

Sistema Recomendador (RecoMMachine)



Aroa Chans
Data Science
Agosto 2024



OBJETIVO

Desarrollo de un sistema de recomendación, en este caso, de cursos.

Basado en datos, utilizando diferentes metodologías de análisis, visualización y evaluación.



APLICACIONES

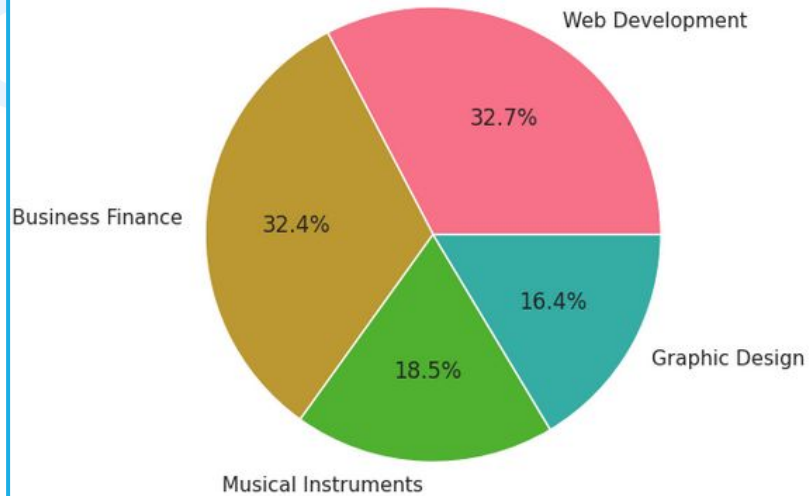
Cursos, programas, entretenimiento, productos, etc.

Nombre	aroachans
Identificación del cliente	bIMgegBn1ApNbD4Xz1Xn6lJPu8Ogzz0tt247Oowf
Secreto de cliente	LkAFjT6baOPXlXwN3sTZyEKZCe6JggMaMko2MxCvQlU0siAdOKVNC6rw2pdTiFFFz0EC91KVi1J80zXchcyqBmCWftz4NGcR97RiiCs7AerwsFTtc6bPoTj

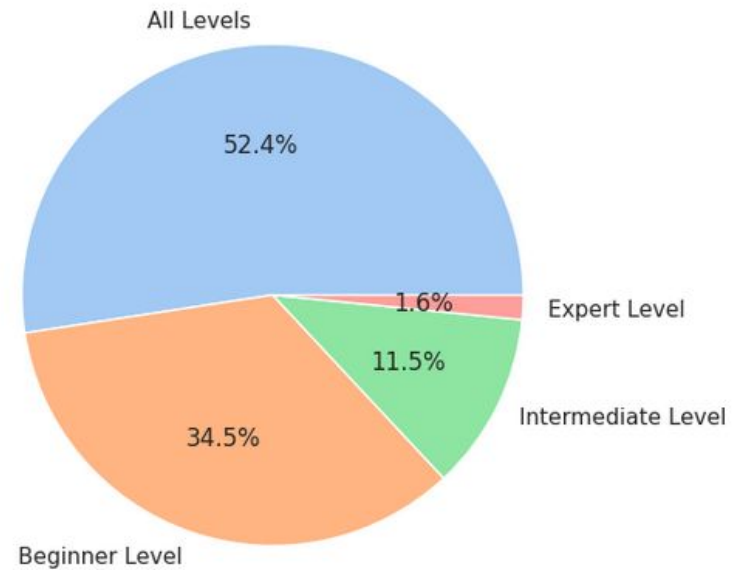


- **Dataset de Udemy:** API de afiliado o Kaggle: ID, título, de pago(o gratuito), precio, nº suscriptores, nº de reseñas, nº lecciones, nivel, duración del contenido, fecha de publicación, tema
- **Librerías de Python:** numpy, pandas, matplotlib, seaborn, sklearn, nltk, workcloud, tkinter...
- **Análisis básico:** descriptivos, limpieza nulos y duplicados
- **Análisis avanzado:** categ, numérico, correlación, etc

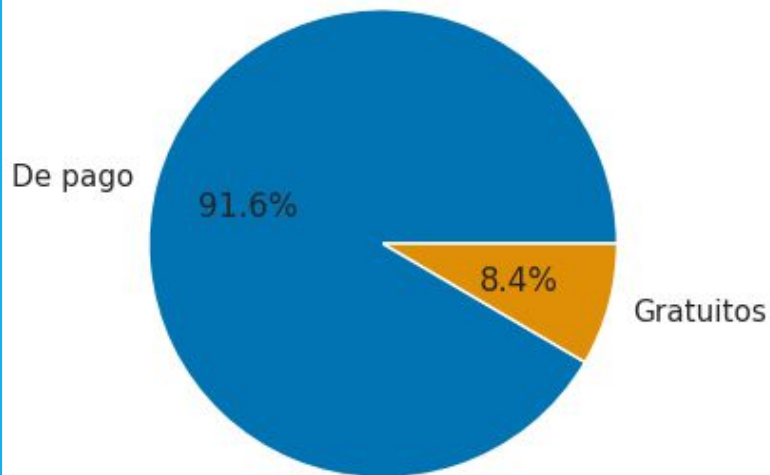
Distribución de los Temas de Cursos



Distribución de los Niveles de Cursos

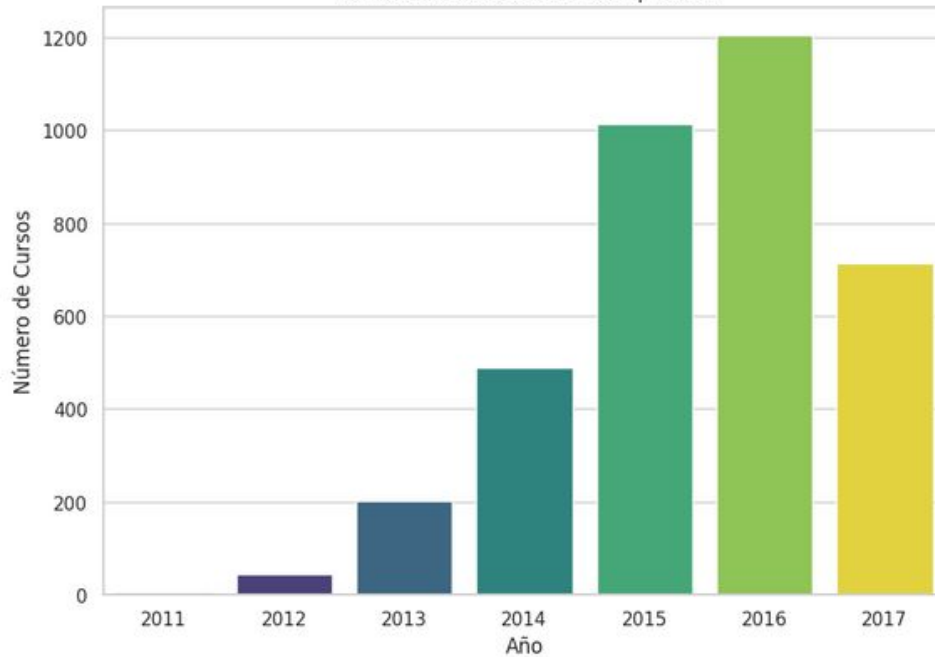


Distribución de Cursos de Pago vs Gratuitos

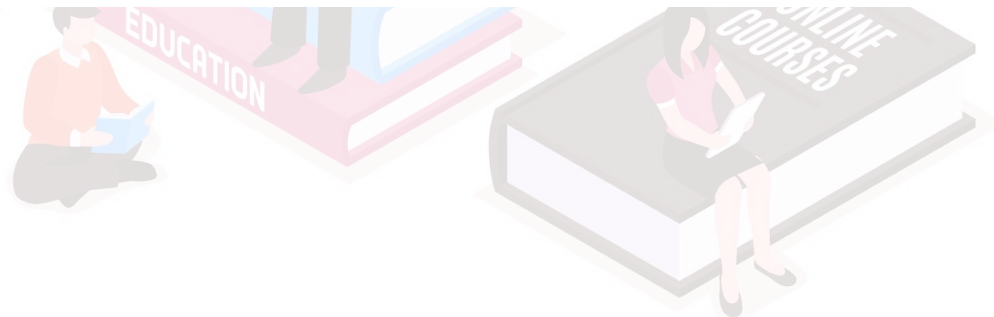
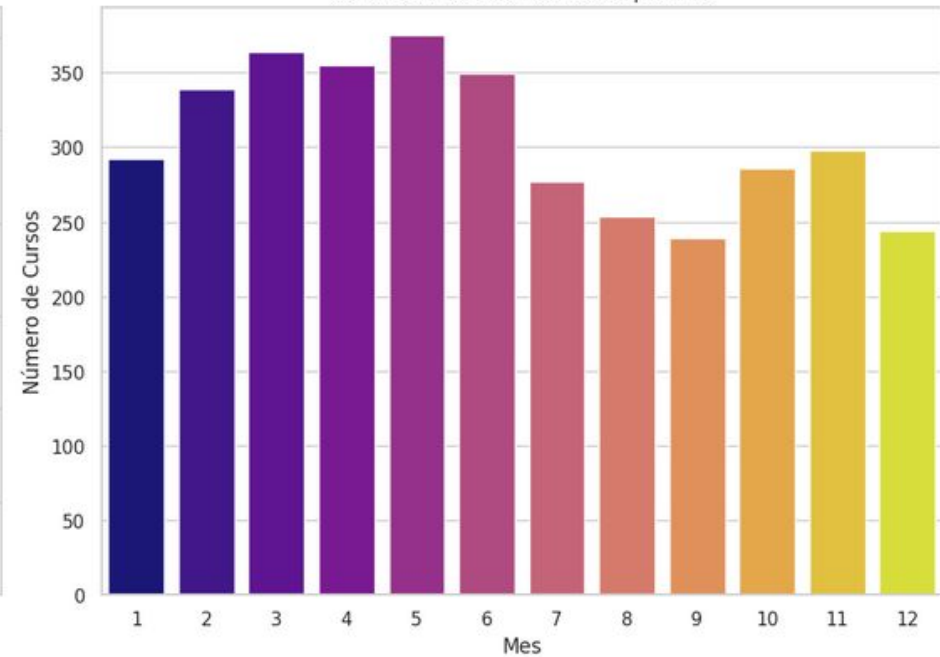


ANÁLISIS NUMÉRICO

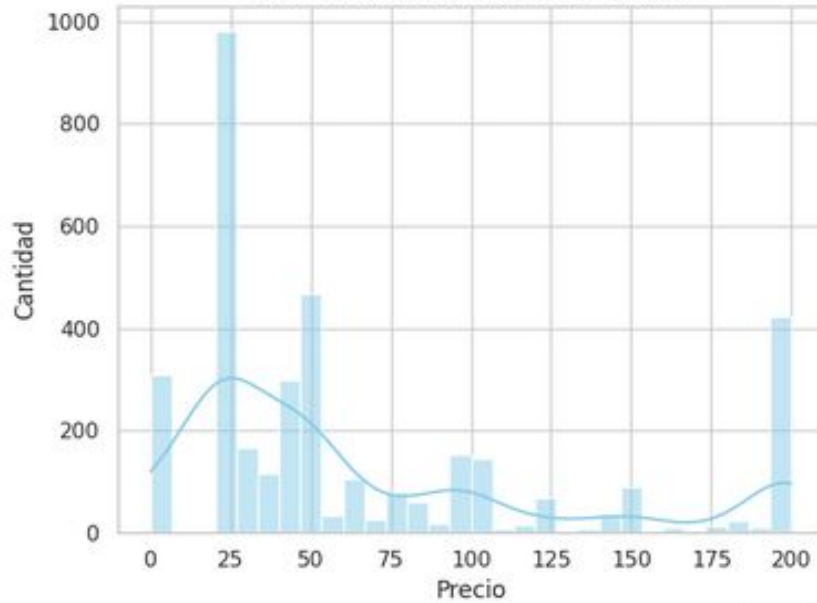
Número de Cursos Publicados por Año



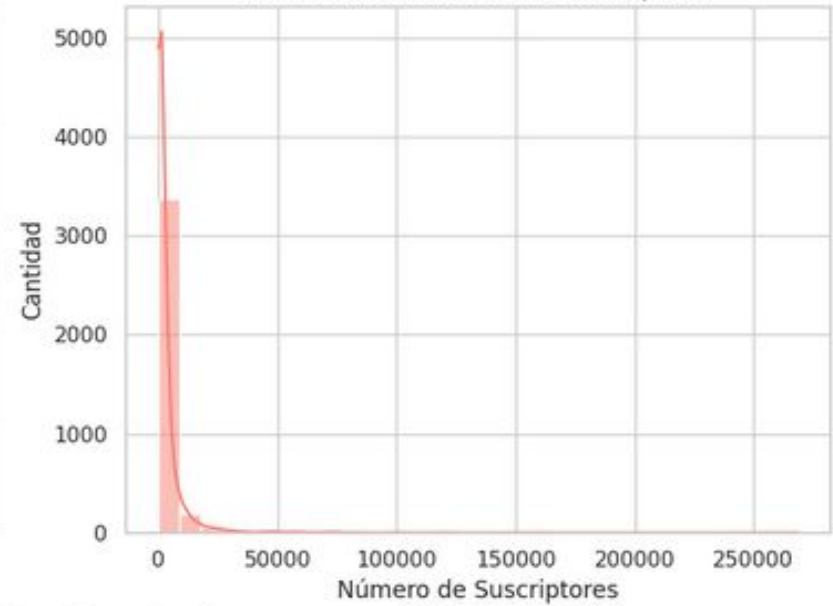
Número de Cursos Publicados por Mes



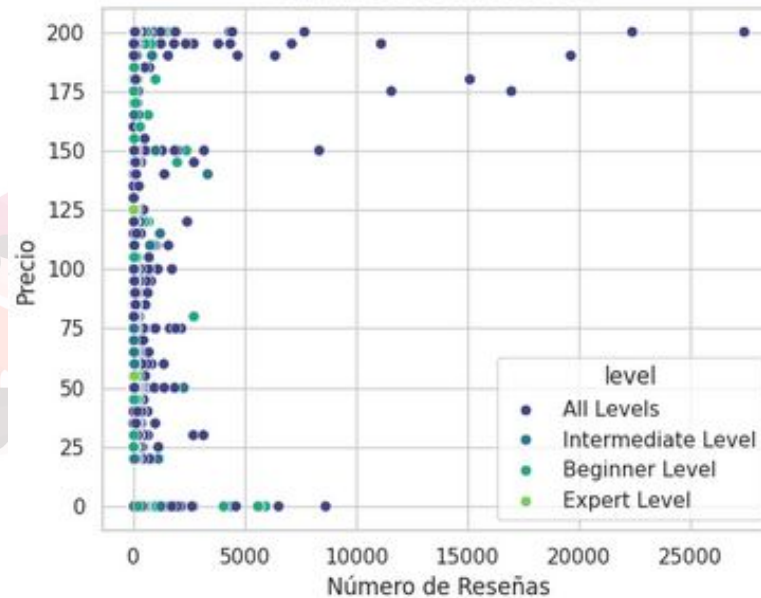
Distribución de Precios de los Cursos



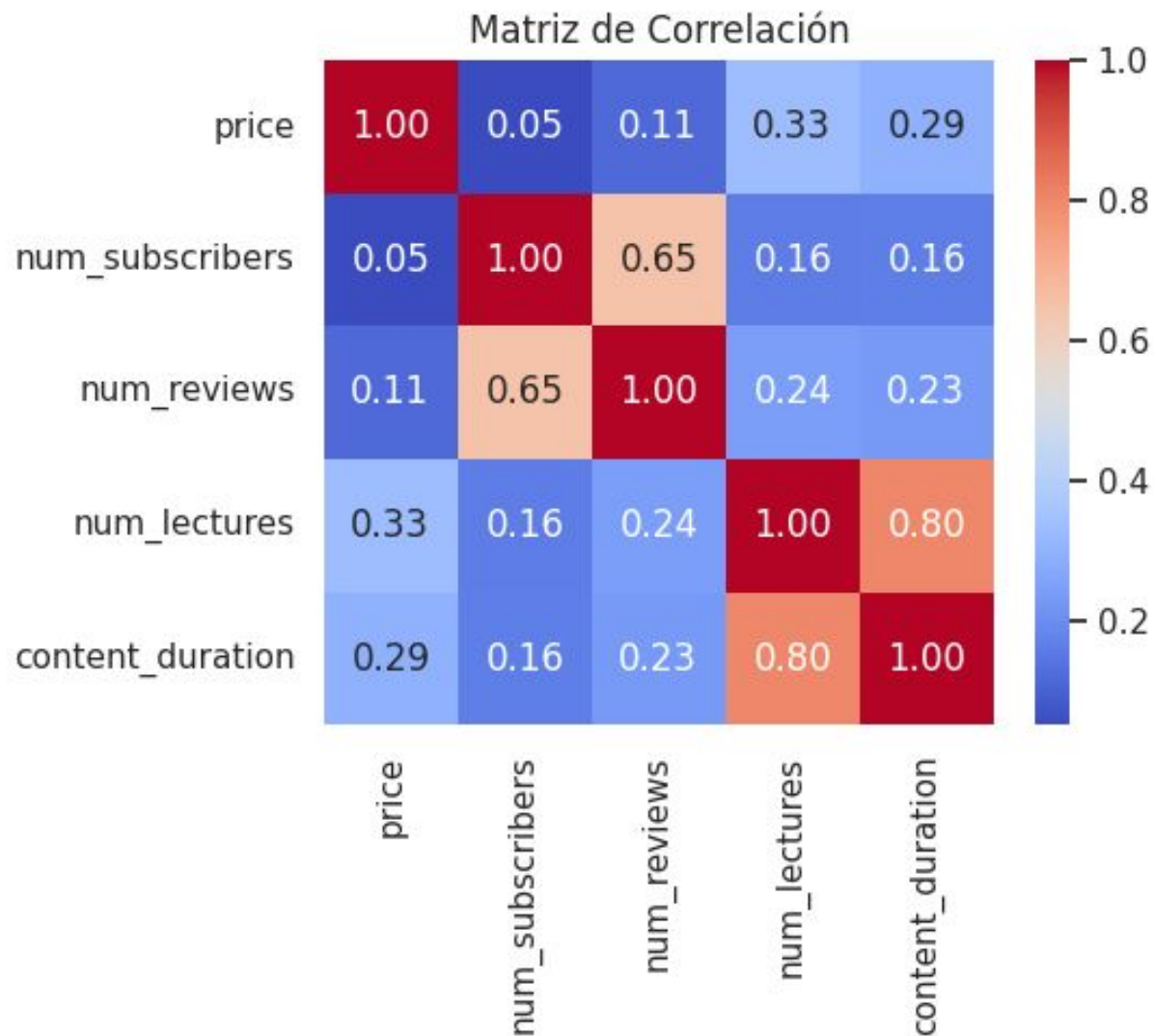
Distribución del Número de Suscriptores



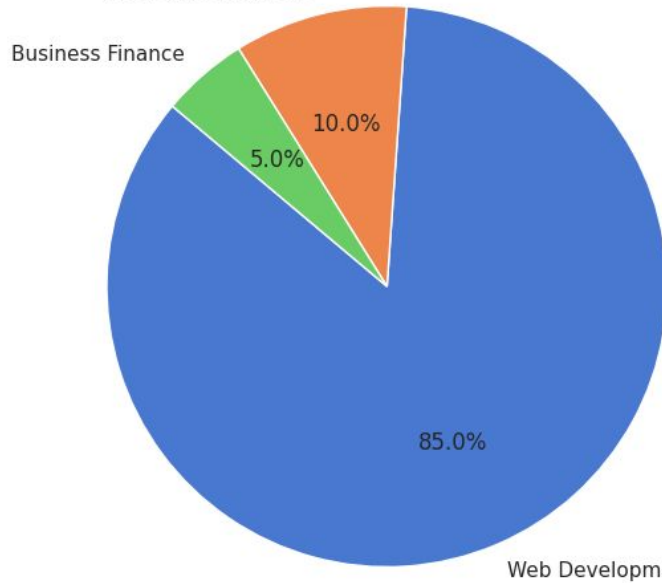
Número de Reseñas vs Precio



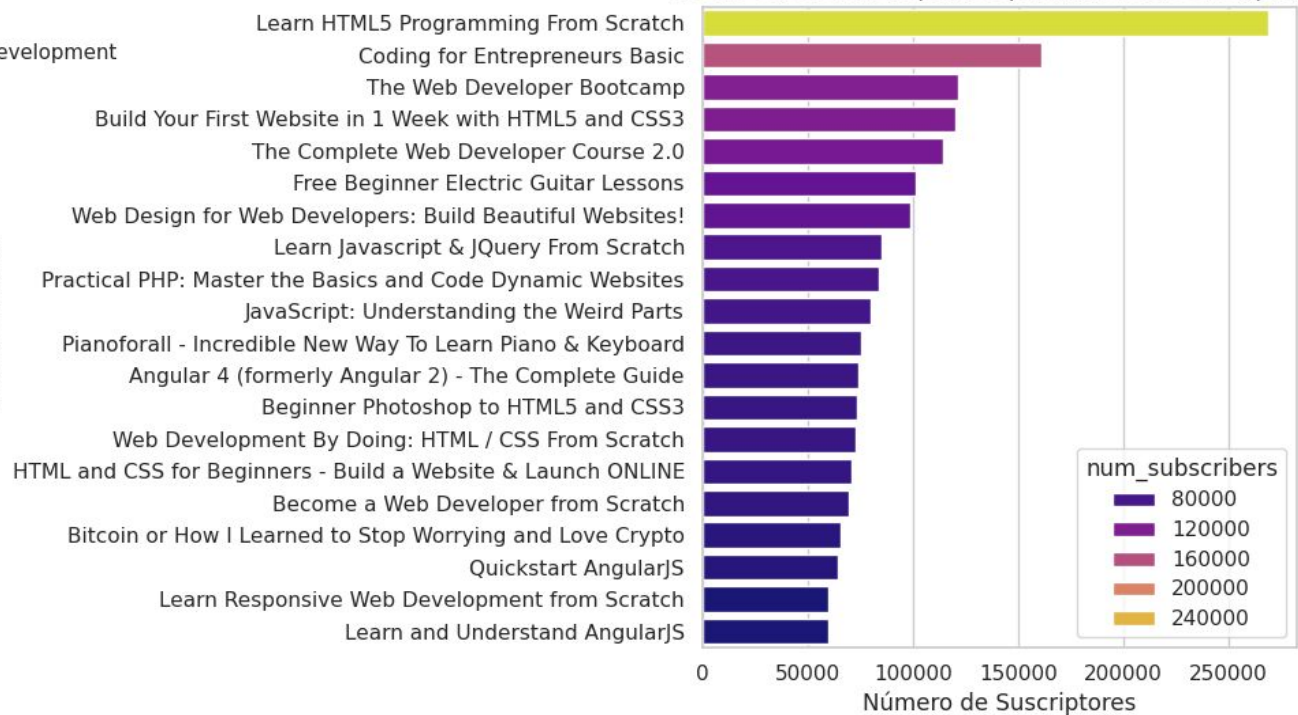
ANÁLISIS NUMÉRICO



Número de Cursos por Tema en el Top 20 de Cursos Más Populares



Los 20 Cursos Más Populares por Número de Suscriptores



Título del Curso

num_subscribers



0 50000 100000 150000 200000 250000

Número de Suscriptores

Natural Language Processing (NLP) & TF-IDF (Term Frequency-Inverse Document Frequency): técnica para evaluar la relevancia de una palabra (o token) dentro de un documento y un conjunto de documentos (corpus).

Tokenización
Vectorización

Similitud del coseno



Basado en

EJEMPLO

“The Web Developer Bootcamp”

1 SOLO ATRIBUTO		PESO DE ATRIBUTOS	
course_title		course_title	
0	The Complete Web Developer Bootcamp - Beginner...	0	The Complete Web Developer Bootcamp - Beginner...
1	Python and Django Full Stack Web Developer Boo...	1	Learn How To Become a Web Developer From Scratch
2	Become a Web Developer from Scratch	2	Become a Web Developer from Scratch
3	The Complete front end web developer Bootcamp ...	3	The Advanced Web Developer Course
4	Learn How To Become a Web Developer From Scratch	4	The Complete Web Developer Course 2.0
3 ATRIBUTOS		REGLAS DE PRECIOS	
course_title		course_title	
0	The Complete Web Developer Bootcamp - Beginner...	0	Become a Web Developer from Scratch
1	Python and Django Full Stack Web Developer Boo...	1	Bootstrap Bootcamp - Responsive web developmen...
2	Become a Web Developer from Scratch	2	HTML/CSS Bootcamp
3	The Complete front end web developer Bootcamp ...		
4	The Complete Web Developer Course 2.0		



DE PALABRAS



Exclude

RUTAS

DE APRENDIZAJE

DE APRENDIZAJE

APLICACIÓN / INTEGRACIÓN

Sistema de Recomendación de Cursos

Ingrese el título del curso:

Selecciona el nivel:

Selecciona el tipo de precio:

Obtener Recomendaciones

Sistema de Recomendación de Cursos

Ingrese el título del curso:

Selecciona el nivel:

Selecciona el tipo de precio:

Obtener Recomendaciones



Sistema de Recomendación de Cursos

Ingrese el título del curso:

Selecciona el nivel:

Selecciona el tipo de precio:

Obtener Recomendaciones

Título: The Complete Web Developer Bootcamp - Beginner to Expert

Enlace: <https://www.udemy.com/job-ready-web-developer/>

Precio: 190

Suscriptores: 14103

Nivel: All Levels

Título: Python and Django Full Stack Web Developer Bootcamp

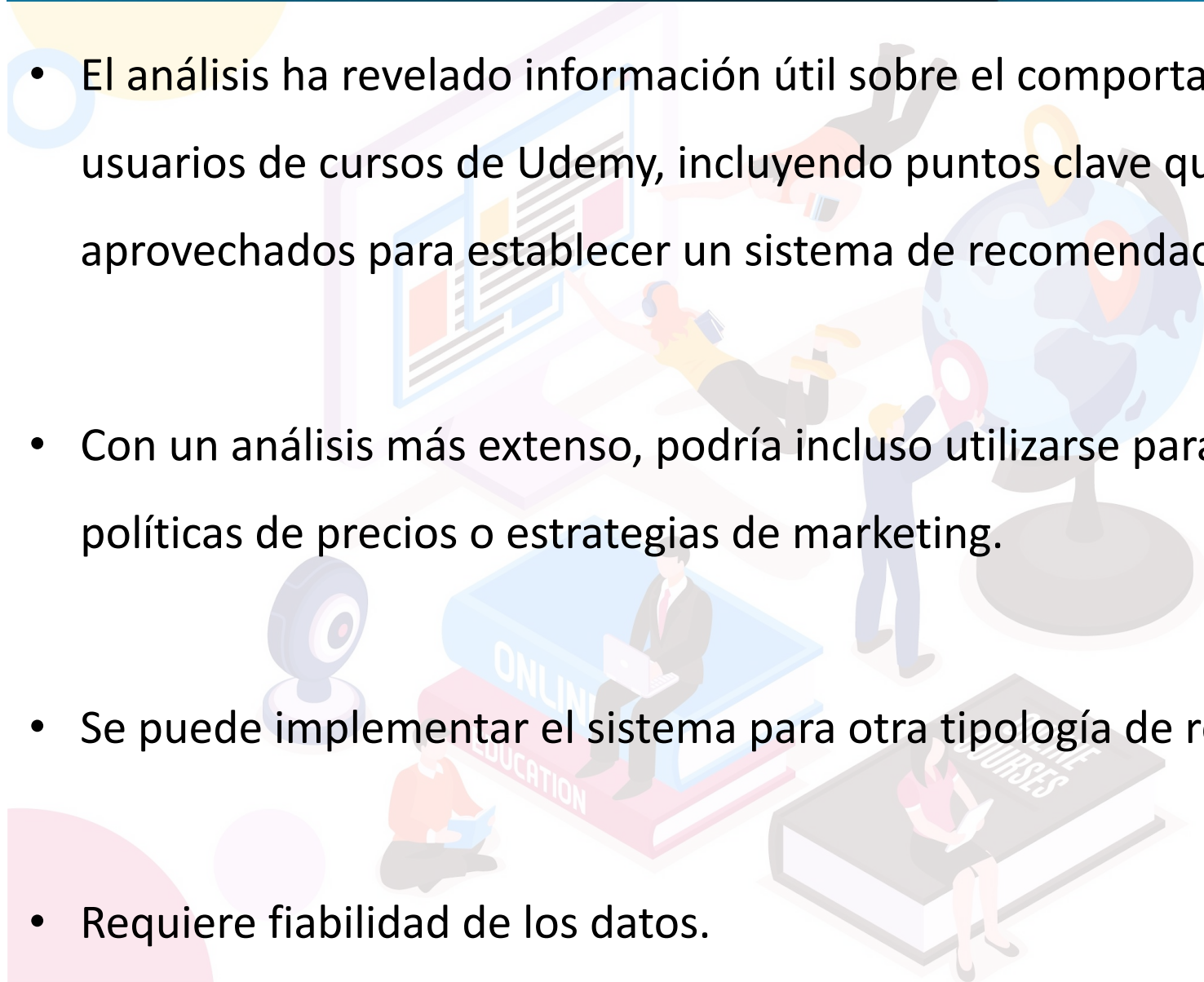
Enlace: <https://www.udemy.com/python-and-django-full-stack-web-developer-bootcamp/>

Precio: 200

Suscriptores: 11832

Nivel: All Levels

- El análisis ha revelado información útil sobre el comportamiento de los usuarios de cursos de Udeemy, incluyendo puntos clave que pueden ser aprovechados para establecer un sistema de recomendación.
- Con un análisis más extenso, podría incluso utilizarse para establecer políticas de precios o estrategias de marketing.
- Se puede implementar el sistema para otra tipología de recursos.
- Requiere fiabilidad de los datos.



DEBILIDADES

Limitación en la precisión

Riesgo de repetición

Dependencia de la vectorización

AMENAZAS

Calidad de datos

Cambios en preferencias

Escalabilidad

FORTALEZAS

Simplicidad de implementación

Accesibilidad a herramientas y bibliotecas estándar en Python

Personalización básica

OPORTUNIDADES

Mejora de la vectorización

Integración de modelos avanzados

Retroalimentación de usuarios

RecoMMachine



GRACIAS POR VUESTRA ATENCIÓN



AROA CHANS