva.

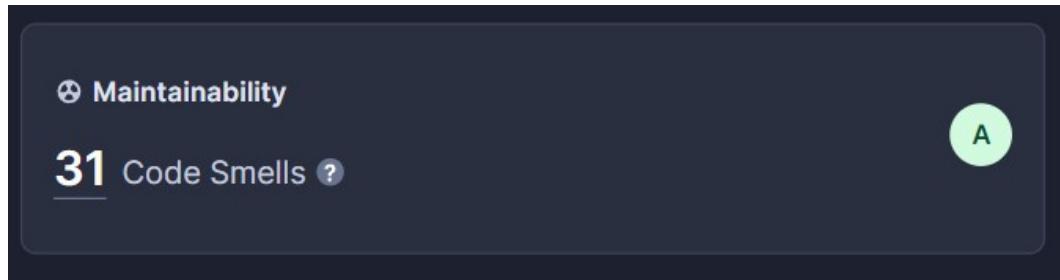


Figura 8.2: Resumen de resultados del análisis Mantenibilidad

## Seguridad

La seguridad del código también ha sido probada con SonarCloud. El análisis ha otorgado la calificación más alta (A) en materias de seguridad (Figuras 8.3 y 8.4).

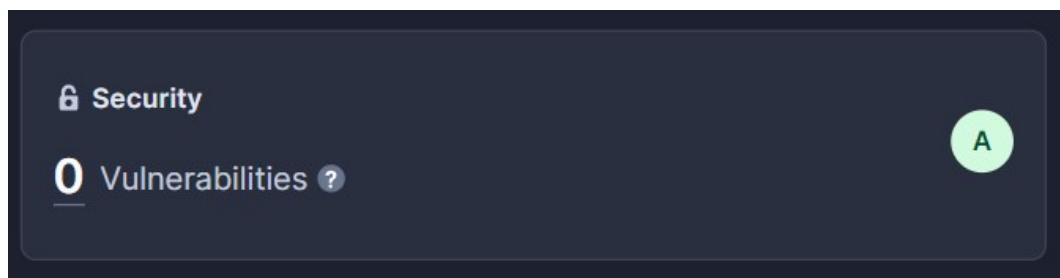


Figura 8.3: Resumen de resultados del análisis de vulnerabilidades

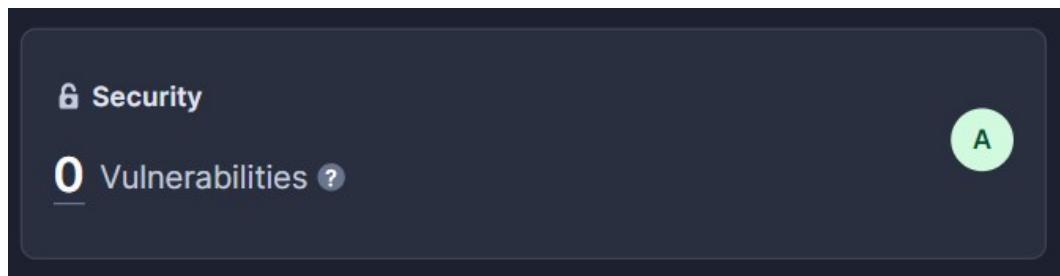


Figura 8.4: Resumen de resultados del análisis de seguridad



# 9. Despliegue del Sistema

---

Este capítulo recoge la arquitectura física planteada para el sistema y las instrucciones para su despliegue.

## 9.1. Arquitectura Física

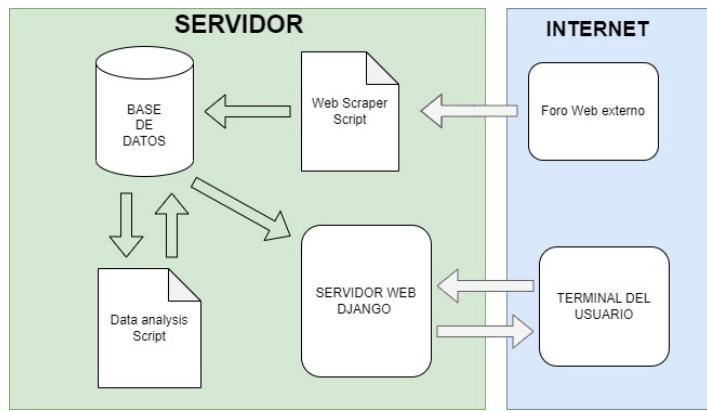


Figura 9.1: Arquitectura Física

La aplicación se ha instalado en un servidor Windows con los siguientes requisitos:

- Procesador: AMD Ryzen 9 4900HS with Radeon Graphics 3.00 GHz
- RAM: 16,0GB
- Tipo de sistema: Sistema operativo de 64 bits, procesador basado en x64
- Edición de Windows: Windows 11 Pro
- Versión de Windows: 22H2

Para ejecutar la aplicación es necesario tener instalado el entorno WampServer (Version 2.2). Adicionalmente también es necesario tener instaladas las librerías especificadas más abajo, con Python 3.8.0.

El sistema es compatible con Linux siempre y cuando se cumplan las especificaciones detalladas.

## 9.2. Instrucciones de despliegue

A continuación, se describen las instrucciones de instalación del sistema sobre la infraestructura física descrita anteriormente.

### 9.2.1. Requisitos previos

- Conexión a internet
- Máquina Windows o Linux
- WampServer (v2.2)
- Python 3.8.0

Adicionalmente hacen falta las librerías:

```
1 asgiref==3.7.2
2 asyncio-timeout==4.0.2
3 attrs==23.1.0
4 autobahn==23.1.2
5 Automat==22.10.0
6 blinker==1.6.2
7 bootstrap4==0.1.0
8 cffi==1.15.1
9 channels==4.0.0
10 channels-redis==4.1.0
11 click==8.1.3
12 colorama==0.4.6
13 constantly==15.1.0
14 crispy-bootstrap4==2022.1
15 cryptography==41.0.1
16 daphne==4.0.0
17 dash==2.9.3
18 dash-bootstrap-components==1.4.1
19 dash-core-components==2.0.0
20 dash-design-kit==0.0.1
21 dash-html-components==2.0.0
22 dash-renderer==1.9.1
23 dash-table==5.0.0
24 Django==4.2.2
25 django-crispy-forms==2.0
26 django-plotly-dash==2.2.1
27 django-redis==5.3.0
28 dpd-components==0.1.0
29 dpd-static-support==0.0.5
30 Flask==2.3.2
31 hyperlink==21.0.0
32 idna==3.4
33 importlib-metadata==6.7.0
34 incremental==22.10.0
35 itsdangerous==2.1.2
36 Jinja2==3.1.2
37 MarkupSafe==2.1.3
38 msgpack==1.0.5
39 numpy==1.24.3
40 packaging==23.1
41 pandas==2.0.2
42 patsy==0.5.3
43 plotly==5.15.0
44 plotly-express==0.4.1
45 pyasn1==0.5.0
46 pyasn1-modules==0.3.0
```

```
47 pycparser==2.21
48 PyMySQL==1.1.0
49 pyOpenSSL==23.2.0
50 python-dateutil==2.8.2
51 pytz==2023.3
52 redis==4.5.5
53 scipy==1.10.1
54 service-identity==23.1.0
55 six==1.16.0
56 sqlparse==0.4.4
57 statsmodels==0.14.0
58 tenacity==8.2.2
59 Twisted==22.10.0
60 txaio==23.1.1
61 typing_extensions==4.6.3
62 tzdata==2023.3
63 Werkzeug==2.3.6
64 zipp==3.15.0
65 zope.interface==6.0
```

## 9.2.2. Inventario de componentes

- Archivos del proyecto
- Copia de la base de datos

## 9.2.3. Procedimientos de instalación

1. Subida de archivo y de la copia de la base de datos al servidor a través del método preferido, por ejemplo por FTP o medio físico.
2. Importación de la copia de la base de datos en MySQL.
3. Copia de archivos de la aplicación en la ruta de instalación deseada
4. Modificación del archivo settings.py de la aplicación web de acuerdo a las credenciales de la base de datos del servidor.
5. Modificación del archivo config.py del web scraper y del procesador de datos de acuerdo a las credenciales de la base de datos del servidor.
6. iniciar la aplicación web situándose en su directorio y ejecutando la siguiente línea:

```
1     python manage.py runserver
2
```

## 9.2.4. Pruebas de implantación

- Comprobar en la terminal la correcta ejecución y la ausencia de errores.
- Comprobar el acceso a la aplicación en el navegador a través de la url

## **Parte III**

### **Epílogo**

# 10. Conclusiones

---

En este último capítulo se detallan las lecciones aprendidas tras el desarrollo del presente proyecto y se identifican las posibles oportunidades de mejora sobre el software desarrollado.

## 10.1. Objetivos alcanzados

Recordemos que el objetivo principal de este proyecto es recoger, clasificar y exponer los problemas y dudas que encuentran los usuarios de App Inventor y presentar los resultados en un dashboard para explorar los datos, ayudarle a realizar comparativas y a obtener recomendaciones, para obtener nueva información.

Como se ha visto a lo largo del proyecto, se han conseguido recoger los datos iniciales mediante web scraping y después se les ha aplicado una serie de transformaciones y de funciones de procesamiento que nos han permitido:

- Clasificación de datos por categorías y temáticas.
- Clasificación mediante Sentiment Analysis.
- Recomendación de preguntas similares mediante clústering de texto.
- Recomendación de respuestas mediante embeddings.

Además gracias a aplicar el clústering a los datos se han descubierto nuevas categorías y temáticas inicialmente ocultas como por ejemplo 'Localización y GPS', 'Uso de Arduino', 'Problemas con emuladores' , 'Uso del componente WebViewer' etc.

#	links	abouts	labels	texts
83299	Tutorials and Guides	8		\nGet your lat/lon location, plus timestamp and many other return datas, all in your webviewer, with a bit of know-how\n\nLocation Services with HTML5
49244	App Showcase	8		Calculate distance between objects on the Canvas using the Pythagorean formula.\n\nWhat distance? The distance can be between the center points of
83079	MIT App Inventor H...	8		Hello good day. I have an activity that needs to get my current location (latitude and longitude). But every time I run my application, it sends me a different one.\nI would like to take the time from the GPS sensor, not from the phone sistem as GSM or WiFi. I need to work with the time that give the satellites.\n\n
81081	MIT App Inventor H...	8		Hi,\nI would like to take the time from the GPS sensor, not from the phone sistem as GSM or WiFi. I need to work with the time that give the satellites.\n\n
30250	Bugs and Other Issues	8		My ai2 creation using Map (roads map) shows unpredictable internet data consumption. This is not really an issue using wifi but it really is a problem using 3G.\n\n
83506	MIT App Inventor H...	8		When GPS changes, a marker should be moved. When save is pressed, a new marker is added. This should be saved without the first marker (user) and load
83468	MIT App Inventor H...	8		someone help me with the programming of the app I'm trying to make. I'd like to create a movit-like app but I'm new to the site.\nI would like to know how to do this.\n\n
82862	MIT App Inventor H...	8		I am sending a latitude and longitude to thingspeak server with the help of NEO-6M GPS Module connected with arduino. I have received all the parameters correctly.\n\n
83062	App Inventor for iOS	8		Hi,\nto anyone who is searching this community.\nI have managed to "fix" what i consider a bug - any app which wants to use the OrientationSensor on a
82588	MIT App Inventor H...	8		How to order the movement of a bus with raspberry card using wifi
83947	MIT App Inventor H...	8		Hello everybody. I want to get local time (for example ?stanbul local time zone) for my app. (I don't want to use device system time). Or get google server time.\n\n
60947081	MIT App Inventor H...	8		\nThis is the value of latitude and longitude in my serial Arduino.\n\nThis is my example design in Mit App inventor.\n\nThis is the blocks of my design in MIT App Inventor.\n\n
60608147	MIT App Inventor H...	8		\nI am working on a scavenger hunt type app for a class project and need to use the location sensor to trigger events during game play. I watched tutorial on how to do this.\n\n
81252	MIT App Inventor H...	8		good morning.\nI've used the locationSensor function to track speed and the BT client for BT communications in separate applications with relatively good
34579153	Tutorials and Guides	8		\nI'm trying to modify MIT app inventor tutorial "Android, Where is my car?", by adding a vibrating function.\nThe phone should vibrate when the stored
36956384	General Discussion	8		\nI basically want to extract the latitude and longitude coordinates sent as sms by gsm module to an android smart phone ...\nI have written an simple app for this.\n\n
84041	MIT App Inventor H...	8		I have a list of chemicals and their corresponding storage locations and they are a lot of chemicals in the list.\nI want when a user opens the app, he should be able to search for a chemical and see its location.\n\n
64883902	Data Storage	8		\nI would like to insert into my TinyDB values like: current_address, current_latitude, current_longitude\nI've created something like this:\n\nhttps://gyazo.c...

Figura 10.1: El clúster descubre una posible nueva categoría para el foro: 'Localización y GPS'

links	about	labels	texts
57019	Tutorials about WebViewer	14	RunJavaScript is a block of the WebViewer component.\nLet's see several examples with this block.\n1.- Get ASCII of \R.\nAppInventor.setWebViewString(\\"...)
31234418	MIT App Inventor Tutorials	14	\nWhen I create an app with App Inventor and I add a webviewer component, it appears an annoying icon to zoom.\nWonder how can I disable this icon.\nIf ...
23329785	Data Storage	14	\nI am creating an app in ai2 that connects to one of my fusion tables as a high score datastore.\nIf I use traditional OAuth2 flow then each user would be pre... ...
81868	Extensions	14	I am doing some work with the sunhttpserver.\nHere is the java (extract) that starts the server and displays Hello World.\n\njava\n@SimpleFunction\npublic ...
82997	MIT App Inventor Tutorials	14	I'm having problems getting some HTML / JavaScript to execute correctly within the WebViewer component. The objective of the code is to render 'abcNotat... ...
81334	MIT App Inventor Tutorials	14	I have the same question. How can I upgrade the webview component?
28869954	MIT App Inventor Tutorials	14	\nI have a webviewer which is the main application and below the webviewer is a button. I would like to open a web page in the user's web browser when the... ...
24959108	Data Storage	14	\nI'm working on an App Inventor based app which uses a lot of local HTML5 and JavaScript within WebViewer components.\nIs it possible to drive App Inventor ... ...
81928	App Show	14	I was reading this topic:\nhttps://community.appinventor.mit.edu/t/tetris-recreation-of-1980s-retro-game/51075/5\nand it occurred to me to look for a code t... ...
43480201	MIT App Inventor Tutorials	14	\nAll right, this is a pretty specific question from a beginner, so bear with me. \nI'm a newbie just learning the ropes. Here's the background (skip to next pa... ...
47276961	MIT App Inventor Tutorials	14	\nMy goal is essentially to upload a video file, like this, but with Ruby on Rails instead of PHP. I can successfully send JSON data to my server, but haven't bee... ...
45761587	Data Storage	14	\nI want to connect my App Inventor App to a website, through the web component with a post request. Now, I want to open a webview of the same webpage... ...
19562460	MIT App Inventor Tutorials	14	\nHello I am building my first App, It's function is to listen to my internet music stream and to play the information about the song what is current playing.\nI ... ...
49189003	General Discussion	14	\nI have made an Application in MIT App Inventor for sending GET request for connecting to WiFi Module [ ESP8266 ].\nThere is a button which when click... ...
26275728	MIT App Inventor Tutorials	14	\nI am trying to integrate MobPartner ads in my app, the ads from MobPartner can be displayed from "Pools" which contain several ads which automatically scro... ...
82630	App Inventor Tutorials	14	How to connect to server socket in MIT App Inventor to control robot programmer by raspberry by wifi ...
38500087	MIT App Inventor Tutorials	14	\nI am doing an assignment to send data to php myadmin database from MIT app inventor 2 through php script. I am using this code I found from youtube in ... ...
82590	MIT App Inventor Tutorials	14	Hi, after following the Alexa Hello World 2 tutorial the only response I get while testing in the browser is "This Skill was created with MIT App Inventor".\nI've... ...
36269336	MIT App Inventor Tutorials	15	\nSo I made two apps until the moment and both have the same problem. One is a complicated user interface and the other is as simple as it gets, only one ... ...

Figura 10.2: El clúster descubre una posible nueva categoría para el foro: 'Uso de Componente WebViewer.'

links	about	labels	texts
82427	MIT App Inventor Tutorials	15	I tried several times and not able to achieve push notification for my app. My goal is to send push notification to installed users.
72738	MIT App Inventor Tutorials	15	Hello,\naiStarter was working fine under Windows 10 Pro. It would open an emulator screen to test an app written here.\nHad to upgrade to Windows 11 Pro. ... ...
82392	General Discussion	15	If we submit our app for app of the season, can the same be re-submitted for appathon later?
83860	MIT App Inventor Tutorials	15	I have checked that the emulator available on MIT App Inventor website is still the old one, i.e. for AI 2.65 version, which is not working since AI has been upd... ...
81584	Extensions	15	Please pardon my ignorance. I'm trying to set up a development environment for the first time..\nDoes it matter that the Extension Template Repository 2 ha... ...
27976467	MIT App Inventor Tutorials	15	\nIs there any way to connect App Inventor 2 to a Genymotion instance?\n
29207717	MIT App Inventor Tutorials	15	\nWhen I run aiStarter.exe as administrator, and the click on connect to the emulator in http://ai2.appinventor.mit.edu, then the emulator starts, the black scre... ...
3243527	MIT App Inventor Tutorials	15	\nCould someone explain how to start Appinventor after installing?\nI downloaded the file and placed in folder as indicated in the provided instructions but a... ...
83041	App Inventor Tutorials	15	I can no longer find anything in the menu to BULD for iOS or Submit.ipk file the Apple connect.\nWhat happened to the menus? How can we publish for iOS... ...
22639294	MIT App Inventor Tutorials	15	\nI'm building and uploading App Inventor 2 to our Google App Engine account.\nI'm following the instructions found here: How to build App Inventor from t... ...
57246884	MIT App Inventor Tutorials	15	\nA module that should be installed via pip (in my case it is selenium) doesn't work on skulpt (app-inventor).\nI'm not a native English speaker. Thanks.\n
83105	MIT App Inventor Tutorials	15	my apps is not actived and it send this message :\nErreur\nCette release ne permet pas d'ajouter ni de supprimer des app bundles.\n(mod translation)\nThis r... ...
27500782	MIT App Inventor Tutorials	15	\nStarted evaluating MIT App Inventor (ver 2.2) on Windows 7 (from Administrator account). Since I don't have Android devices, I have to use an emulator.\n... ...
28607658	MIT App Inventor Tutorials	15	\nI'm having an issue when trying to Call Twitter.Authorize in App Inventor 2. I have the correct consumer key and secret set when the screen initialized but I... ...
27558102	MIT App Inventor Tutorials	15	\nWhen I try the following (i.e., spoken Hebrew) in App Inventor 2, I hear nothing.\nHowever, the following (switching parameters to spoken English) works j... ...
34615585	MIT App Inventor Tutorials	15	\nWhen running aiStarter, I receive the following error:\nBottle server starting up (using WSGIRefServer()).\nListening on http://127.0.0.1:8004\nHit Ctrl-C to... ...
28375	App Inventor Tutorials	15	I am attempting to create a basic Sign/Log in system. Previously I had a few issues with lists duplicating when communicating with a CloudDB. However, when... ...
82091	MIT App Inventor Tutorials	15	Bonjour à nouveau,\nje galère vraiment ce matin.\nPremier problème : l'éulateur ne veut plus se lancer après une mise à jour de l'appli sur mon PC.\nJ'ai ... ...
82028	MIT App Inventor Tutorials	15	Hi - I'm trying to do the Alexa Skill tutorial. My Amazon developer account is active and it has one (old) skill in it that I started before and abandoned. I get a... ...

Figura 10.3: El clúster descubre una posible nueva categoría para el foro: 'Problemas con Emuladores'

Todos estos datos mencionados antes están disponibles para visualizar en forma de tablas y gráficas en la aplicación web para que el usuario final pueda explorarlas con más detalle.

Con todo esto quedan cumplidos los objetivos iniciales del presente proyecto.

## 10.2. Lecciones aprendidas

Durante el desarrollo de este proyecto, me he aproximado por primera vez a técnicas de análisis de datos como la clasificación y el clústering, por lo que he tenido que dedicar tiempo a aprender sobre el Procesamiento del Lenguaje Natural y sobre diferentes modelos.

Además he mejorado mis conocimientos acerca del minado de datos y de web scraping que había aprendido anteriormente cursando el Grado. También he aprendido la disciplina y el trabajo necesario para desarrollar un proyecto completo en solitario.

Finalmente, gracias a este trabajo, he reconocido la importancia y el peso de los datos y la información hoy en día y del potencial de su análisis y he descubierto un gusto por este campo de la informática que desconocía que tenía.

## 10.3. Trabajo futuro

- Implementar sesiones de usuario en la web con guardado de búsquedas.  
La aplicación no cuenta con sesiones de usuario, sería interesante para un futuro gestionar el acceso mediante usuarios para así poder guardar sus consultas y sus búsquedas.
- Posibilidad de seleccionar y comparar múltiples respuestas a la vez en los mismos gráficos.  
En la pantalla 'Questions' de la web, solo podemos seleccionar una única pregunta para ver sus datos en las gráficas. Podría ser útil implementar la selección múltiple de preguntas para así poder comparar los datos de ambas de un vistazo en las gráficas sin tener que cambiar la selección entre ellas.
- Adaptar la web a dispositivos móviles.  
Actualmente la web y sus tablas y gráficas no están preparadas para mostrarse en pantallas de dispositivos móviles o de menor tamaño debido a sus características.
- Añadir más gráficas adicionales que se puedan considerar importantes.
- Adaptar el análisis de los datos a otros idiomas aparte del inglés. Aunque la mayoría de interacciones en los foros están escritas en inglés, hay algunas en otros idiomas como por ejemplo español, portugués e italiano.



## **11. Información sobre Licencia**

---

# GNU Free Documentation License

---

Version 1.3, 3 November 2008

Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.

<<https://fsf.org/>>

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

## Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document “free” in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of “copyleft”, which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

## 1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The “**Document**”, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as “**you**”. You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A “**Modified Version**” of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “**Secondary Section**” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains

nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “**Invariant Sections**” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “**Cover Texts**” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “**Transparent**” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “**Opaque**”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “**Title Page**” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “**publisher**” means any person or entity that distributes copies of the Document to the public.

A section “**Entitled XYZ**” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “**Acknowledgements**”, “**Dedications**”, “**Endorsements**”, or

**“History”.)** To “**Preserve the Title**” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

## **2. VERBATIM COPYING**

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

## **3. COPYING IN QUANTITY**

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document’s license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers)

of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

## 4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published

at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

- K. For any section Entitled “Acknowledgements” or “Dedications”, Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled “Endorsements”. Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled “Endorsements” or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version’s license notice.

These titles must be distinct from any other section titles.

You may add a section Entitled “Endorsements”, provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

## 5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant

Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled “History” in the various original documents, forming one section Entitled “History”; likewise combine any sections Entitled “Acknowledgements”, and any sections Entitled “Dedications”. You must delete all sections Entitled “Endorsements”.

## **6. COLLECTIONS OF DOCUMENTS**

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

## **7. AGGREGATION WITH INDEPENDENT WORKS**

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an “aggregate” if the copyright resulting from the compilation is not used to limit the legal rights of the compilation’s users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document’s Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

## **8. TRANSLATION**

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original

versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled “Acknowledgements”, “Dedications”, or “History”, the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

## 9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

## 10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

See <https://www.gnu.org/licenses/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License “or any later version” applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy’s public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

## 11. RELICENSING

“Massive Multiauthor Collaboration Site” (or “MMC Site”) means any World Wide

Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A “Massive Multiauthor Collaboration” (or “MMC”) contained in the site means any set of copyrightable works thus published on the MMC site.

“CC-BY-SA” means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

“Incorporate” means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is “eligible for relicensing” if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

# GNU GENERAL PUBLIC LICENSE

Version 3, 29 June 2007

Copyright © 2007 Free Software Foundation, Inc. <https://fsf.org/>

Everyone is permitted to copy and distribute verbatim copies of this  
license document, but changing it is not allowed.

## Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed to take away your freedom to share and change the works. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change all versions of a program—to make sure it remains free software for all its users. We, the Free Software Foundation, use the GNU General Public License for most of our software; it applies also to any other work released this way by its authors. You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for them if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs, and that you know you can do these things.

To protect your rights, we need to prevent others from denying you these rights or asking you to surrender the rights. Therefore, you have certain responsibilities if you distribute copies of the software, or if you modify it: responsibilities to respect the freedom of others.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must pass on to the recipients the same freedoms that you received. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

Developers that use the GNU GPL protect your rights with two steps: (1) assert copyright on the software, and (2) offer you this License giving you legal permission to copy, distribute and/or modify it.

For the developers' and authors' protection, the GPL clearly explains that there is no warranty for this free software. For both users' and authors' sake, the GPL requires that modified versions be marked as changed, so that their problems will not be attributed erroneously to authors of previous versions.

Some devices are designed to deny users access to install or run modified versions of the software inside them, although the manufacturer can do so. This is fundamentally incompatible with the aim of protecting users' freedom to change the software. The systematic pattern of such abuse occurs in the area of products for individuals to use, which is precisely where it is most unacceptable. Therefore, we have designed this version of the GPL to prohibit the practice for those products. If such problems arise substantially in other domains, we stand ready to extend this provision to those domains in future versions of the GPL, as needed to protect the freedom of users.

Finally, every program is threatened constantly by software patents. States should not allow patents to restrict development and use of software on general-purpose computers, but in those that do, we wish to avoid the special danger that patents applied to

a free program could make it effectively proprietary. To prevent this, the GPL assures that patents cannot be used to render the program non-free.

The precise terms and conditions for copying, distribution and modification follow.

# TERMS AND CONDITIONS

## 0. Definitions.

“This License” refers to version 3 of the GNU General Public License.

“Copyright” also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

“The Program” refers to any copyrightable work licensed under this License. Each licensee is addressed as “you”. “Licensees” and “recipients” may be individuals or organizations.

To “modify” a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a “modified version” of the earlier work or a work “based on” the earlier work.

A “covered work” means either the unmodified Program or a work based on the Program.

To “propagate” a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To “convey” a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays “Appropriate Legal Notices” to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

## 1. Source Code.

The “source code” for a work means the preferred form of the work for making modifications to it. “Object code” means any non-source form of a work.

A “Standard Interface” means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The “System Libraries” of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in

source code form. A “Major Component”, in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The “Corresponding Source” for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to control those activities. However, it does not include the work’s System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

## 2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

## 3. Protecting Users’ Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by

exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

#### 4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

#### 5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".
- c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work, and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

#### 6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4

and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

- a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.
- b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.
- e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A “User Product” is either (1) a “consumer product”, which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, “normally used” refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a

consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

“Installation Information” for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

## 7. Additional Terms.

“Additional permissions” are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a

covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or
- b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or
- c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or
- d) Limiting the use for publicity purposes of names of licensors or authors of the material; or
- e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or
- f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered “further restrictions” within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

## 8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the

copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

#### 9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

#### 10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An “entity transaction” is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party’s predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

#### 11. Patents.

A “contributor” is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor’s “contributor version”.

A contributor’s “essential patent claims” are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, “control” includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor’s essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a “patent license” is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To “grant” such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. “Knowingly relying” means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient’s use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is “discriminatory” if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that

patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License “or any later version” applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.





# Bibliografía

---

- [1] (2022). STING – Statistical Information Grid in Data Mining . <https://www.geeksforgeeks.org/sting-statistical-information-grid-in-data-mining/>.
- [2] (2023). The Current State of the Art in Natural Language Processing (NLP). <https://aicromo.com/ai-blogs/the-current-state-of-the-art-in-natural-language-processing-nlp>.
- [3] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever (2018). How BERT and GPT models change the game for NLP. [https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf).
- [4] Ali, A., Bin Faheem, Z., Waseem, M., Draz, U., Safdar, Z., Hussain, S., y Yaseen, S. (2020). Systematic review: A state of art ml based clustering algorithms for data mining. In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pages 1–6.
- [5] Brown, S. (2022). The c4 model for visualising software architecture.
- [6] de Oviedo, U. (2020). K Means. [https://www.unioviedo.es/compnum/laboratorios\\_py/kmeans/kmeans.html](https://www.unioviedo.es/compnum/laboratorios_py/kmeans/kmeans.html).
- [7] Hartmann, J. (2023). emotion-english-distilroberta-base . <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base/blob/main/README.md>.
- [8] Khder, M. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and its Applications*, 13:145–168.
- [9] Nazeri, S. (2023). Comparing The-State-of-The-Art Clustering Algorithms. <https://medium.com/@sina.nazeri/comparing-the-state-of-the-art-clustering-algorithms-1e65a08157a1>.
- [10] of Technology, M. I. (2010). MIT App Inventor. <https://appinventor.mit.edu/about-us>.
- [11] OpenAI (2020). OpenAI API Documentation . <https://platform.openai.com/docs/introduction/overview>.
- [12] OpenAI (2023a). Embedding models. <https://platform.openai.com/docs/guides/embeddings/embedding-models>.
- [13] OpenAI (2023b). Models: gpt-3.5-turbo. <https://platform.openai.com/docs/models/gpt-3-5>.

- [14] Otten, N. V. (2023). The best text clustering algorithms. [https://spotintelligence.com/2023/01/17/text-clustering-algorithms/#The\\_best\\_text\\_clustering\\_algorithm](https://spotintelligence.com/2023/01/17/text-clustering-algorithms/#The_best_text_clustering_algorithm).
- [15] Python (2023). Python 3.11.5 documentation. <https://docs.python.org/3/>.
- [16] Rouse, M. (2023). Web Scraping. <https://www.techopedia.com/definition/5212/web-scraping>.
- [17] Shamala Gallagher, Anna Rafferty, A. W. (2004). Natural Language Processing. [https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview\\_history.html](https://cs.stanford.edu/people/eroberts/courses/soco/projects/2004-05/nlp/overview_history.html).
- [18] Silgado, S. (2022). Teoría de las emociones de Paul Ekman. <https://www.psicologia-online.com/teoria-de-las-emociones-de-paul-ekman-5391.html>.





## **Parte IV**

### **Anexos**

# A. Manual del desarrollador

---

## A.1. Introducción

Este software tiene como principal objetivo recoger, clasificar y exponer los problemas o dudas que encuentran los usuarios de AppInventor y presentar los resultados en forma de infografías en un cuadro de mandos web que el usuario podrá explorar con detalle. Para ello aplica técnicas de web scraping, procesamiento del lenguaje natural, clústering etc.

## A.2. Preparación del entorno de trabajo

- Conexión a internet
- Máquina Windows o Linux
- WampServer (v2.2)
- Python 3.8.0

Adicionalmente hacen falta las librerías incluidas en el archivo *requirements.txt*:

```
1 asgiref==3.7.2
2 asyncio==4.0.2
3 attrs==23.1.0
4 autobahn==23.1.2
5 Automat==22.10.0
6 blinker==1.6.2
7 bootstrap4==0.1.0
8 cffi==1.15.1
9 channels==4.0.0
10 channels-redis==4.1.0
11 click==8.1.3
12 colorama==0.4.6
13 constantly==15.1.0
14 crispy-bootstrap4==2022.1
15 cryptography==41.0.1
16 daphne==4.0.0
17 dash==2.9.3
18 dash-bootstrap-components==1.4.1
19 dash-core-components==2.0.0
20 dash-design-kit==0.0.1
21 dash-html-components==2.0.0
22 dash-renderer==1.9.1
23 dash-table==5.0.0
24 Django==4.2.2
25 django-crispy-forms==2.0
26 django-plotly-dash==2.2.1
27 django-redis==5.3.0
```

```

28 dpd-components==0.1.0
29 dpd-static-support==0.0.5
30 Flask==2.3.2
31 hyperlink==21.0.0
32 idna==3.4
33 importlib-metadata==6.7.0
34 incremental==22.10.0
35 itsdangerous==2.1.2
36 Jinja2==3.1.2
37 MarkupSafe==2.1.3
38 msgpack==1.0.5
39 numpy==1.24.3
40 packaging==23.1
41 pandas==2.0.2
42 patry==0.5.3
43 plotly==5.15.0
44 plotly-express==0.4.1
45 pyasn1==0.5.0
46 pyasn1-modules==0.3.0
47 pycparser==2.21
48 PyMySQL==1.1.0
49 pyOpenSSL==23.2.0
50 python-dateutil==2.8.2
51 pytz==2023.3
52 redis==4.5.5
53 scipy==1.10.1
54 service-identity==23.1.0
55 six==1.16.0
56 sqlparse==0.4.4
57 statsmodels==0.14.0
58 tenacity==8.2.2
59 Twisted==22.10.0
60 txai==23.1.1
61 typing_extensions==4.6.3
62 tzdata==2023.3
63 Werkzeug==2.3.6
64 zipp==3.15.0
65 zope.interface==6.0

```

### A.3. Consideraciones generales sobre el desarrollo

Aspectos importantes a tener en cuenta a la hora de modificar y extender el código fuente, guías de estilo, etc. Asimismo, se detallarán las directrices que sean de aplicación a la hora de realizar pruebas sobre las nuevas mejoras introducidas.

Tenga en cuenta que este proyecto está dividido en tres subproyectos, bien diferenciados en su función y ámbito. Es importante respetar esta estructura.

Para establecer o modificar variables globales y de configuración use los archivos **config.py** y **settings.py**.

Si desea añadir más secciones a la aplicación web, cree una nueva app en django mediante

```
1 $ python manage.py startapp nombredeapp
```

siga de ejemplo la estructura que tienen el resto de aplicaciones respetando el framework Django:

```
1 polls/
2     __init__.py
3     admin.py
4     apps.py
5     migrations/
6         __init__.py
7     models.py
8     tests.py
9     views.py
```

## A.4. Instrucciones para construcción

- a) Subida de archivo y de la copia de la base de datos al servidor a través del método preferido, por ejemplo por FTP o medio físico.
- b) Importación de la copia de la base de datos en MySQL.
- c) Copia de archivos de la aplicación en la ruta de instalación deseada
- d) Modificación del archivo settings.py de la aplicación web de acuerdo a las credenciales de la base de datos del servidor.
- e) Modificación del archivo config.py del web scraper y del procesador de datos de acuerdo a las credenciales de la base de datos del servidor.
- f) Iniciar la aplicación web situándose en su directorio y ejecutando la siguiente línea:

```
1 python manage.py runserver
2
```

## A.5. Ejecución de los scripts de web scraping y de análisis de datos

Para ejecutar el script de recogida de datos **sítuese en el directorio del proyecto 'web\_scraping'** y ejecute en la terminal de comandos el siguiente script.

```
1 python main.py
```

Para ejecutar el script de recogida de datos **sítuese en el directorio del proyecto 'data\_processing\_and\_analysis'** y ejecute en la terminal de comandos el siguiente script.

```
1 python main.py
```





# B. Manual de usuario

---

## B.1. Introducción

Este manual se dirige al usuario. A continuación se detallan las instrucciones de navegación y uso general de la herramienta de visualización de datos.

## B.2. Ejecutar la recogida de los datos

En la terminal de comandos, sitúese en el directorio de la carpeta `web_scraper` del proyecto y ejecute el siguiente script:

```
1 python main.py
```

## B.3. Ejecutar el análisis de los datos

En la terminal de comandos, sitúese en el directorio de la carpeta `data_processing_and_analysis` del proyecto y ejecute el siguiente script:

```
1 python main.py
```

## B.4. Uso de la aplicación web

A continuación se describen las pantallas que forman la web y como hacer uso de ellas para visualizar todos los datos recogidos y procesados previamente.

### B.4.1. Menú de navegación

A la izquierda de la pantalla usted tendrá siempre visible el menú de navegación. Si desea comprimirlo, pulse la flecha en la parte baja del menú.

### B.4.2. Home

Al acceder a AppInventor Dashboard nos encontraremos en la pantalla principal o 'Home'. Esta pantalla muestra la información más general del sitio.



Figura B.1: Pantalla Questions

### B.4.3. Questions

Al acceder a la sección 'Questions' en el menú de navegación, se le mostrará una tabla que muestra información sobre todas las preguntas del sitio. Si desea saber más sobre una pregunta, selecciónela en la tabla. Una vez seleccionada, más abajo aparecerá información adicional sobre ella, como por ejemplo: sus respuestas, el análisis de sentimientos, tags sugeridos y preguntas similares.

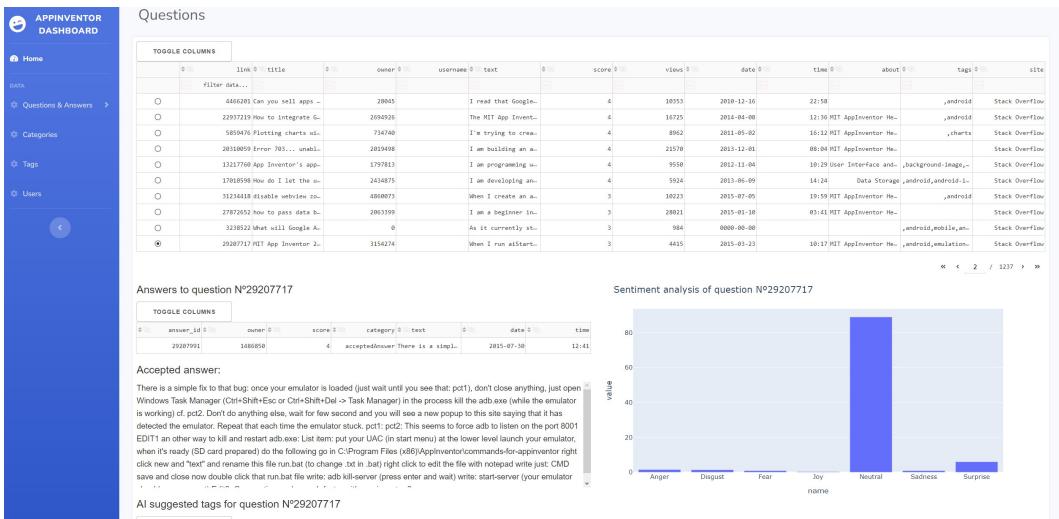


Figura B.2: Pantalla Home

### B.4.4. Answers

Al acceder a la sección 'Answers' en el menú de navegación, se le mostrará una tabla que muestra información sobre todas las respuestas del sitio. Si desea saber más sobre una respuesta, selecciónela en la tabla. Una vez seleccionada, más abajo aparecerá información adicional sobre ella.

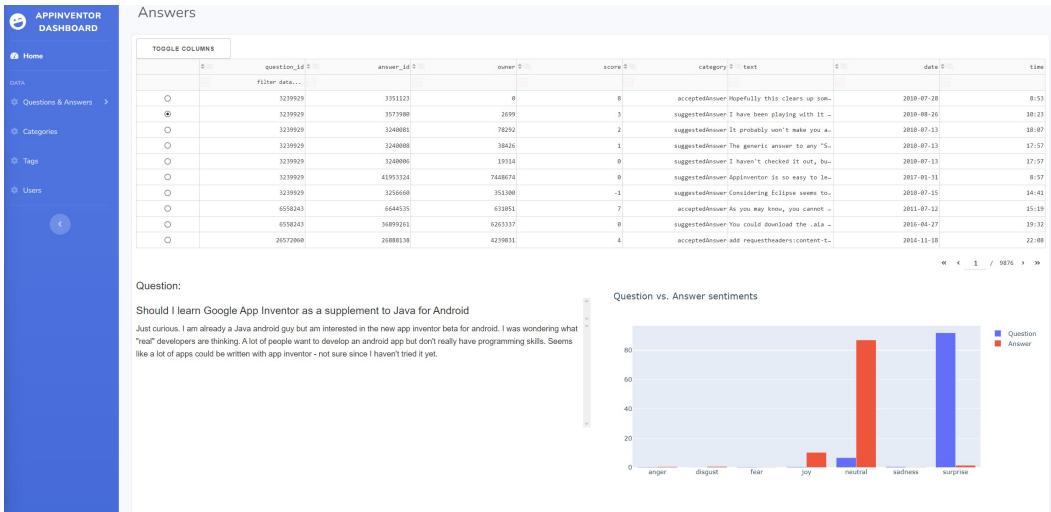


Figura B.3: Pantalla Answers

#### B.4.5. Categories

Al acceder a la sección 'Categories' en el menú de navegación, se le mostrará una tabla que muestra información sobre todas las categorías del sitio. Si desea saber más sobre una categoría, selecciónela el desplegable que hay más abajo. Una vez seleccionada, aparecerá información adicional sobre ella, como por ejemplo: tags mas relevantes, usuarios más activos en esa categoría, mapa etc.

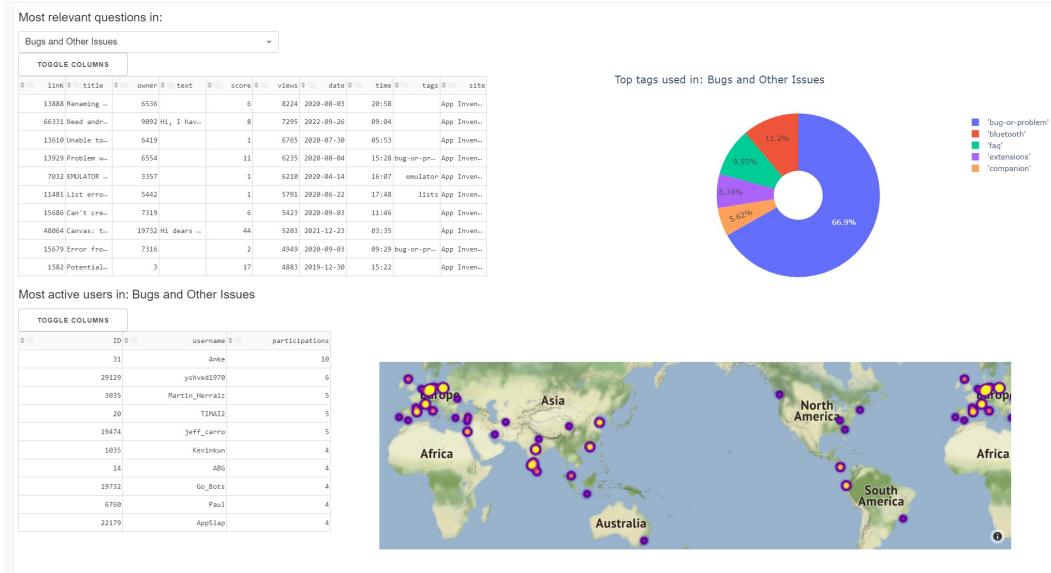


Figura B.4: Pantalla Categories

#### B.4.6. Tags

Al acceder a la sección 'Tags' en el menú de navegación, se le mostrará una tabla que muestra información sobre todos los tags. Si desea saber más sobre un tag,

selecciónelo en la tabla. Una vez seleccionado, más abajo aparecerá información adicional sobre el mismo.

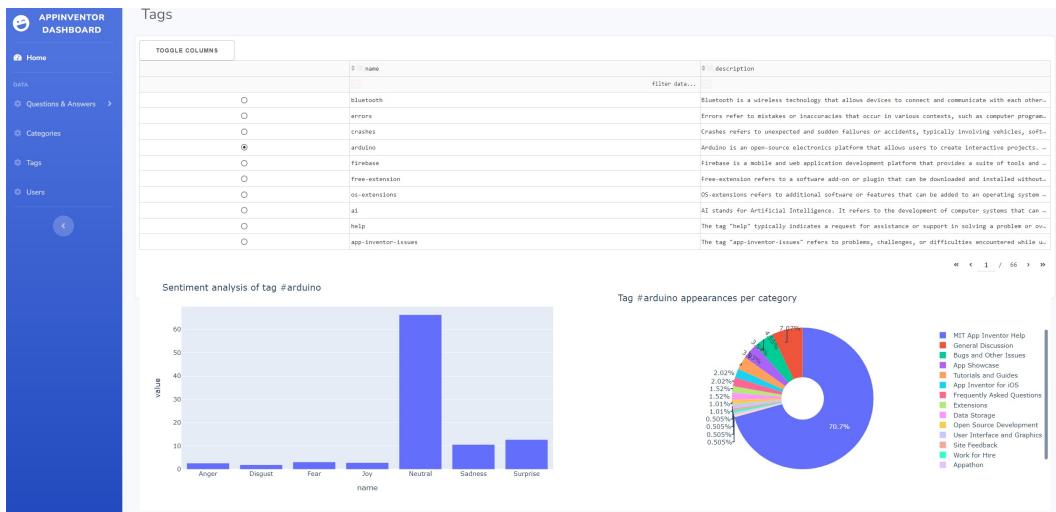


Figura B.5: Pantalla Tags

### B.4.7. Users

Al acceder a la sección 'Users' en el menú de navegación, se le mostrará una tabla que muestra información sobre todos los usuarios. Si desea saber más sobre un usuario, selecciónelo en la tabla. Una vez seleccionado, más abajo aparecerá información adicional sobre el mismo.

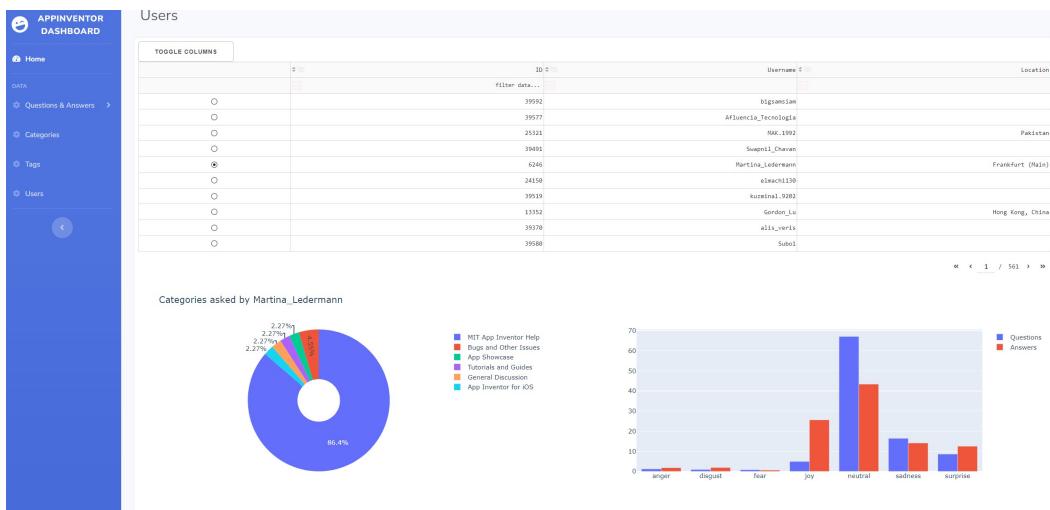


Figura B.6: Pantalla Users

### B.4.8. Uso de tablas y gráficas

- Todas las gráficas generadas en sitio son descargables. Para descargar una gráfica pase el ratón sobre ella y pulse en el icono de la cámara de fotos.
- Todas las tablas poseen la opción de ocultar o mostrar más columnas según usted lo desee mediante el botón 'toggle columns'.
- Todas las gráficas y tablas tienen filtros para mostrar solo los datos que le interesen.



# C. Extractos de código fuente

---

## C.1. Introducción

Esta sección muestra extractos del código fuente considerados de especial interés.

## C.2. Extractos de código

A continuación se muestran los extractos de código citados en la memoria.

```
1 #Funcion que nos ayuda a elegir la K que vamos a aplicar a kmeans
2 def get_k():
3
4     X = ready_data()
5
6     Nc = range(1, 100)
7     kmeans = [KMeans(n_clusters=i) for i in Nc]
8
9     score = [kmeans[i].fit(X).score(X) for i in range(len(kmeans))]
10
11    plt.plot(Nc, score)
12    plt.xlabel('Number of Clusters')
13    plt.ylabel('Score')
14    plt.title('Elbow Curve')
15    plt.show()
16
```

Figura C.1: Código de get\_k()

```
1 #Llamada al modelo de embedding
2 def get_embedding(text, model="text-embedding-ada-002"):
3     text = text.replace("\n", " ")
4     embedding = openai.Embedding.create(input = [text], model=model)['data'
5     ''][0]['embedding']
6     time.sleep(20)
7     return embedding
```

Figura C.3: Código de get\_embedding()

```

1 def apply_kmeans():
2
3     X = ready_data()
4
5     #Obtener los centroides
6     kmeans = KMeans(n_clusters=20, max_iter=1000, random_state=3).fit(X)
7     centroids = kmeans.cluster_centers_
8     print(centroids)
9
10    #Ejecutar kmeans
11    labels = kmeans.predict(X)
12    C = kmeans.cluster_centers_
13
14    colores=['red','green','blue','cyan','yellow','orange','indigo','
15    coral','slategray','palegreen','lime','navy','turquoise','tan','olive',
16    'pink','cadetblue','deeppink','palevioletred','darkslateblue']
17    asignar=[]
18    for row in labels:
19        asignar.append(colores[row])
20
21    plt.scatter(X[:,0], X[:,1], c=asignar, s=70)
22    plt.scatter(C[:, 0], C[:, 1], marker='*', c=colores, s=1000)
23    plt.show()
24
25    return(labels)
26
```

Figura C.2: Código de apply\_kmeans()

```

1 # Funcion que analiza y guarda los sentimientos de las respuestas en la
2 # BBDD
3 def answers_diagnosis(answers):
4
5     # las pasamos al modelo de sentiment analysis y las guardamos el
6     # resultado en la BBDD
7     for answer in answers:
8         classifier = pipeline("text-classification", model = "j-hartmann/
9         emotion-english-distilroberta-base", return_all_scores=True,
10         truncation = True)
11         sentiments = classifier(answer[1])
12
13         diagnosis = ""
14         for sentiment in sentiments[0]:
15             diagnosis = diagnosis + sentiment['label'] + ":" + str(
16             sentiment['score']) + ", "
17
18         query = "UPDATE answers SET sentiments = %s WHERE answer_id = %s
19 AND site = %s"
20         params = [diagnosis, answer[0], answer[2]]
21         cursor.execute(query, params)
22         conn.commit()
23
```

Figura C.4: Código de answers\_diagnosis()

```
1 # tarea para chatGPT
2 duty_explanation = '''Give a short description to this tag in a
3     programming context: ''' + tag[0]
4 completion = openai.ChatCompletion.create(model="gpt-3.5-
5     turbo", temperature=0,
6         messages=[
7             {"role": "user", "content": duty_explanation},
8         ]
```

Figura C.5: Código de get\_tags\_description()



# D. Test de usabilidad

---

## D.1. Introducción

A continuación se recogen las preguntas y los resultados de la encuesta de usabilidad que han realizado los usuarios.

## D.2. Resultados

Edad

5 respuestas

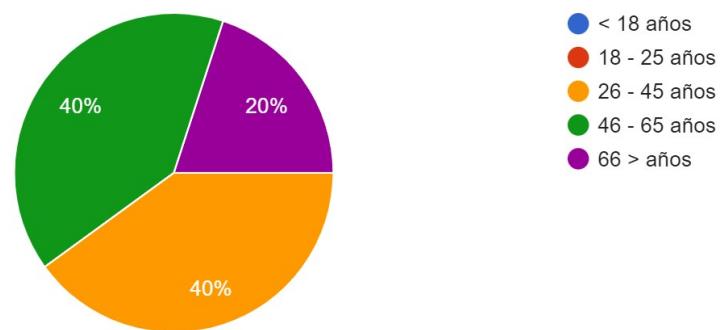


Figura D.1: Edades de los usuarios encuestados

Valore del 1 al 5 sus conocimientos informáticos.

5 respuestas

Copiar

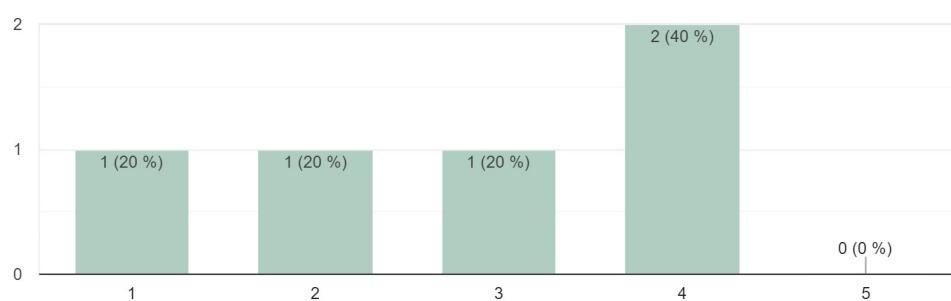


Figura D.2: Conocimientos informáticos de los encuestados

### Nivel de inglés

5 respuestas

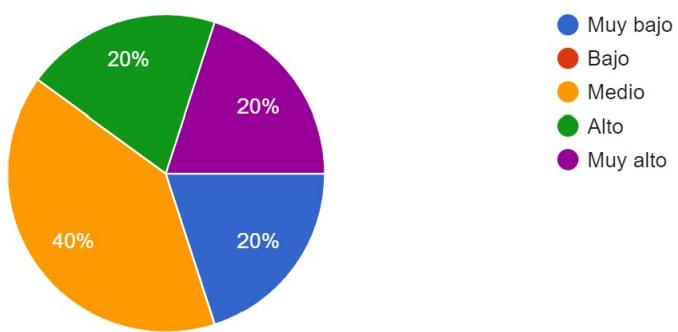


Figura D.3: Nivel de inglés de los encuestados

### Valore la dificultad de acceder al listado de preguntas

5 respuestas

Copiar

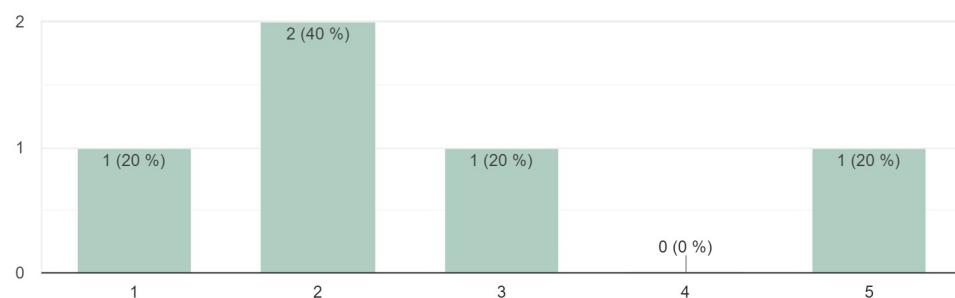


Figura D.4: Pregunta 1

Valore la dificultad de acceder a la información adicional y gráficas de una categoría concreta

Copiar

5 respuestas

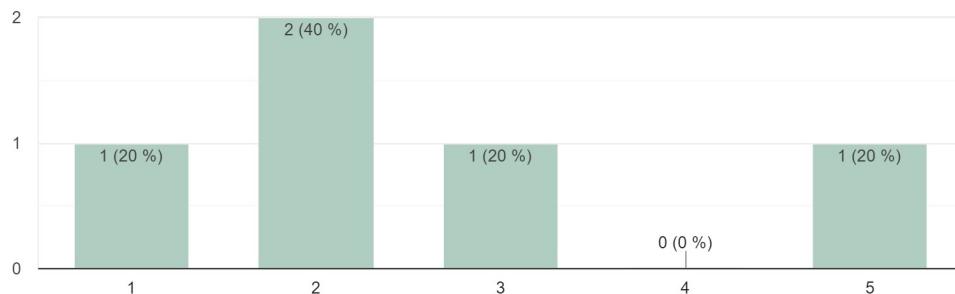


Figura D.5: Pregunta 2

Valore la dificultad del uso de las tablas (selección, filtrado, ordenación etc)

Copiar

5 respuestas

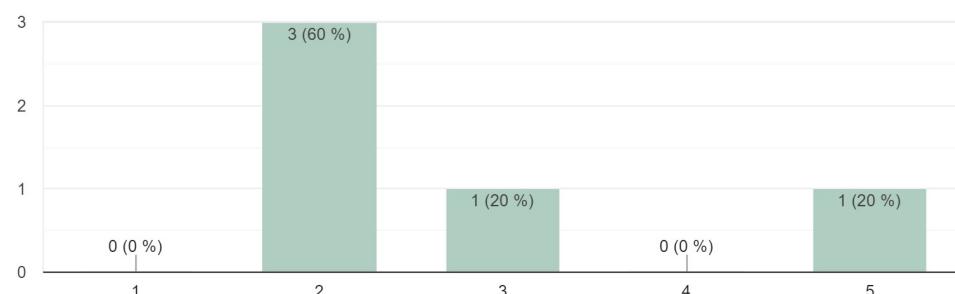


Figura D.6: Pregunta 3

Por último, ¿Alguna sugerencia?

1 respuesta

Me gustaria que se pudiese poner la pagina en español

Figura D.7: Sugerencia de los usuarios



# E. Test de portabilidad

## E.1. Introducción

A continuación se muestran los resultados de los análisis de LightHouse sobre la web, tanto en escritorio ([E.1](#)) como en dispositivos móviles ([E.2](#)).

## E.2. Resultados

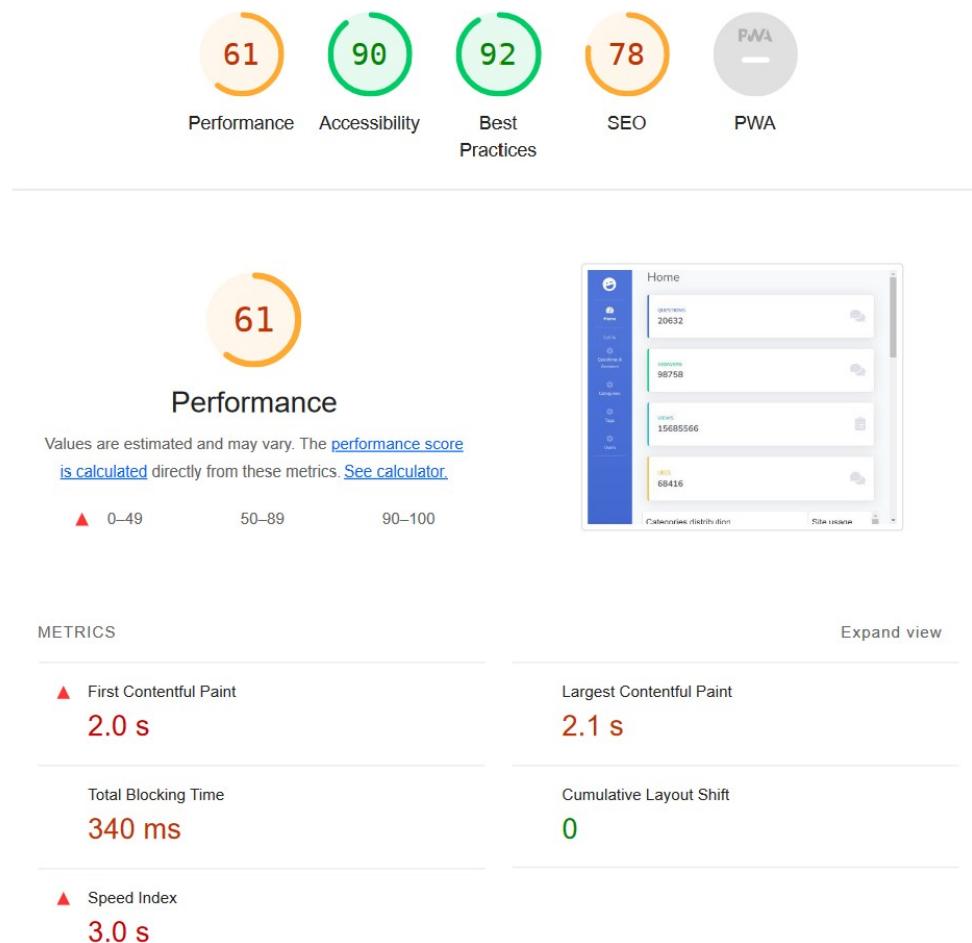


Figura E.1: Resultados portabilidad web (Escritorio)

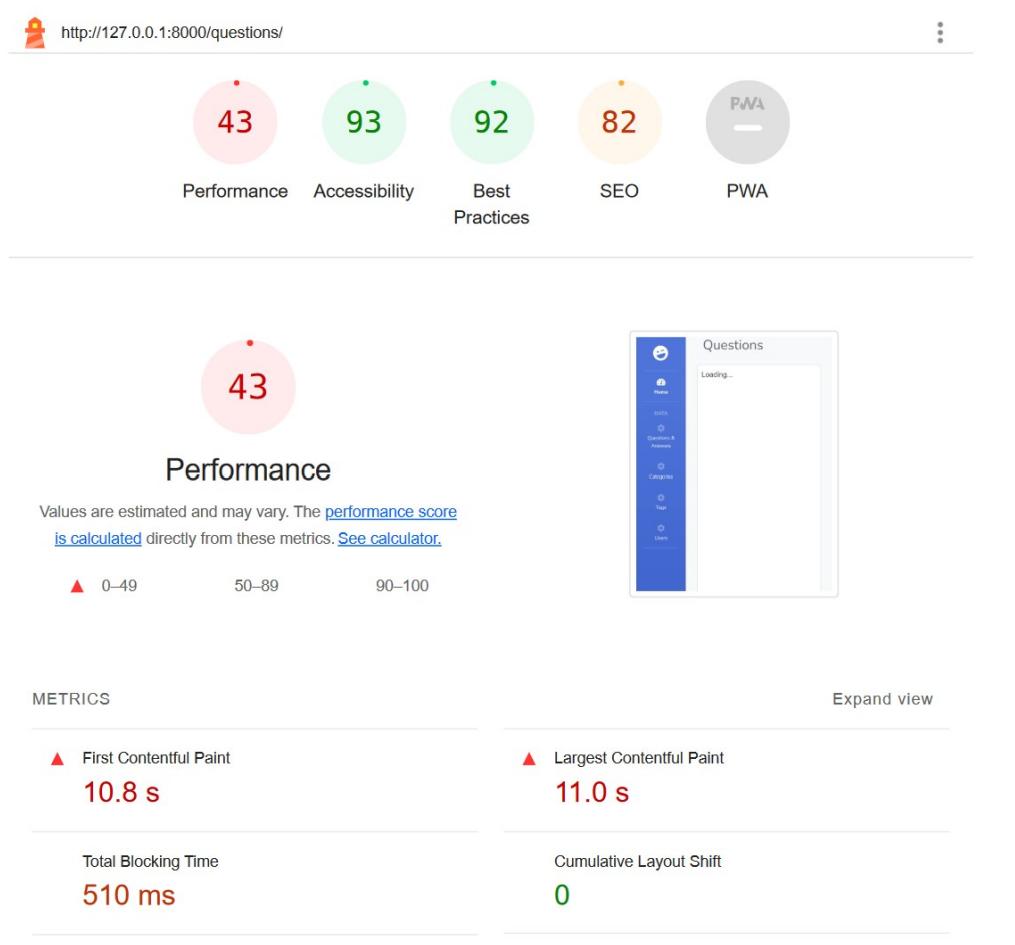


Figura E.2: Resultados portabilidad web (Móvil)