

Análisis de las competencias alcanzadas en nuestro curso con portales de empleo que demandan DS

Ricardo Guevara

It Academy

Barcelona 2023

IT ACADEMY

Introducción

- El propósito del siguiente proyecto es comparar todas las competencias alcanzadas en el curso de **Data Science** con portales de empleo en los cuales se demandan personas con competencias en Data Science. Para tal propósito se realizará web scraping de los siguientes portales:

Introducción

- El propósito del siguiente proyecto es comparar todas las competencias alcanzadas en el curso de **Data Science** con portales de empleo en los cuales se demandan personas con competencias en Data Science. Para tal propósito se realizará web scraping de los siguientes portales:
- reed.co.uk
- itjobs.ca
- [jobscube](https://jobscube.com)

Objetivos

- Analizar todo el texto de los archivos HTML descargados.

Objetivos

- Analizar todo el texto de los archivos HTML descargados.
- Explorar los resultados analizados para saber cómo se describen las competencias laborales en los anuncios. Prestaremos especial atención a si determinadas etiquetas HTML están más asociadas a las descripciones de competencias.

Objetivos

- Analizar todo el texto de los archivos HTML descargados.
- Explorar los resultados analizados para saber cómo se describen las competencias laborales en los anuncios. Prestaremos especial atención a si determinadas etiquetas HTML están más asociadas a las descripciones de competencias.
- Intentar filtrar los anuncios de empleo relevantes de nuestro conjunto de datos.

Objetivos

- Analizar todo el texto de los archivos HTML descargados.
- Explorar los resultados analizados para saber cómo se describen las competencias laborales en los anuncios. Prestaremos especial atención a si determinadas etiquetas HTML están más asociadas a las descripciones de competencias.
- Intentar filtrar los anuncios de empleo relevantes de nuestro conjunto de datos.
- Agrupar las competencias laborales en función de su similitud textual.

Objetivos

- Analizar todo el texto de los archivos HTML descargados.
- Explorar los resultados analizados para saber cómo se describen las competencias laborales en los anuncios. Prestaremos especial atención a si determinadas etiquetas HTML están más asociadas a las descripciones de competencias.
- Intentar filtrar los anuncios de empleo relevantes de nuestro conjunto de datos.
- Agrupar las competencias laborales en función de su similitud textual.
- Visualizar los clusters mediante nubes de palabras.

Objetivos

- Analizar todo el texto de los archivos HTML descargados.
- Explorar los resultados analizados para saber cómo se describen las competencias laborales en los anuncios. Prestaremos especial atención a si determinadas etiquetas HTML están más asociadas a las descripciones de competencias.
- Intentar filtrar los anuncios de empleo relevantes de nuestro conjunto de datos.
- Agrupar las competencias laborales en función de su similitud textual.
- Visualizar los clusters mediante nubes de palabras.
- Comparar las competencias agrupadas con la tabla de contenidos de nuestro curso de Data Science para descubrir las competencias que faltan.

Metodología

title_html	body_html	li_body	non_li_body
Data Science Manager Leeds - Reed.co.uk	Data Science Manager Data Science Manager Up t...	[Leading data science projects under the direc...	Data Science Manager Data Science Manager Up t...
Data Science Engineer London - Reed.co.uk	Data Science Engineer Data Science Engineer LO...	[5 Years' experience im- plementing within Data ...	Data Science Engineer Data Science Engineer LO...
Head of Data Science London - Reed.co.uk	Head of Data Science Head of Data Sci- enceLondo...	[Manage a small team of 3 Data Scientists incl...	Head of Data Science Head of Data Sci- enceLondo...
Python Quantitative Researcher - London- Globa...	Python Quantitative Researcher - London- Globa...	[Minimum 2+ years' ex- perience, Hands-on pro- gra...	Python Quantitative Researcher - London- Globa...
Engineering Lead London - Reed.co.uk	Engineering Lead Engi- neering Manager Fin- Tec...	[Lead/manage a diverse team of up to 30+ engin...	Engineering Lead Engi- neering Manager Fin- Tec...

Table 1: Data frame con los datos de los 4308 HTML obtenidos

Analizar datos

	title_html	body_html
count	4308	4308
unique	3889	3929
top	Functional Analyst, HRMS Job in Edmonton	Product Specialist Description JOB SUMMA-
freq	for S...	RY The...
	9	23

Table 2: Estadística descriptiva

3929 cuerpos (body_html) son únicos, por lo que 379 son duplicados. Por lo tanto, debemos eleiminar los duplicados de la columna body_html aunque algunas ofertas compartan un título genérico común.

Analizar viñetas

Palabras	Suma TFIDF
experience	1958.677833
data	1757.298992
skills	1044.849553
work	992.226009
science	893.043257

(a) Columna **body_li**

Palabras	Suma TFIDF
data	246.002726
team	122.704533
experience	116.499467
work	110.406079
science	108.221937

(b) Columna **body_non_li**

Table 3: Palabras mejor clasificadas

Filtrar empleos por relevancia

title_html	body_html	li_body	non_li_body	relevancy
Amazon.com Services LLC - Data Science Manager...	Data Science ManagerDescriptionOperations Tech...	[]	Data Science ManagerDescriptionOperations Tech...	0.219130
Junior Data Scientist Manchester - Reed.co.uk	Junior Data Scientist Title: Junior Data Scien...	[Collect, clean, and preprocess data to make i...	Junior Data Scientist Title: Junior Data Scien...	0.177863
Data Scientist East London - Reed.co.uk	Data Scientist JD For Data Scientist Responsib...	[4+ years of relevant experience specialized i...	Data Scientist JD For Data Scientist Responsib...	0.177429
Middlebury College - Data Services Librarian j...	Data Services LibrarianJob DescriptionThe Data...	[Develop strong working relationships with fac...	Data Services LibrarianJob DescriptionThe Data...	0.174417
Affinity - Data Scientist job in Cleveland OH ...	Data ScientistDescriptionThe primary focus wil...	[Lead discovery processes with business stakeh...	Data ScientistDescriptionThe primary focus wil...	0.152593

Table 4: Data frame con la columna **relevancy** en orden descendente

Empleos por la relevancia del título

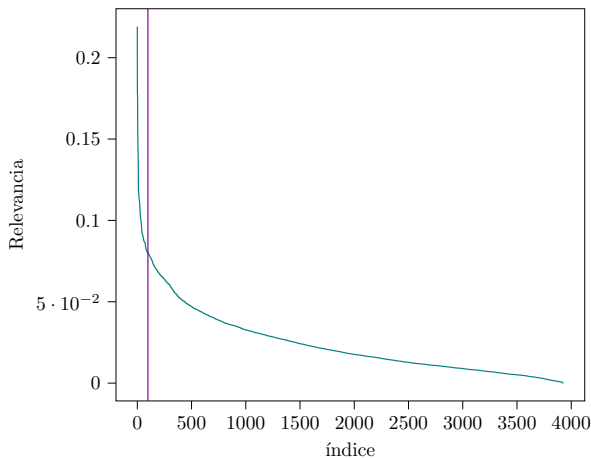


Figure 1: Gráfico con una línea vertical en 98. Los índices menores a 98 son títulos de empleos con relevancia mucho más elevados.

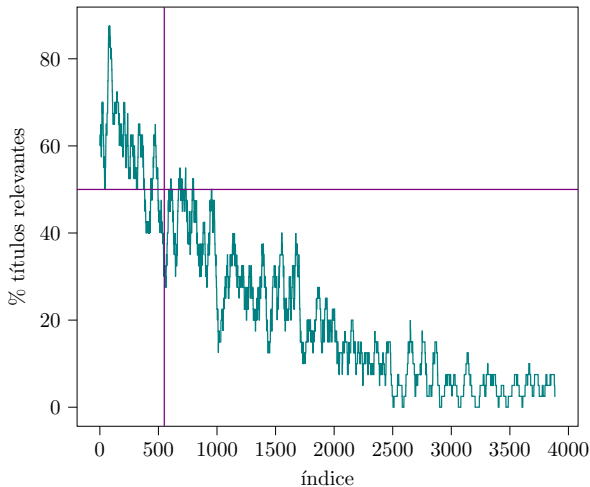


Figure 2: Una línea horizontal marca el 50% de relevancia, la línea vertical marca un índice de 550.

Agrupación de los 98 empleos relevantes

¿Qué hacer?

- Agrupar los 98 textos en `ds_jobs[:98].body_html`.
- Agrupar los cientos de viñetas individuales en `ds_jobs[:98].li_body`.

Agrupación de los 98 empleos relevantes

¿Qué hacer?

- Agrupar los 98 textos en `ds_jobs[:98].body_html`.
- Agrupar los cientos de viñetas individuales en `ds_jobs[:98].li_body`.

La segunda opción es preferible por las siguientes razones:

- Las viñetas se centran más en las competencias individuales que en el cuerpo heterogéneo de cada puesto.
- Las viñetas cortas son fáciles de imprimir y leer. No ocurre lo mismo con los textos más extensos.

98 empleos relevantes

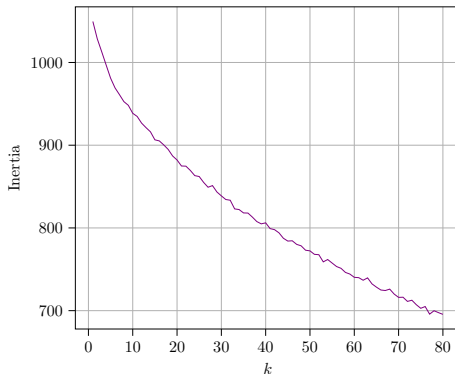


Figure 3: Gráfico de codo generado mediante K-means, con valores de k comprendidos entre 1 y 80. La ubicación exacta del codo es difícil de determinar. Utilizando el método KneeLocator encontramos que $k = 27$.

550 empleos relevantes

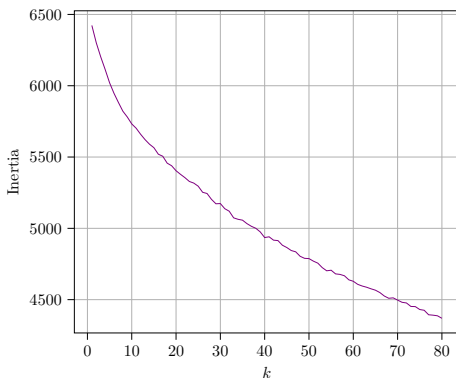
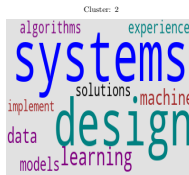
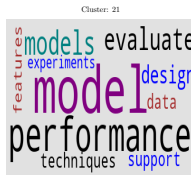
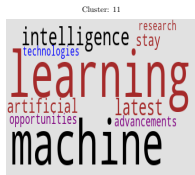
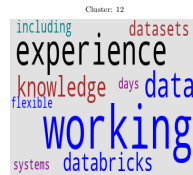
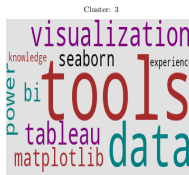
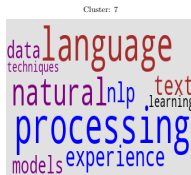
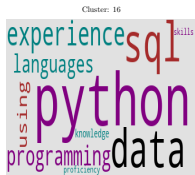
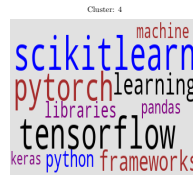
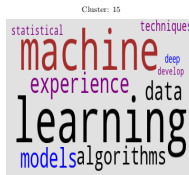
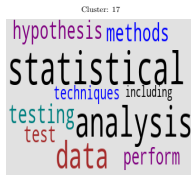
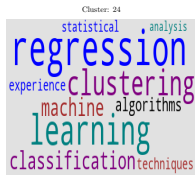


Figure 4: Gráfico de codo generado mediante K-means, con valores de k comprendidos entre 1 y 80. La ubicación exacta del codo es difícil de determinar. Utilizando el método KneeLocator encontramos que $k = 29$.

12 primeros clusters técnicos



15 clusters no técnicos



Clusters de los 550 empleos más relevantes



Conclusiones

En primer lugar, podemos ver lo importante que son nuestros conocimientos en Python como científicos de datos. Además, podemos destacar con orgullo nuestras habilidades de aprendizaje automático alcanzadas en nuestro curso.

Conclusiones

En primer lugar, podemos ver lo importante que son nuestros conocimientos en Python como científicos de datos. Además, podemos destacar con orgullo nuestras habilidades de aprendizaje automático alcanzadas en nuestro curso.

Competencias por alcanzar

- 1 Herramientas de Business Intelligence tales como Tableau y PowerBi

Conclusiones

En primer lugar, podemos ver lo importante que son nuestros conocimientos en Python como científicos de datos. Además, podemos destacar con orgullo nuestras habilidades de aprendizaje automático alcanzadas en nuestro curso.

Competencias por alcanzar

- 1 Herramientas de Business Intelligence tales como Tableau y PowerBi
- 2 Frameworks tales como Tensorflow o Pytorch

Conclusiones

En primer lugar, podemos ver lo importante que son nuestros conocimientos en Python como científicos de datos. Además, podemos destacar con orgullo nuestras habilidades de aprendizaje automático alcanzadas en nuestro curso.

Competencias por alcanzar

- 1 Herramientas de Business Intelligence tales como Tableau y PowerBi
- 2 Frameworks tales como Tensorflow o Pytorch
- 3 Herramientas para obtener y almacenar datos remotos. Estas herramientas incluyen bases de datos y servicios de alojamiento web

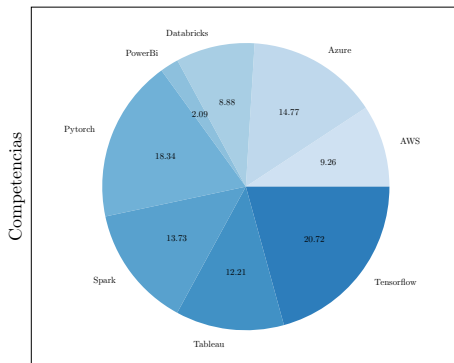
Conclusiones

En primer lugar, podemos ver lo importante que son nuestros conocimientos en Python como científicos de datos. Además, podemos destacar con orgullo nuestras habilidades de aprendizaje automático alcanzadas en nuestro curso.

Competencias por alcanzar

- 1 Herramientas de Business Intelligence tales como Tableau y PowerBi
- 2 Frameworks tales como Tensorflow o Pytorch
- 3 Herramientas para obtener y almacenar datos remotos. Estas herramientas incluyen bases de datos y servicios de alojamiento web
- 4 Habilidades comunicativas
 - Colaborar con compañeros de equipo en un problema de datos difícil
 - Transmitir resultados complejos, de forma oral o escrita, a un público no técnico

Competencias por alcanzar



Gracias!