
INFORME DEL ANÁLISIS PARA ADALAB

¿ES RENTABLE Y VIABLE UN BOOTCAMP DE MACHINE LEARNING EN ADALAB?

Informe realizado por el equipo Data Paired compuesto por Chloe Aroca, Ariana Rosales, Verónica Bellido y Lara Sancho.

PLANTEAMIENTO DE NUESTRO ANÁLISIS

El mercado laboral necesita profesionales capaces de adaptarse a las infinitas posibilidades de todas las innovaciones tecnológicas de una empresa. Prepararse en formaciones intensivas como los *bootcamps*, permite a los alumnos adquirir las competencias que necesitan para comenzar a trabajar de inmediato.

Por esta razón, en base a la información de las variables y ficheros que nuestro cliente nos ha proporcionado, quisimos orientar el estudio hacia la pregunta de si sería una buena idea para nuestro cliente implementar un *bootcamp* de Machine Learning (ML) en el futuro.

Para poder responder a esta pregunta es importante conocer la importancia y situación de ML en nuestra sociedad, llevando a cabo una metodología de limpieza de los ficheros para finalmente rescatar la información más relevante y demostrar gráficamente los resultados obtenidos con sus respectivas conclusiones y comprobar si de verdad ML es tendencia.

Pero ¿Qué es el Machine Learning?

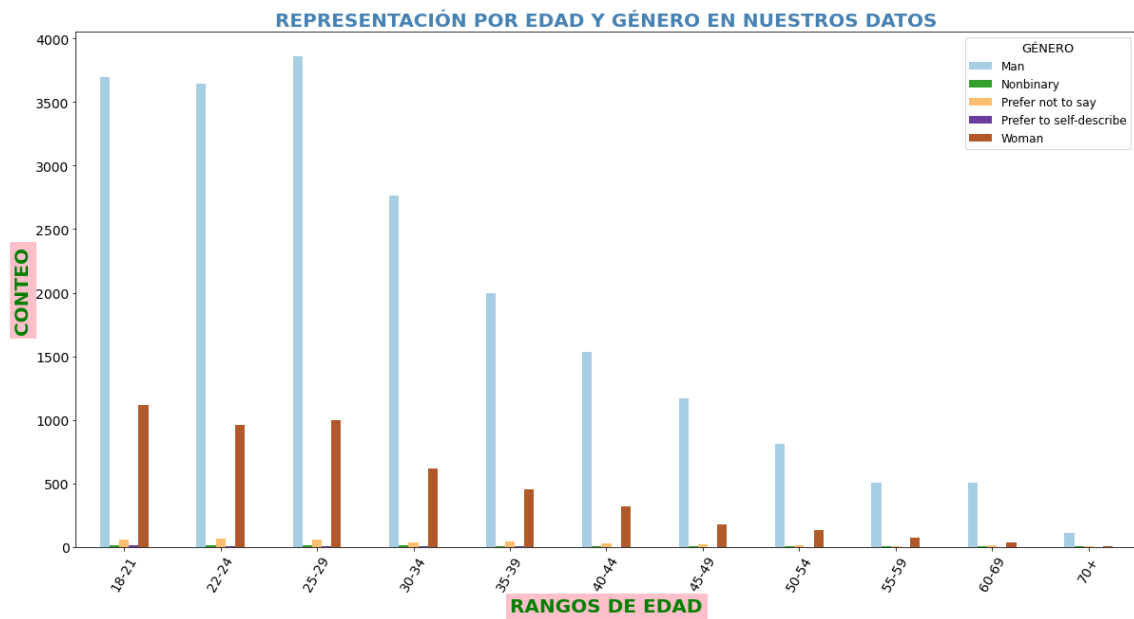
Machine Learning (ML) o aprendizaje automático, es un campo de estudio que brinda a las computadoras la capacidad de aprender sin ser programadas explícitamente. El aprendizaje automático se aplica utilizando algoritmos para poder procesar los datos y capacitarse para entregar predicciones futuras sin intervención humana. El aprendizaje automático es utilizado ampliamente por empresas como Facebook, Google, etc.

La sociedad en conjunto puede beneficiarse de las oportunidades que ofrece el aprendizaje automático como por ejemplo en la detección de fraudes o la recomendación en la línea de ofertas de Amazon y Netflix.

CONTEXTO DE LOS DATOS

Antes de comenzar con nuestro análisis, queríamos poner nuestros datos en contexto, ya que es importante tenerlo en cuenta a la hora de hacer interpretaciones.

1. Género y edad:

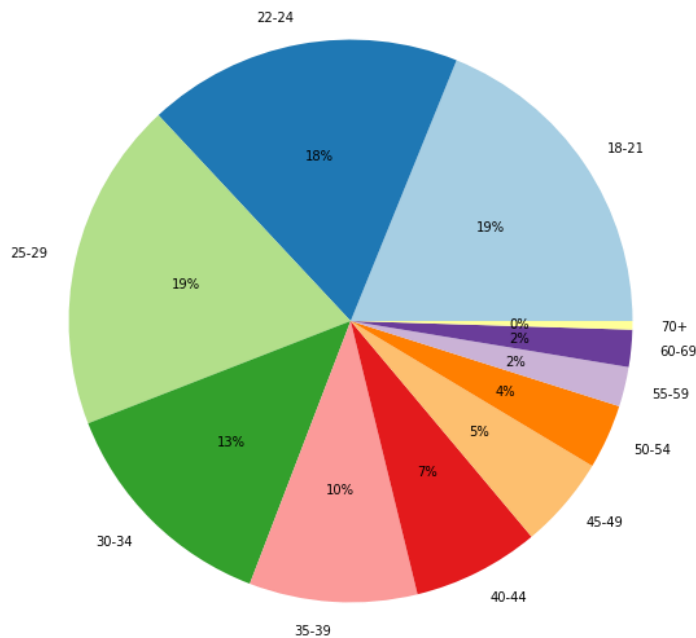


Hemos podido comprobar que las personas que más han respondido a las preguntas de nuestro estudio han sido hombres de entre 25 y 29 años.

El género masculino es el que más representado está, por lo que los análisis que tengan el factor género deberán hacerse separando por género y comparando las mediciones con respecto a sí mismos, es decir, no podemos comparar resultados de conteo o frecuencia directamente entre hombres y mujeres.

2. Edades:

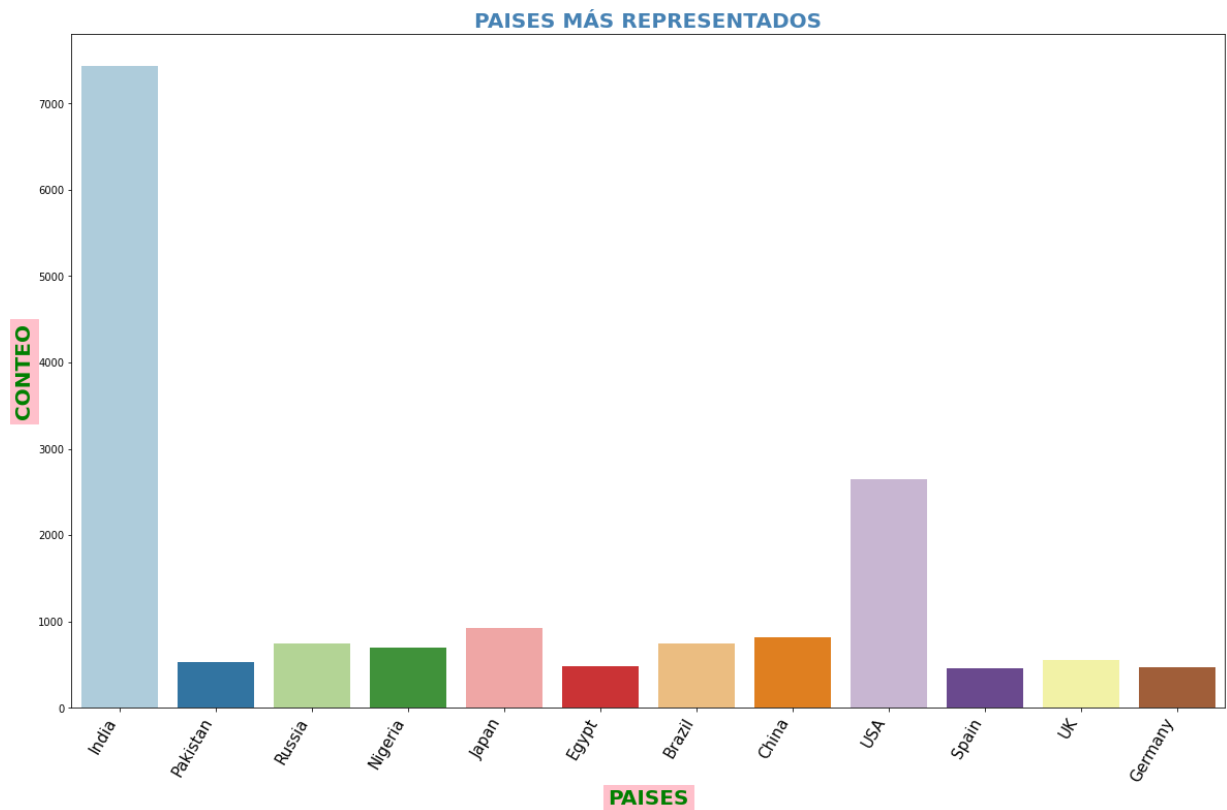
REPRESENTACIÓN DE EDADES EN NUESTROS DATOS



Independientemente del género, las edades más representadas están en el rango de entre los 18 y los 30 años.

Esto también se debe tener en cuenta en los análisis que hagamos, ya que no tenemos muestras igual de representadas para cada grupo y no podemos comparar resultados de conteo entre ellas, es decir, las menos representadas siempre saldrán más bajas en el conteo.

3. Nacionalidad:

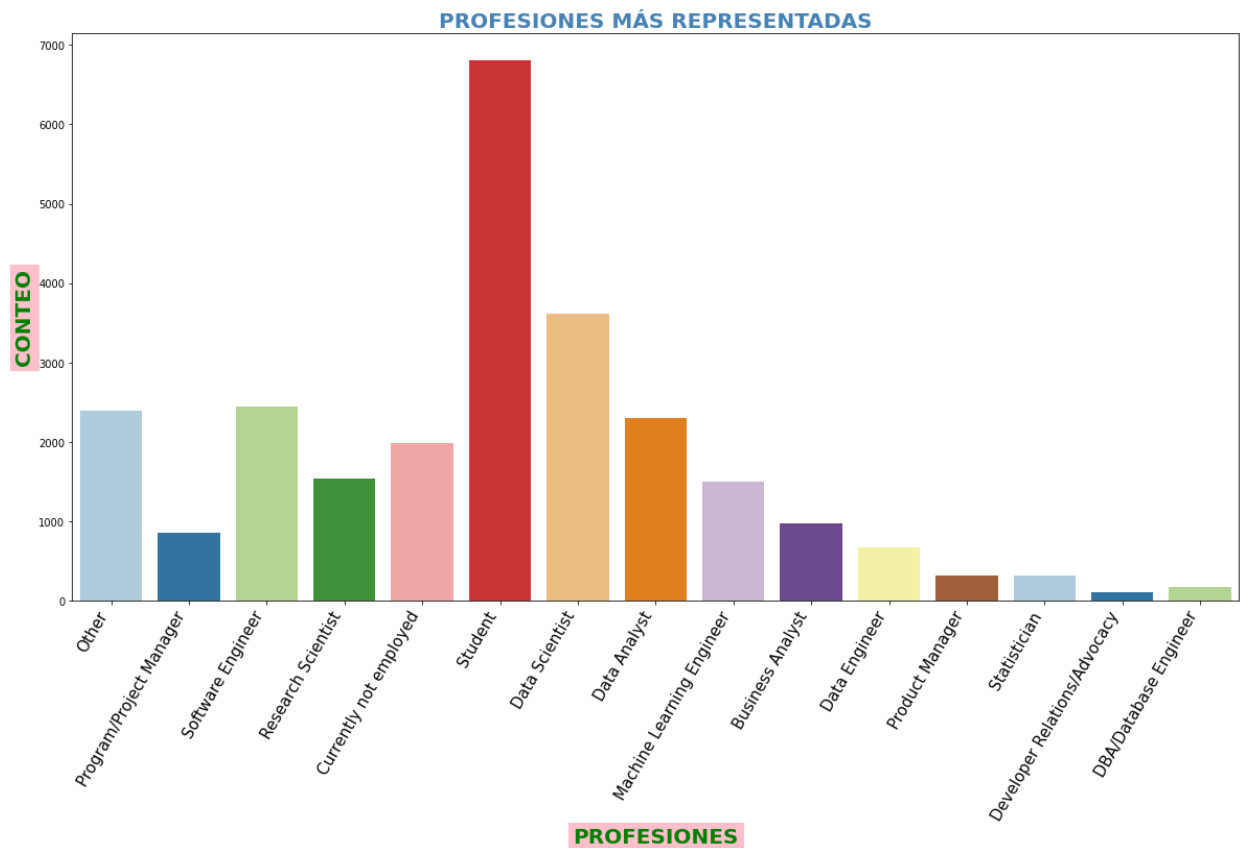


Las nacionalidades más frecuentes de este estudio van a ser muy relevantes a la hora de analizar datos relacionados con los salarios, las inversiones de las empresas, los niveles de educación, y los puestos de trabajo.

Vemos que, aunque la mayoría de los datos recogidos son de la India y EEUU, estamos frente a un estudio global sobre el uso y la aplicación del Data Science y Machine Learning, ya que todos estos países no son más que una selección del total de la muestra.

A la hora de analizar temas en los que el país sea relevante, habrá que hacer un filtrado previo para que los resultados del análisis sean más concluyentes.

4. Roles profesionales más representados:



La ocupación más frecuente en este estudio es la de estudiante, algo que nos parece que guarda cierta lógica con que los rangos de edades más frecuentes sean de 18 a 30 años.

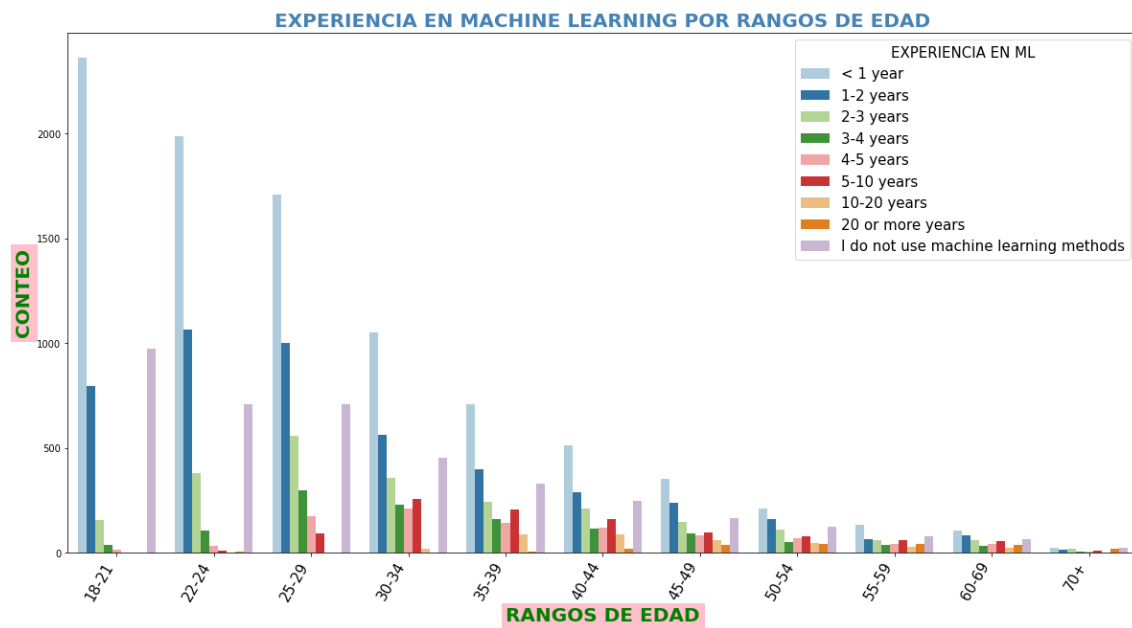
De nuevo tendremos en cuenta estas conclusiones a la hora de estudiar nuestros datos, y nos podremos centrar en las ocupaciones que más están representadas en este estudio.

ANÁLISIS PRINCIPAL DE LOS DATOS

1- ¿Cuál es la tendencia para los nuevos empleos?

Para contestar a esta pregunta nos hemos basado en el análisis de la experiencia en Machine Learning que tienen las personas de nuestro estudio, su edad y su género.

- Tendencias en Machine Learning por rangos de edad:



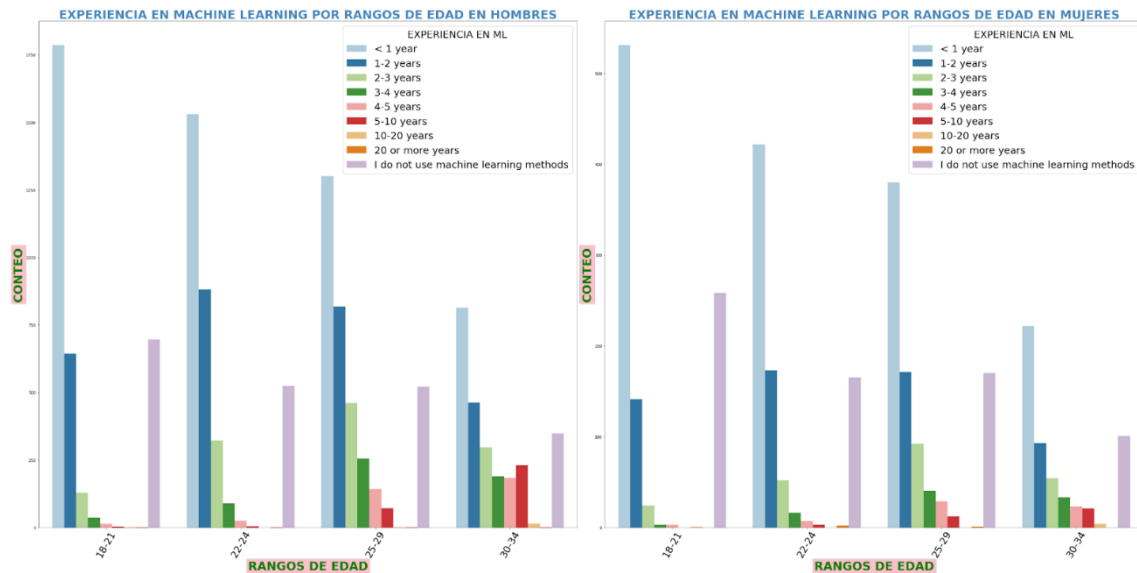
Interpretamos que la tendencia es que las personas más jóvenes están empezando a formarse en ML, siendo las que más lo usan, independientemente de su experiencia.

Los grupos de edades de entre 18-30 años son la muestra más abundante, por lo que los datos pueden estar sesgados, sin embargo, vemos la tendencia de que a menor edad, más uso de técnicas de Machine Learning.

Como ya nos podríamos imaginar, podríamos decir que las técnicas de ML son tendencia actual.

- Tendencias en Machine Learning por rangos de edad y género:

Nos vamos a centrar en esas edades más frecuentes para hacer un análisis por edad, género y experiencia en técnicas de ML, ya que Adalab es una escuela para mujeres, y nos puede interesar ese cálculo.

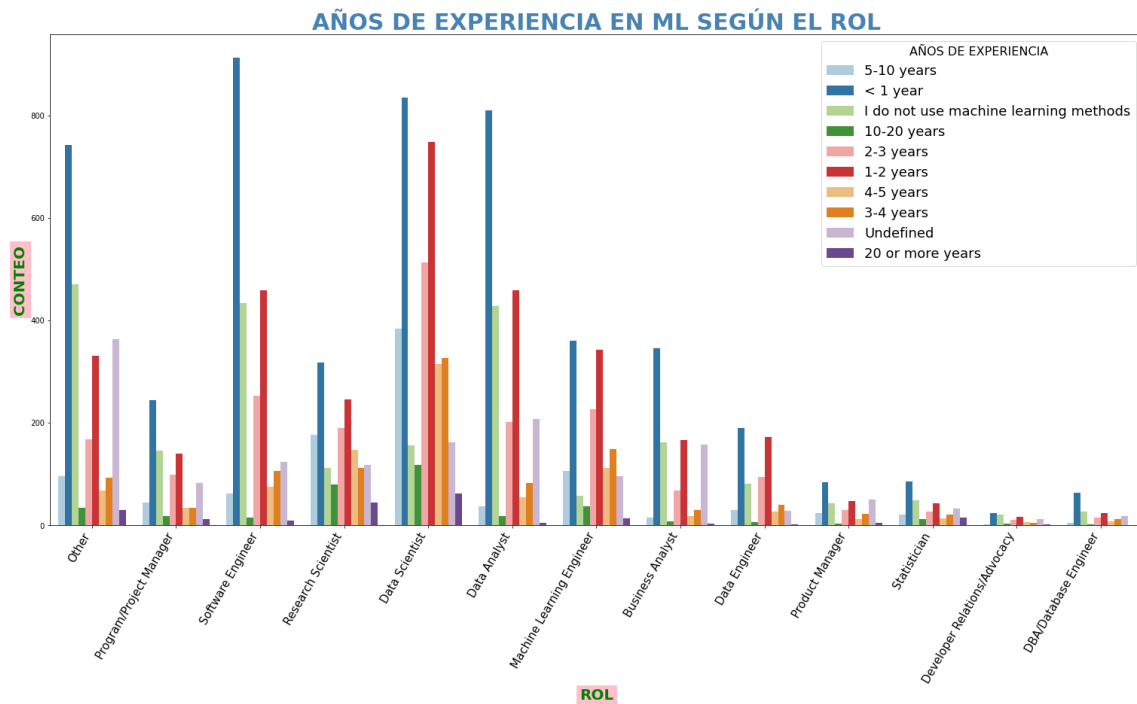


De nuevo, se ve que los hombres están mucho más representados en este estudio, pero se ve una tendencia creciente a medida que las edades descienden en ambos géneros.

El hecho de que haya menos mujeres aprendiendo ML o con experiencia, se podría interpretar como una ventana de oportunidad para las mujeres, ya que saber técnicas de ML es un punto diferenciador, especialmente para ellas.

Los datos nos indican en general que el tiempo que llevan trabajando con métodos de ML es menor a un año

2. ¿Cuánta experiencia en ML tienen los profesionales del sector?



Los datos nos indican en general que el tiempo que llevan trabajando con métodos de ML es menor a un año con mayor tendencia o un promedio entre 1 y 3 años, lo que demostraría que todos los roles profesionales no tienen muchos años de experiencia empleando métodos de ML

3. ¿Cómo se han formado los roles que más usan métodos de Machine Learning?

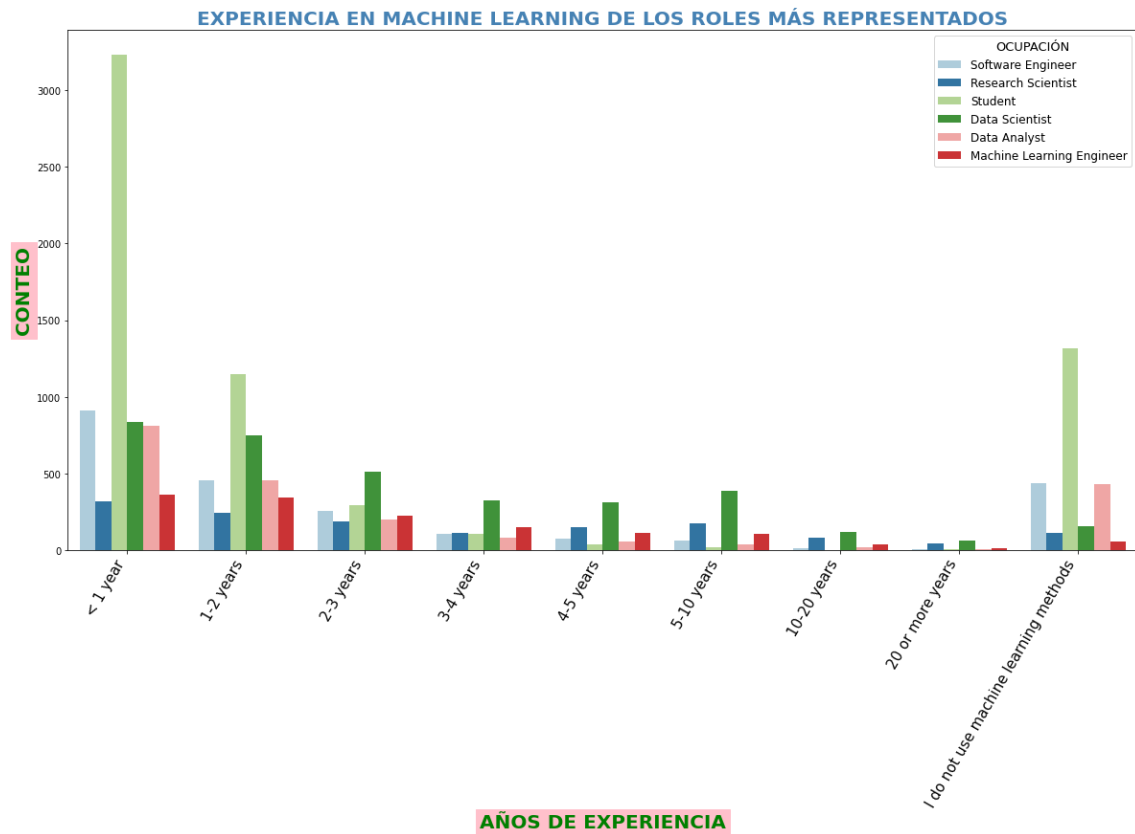
- Roles que más usan Machine Learning:

Antes de nada, queremos extraer los roles que más usan métodos ML, relacionando los datos sobre profesión y experiencia en métodos de Machine Learning.

q_15_years_using_ml	1-2 years	10-20 years	2-3 years	20 or more years	3-4 years	4-5 years	5-10 years	< 1 year	I do not use machine learning methods
q_5_role									
Business Analyst	166	7	67	3	29	18	15	345	161
DBA/Database Engineer	24	1	15	0	12	7	5	63	27
Data Analyst	458	17	201	5	83	55	37	810	428
Data Engineer	172	6	94	1	40	26	30	190	81
Data Scientist	748	117	513	62	326	314	384	835	156
Developer Relations/Advocacy	16	3	10	2	4	6	2	24	21
Machine Learning Engineer	342	37	227	13	149	111	106	360	58
Other	331	34	168	29	93	68	95	742	470
Product Manager	47	3	30	5	22	12	23	84	43
Program/Project Manager	139	18	98	11	34	34	44	244	145
Research Scientist	245	80	190	44	111	147	176	317	111
Software Engineer	458	15	253	9	106	75	62	913	434
Statistician	42	12	26	15	20	13	20	85	48

Los datos nos dicen que los profesionales con más experiencia en Machine Learning son los Data Scientists, Data Analysts y Software Engineers, seguidos de los ML Engineers y los Research Scientists.

Estos roles coinciden también con las profesiones más frecuentes en el estudio. Por ello hemos elaborado una gráfica con la experiencia en métodos de ML en función de las ocupaciones más frecuentes.

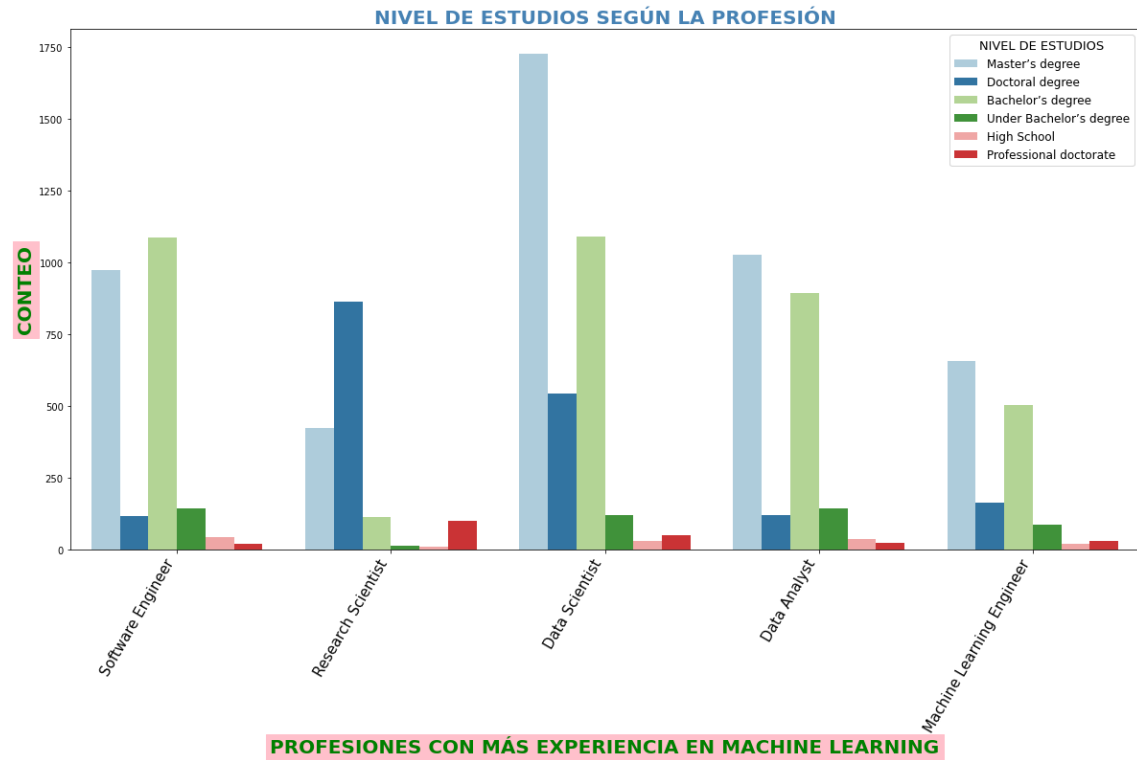


Observamos cómo los Data Scientists son constantes en todos los niveles de experiencia.

Además, vemos que los estudiantes, los Software Engineers y los Data Analysts son de los que más hay en el rango de experiencia más bajo. Esto nos lleva a interpretar que son estos roles los que tienen más interés por formarse en Machine Learning.

- Nivel de estudios de los roles que más usan Machine Learning:

Lo siguiente que nos interesa saber es qué estudios previos tienen estos roles que más se manejan con el machine learning, para ver cómo podríamos enfocar un Bootcamp sobre este tema.

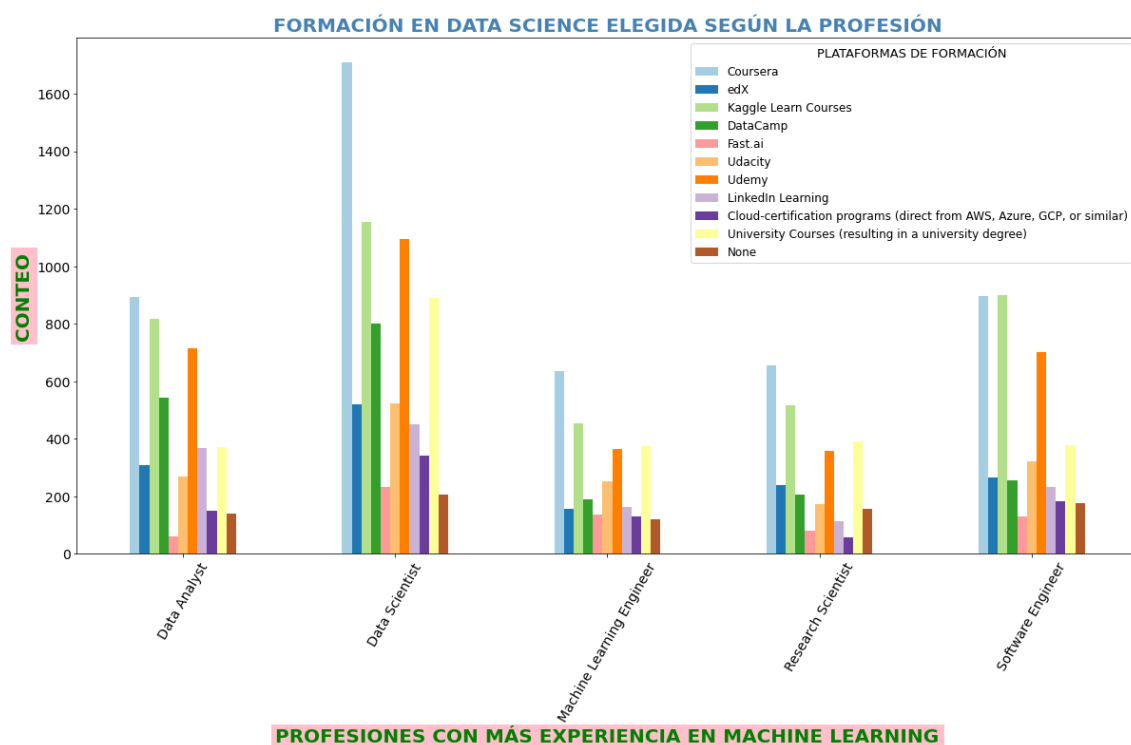


Desconocemos los detalles o especialidades de la formación, pero vemos que los niveles de estudios más frecuentes son grados universitarios y de postgrado.

4. ¿Se eligen los BootCamps para aprender ML?

Seleccionando los puestos que más usan o están aprendiendo ML, vemos qué tipo de formaciones eligen:

	Coursera	edX	Kaggle Learn Courses	DataCamp	Fast.ai	Udacity	Udemy	LinkedIn Learning	Cloud-certification programs (direct from AWS, Azure, GCP, or similar)	University Courses (resulting in a university degree)	None
q_5_role											
Data Analyst	892	308	818	543	60	269	714	367	149	371	139
Data Scientist	1710	519	1155	800	233	524	1094	451	342	891	206
Machine Learning Engineer	636	155	454	189	137	253	363	163	130	373	120
Research Scientist	656	238	517	206	80	173	359	115	57	390	155
Software Engineer	895	265	900	255	129	323	701	233	183	378	176



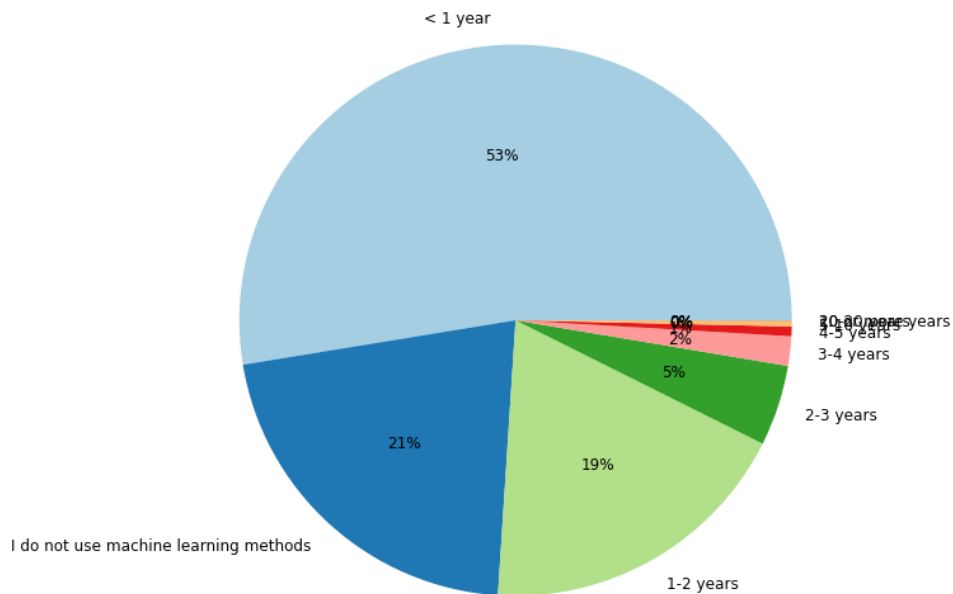
No sabemos si lo que buscan es aprender Machine Learning, pero vemos que el formato *bootcamp* está entre sus **5 primeras opciones** y que hay una tendencia general por parte de estos profesionales a iniciarse en el uso de técnicas de ML.

Vemos que los tipos de cursos más elegidos son (por orden): Coursera, Kaggle Learn, Udemy y, en cuarta posición, los Datacamps para los Data Analysts y Data Scientists, y los cursos universitarios para el resto.

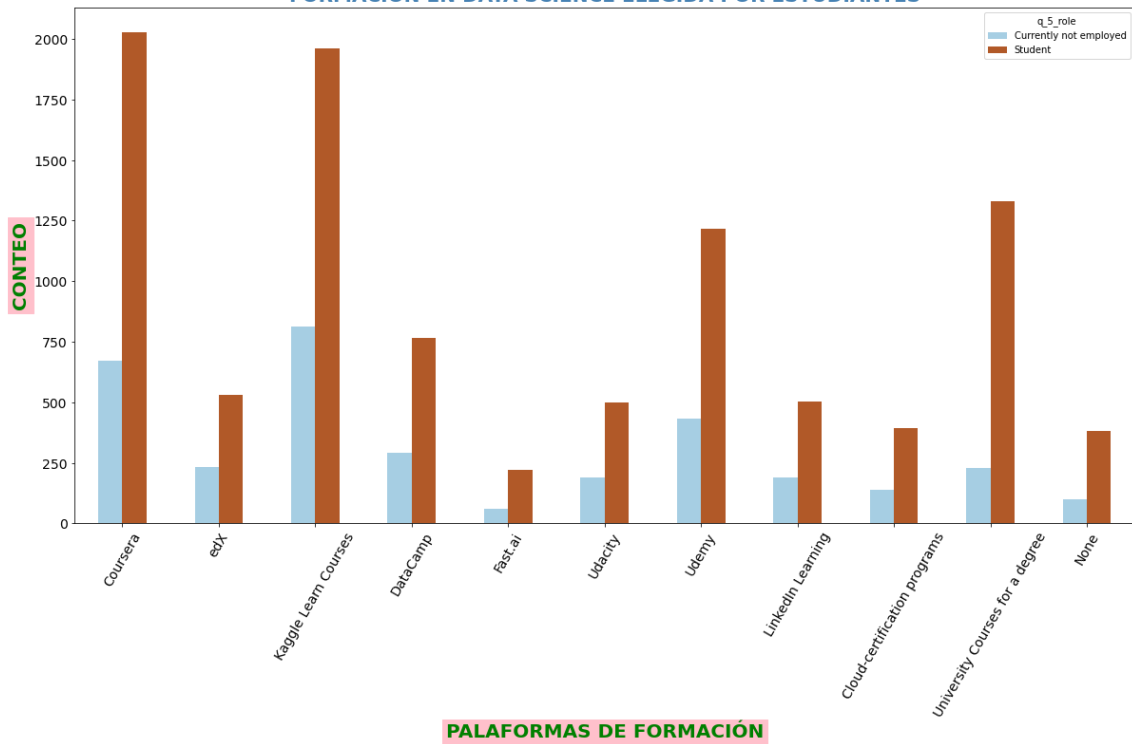
Aunque no es puntera, es mucho más clara la tendencia de hacer bootcamps entre los Data Analysts y Data Scientists.

- ¿Y los estudiantes y desempleados?:

EXPERIENCIA EN ML DE LOS ESTUDIANTES



FORMACIÓN EN DATA SCIENCE ELEGIDA POR ESTUDIANTES



Más de la mitad de los estudiantes encuestados se han iniciado en técnicas de ML, lo que demuestra la tendencia a aprender ML en nuestro grupo estudiado.

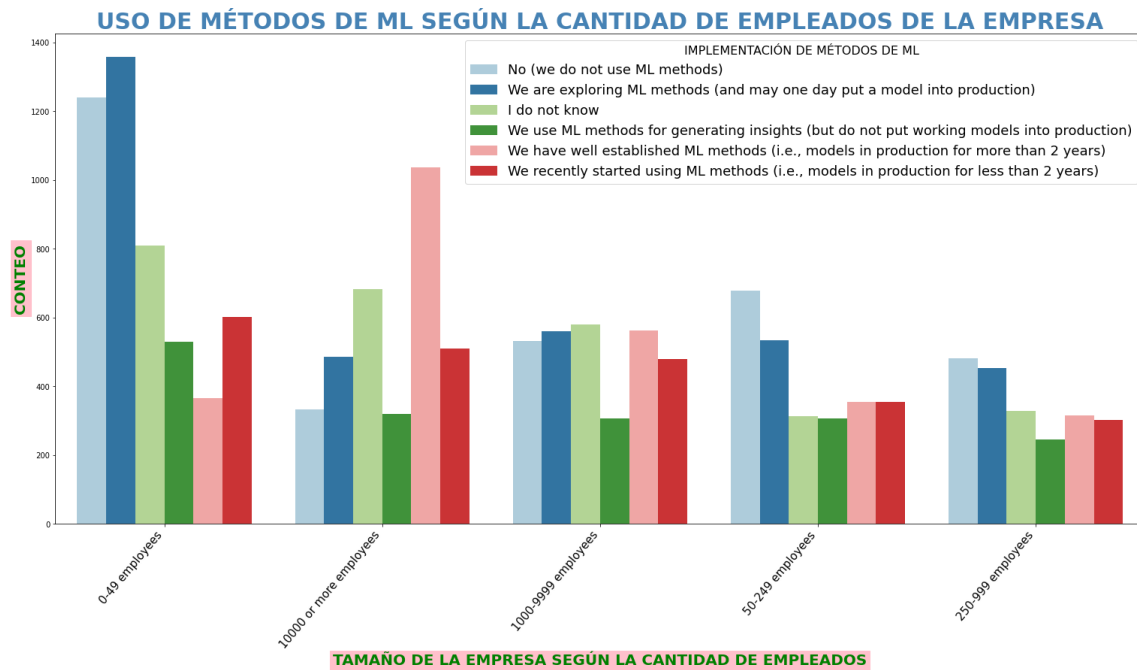
En cuanto a la formación que eligen, es parecida a la que hemos visto en los profesionales.

Interpretamos que las más elegidas, como Coursera, Kaggle y Udemy, pueden serlo por ser gratuitas o muy asequibles.

Entre las que requieren un compromiso económico más grande, los *Datacamp*s sólo están por detrás de los cursos universitarios.

En conclusión, interpretamos que los *Datacamp*s son bastante elegidos para aprender temas relacionados con *Data Science*.

5. ¿Se utilizan en las grandes empresas los métodos de ML?



Vemos que los resultados arrojan en general que las empresas no utilizan métodos de ML.

En las empresas que utilizan métodos de ML, los datos sugieren que:

- En las pequeñas empresas hay tendencia a explorar métodos de ML.
- En las medianas y grandes empresas tienen métodos de ML bien establecidos.

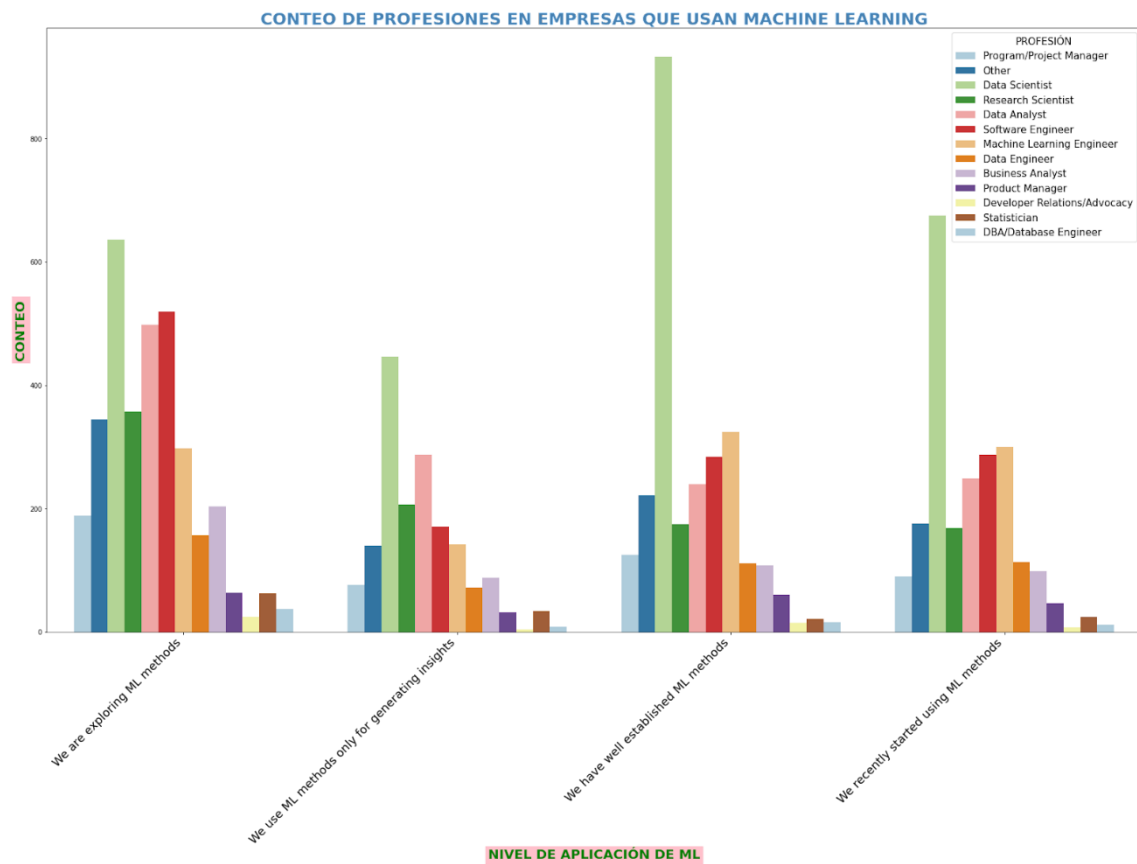
6. ¿En qué puestos se demanda el uso de ML?

Por un lado, ya conocemos los profesionales que más usan métodos de ML:

Data Analysts, Data Scientists, ML Engineers, Research Scientists y Software Engineers.

Por otro lado, nos queda responder:

Las empresas que usan ML, ¿Qué tipo de profesionales tienen en sus equipos?

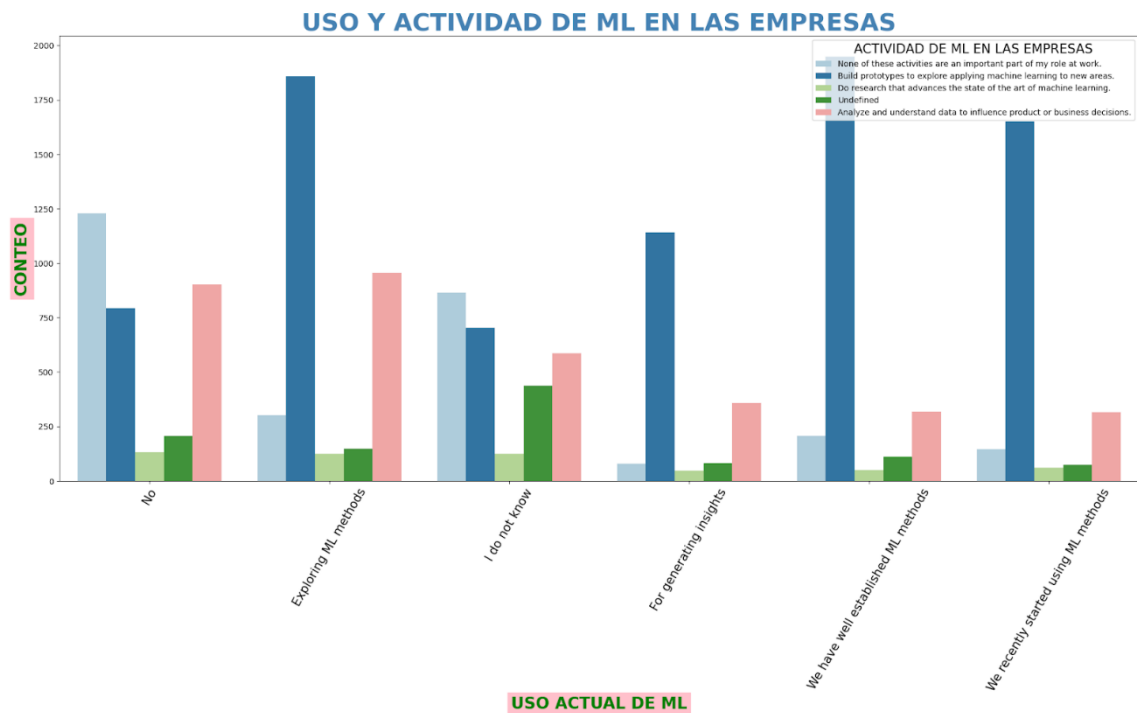


Filtrando por las empresas que usan métodos Machine Learning, vemos que en las compañías donde han establecido el ML durante más tiempo son las que tienen más Data Scientists en su equipo, seguidos de Data Analysts y Software Engineers y ML Engineers.

ANÁLISIS SECUNDARIO DE LOS DATOS:

Una vez hecho el análisis sobre la viabilidad de crear un bootcamp de ML, desde el equipo aportamos información extra que consideramos relevante en caso de que se consideren favorables las conclusiones del estudio previo.

7. ¿Hacia dónde se puede orientar el bootcamp de ML?



Esta gráfica muestra la relación entre las diferentes actividades específicas que se realizan en los diferentes sectores y el uso actual de *Machine Learning* en las empresas.

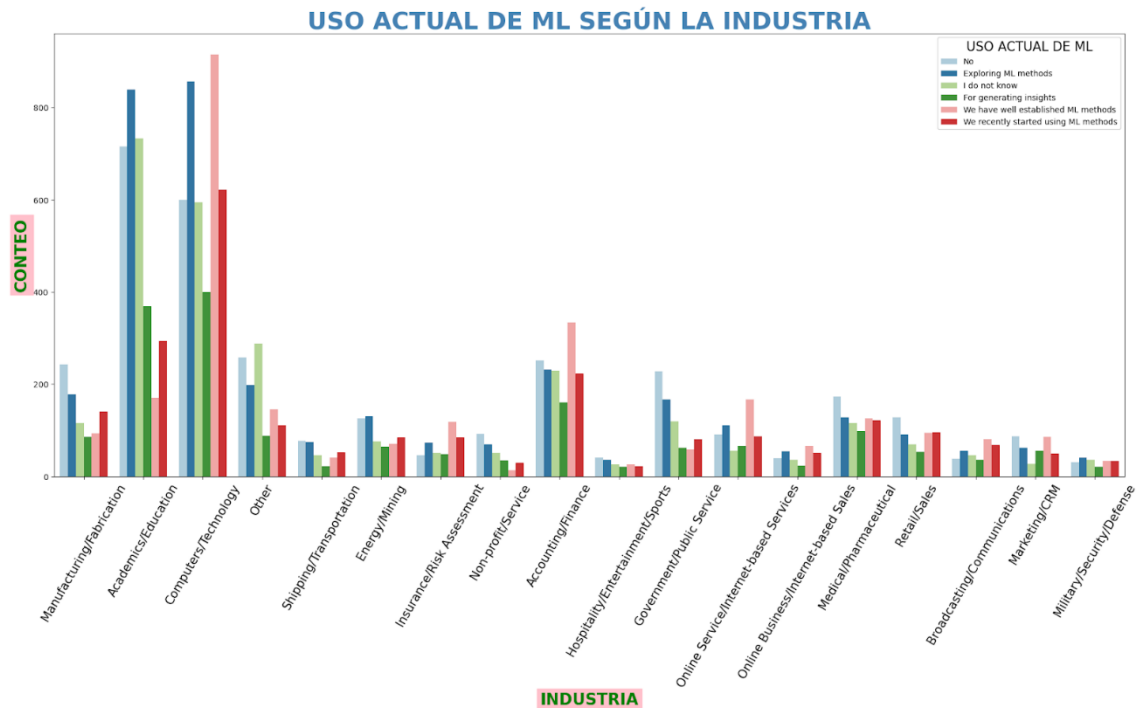
Aún el desconocimiento o el uso de Machine Learning en muchos sectores es mínimo, aunque, de manera genérica, independientemente de la experiencia o el uso que estén dando las diferentes industrias a *Machine Learning*, la actividad más repetida es construir prototipos para ver cómo el *Machine Learning* puede ayudar en diferentes sectores. Buscar una utilidad real para sus propios datos.

La siguiente actividad más común sería el análisis de los datos a través del *Machine Learning*, esto se repite sobre todo en aquellas industrias que están explorando los métodos propios de *Machine Learning*.

Ambas actividades potencian los posibles beneficios de las empresas o, al menos, ayudan a tener un mayor conocimiento de los datos que están generando. Esto es una buena noticia para *Adalab*, ya que, junto con otras gráficas, la tendencia es positiva para esta tecnología. Si se apostara por la creación de un nuevo Bootcamp sobre Machine Learning, los contenidos podrían albergar parte de los que se contemplan en el actual Bootcamp de Data Analytics, ya que estos van orientados sobre todo al estudio y entendimiento de los datos. Más adelante, se deberían introducir contenidos que

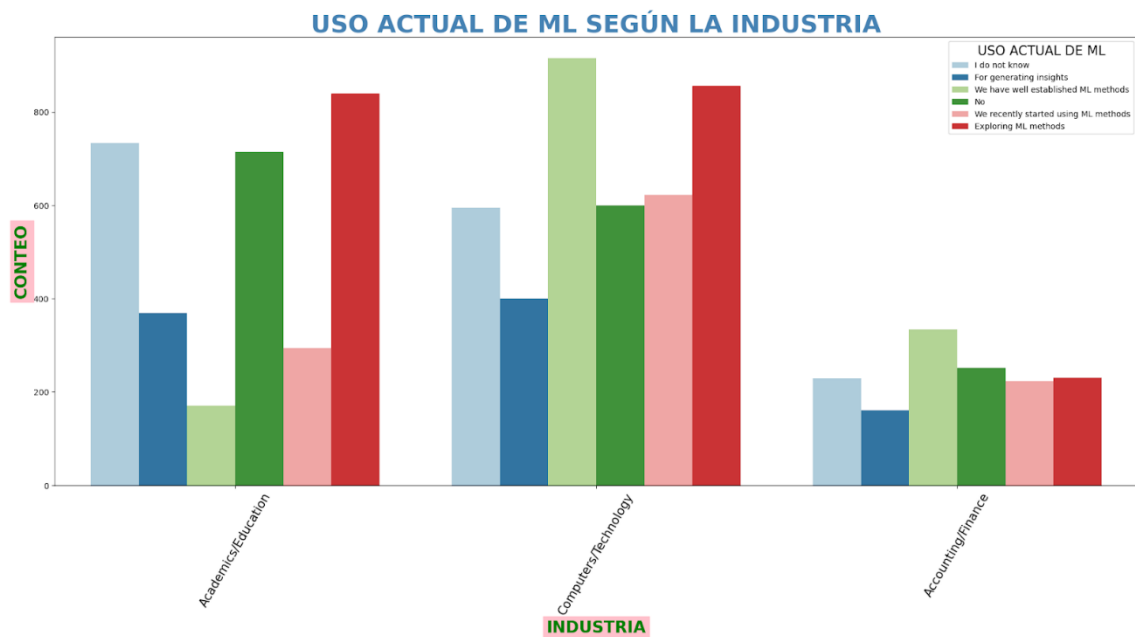
recogieran la creación de *Insights* de forma más exhaustiva, además de la creación de programas para poder automatizar el conocimiento de dichos datos.

8. ¿Cuál es el uso de ML que tiene lugar en cada industria diferente?



Con estas gráficas, analizamos cuál es el uso actual de *Machine Learning* en las diferentes industrias registradas. Esto es interesante, ya que aporta información sobre cómo están trabajando las empresas con esta tecnología, si están desarrollando nuevos métodos porque su experiencia es amplia, o si están en pleno desarrollo y formación.

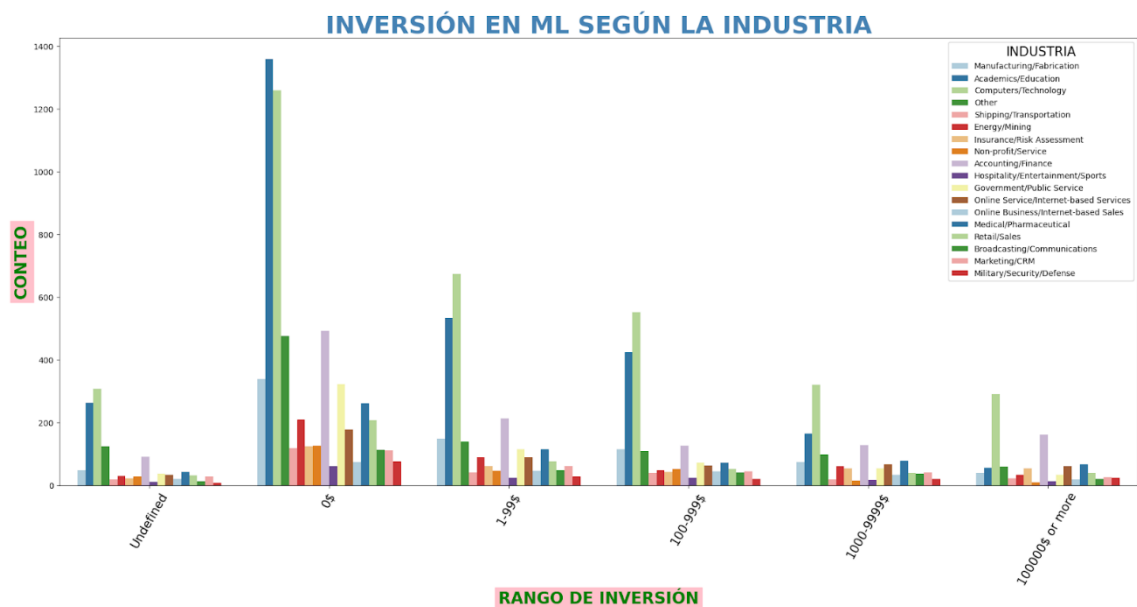
De manera general, aplicado a casi todas las industrias, se desconoce cuál es el uso de *Machine Learning*. Los valores que más se repiten son el “No” y el “I do not know”. Esto puede indicar que aún no se ha implementado lo suficiente, lo cual no significa que no esté en pleno desarrollo. Por este motivo, analizamos más datos sobre la encuesta para investigar si esta metodología de trabajo tiene previsión de futuro o es solo una moda más.



Por otro lado, los sectores tecnológicos y financieros sí tienen un plan establecido de uso y trabajo de esta metodología.

El sector académico está empezando a desarrollarse en esta tecnología, usando recientemente el *Machine Learning* en sus empresas dentro de la industria educativa. Esto puede indicar que este sector puede estar orientándose a la enseñanza de la misma, por lo que puede ser un punto interesante para nuestro cliente *Adalab* a la hora de valorar un posible Bootcamp.

9. ¿Cuál es la inversión en ML por industria?



Podemos observar que, generalizado a casi todas las industrias estudiadas, el valor destacado como inversión, es el rango menor. El mayor número de empresas de cada industria invierte aún poco dinero en esta tecnología.

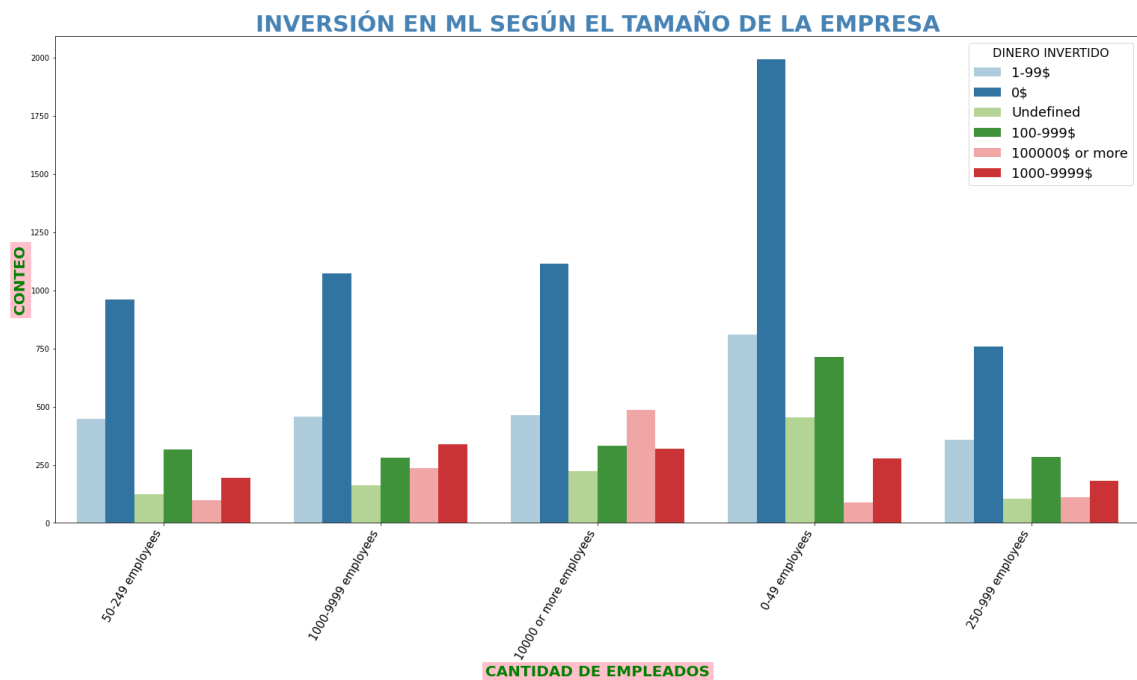
Aun así, si vemos el crecimiento según el rango de inversión, destacamos la industria *Computers/Technology*, que tiene una gran presencia en todos los rangos. Esta industria es la que más apuesta por esta metodología de trabajo.

La industria *Accounting/Finance*, también invierte bastante en Machine Learning, seguido de *Academics/Education*, la cual no es de las más destacadas en el rango más elevado, pero sí en los demás.

Estas tres industrias son las que más nos interesan para conocer si el mercado en el que van a moverse las alumnas, una vez finalicen el posible BootCamp de *Machine Learning*, está interesado en esta tecnología.

La respuesta que definimos una vez estudiada la gráfica es positiva para la empresa *Adalab*, a la hora de formalizar un nuevo curso.

10. ¿Cuál es la relación entre la inversión en ML y el tamaño de la empresa?

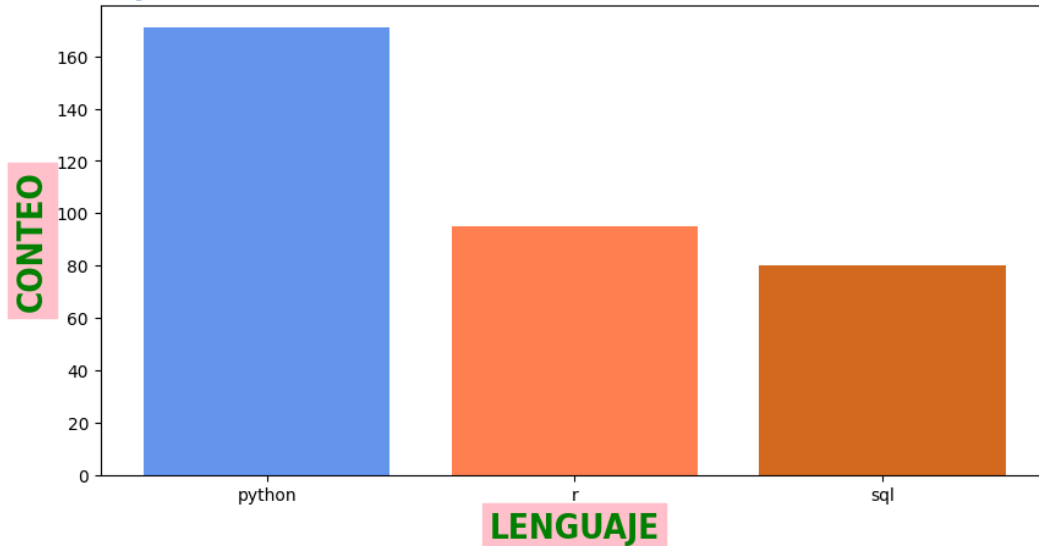


En una visión general, los datos muestran tendencia a que las empresas no invierten mucho dinero en métodos de ML.

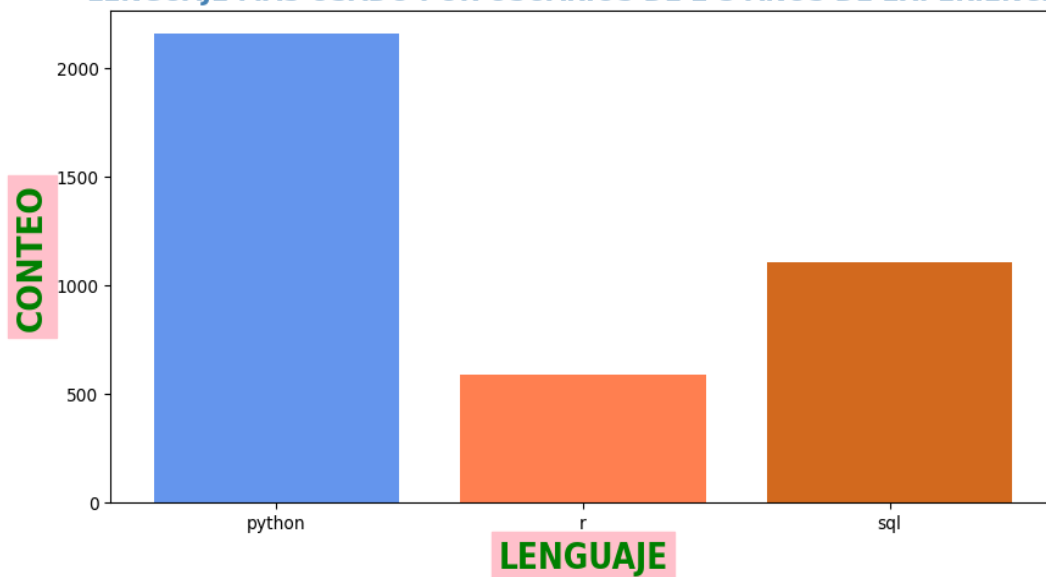
Dentro de las empresas que más invierten en ML, tenemos a las grandes empresas, es decir, las que tienen más de 1000 empleados en adelante.

11. Dependiendo de la experiencia, ¿cuál es el lenguaje más utilizado?

LENGUAJE MÁS USADO POR USUARIOS DE MÁS DE 20 AÑOS DE EXPERIENCIA



LENGUAJE MÁS USADO POR USUARIOS DE 2-3 AÑOS DE EXPERIENCIA



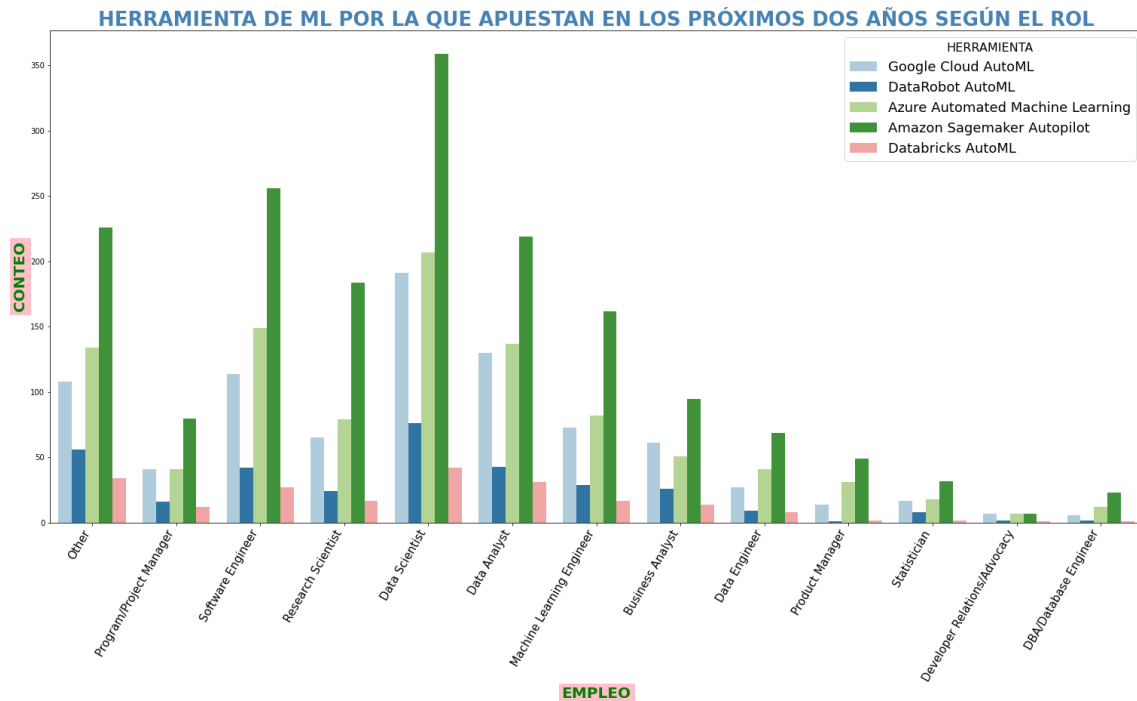
Como se puede observar en las gráficas, Python es el lenguaje más utilizado por todos los usuarios independientemente de su experiencia.

Al analizar y comparar la relación de ambas variables pudimos observar cómo, a medida que avanzaban los años de experiencia, SQL iba adquiriendo más importancia pero 'R' era mucho más popular que SQL en los usuarios con menos experiencia.

Estos últimos tienen los estudios más recientes y empiezan a formarse en lo que las empresas están demandando en la actualidad, por lo que su opinión podría tener más peso que la de usuarios más experimentados que podrían haberse quedado atrás en este mundo que, como bien sabemos, está en constante cambio y evolución.

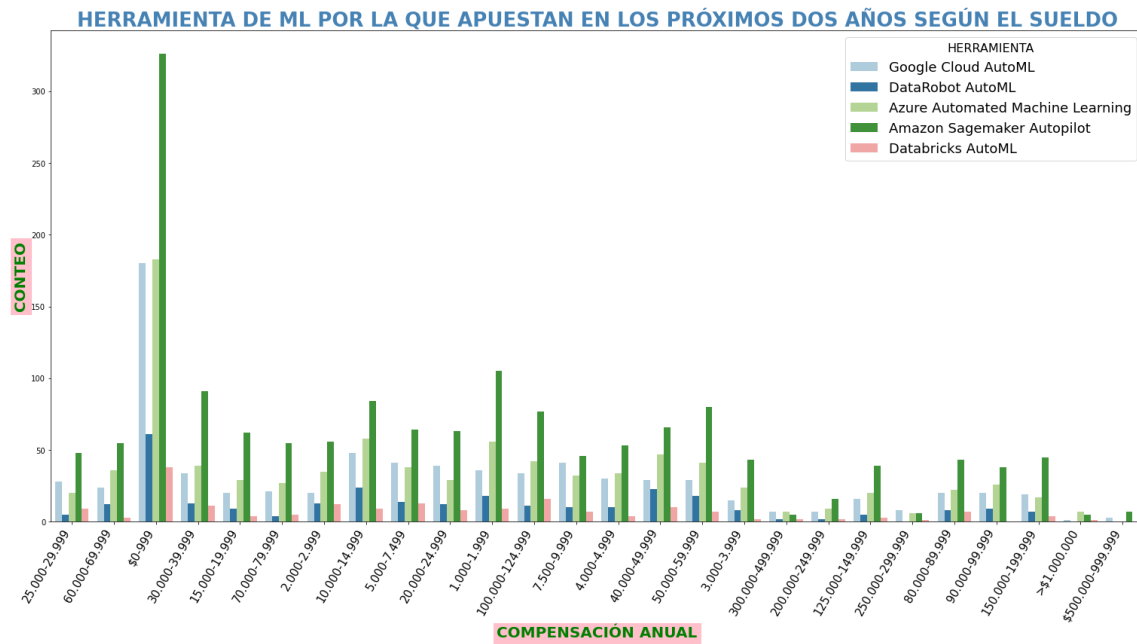
Podríamos concluir que 'R' está siendo el segundo lenguaje más demandado por las empresas actualmente y se recomienda su enseñanza en este nuevo proyecto de Bootcamp.

12. ¿Cuál es la herramienta por la que los profesionales apuestan que va a despuntar en 2 años?



Los datos sugieren que las herramientas de ML que más van a despuntar para los siguientes dos años son las de Amazon SageMaker Autopilot y Azure Automated Machine Learning respectivamente.

También se observa según los datos obtenidos, que serán los Data Scientist, el Software Engineer, Machine Learning Engineer, y el Data Analyst podrán ser los roles que más apostarán por Amazon SageMaker Autopilot y Azure Automated Machine Learning respectivamente, y como tercera opción DataRobot AutoML como preferencia a usar para los próximos dos años.



CONCLUSIONES

Responder a la pregunta que nos planteamos no es sencillo, ya que depende de varios factores a la vez.

Lo que sí hemos podido comprobar es que la demanda de formación en Machine Learning por parte de los profesionales de Data Science y estudiantes, parece haber aumentado durante los últimos años, especialmente entre los más jóvenes.

También hemos visto que, de acuerdo a los datos, de entre las formaciones de pago, la formación en formato bootcamp es bastante elegida por varios profesionales del mundo de los datos, como son los Data Scientists y los Data Analysts.

Sin embargo, hemos interpretado que para aprender Machine Learning es necesario partir de unos conocimientos previos.

Por ello, si Adalab lo viera interesante, veríamos más viable que fuera una formación en formato bootcamp o curso adicional, que se ofreciera a las antiguas alumnas de otros bootcamps, como el de Data Analytics.

Solo en caso de que hubiera una alta demanda de nuevas alumnas para Machine Learning, sería planteable crear otro tipo de formación más extensa que incluyera las bases de todo lo necesario.

En conclusión, pensamos que el análisis extraído del estudio realizado son sobre todo orientativas, ya que pueden no ser del todo significativas por tratarse de unos datos recogidos a nivel global y no tan enfocados al mercado al que se dirige Adalab.

Para obtener resultados más fiables, recomendamos profundizar el análisis, quizás mediante una encuesta propia a antiguas alumnas o con datos de estudios más enfocados a España o Europa.