

**"STOP DREAMING
AND START
DOING"**



ADALAB: ESTUDIO DE MERCADO





Índice:

1. Objetivos
2. Proyecto
3. Pasos realizados
4. Gráficas marketing
5. Gráficas empleabilidad
6. Gráficas formacion
7. Conclusiones
 - 7.1. Próximos pasos

1. Objetivo:

Nuestro cliente, Adalab, es una escuela tecnológica dirigida exclusivamente a personas que se identifican como mujeres, y que tiene un objetivo muy claro: reducir la brecha de género que se puede encontrar actualmente en el sector tecnológico.

Adalab nos ha contratado porque desea obtener un conocimiento profundo del sector de *Data Analytics*, ya que tiene interés en saber si está ofreciendo a sus alumnas la formación más actualizada y adaptada a lo que se está empleando actualmente en dicho sector. De esta manera, podría adaptarse a las nuevas corrientes dentro del sector, si fuera necesario, para poder seguir proporcionando a sus alumnas la formación de mayor calidad.

2. Proyecto

Nuestro equipo ha decidido enfocar el análisis en tres vertientes:

- Marketing: Hemos realizado un estudio del mercado para comprobar que el objetivo de Adalab es el correcto. Además, hemos estudiado lo que podría ser interesante para incluir en sus campañas de Marketing.
- Herramientas: Hemos realizado un estudio de las herramientas, lenguajes de programación, etc. que se están empleando actualmente en el sector, para comprobar si Adalab está ofreciendo a sus alumnas la formación más actualizada posible.
- Empleabilidad: Hemos realizado un estudio sobre el tipo de empresa en el que se encuentran trabajando las personas de este sector, ya que es relevante para el enfoque de Adalab.

Como resultado del análisis de estados vertientes, hemos podido realizar un estudio muy completo de la situación actual del mercado para poder ofrecer a nuestro cliente la mayor información posible en base a los datos que nos han entregado.

3. Pasos realizados

1. Análisis exploratorio de los datos: Hemos realizado un estudio de los datos proporcionados por el cliente para conocer su estructura y la relación existente entre los diferentes archivos. Además, nos permitió conocer qué tipo de limpieza iba a ser necesaria.

2. Unión de los diferentes archivos: El cliente entregó los datos separados en diferentes archivos para una misma encuesta, así que los unificamos para poder trabajar mejor con ellos.

3. Limpieza de los datos: Eliminamos símbolos raros, imputamos nulos, etc., dejando los datos lo más legibles posible.

3. Pasos realizados

4. Planteamiento de cuestiones para realizar el estudio: Tras analizar los datos y limpiarlos, planteamos diferentes preguntas para poder enfocar y realizar el estudio posteriormente.

5. Realización de gráficas: Tras plantearnos diferentes preguntas, realizamos gráficas para poder comprobar, o refutar, nuestras hipótesis propuestas en el anterior punto.

6. Establecimiento de conclusiones: Después de realizar las gráficas, pudimos sacar las conclusiones necesarias en base a nuestro análisis.

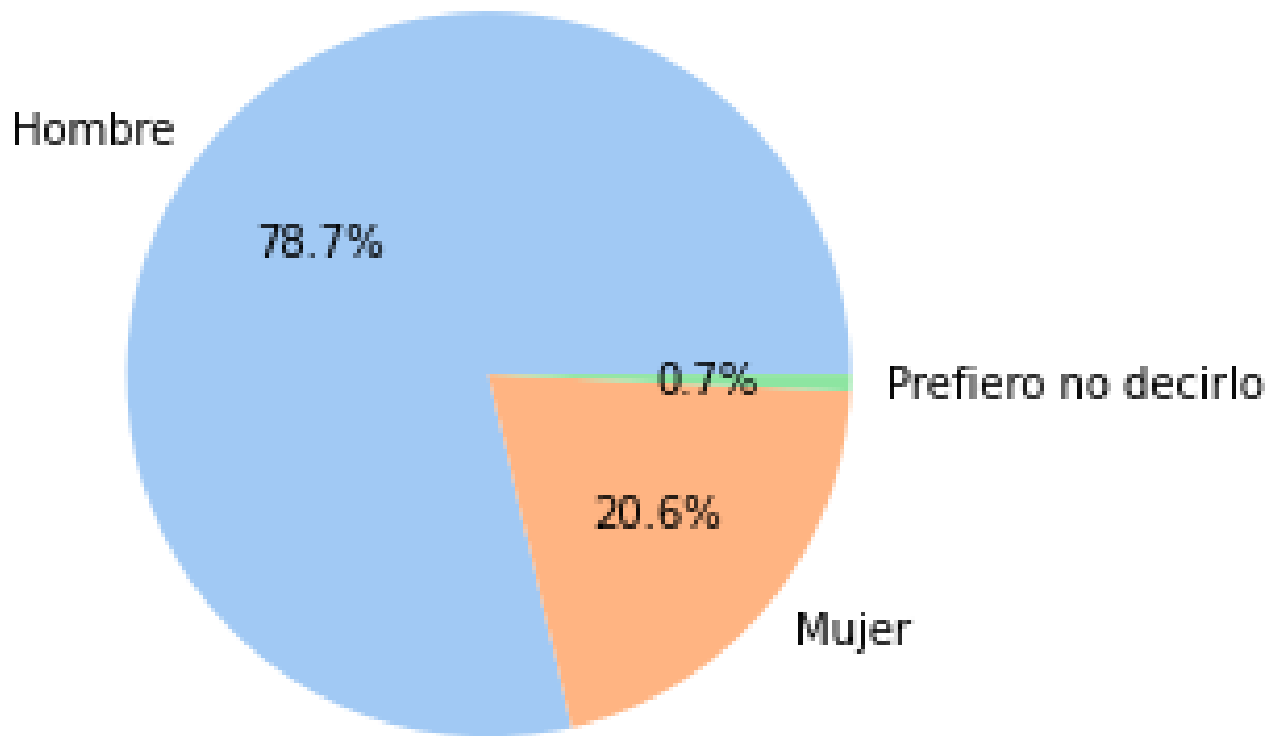


GRÁFICAS

MARKETING

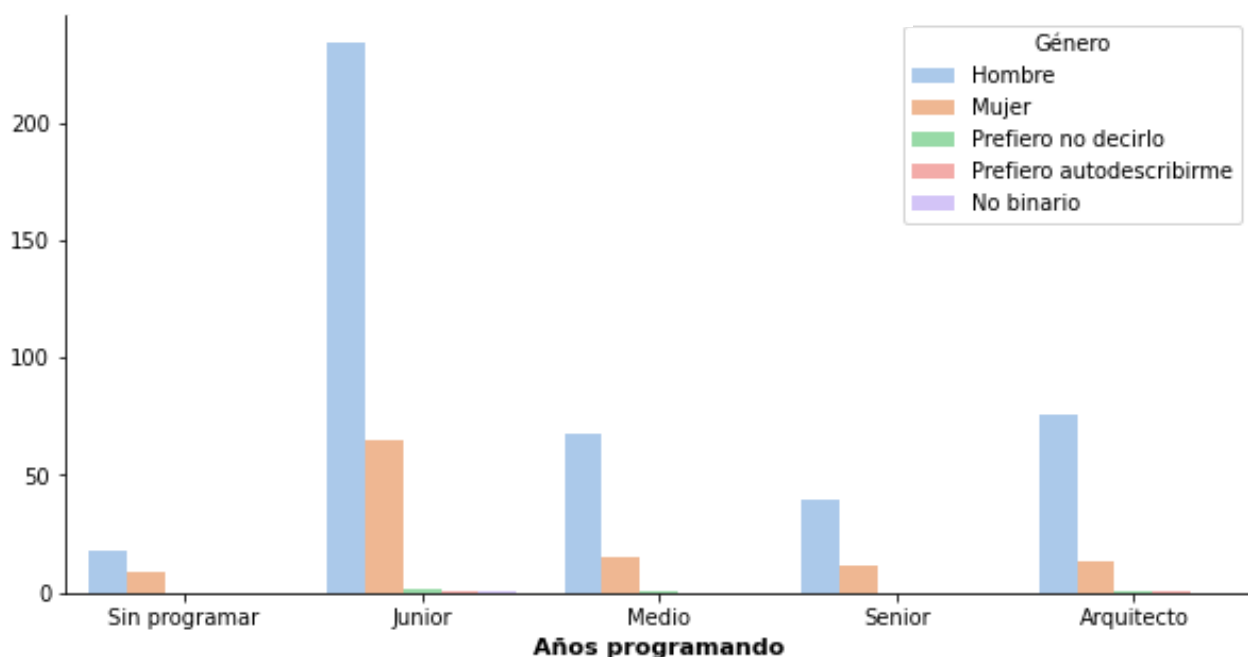


BRECHA DE GÉNERO EN EL SECTOR



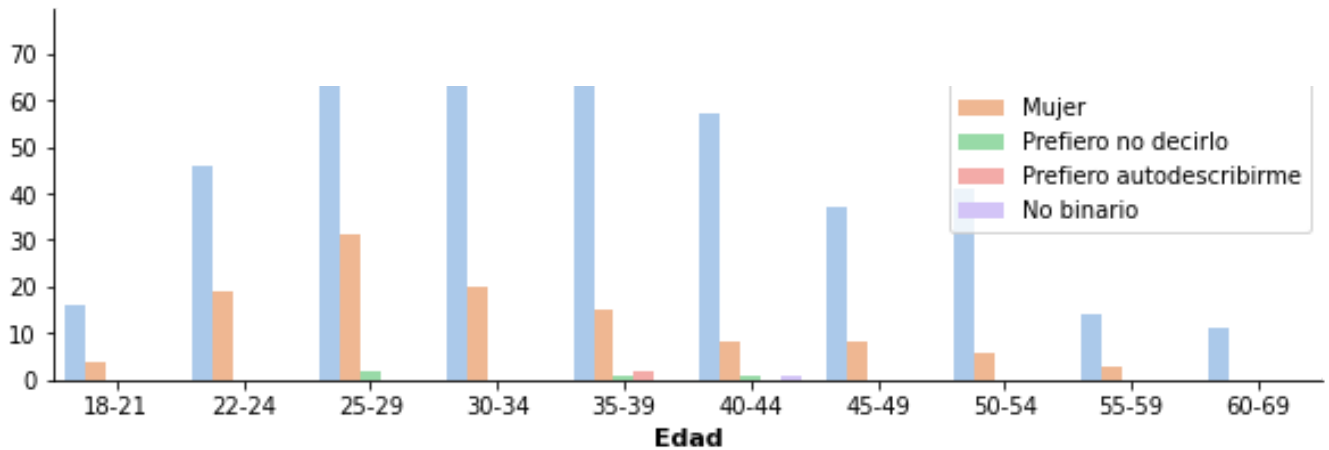
Se puede apreciar que hay una enorme brecha de género en el sector, ya que casi el 80% de los encuestados se identifican como hombres, mientras que apenas el 20% se identifican como mujeres, y hay un porcentaje minúsculo de personas que se identifican como no binarias. Por tanto, la misión que tiene Adalab, cuyo objetivo es reducir la brecha de género en este sector, está muy bien encaminada, ya que esta brecha está todavía muy latente.

AÑOS PROGRAMANDO POR GÉNERO



En este gráfico se puede observar que la mayoría de personas encuestadas, aparte de identificarse como hombres, tienen una experiencia junior dentro del sector, es decir, llevan programando entre 0 y 3 años. Esto podría indicar que es un sector relativamente joven, y que está empezando a coger especial fuerza en estos momentos. Por tanto, parece que la apuesta de Adalab por formar a futuras *Data Analyst* es apropiada.

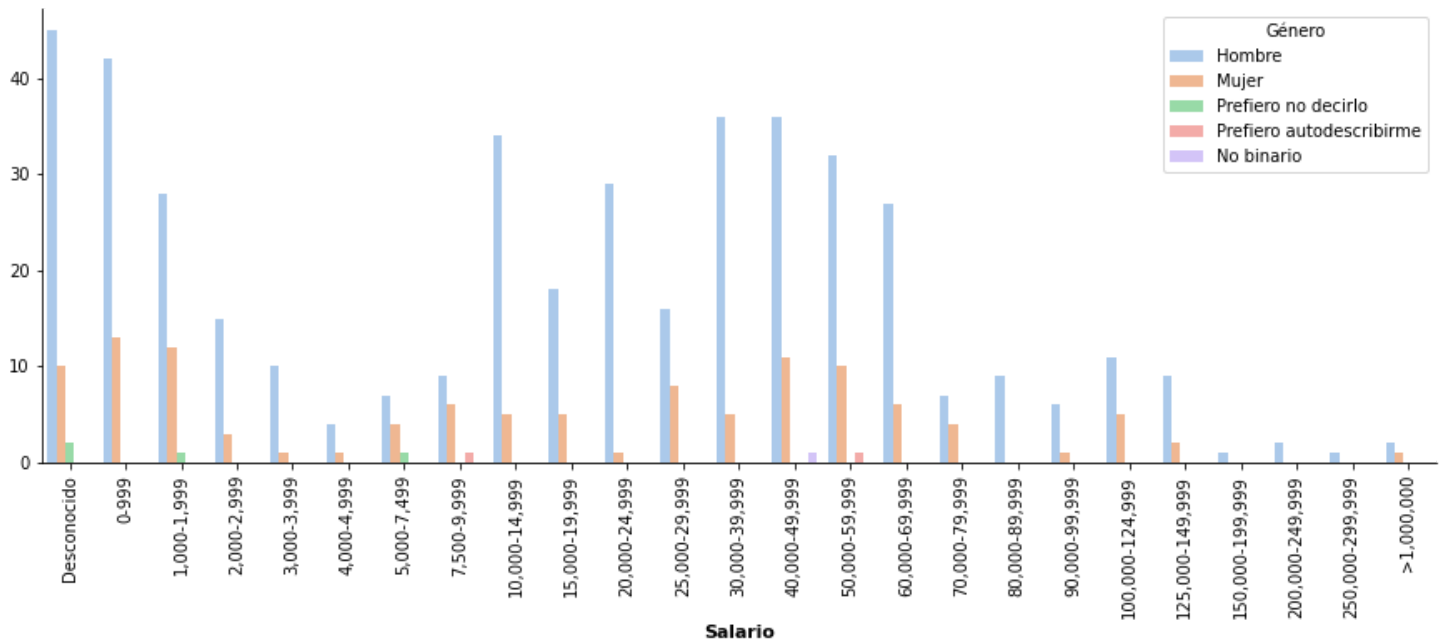
EDADES POR GÉNERO EN EL SECTOR



Este gráfico permite ver que el sector es un sector joven, ya que el grueso de los encuestados se encuentra entre los 25 y los 45 años. Además, es interesante que se mantenga una tendencia similar en edad entre hombres y mujeres, lo que permite sacar la conclusión de que, en este sector, las mujeres no se ven obligadas a abandonar su puesto de trabajo para dedicarse a tareas correspondientes a los cuidados, como el cuidado de los hijos o personas dependientes, como es la tendencia usual.

Además, hemos realizado un breve estudio de ofertas actual en el que se puede ver que este sector ofrece buenas oportunidades de conciliación laboral, como el teletrabajo o trabajo híbrido, o ayudas para el cuidado de los hijos.

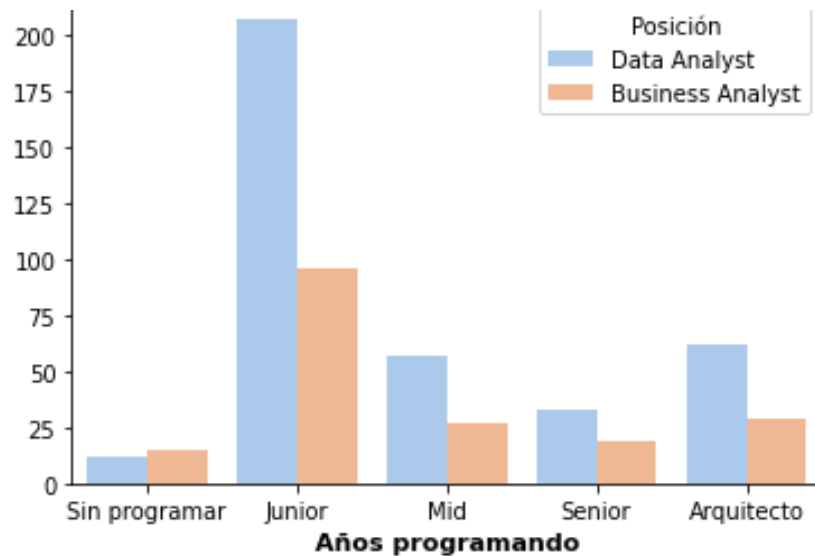
SALARIO POR GÉNERO EN EL SECTOR



Este gráfico muestra que el salario medio en Europa en el sector parece estar entre los 10,000 y los 60,000\$ anuales. Hay salarios anuales muy bajos, que podrían corresponder a salarios de prácticas o becarios, aunque también podría corresponderse a una equivocación del encuestado al considerar que se le preguntaba por salario mensual en vez de anual. Además, también hay salarios muy altos, entre los 150,000 y más de un millón de dólares, aunque sólo se benefician de estos sueldos un porcentaje muy pequeño de personas. Como dato relevante, el salario medio de la UE se encuentra en 2,194€ mensuales en 2022, pero varía mucho por países, ya que el mínimo es Bulgaria, con 650€ mensuales, y el máximo es el de Luxemburgo, con 3,713€ al mes.

No parece que haya una diferencia significativa en el salario entre hombres y mujeres en este sector, ya que parece las curvas tienen más o menos la misma tendencia en ambos géneros.

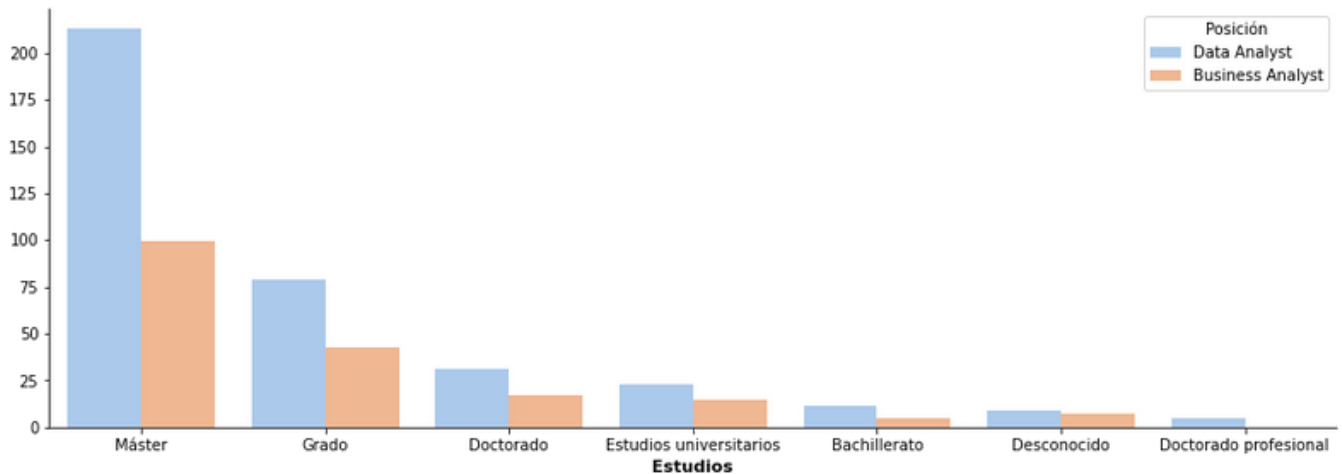
EXPERENCIA POR PUESTO



Este gráfico muestra los años programando por puesto de *Data Analyst* y de *Business Analyst*. De nuevo, se puede ver que es un sector relativamente joven y que está empezando a coger ahora mucha fuerza ya que la mayoría de encuestados llevan programando menos de tres años. En este caso, hemos dividido por posición para saber si ambos puestos son igual de jóvenes, y sí que parece que ambos puestos siguen la misma tendencia.

Aunque somos conscientes de que en esta categorización puede resultar simplista, hemos considerado que menos de 3 años es junior, menos de 6 años es Mid, menos de 10 es Senior y mas de 10 es Arquitecto.

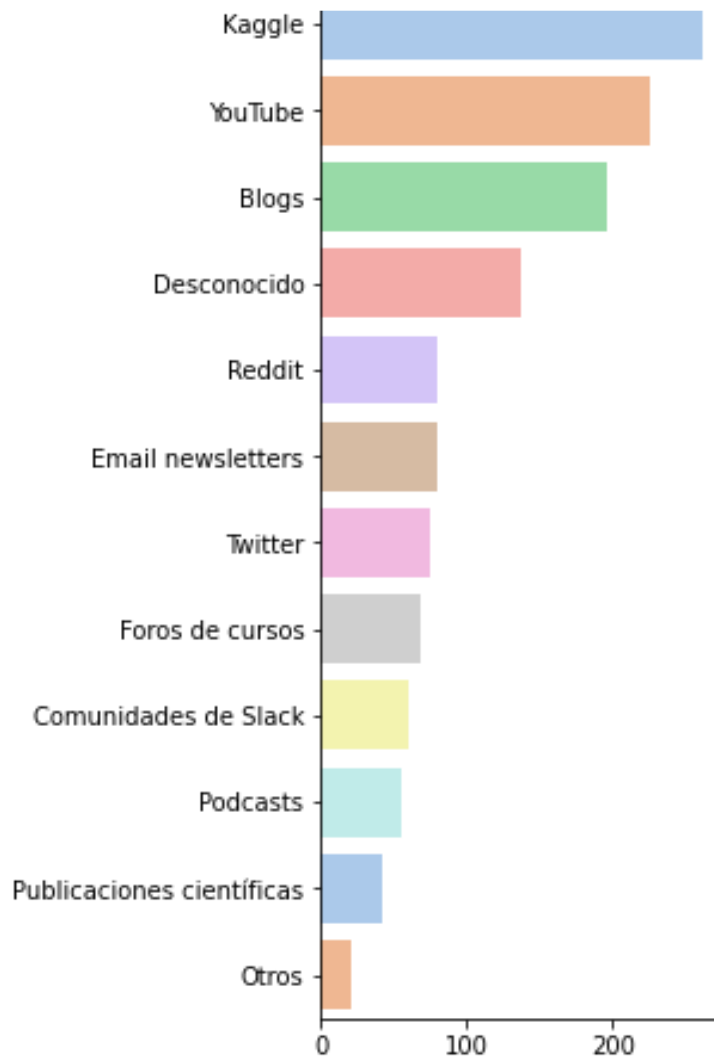
ESTUDIOS POR PUESTO



En este gráfico se puede ver que la mayoría de personas en el sector han alcanzado un nivel de estudios de Máster o similar, seguido del grado universitario.

Sería interesante ver si los estudios de máster y/o grado están relacionados con el sector o si, por el contrario, no están relacionados y simplemente han reinventado su carrera profesional.

REDES SOCIALES



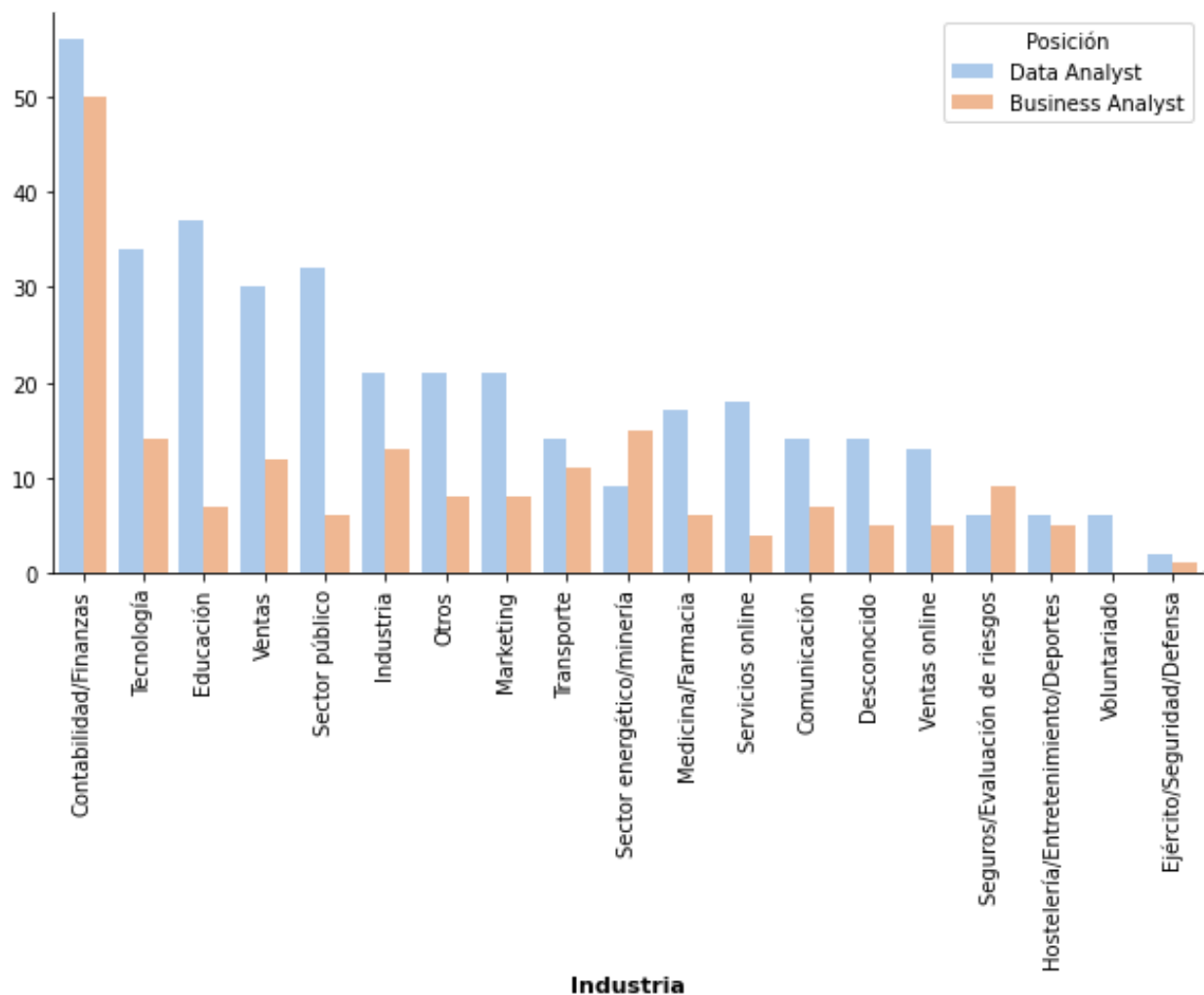
Los lugares más comúnmente empleados para mantenerse al día de las actualizaciones del sector son Kaggle, Youtube y blogs. Esto podría ser empleado por Adalab para centrar sus estrategias de marketing en estas redes sociales. Así, las personas que ya se encuentran dentro del sector podrían recomendar este *bootcamp* a personas que quieren experimentar esa reinversión laboral. Se utilizan estas redes sociales porque ofrecen contenido de alta calidad y en diferentes formatos y tienen una comunidad activa. Además, están en constante actualización y son de fácil acceso.



GRÁFICAS EMPLEABILIDAD

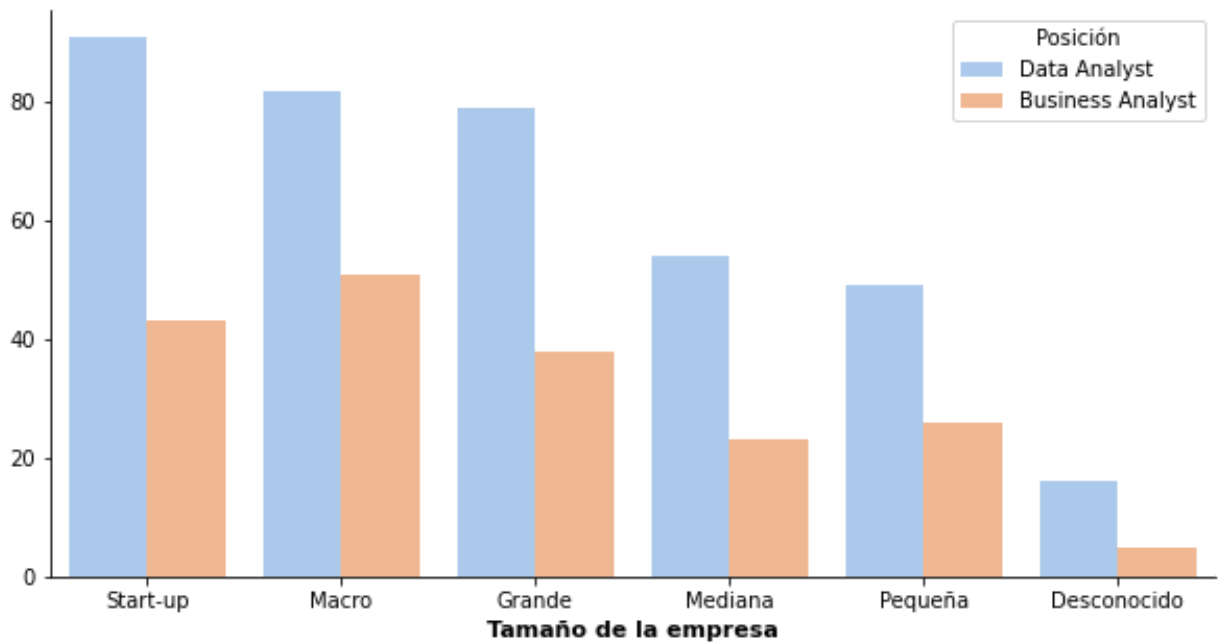


PERSONAS POR TIPO DE INDUSTRIA



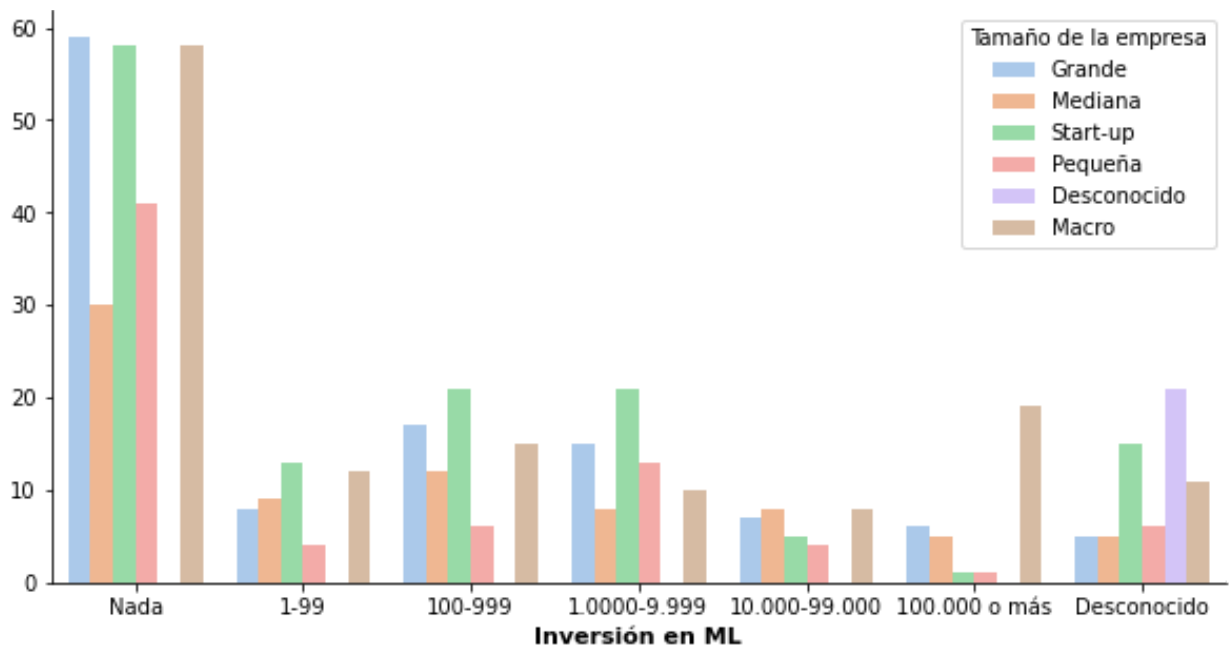
En esta gráfica, se puede ver que las posiciones de *Data Analyst* y *Business Analyst* se concentran, sobre todo, en los sectores de Contabilidad/Finanzas y Tecnología, por lo que es interesante tenerlo en cuenta a la hora de captar mentores en estos sectores, ya que tendrán más contactos para que las alumnas puedan hacer *networking*. Estos sectores pueden tener más peso ya que se basan mucho más en los datos y/o tecnología para el desarrollo de su actividad principal.

PERSONAS POR TAMAÑO DE LA EMPRESA



Las start-ups son las empresas donde más trabajan estas posiciones, seguido de las macro empresas y las grandes empresas. Las macro y grandes empresas son las que tienen la capacidad de invertir en este tipo de servicios para poder mejorar su negocio y tienen equipos más grandes, mientras que las start-up son aquellas que entran en el sector y necesitan personas que estén cualificadas para estos puestos.

INVERSIÓN EN ML POR TAMAÑO DE LA EMPRESA



Se puede observar que la mayoría de las empresas, por lo general, no están invirtiendo en Machine Learning, independientemente del tamaño de la empresa. No obstante, también se puede ver que las start-up destacan en las inversiones de menor cantidad, pero las empresas que más invierten en ML son las macro empresas, ya que son las que pueden permitírselo. Esto tiene lógica, ya que no todas las empresas tienen la capacidad de realizar grandes inversiones en ello. Parece que sí que hay una tendencia a invertir en Machine Learning, por lo cual el hecho de que Adalab enseñe a sus alumnas estos procesos es altamente positivo para su futuro.

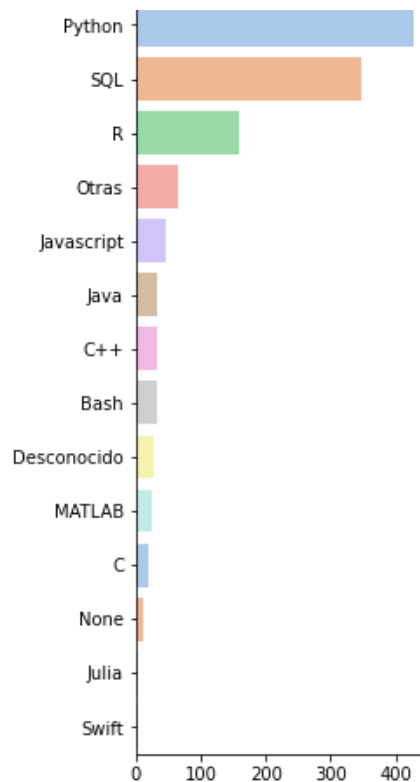


GRÁFICAS

HERRAMIENTAS



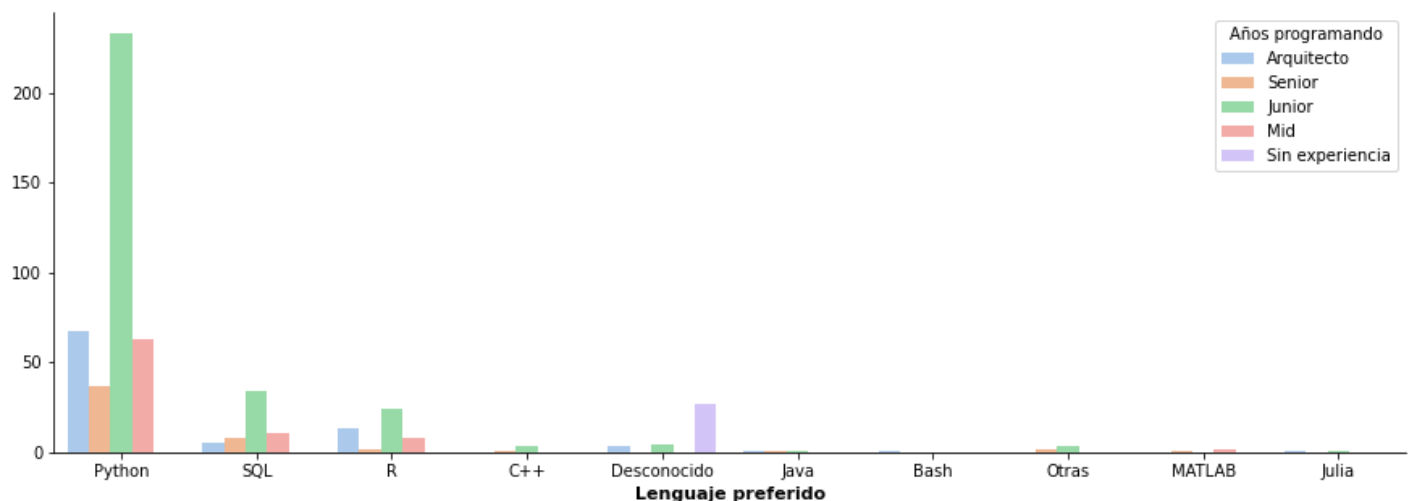
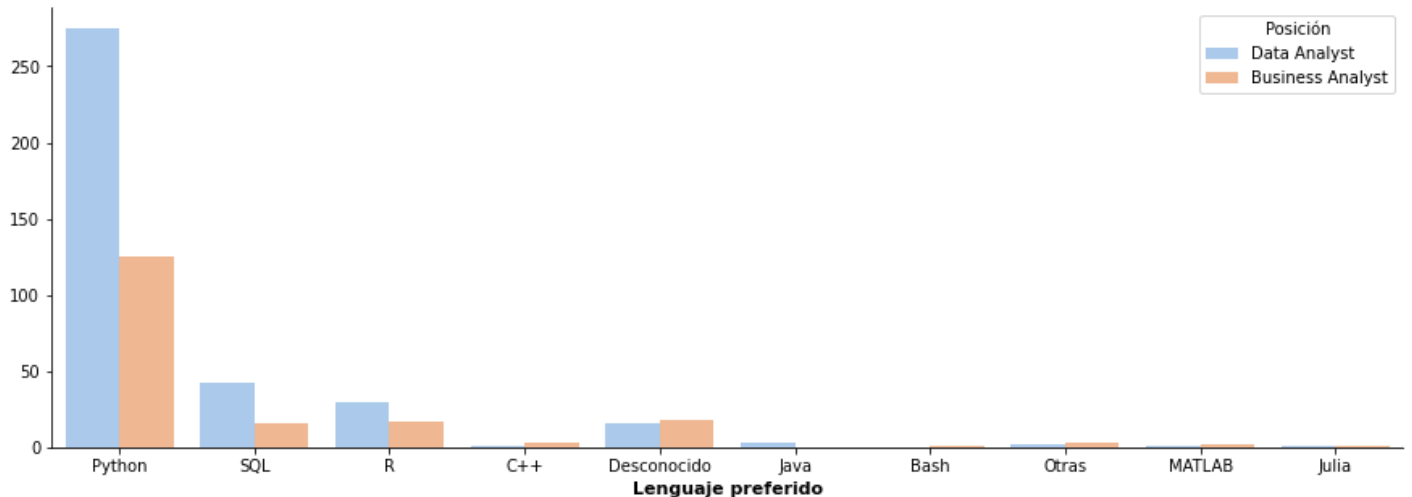
LENGUAJES DE PROGRAMACIÓN MÁS USADOS



En esta gráfica, se ve que Python y SQL son los lenguajes más empleados en el sector. Por tanto, Adalab está ofreciendo a sus alumnas los conocimientos que se están empleando actualmente en este mercado.

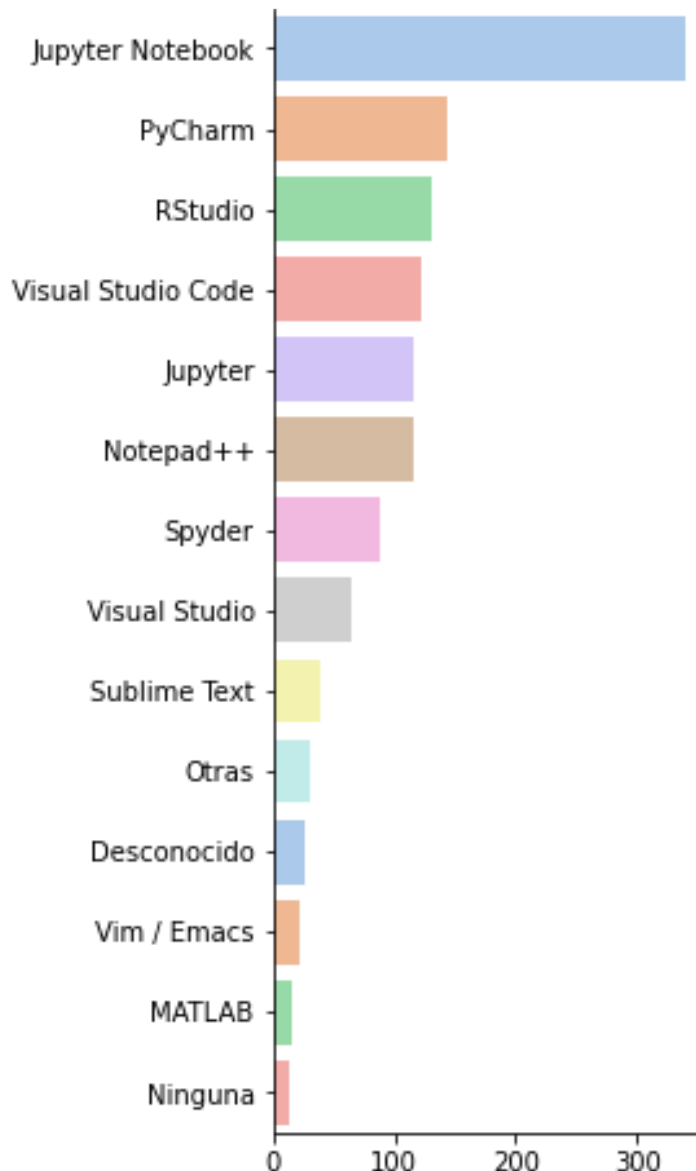
Python es muy popular porque emplea una sintaxis sencilla y eficiente, dispone de una gran cantidad de librerías y herramientas especializadas y permite trabajar con grandes conjuntos de datos. Por otro lado, SQL se integra bien con diferentes lenguajes de programación, es fácil de aprender y permite acceder de manera sencilla a la información almacenada. Además, permite el almacenamiento y administración de grandes cantidades de datos gracias a su interfaz visual, preserva la integridad de éstos protegiéndolos de posibles corrupciones provocadas por fallos del sistema o actualizaciones inconscientes y reduce el tiempo de gestión.

LENGUAJES PREFERIDO POR PUESTO Y EXPERIENCIA



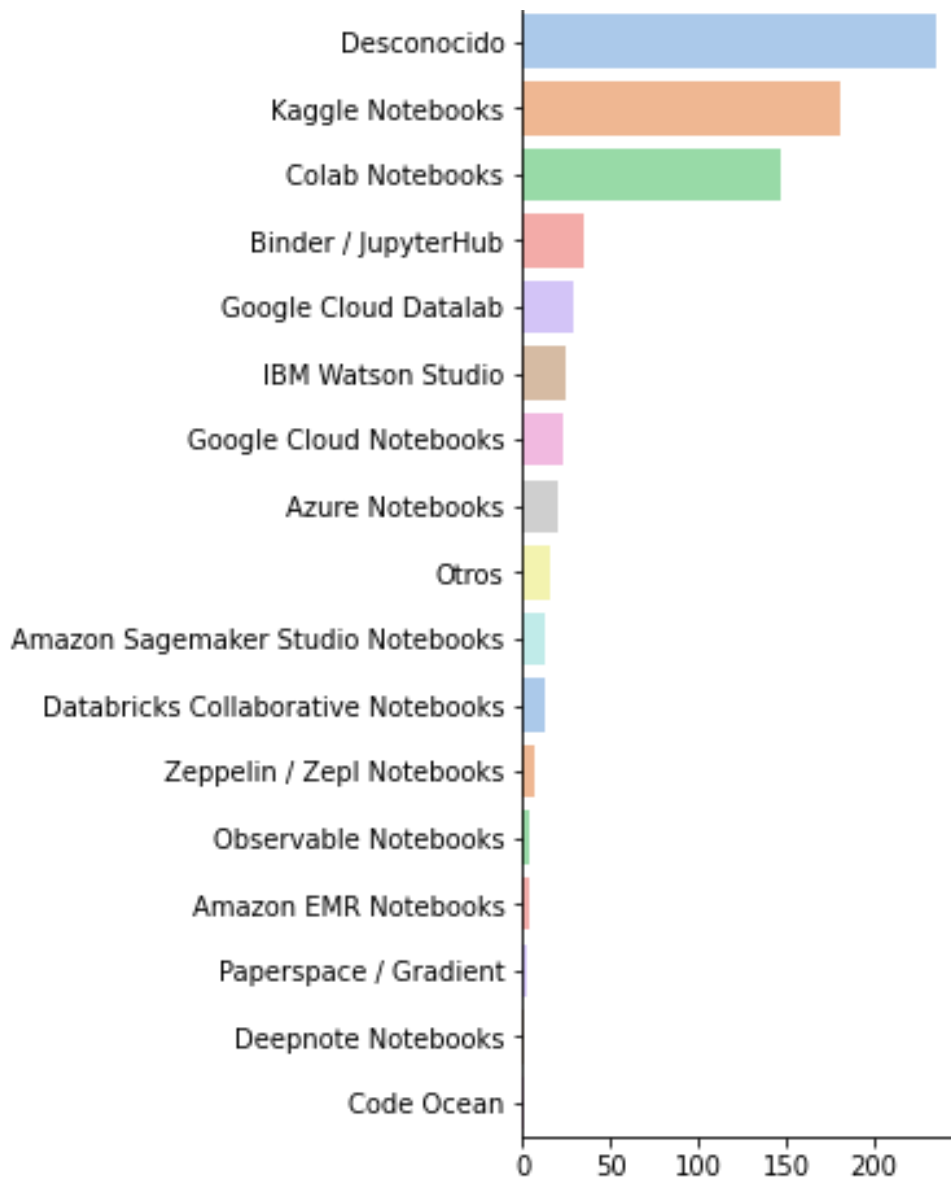
Esta gráficas muestra que los lenguajes de programación preferidos por los *Data Analyst* y *Business Analyst* son Python y SQL. Por tanto, Adalab está ofreciendo a sus alumnas lo que ahora mismo se está empleando más en el sector, ya que son los dos lenguajes de programación que está enseñando. Deberían observar el posible crecimiento del lenguaje R, ya que se puede apreciar que hay juniors que tienen este lenguaje como preferido, aunque Python sigue siendo el favorito de la mayoría de personas que entran en este rango de experiencia.

ENTORNO DE DESARROLLO INTEGRADO



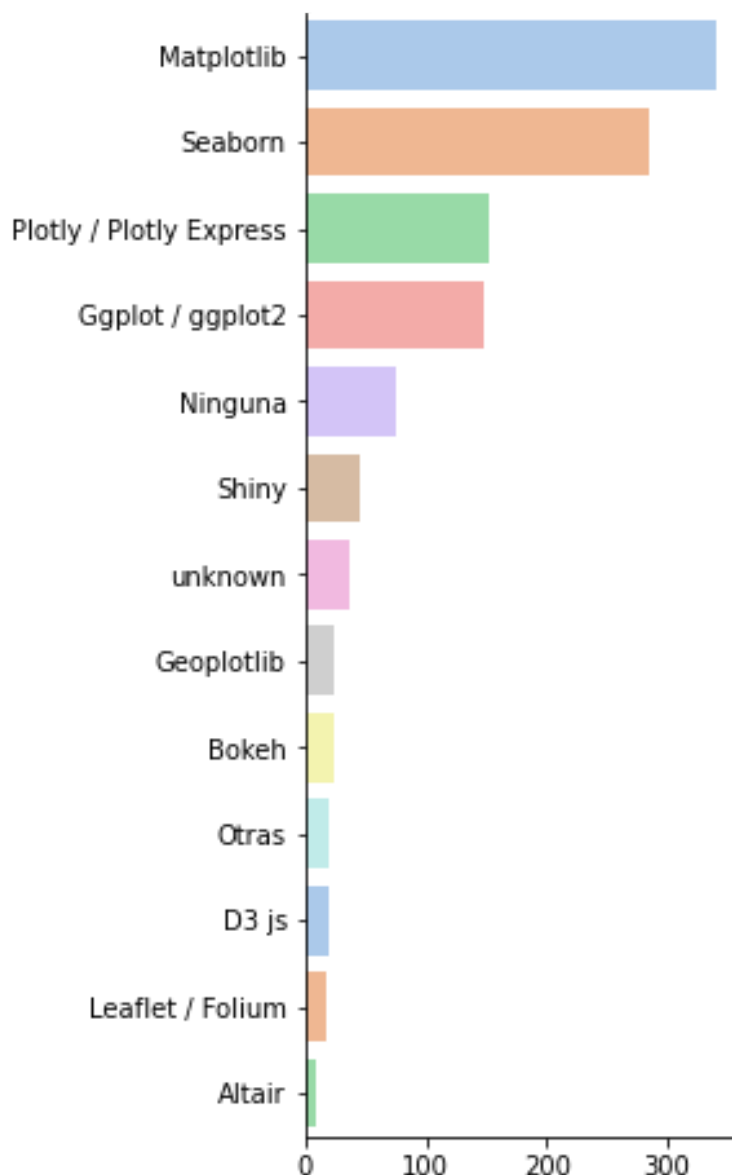
La herramienta de EDA más empleada es Jupyter Notebook, la cual es, a su vez, la más utilizada en el bootcamp de *Data Analytics* de Adalab. Además, Visual Studio Code se encuentra entre las cuatro más empleadas, y es una herramienta que también se enseña en el *bootcamp*. No obstante, hay dos herramientas, PyCharm y RStudio, que no se enseñan, y quizá habría que ver su tendencia por si fuera necesario comenzar a enseñarla en función de lo que pide el mercado actual. Pese a esto, es lógico no enseñar PyCharm, puesto que es una herramienta de pago, y RStudio está enfocada al lenguaje de programación R.

NOTEBOOKS EN LA NUBE



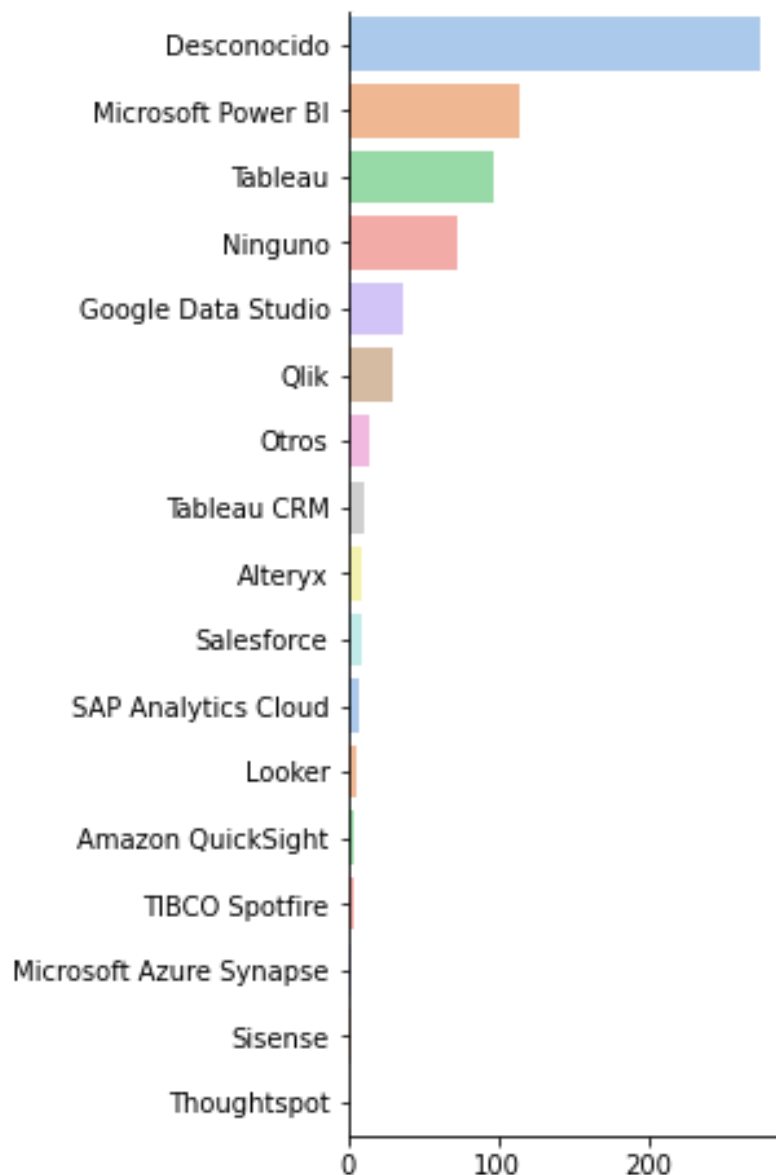
Hay una mayoría de personas que no ha respondido a esta pregunta. No obstante, se puede concluir que, por lo general, Kaggle Notebooks y Google Colab son las favoritas en el sector. Al trabajarse con Jupyter Notebooks, estas dos opciones no se emplean. No obstante, quizá deberían ser tenidas en cuenta, ya que son herramientas gratuitas y fáciles de usar, no requieren experiencia previa. Además, Kaggle destaca por tener una gran comunidad de usuarios que están dispuestos a ayudarse unos a otros.

LIBRERÍAS DE VISUALIZACIÓN



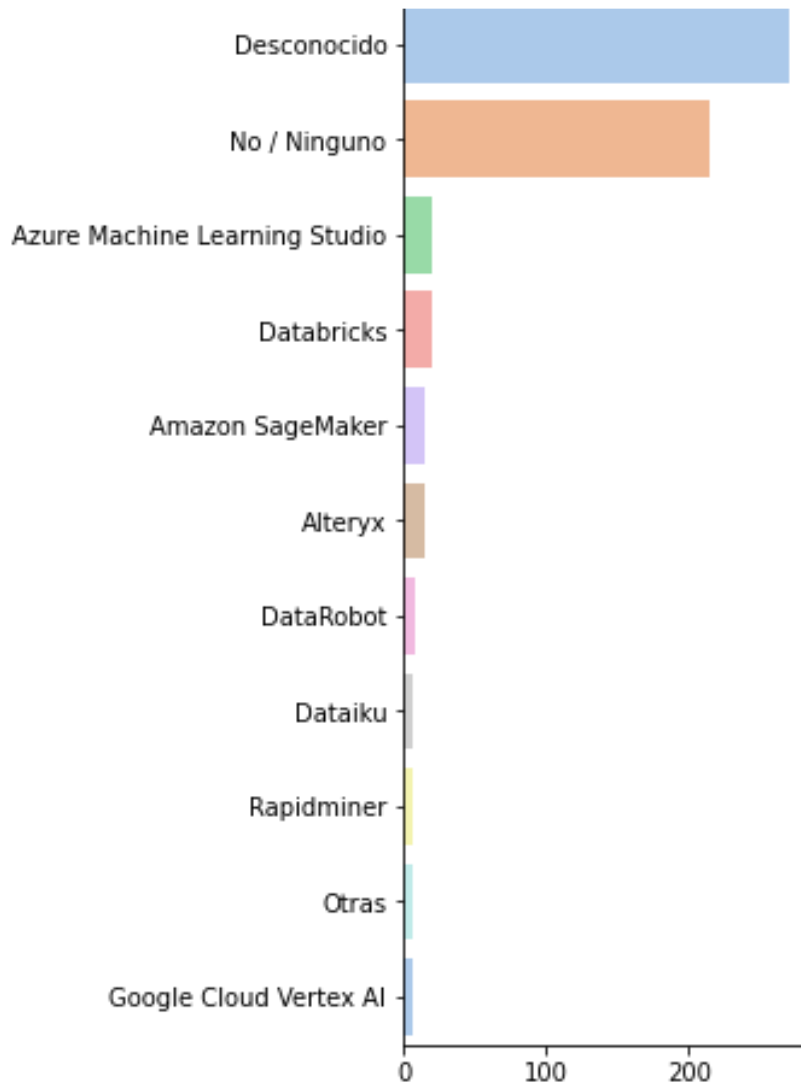
Las librerías de visualización más empleadas en el sector actualmente son Matplotlib y Seaborn. Por tanto, Adalab sí que está al día con las tendencias del momento, ya que se trabaja con ambas librerías a lo largo de su *bootcamp*. Por si acaso, se recomienda estar pendiente de otras librerías que también tienen popularidad, como Plotly y Ggplot, ya que podrían empezar a coger también mayor relevancia.

HERRAMIENTAS DE BI



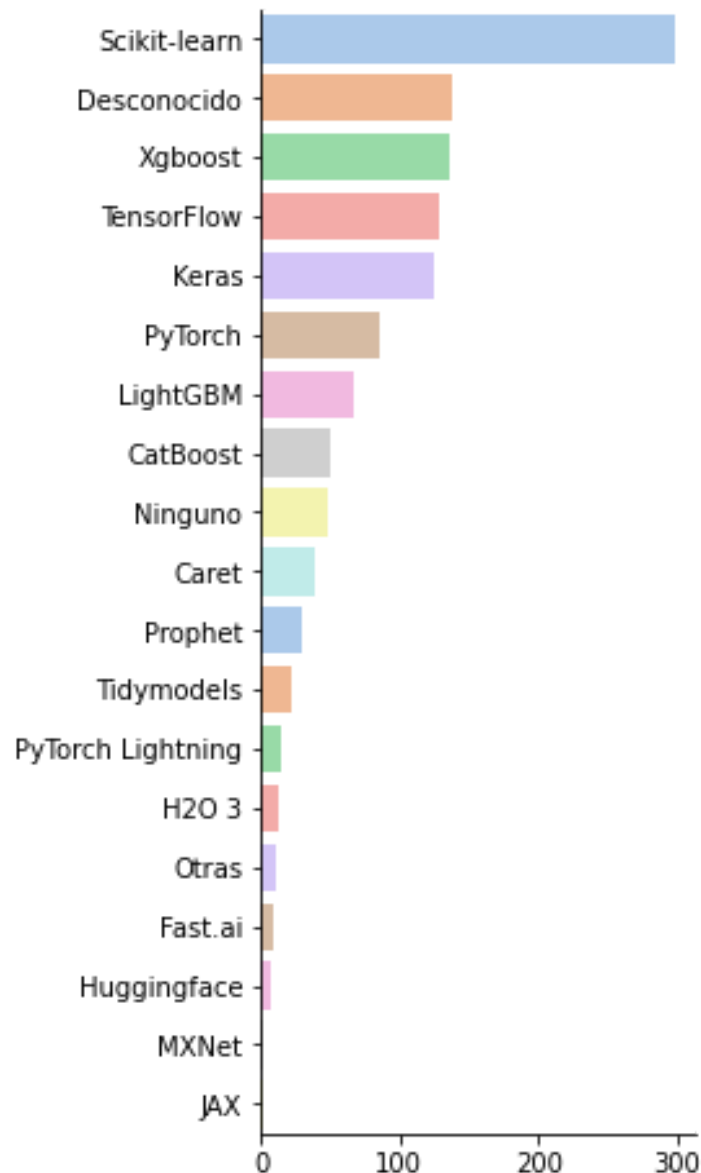
Power BI es la herramienta más empleada, seguida de Tableau. Ambas son herramientas gratuitas y que pueden manejar una gran cantidad de datos, y son relativamente sencillas de usar, si bien Power BI es algo más sencilla. Por tanto, Adalab quizá debería plantearse formar a sus alumnas en esta herramienta.

PRODUCTOS DE MACHINE LEARNING



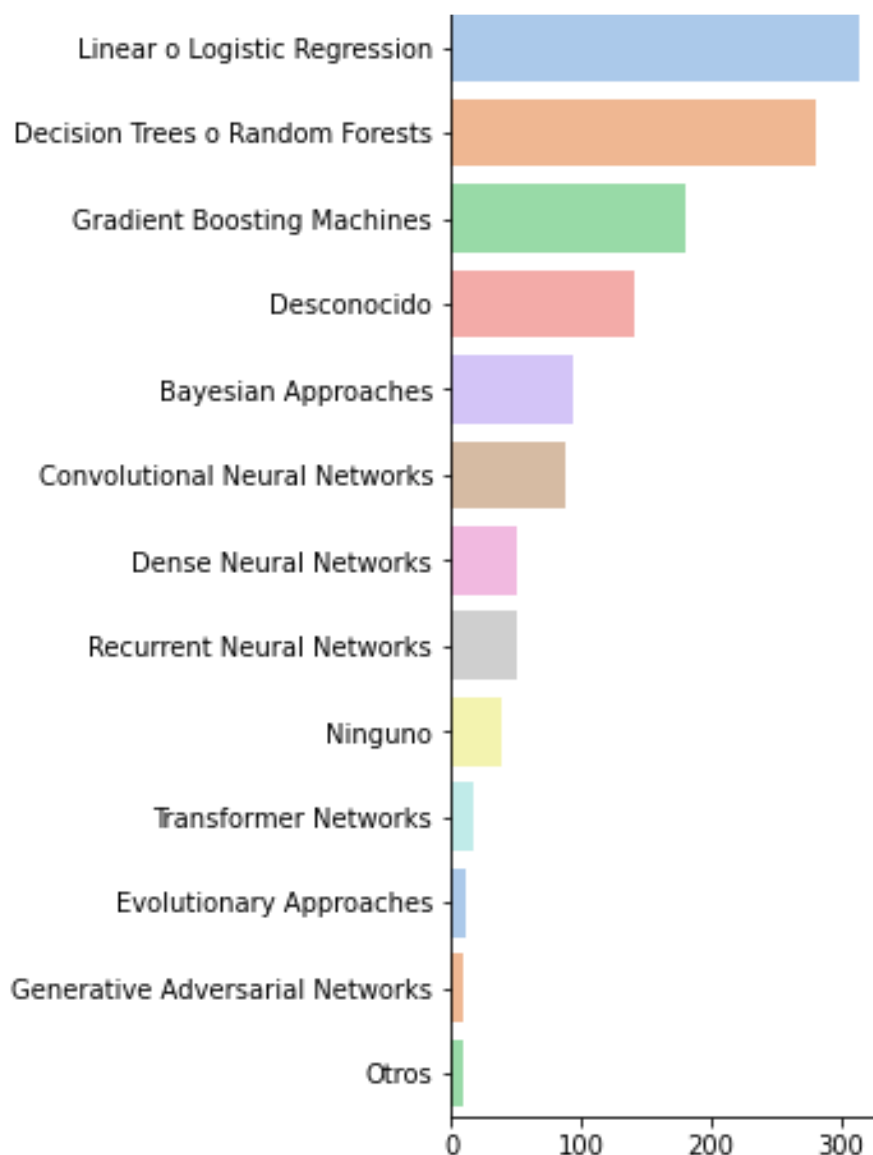
Se puede observar que, por lo general, las personas que ocupan estos puestos no utilizan productos de Machine Learning, por lo que no es necesario incluirlo en su formación.

HERRAMIENTAS DE MACHINE LEARNING



La herramienta más empleada de Machine Learning es Scikit-learn, por lo que Adalab incluye en su formación la herramienta con más relevancia actualmente en el mercado. Por si acaso, debería estar pendiente de la tendencia de otras herramientas como Xgboost o TensorFlow, por si empezaran a cobrar más importancia en el futuro.

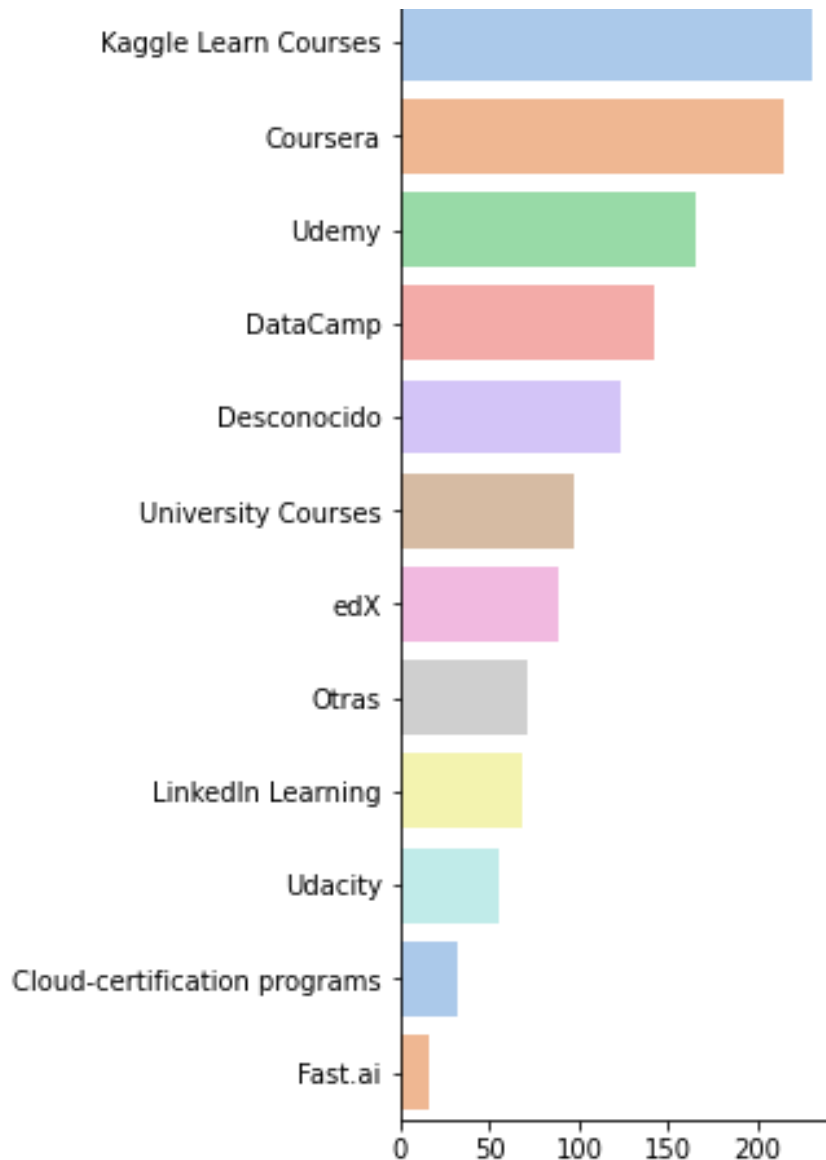
ALGORITMOS DE MACHINE LEARNING



Los algoritmos de ML más empleados son Linear o Logistic Regression y Decision Trees o Random Forests, lo cual corresponde a lo que enseña Adalab durante su bootcamp, por lo que está actualizada en este sentido.

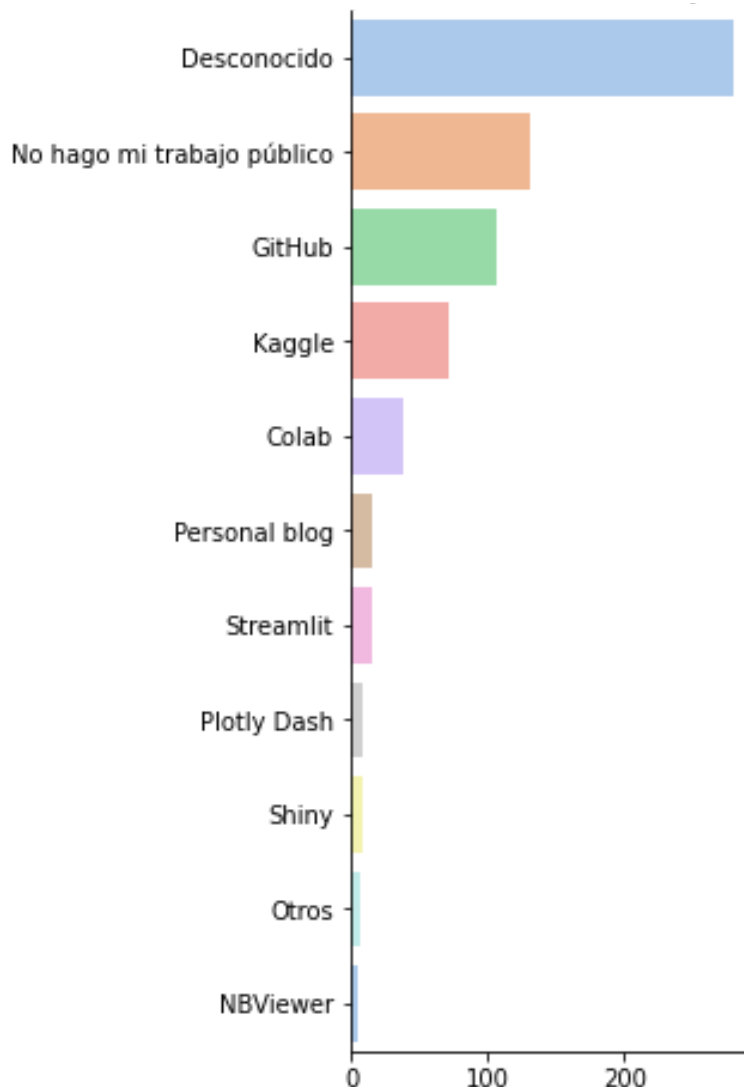
Ambos son modelos altamente interpretables, flexibles y que permiten manejar datos tanto numéricos como categóricos. Además, son algoritmos robustos. Todas estas técnicas están ampliamente disponibles en bibliotecas de aprendizaje automático populares como Scikit-learn en Python, lo que facilita su implementación y uso. Estas técnicas han demostrado su eficacia en una amplia variedad de problemas del mundo real, lo que las convierte en elecciones sólidas para muchos proyectos de aprendizaje automático.

PLATAFORMAS DE CURSOS



En esta gráfica no aparece Platzi, que es la plataforma empleada por Adalab para realizar el módulo 0. Quizá sería conveniente considerar Kaggle Learn o Coursera pero, sobre todo, Udemy, ya que tiene mucha relevancia en este sector. Estas plataformas ofrecen una amplia gama de cursos sobre una variedad de temas, y los cursos son impartidos por instructores expertos. Además son, o bien gratuitos, o bien tienen precios muy asequibles.

PLATAFORMAS DE USO COMPATIDO



Muchas personas no comparten su trabajo de manera pública, pero aquellas que sí lo hacen lo realizan en Github y Kaggle. GitHub sí es enseñada en Adalab, aunque no Kaggle, por lo que sería interesante mostrar también estas herramientas.



CONCLUSIONES



7. Conclusiones:

A modo de conclusión, parece que Adalab sí que está al corriente de las tendencias del sector, lo que le permite ofrecer a sus alumnas la formación más actualizada. En general, consideramos que Adalab tiene un objetivo muy adecuado a la realidad actual, puesto que hay una brecha de género muy latente. Además, este sector parece ofrecer muy buenas condiciones, tanto en salario como en la conciliación familiar, lo que es llamativo para las mujeres que desean realizar su reinversión laboral.

En cuanto a empleabilidad, creemos que Adalab debería centrar sus esfuerzos en encontrar mentores en el sector de la Contabilidad y Finanzas, además del sector tecnológico, ya que es donde se encuentra el mayor número de *Data Analyst* y *Business Analyst* en Europa. Con respecto al tamaño de las empresas, creemos que centrarse en start-ups y macro-empresas sería un acierto para una mejor incorporación al mundo laboral para las Adalabers.

No obstante, recomendamos hacer estos estudios de una manera periódica, ya que este sector es muy dinámico y está en constante cambio, por lo que es previsible una transformación continua y a corto plazo. El equipo Data Dream Team ya tiene los procesos automatizados para poder continuar realizando estos estudios en un futuro, de una manera más ágil, rápida y efectiva.

7.1. Próximos pasos:

En función del trabajo realizado, consideramos que habría diferentes pasos que podrían llevarse a cabo para realizar un estudio aún más completo. Entre estos pasos, se encuentra:

- Sólo se han empleado 23 de las 53 columnas, por lo que podría también realizarse un estudio de las columnas restantes.
- Los datos estaban recogidos en categorías y rangos, lo que imposibilita realizar ciertos estudios como media salarial, etc., así como ciertas representaciones gráficas para llegar a conclusiones más concretas.
- Realizar un estudio más centrado en función del salario y del país, ya que los salarios base de cada país varían enormemente.
- Actualizar los datos y realizar un estudio en base a información más reciente, ya que consideramos que podría haber cambios significativos. Este proceso podría hacerse de manera periódica, para que Adalab siempre se mantenga al día en su formación.

