

AGOSTO , 2020

Ciencia de Datos. Mercado laboral

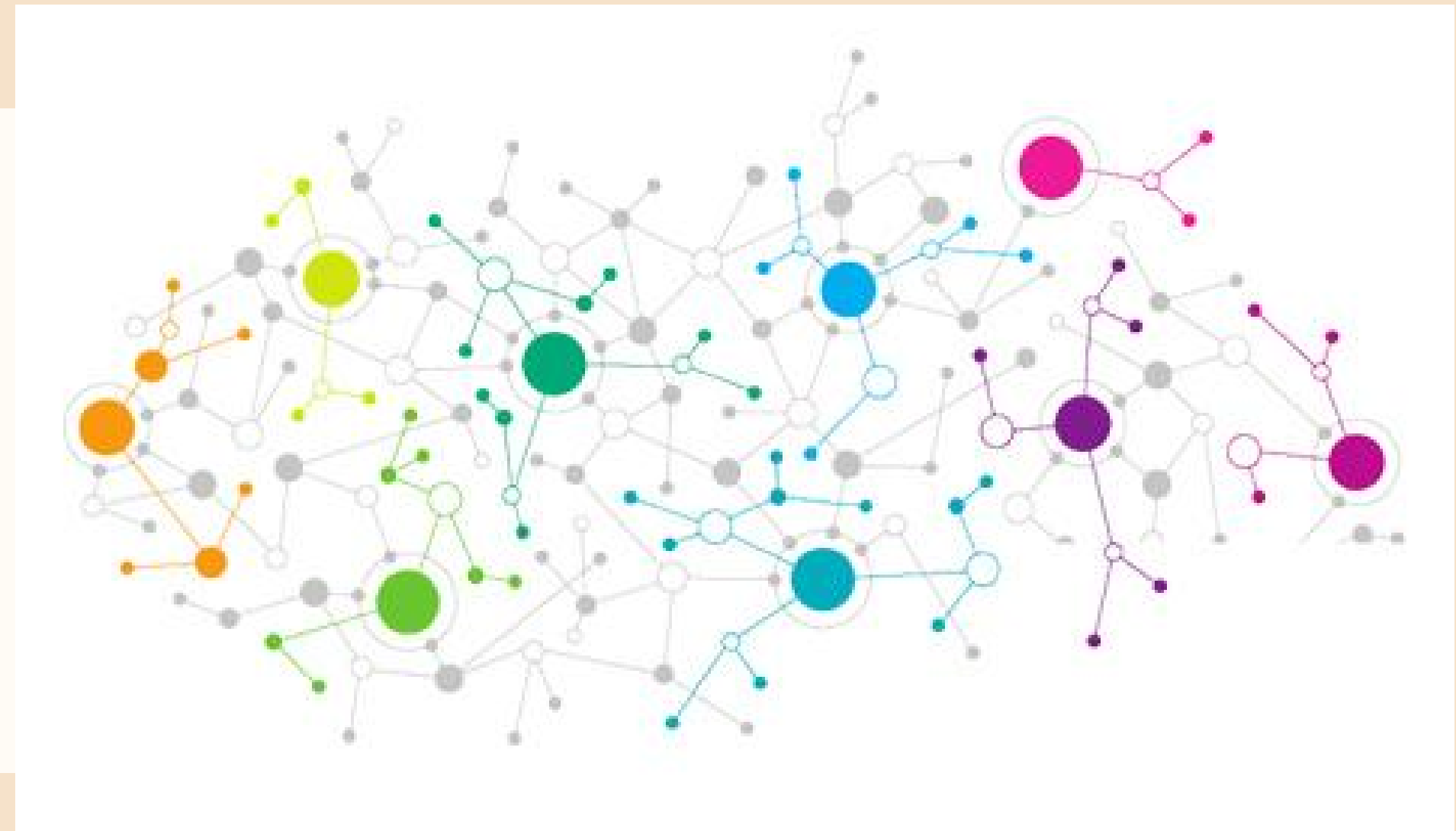
THE BRIDGE - DATA SCIENCE BOOT CAMP



Ciencia de Datos Mercado Laboral

Presentación

- Objetivo
- Contexto
- Hipótesis
- Data wrangling & Data cleaning
- EDA
- Conclusión



Etapas del proyecto

Agosto 2020
102 horas

DEFINICIÓN

8h

PLANIFICACIÓN Y DISEÑO

4h

DATA COLLECTION

4h

DATA WRANGLING & DATA CLEANING

32h

EDA

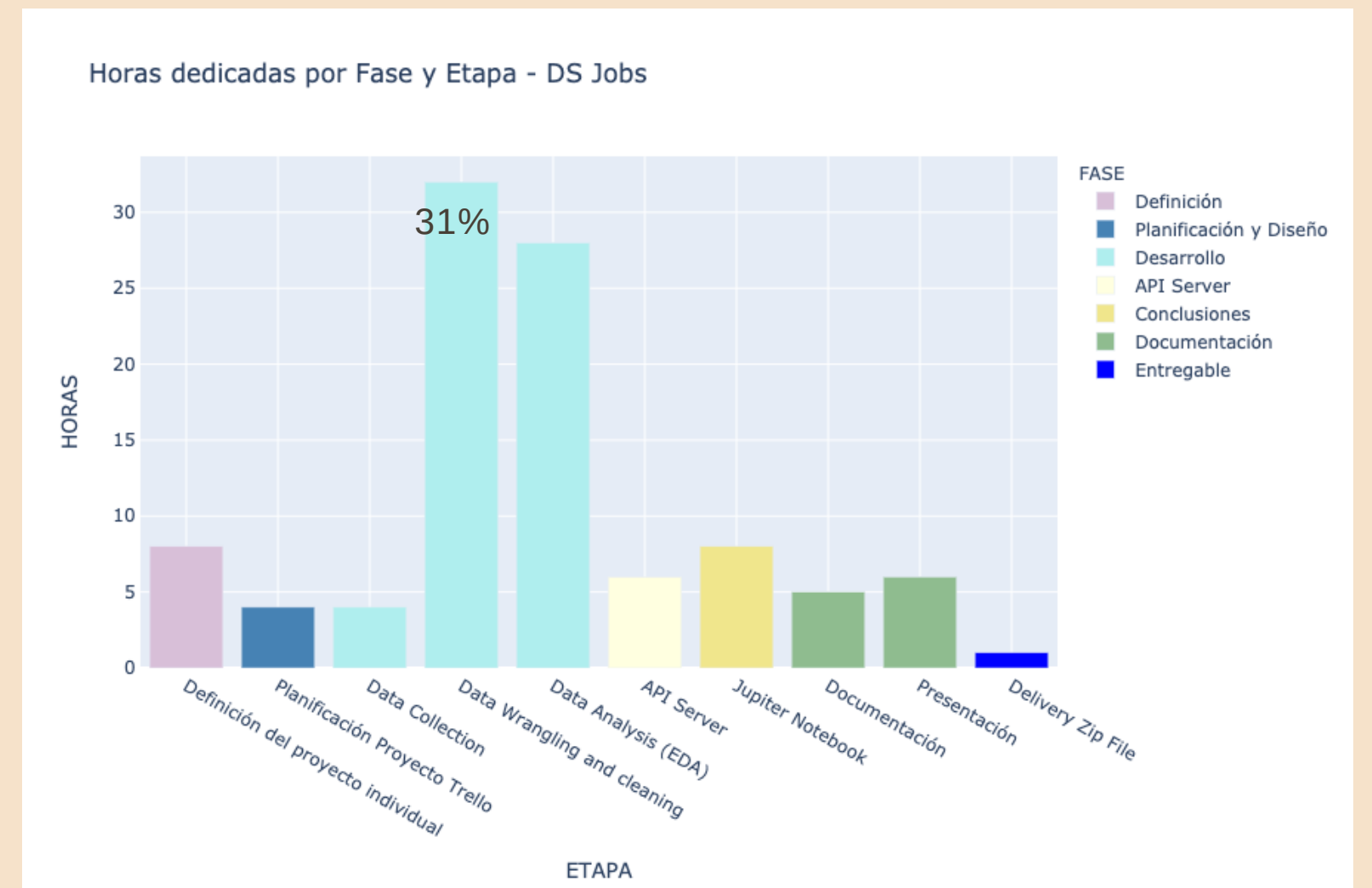
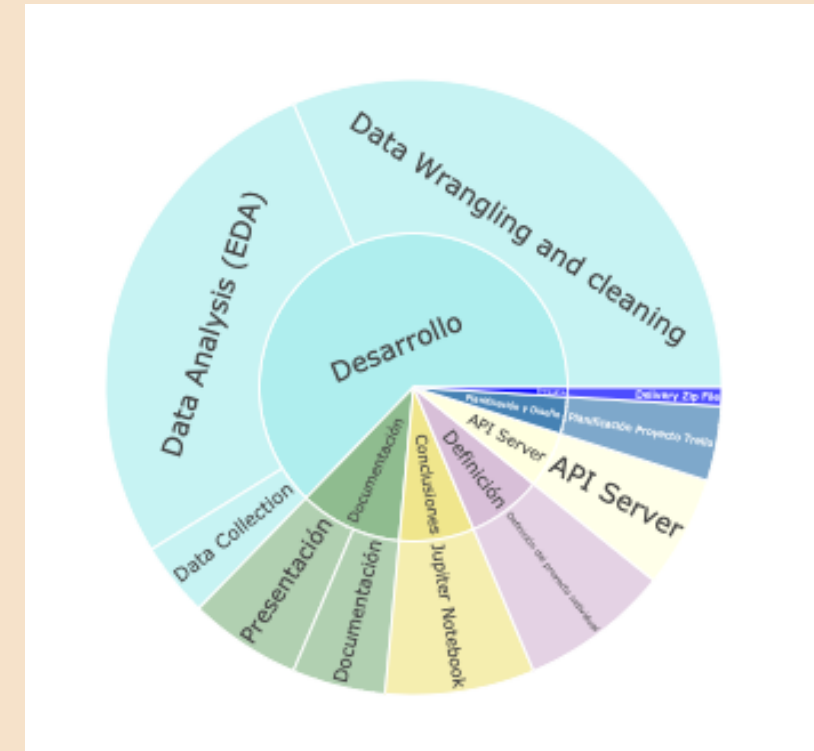
28h

API - SERVER

6h

CONCLUSIÓN

20h



OBJETIVO

¿Por qué este proyecto?



FUTURO PROFESIONAL

Necesidades actuales

Perfil e Intereses

Transformación digital - Big Data

FORMACIÓN INTENSIVA EN DATA SCIENCE

¿HAY OPORTUNIDADES?

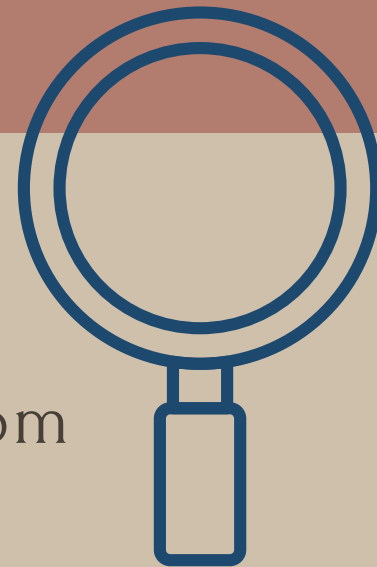
¿Cómo está el mercado en términos de oferta? ¿Qué me espera el día después?

Contexto

INDEED.com
GLASSDOOR.com
LINKEDIN.com

UNA DE LAS COMPAÑÍAS MÁS
GRANDES DEL MUNDO DE
BÚSQUEDA DE EMPLEO Y
RECLUTAMIENTO

Dataset Kaggle. Noviembre 2019



ROLES

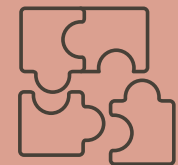
ANALISTA DE
DATOS



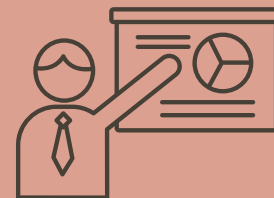
INGENIERO DE
DATOS



CIENTÍFICO DE
DATOS



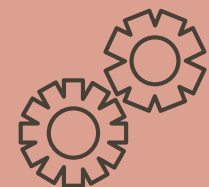
ANALISTA DE
NEGOCIO



GERENTE DE
PROYECTOS



MACHINE LEARNING
AI





Hipótesis

Al menos la mitad - 50% - de la oferta laboral en el área de Ciencia de Datos, corresponde a los perfiles de Análisis (Datos y Negocio)

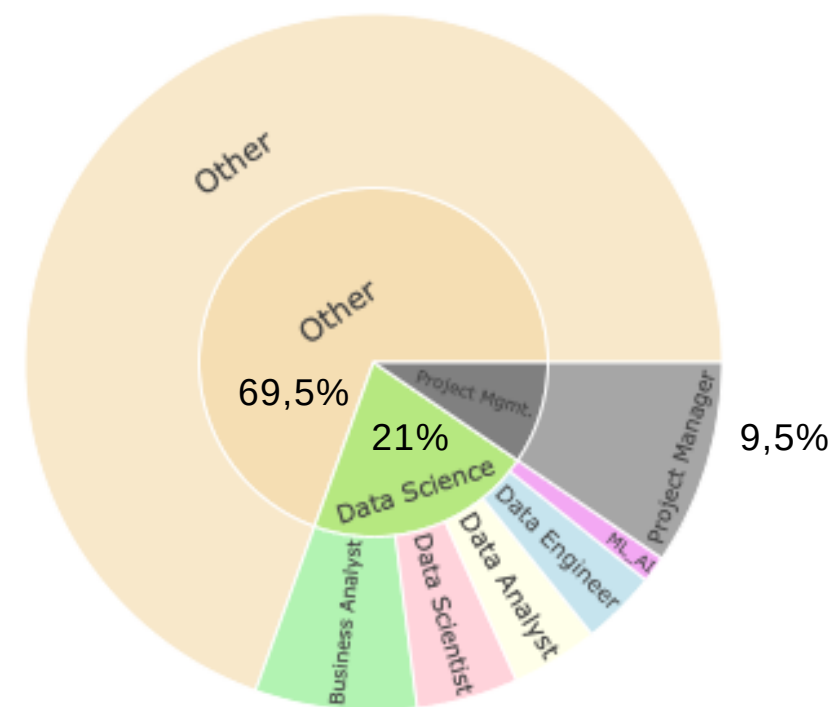
Data Wrangling & Data Cleaning



- ELIMINACIÓN INFORMACIÓN NO RELEVANTE O REDUNDANTE
- ESTANDARIZACIÓN CÓDIGOS DE PAÍS
- ELIMINACIÓN DE DUPLICADOS
- REVISIÓN - VALIDACIÓN DE VALORES NULOS

1. De las 165.290 filas del DF original se eliminaron (por inconsistencias) alrededor del 35%, quedando 104.486 registros
2. Se redujo la memoria utilizada por el dataset en más de 90%% (De 193 MB a 19 MB)

OFERTA DE EMPLEO - DATA SCIENCE - POR ROL (total)



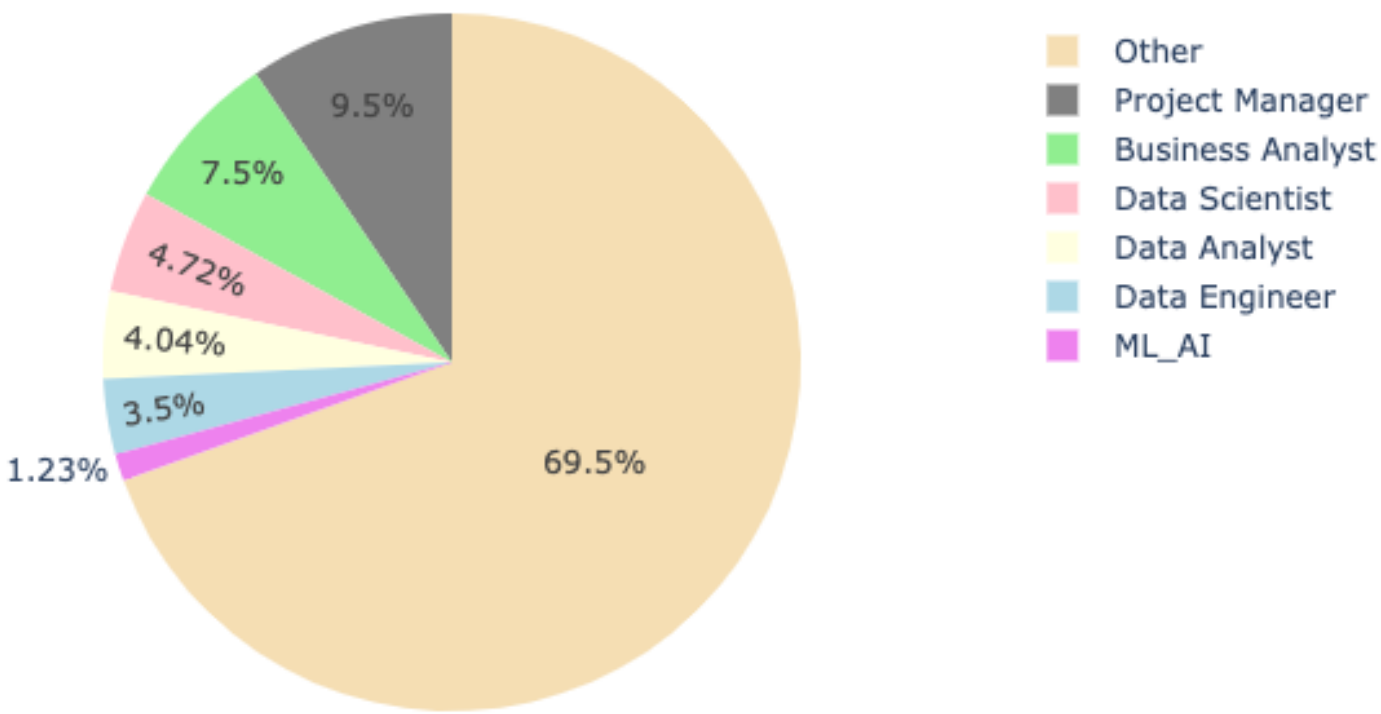
EDA

Data-science por rol

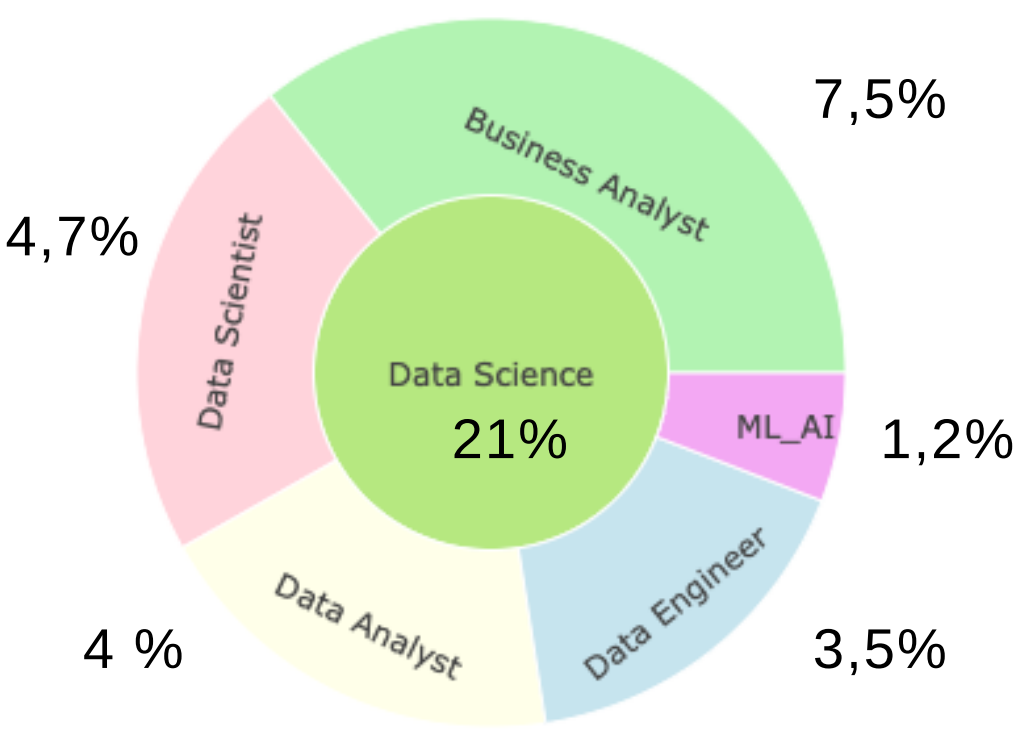
Al menos la mitad - 50% - de la oferta laboral en el área de Ciencia de Datos, corresponde a los perfiles de Análisis (Datos y Negocio)

Hipótesis confirmada. El 55% corresponde a perfiles analíticos (Datos y Negocio)

OFERTA DE EMPLEO - DATA SCIENCE - POR ROL

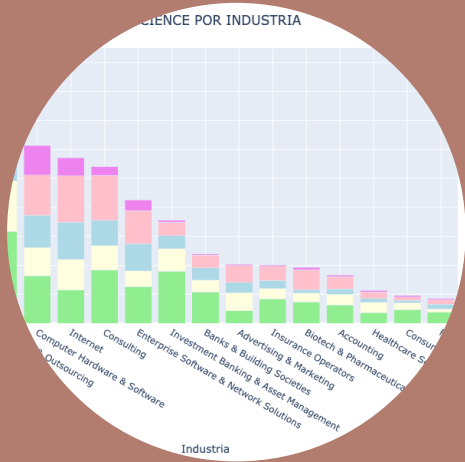


OFERTA DE EMPLEO - DATA SCIENCE - POR ROL (%)

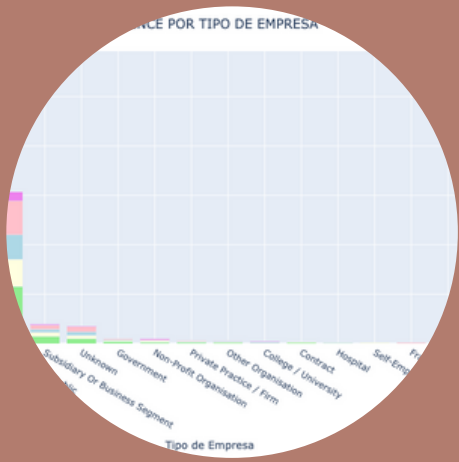


Exploración de los datos - EDA

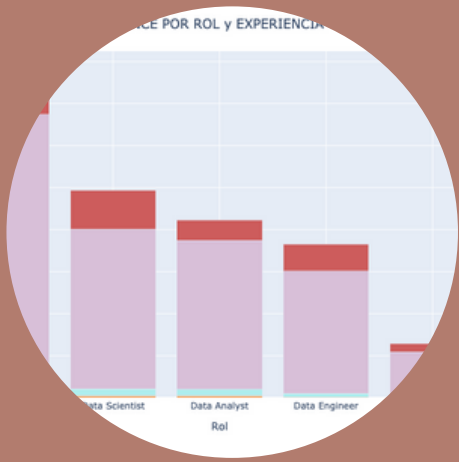
TENDENCIAS



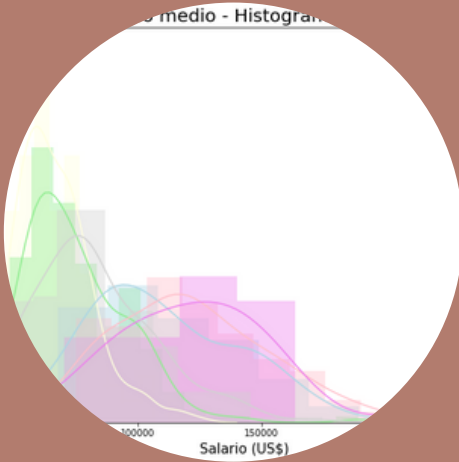
INDUSTRIA



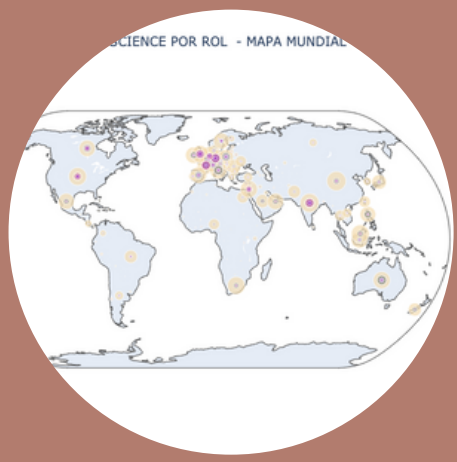
SECTOR



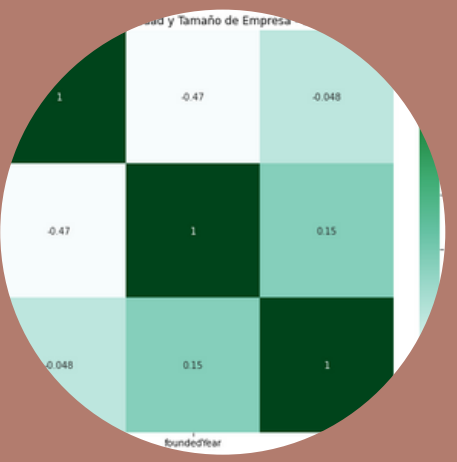
EXPERIENCIA



SALARIOS



DS EN EL MUNDO



CORRELACIÓN

EMPRESAS



EMPRESA



TIPO DE EMPRESA

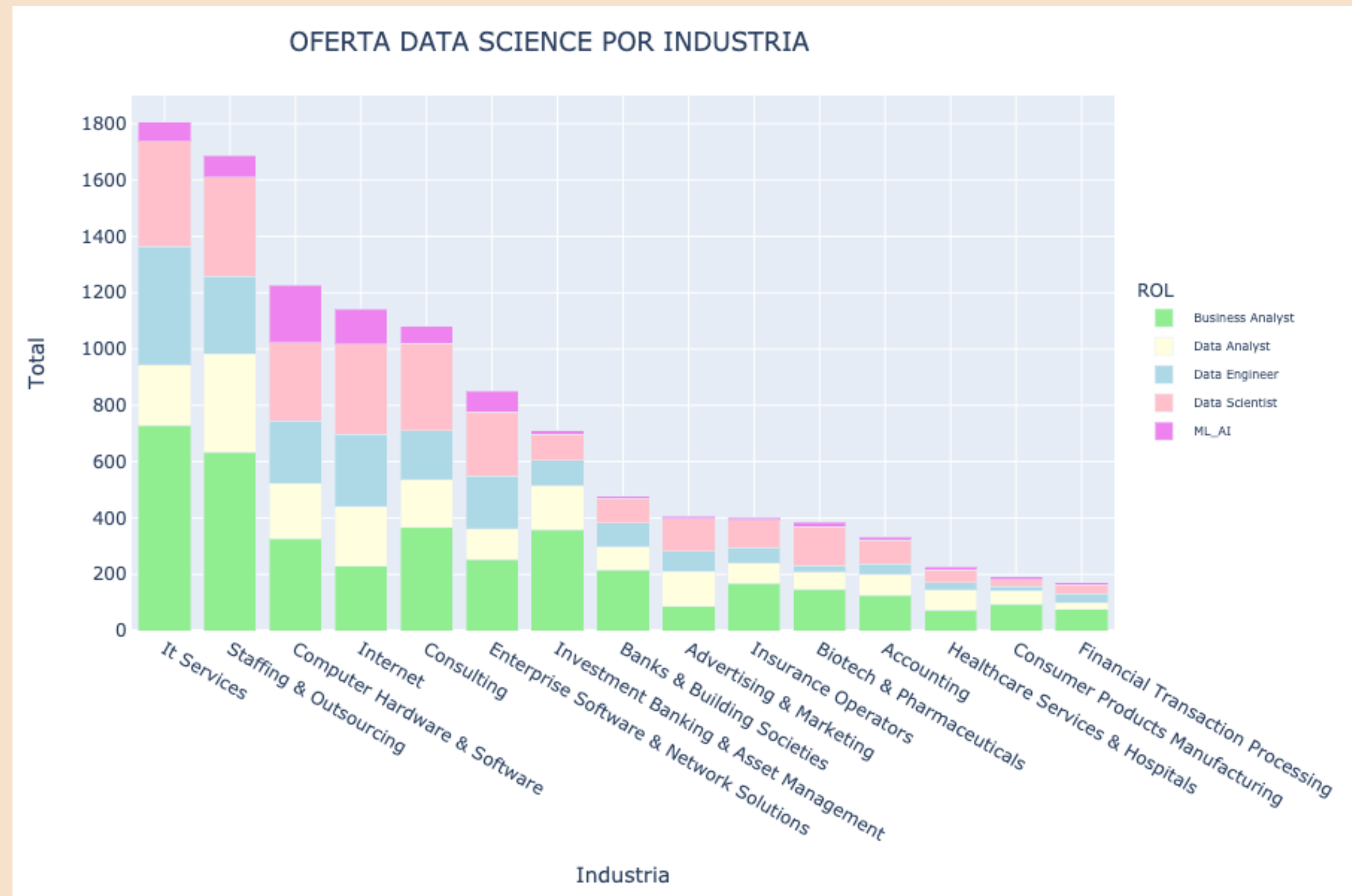


TAMAÑO Y ANTIGÜEDAD

Industria

El 45% de la oferta se concentra en la industria tecnológica y de consultoría.

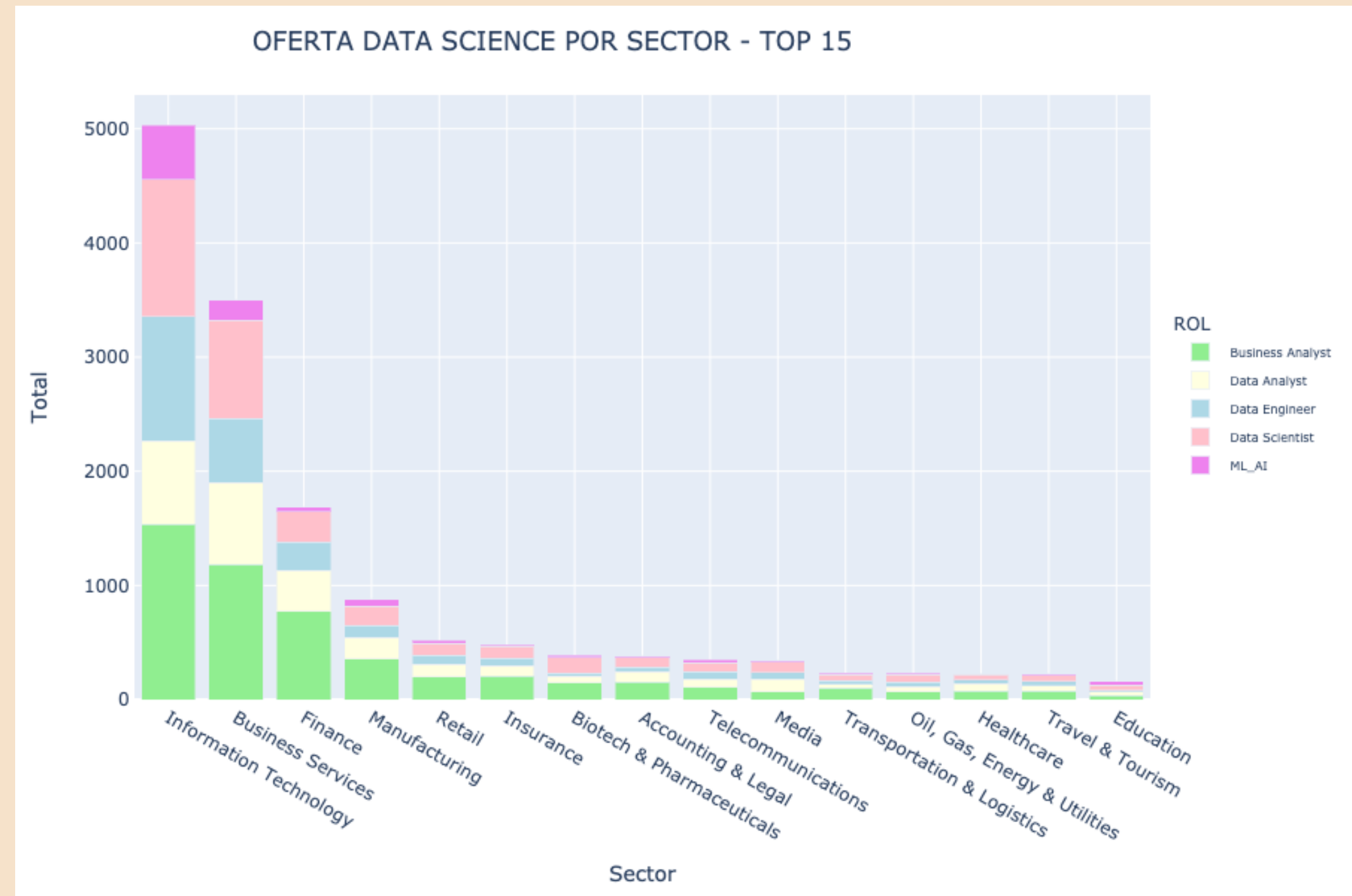
Para el rol de Business Analyst, el sector de Banca e Inversión desplaza a Internet y se coloca entre los 5 primeros



Sector

Los 3 primeros sectores:
Tecnología de Información,
Servicios de negocio y
Finanzas
concentran el 60% de la oferta

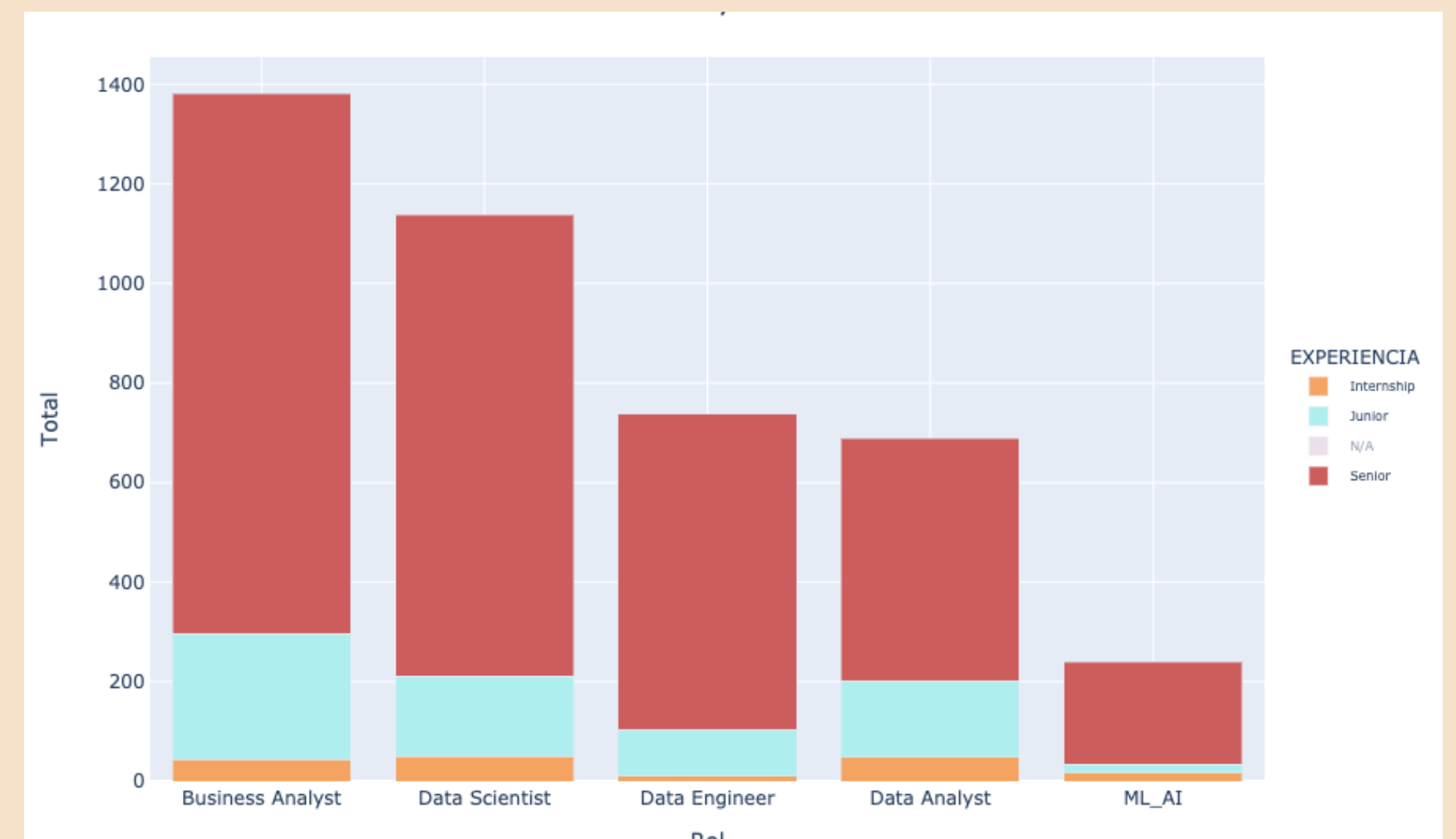
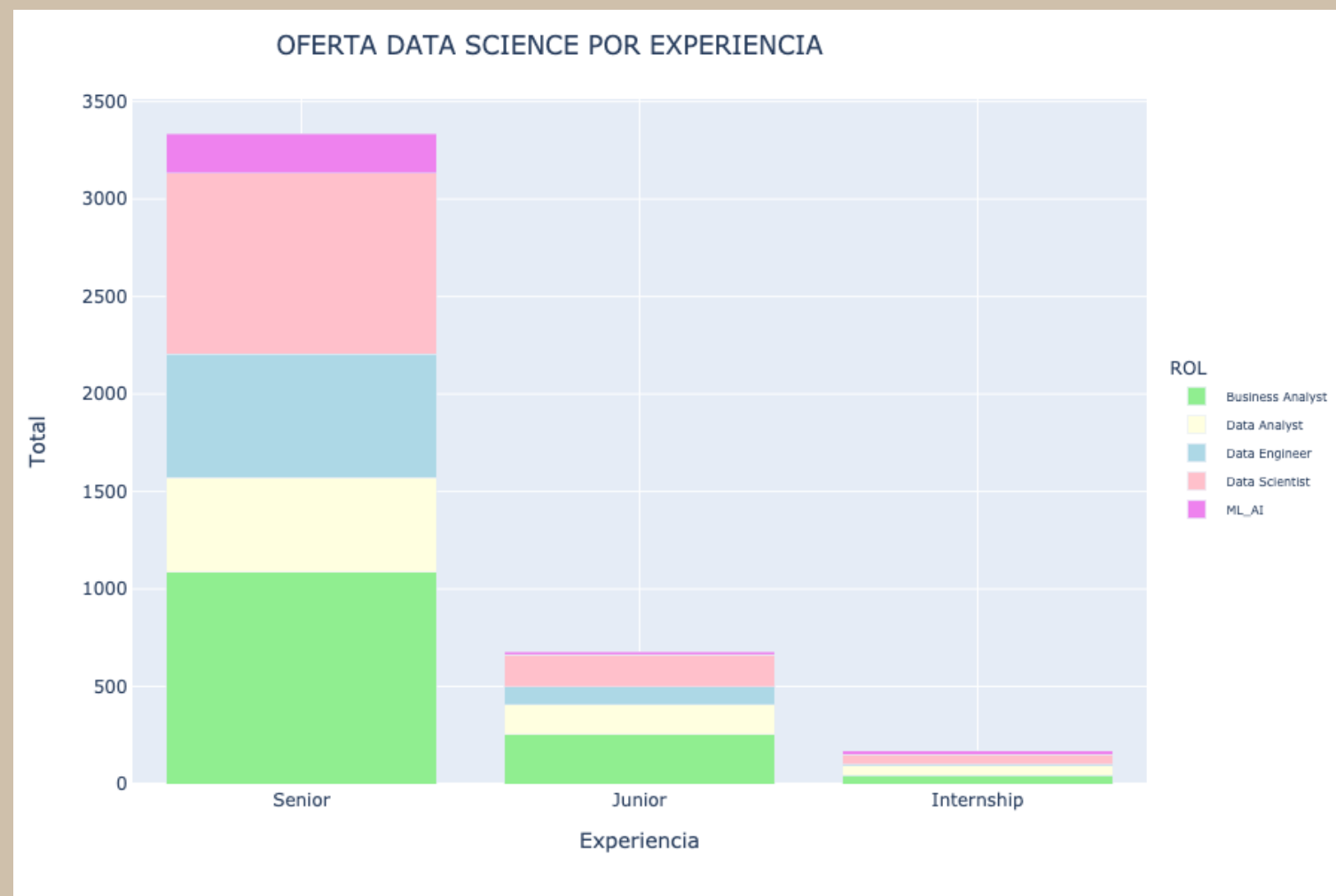
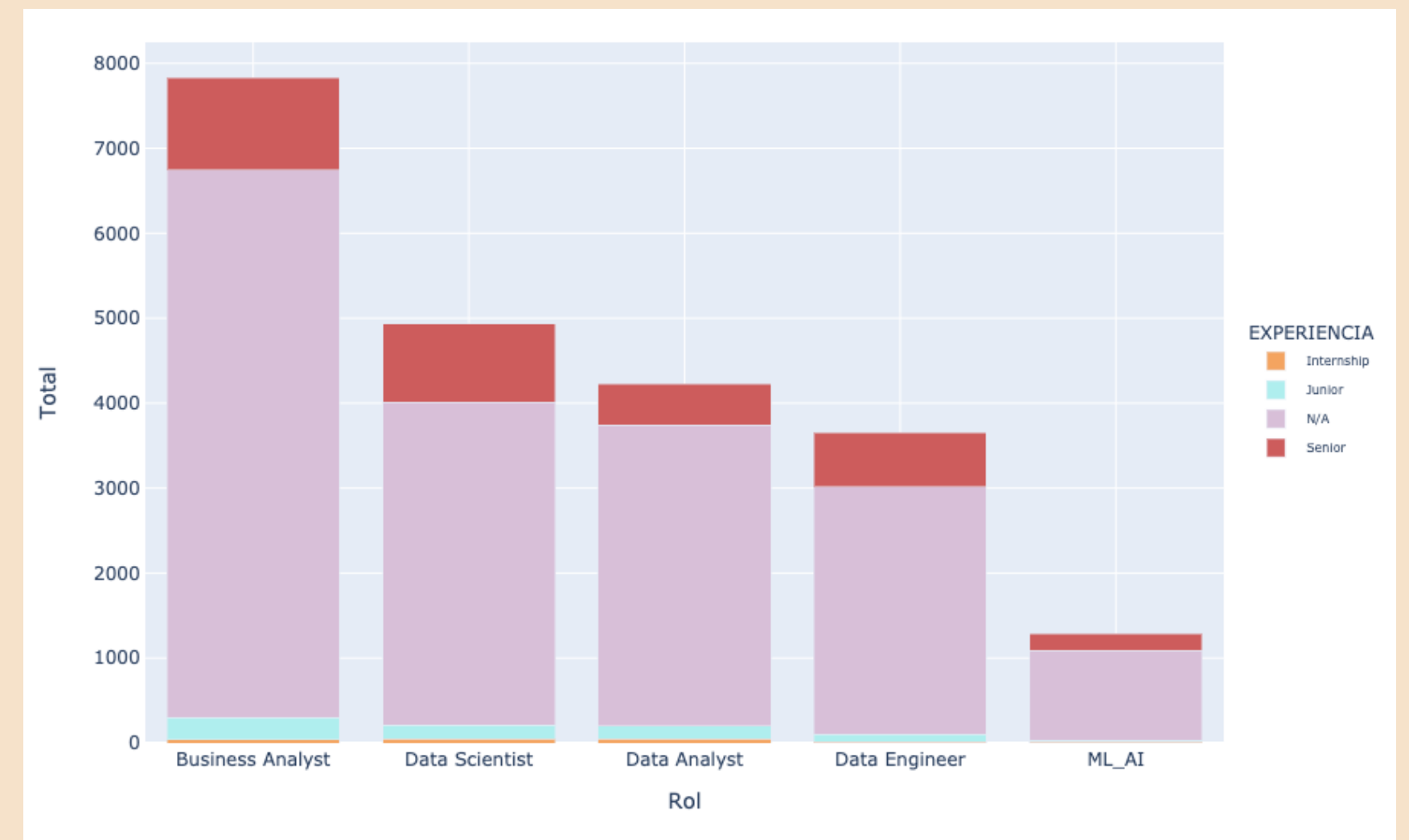
Esta proporción se mantiene para todos los
roles de DS analizados.



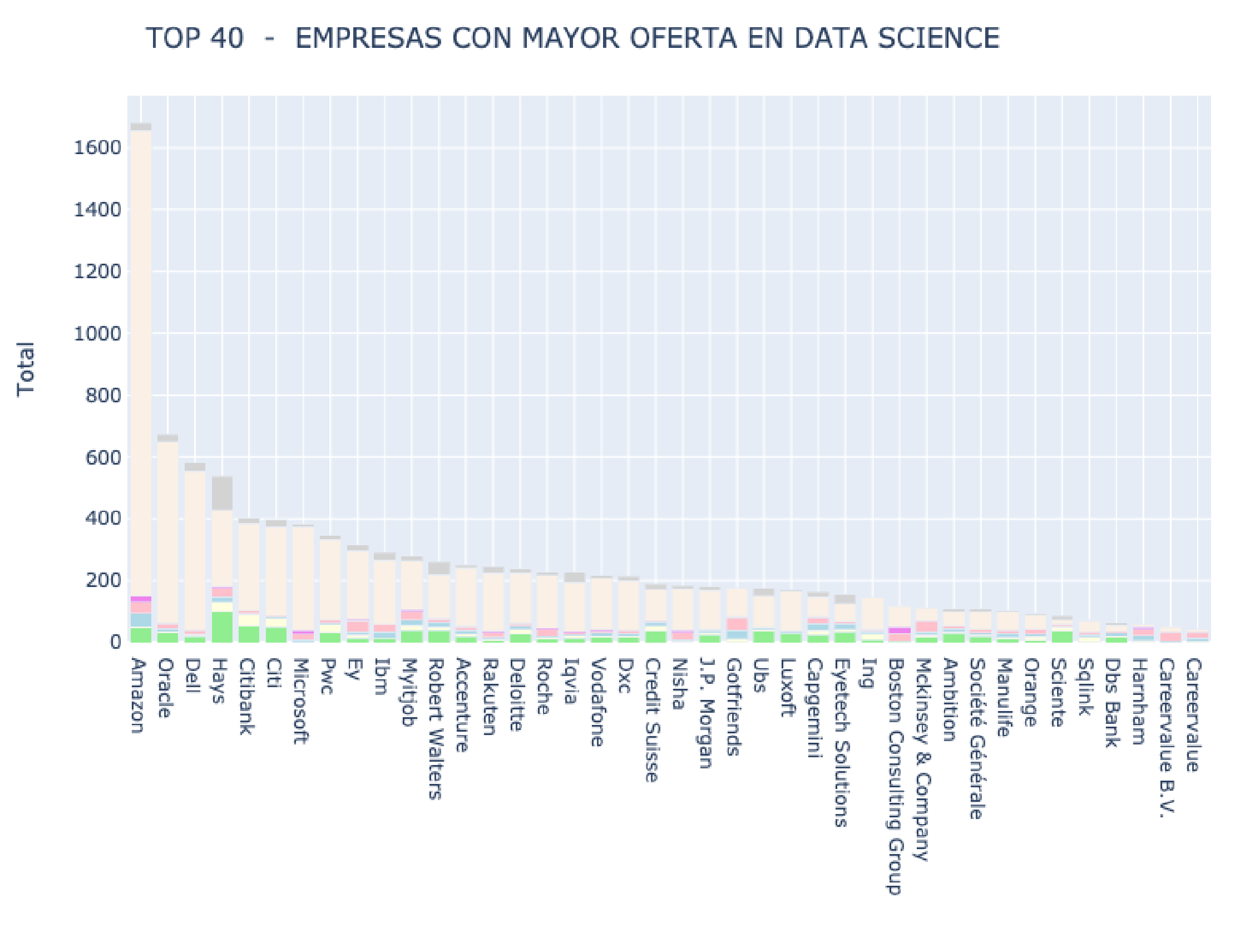
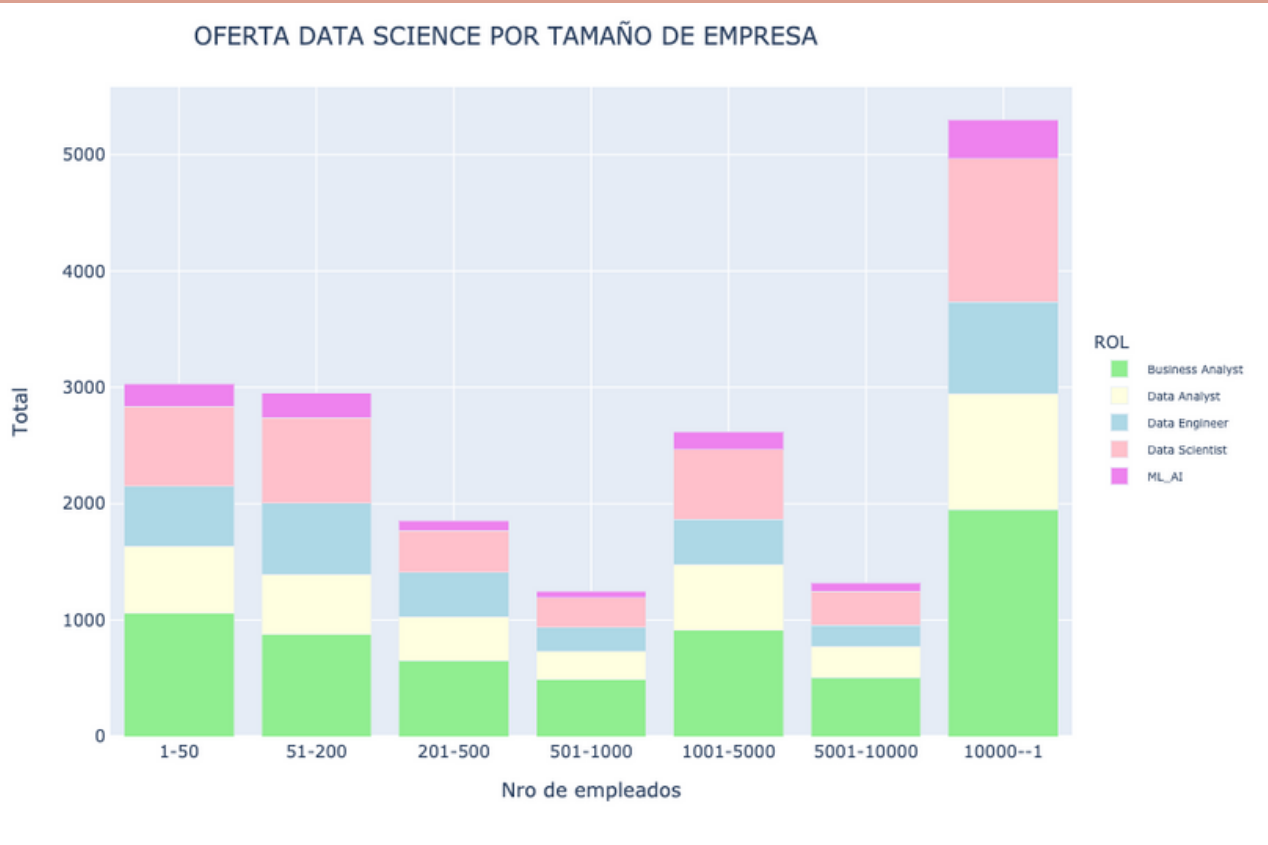
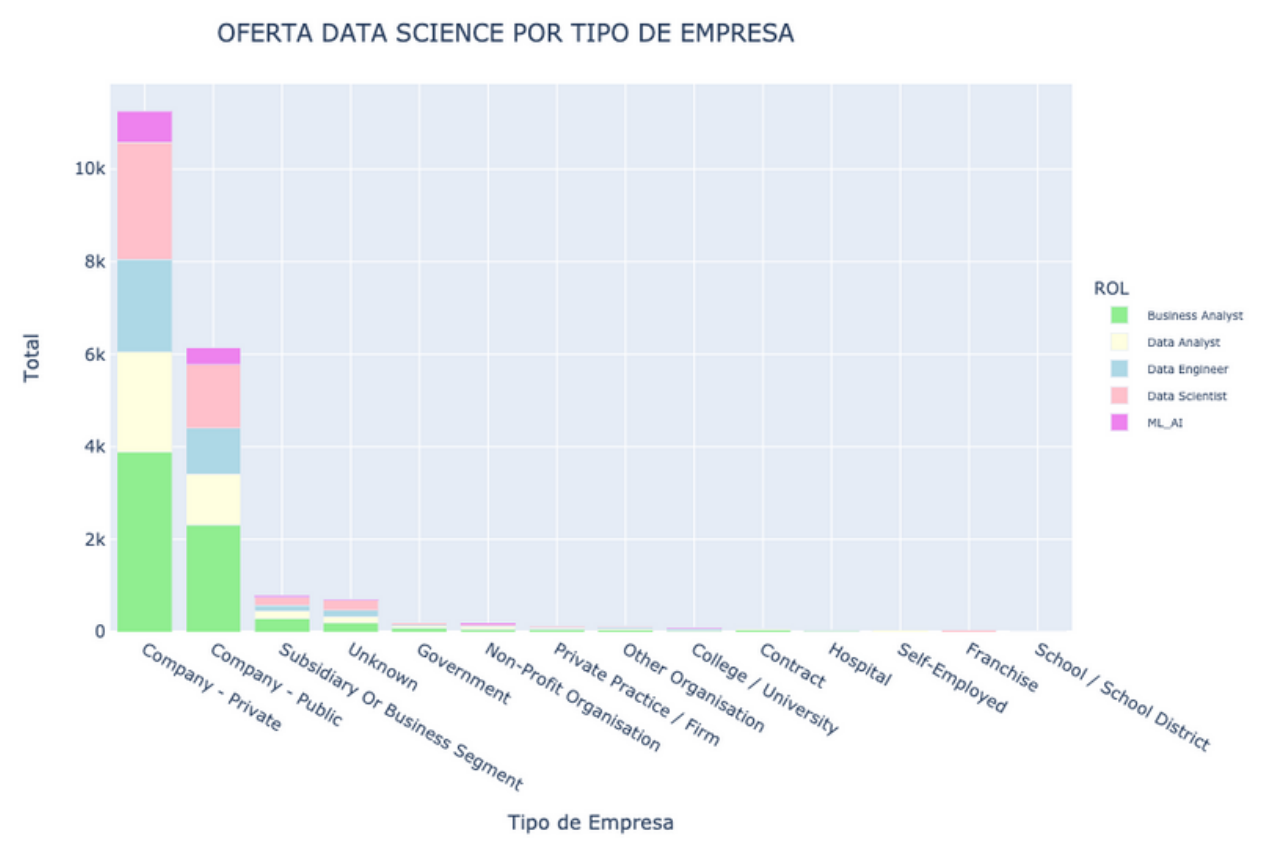
Experiencia

Solo el 25% de las ofertas en Data Science especificaban el nivel de experiencia requerida. 19.6% Senior, 4.2% Junior y 1.2 recién titulados.

Esto es más notable para el caso de los roles más especializados

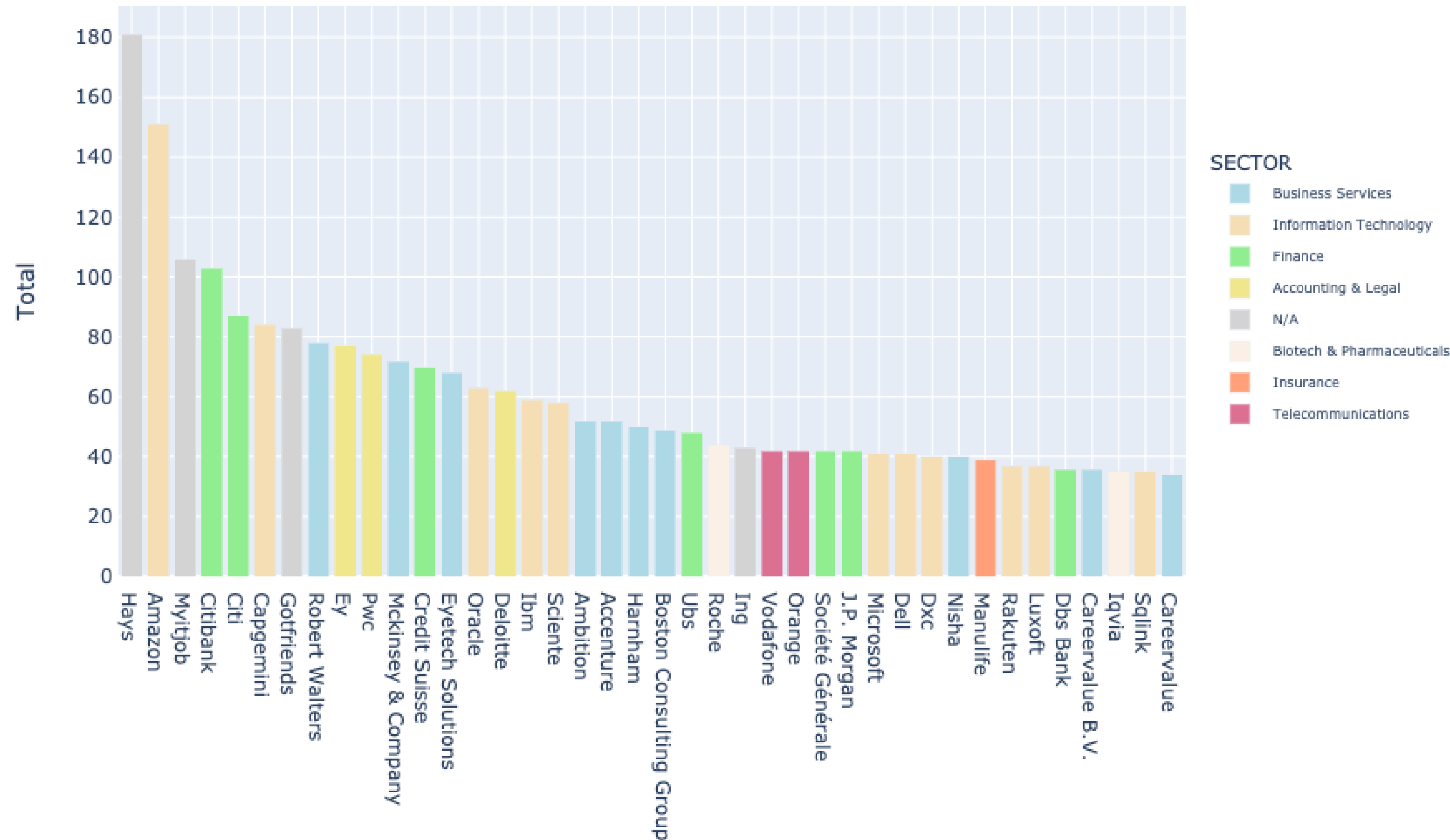


Empresa



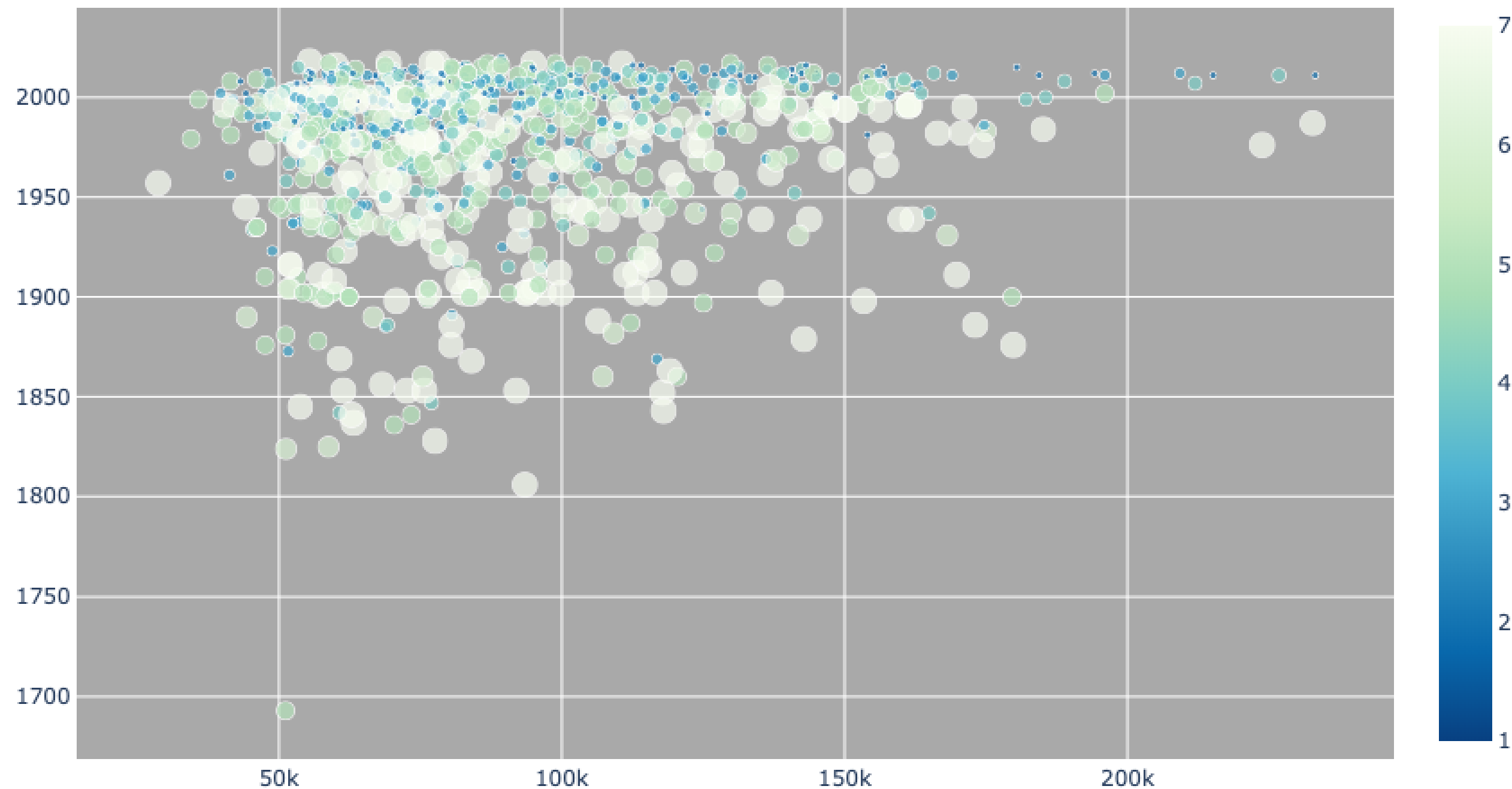
Empresa

TOP 40 - EMPRESAS CON OFERTA EN DATA SCIENCE POR SECTOR



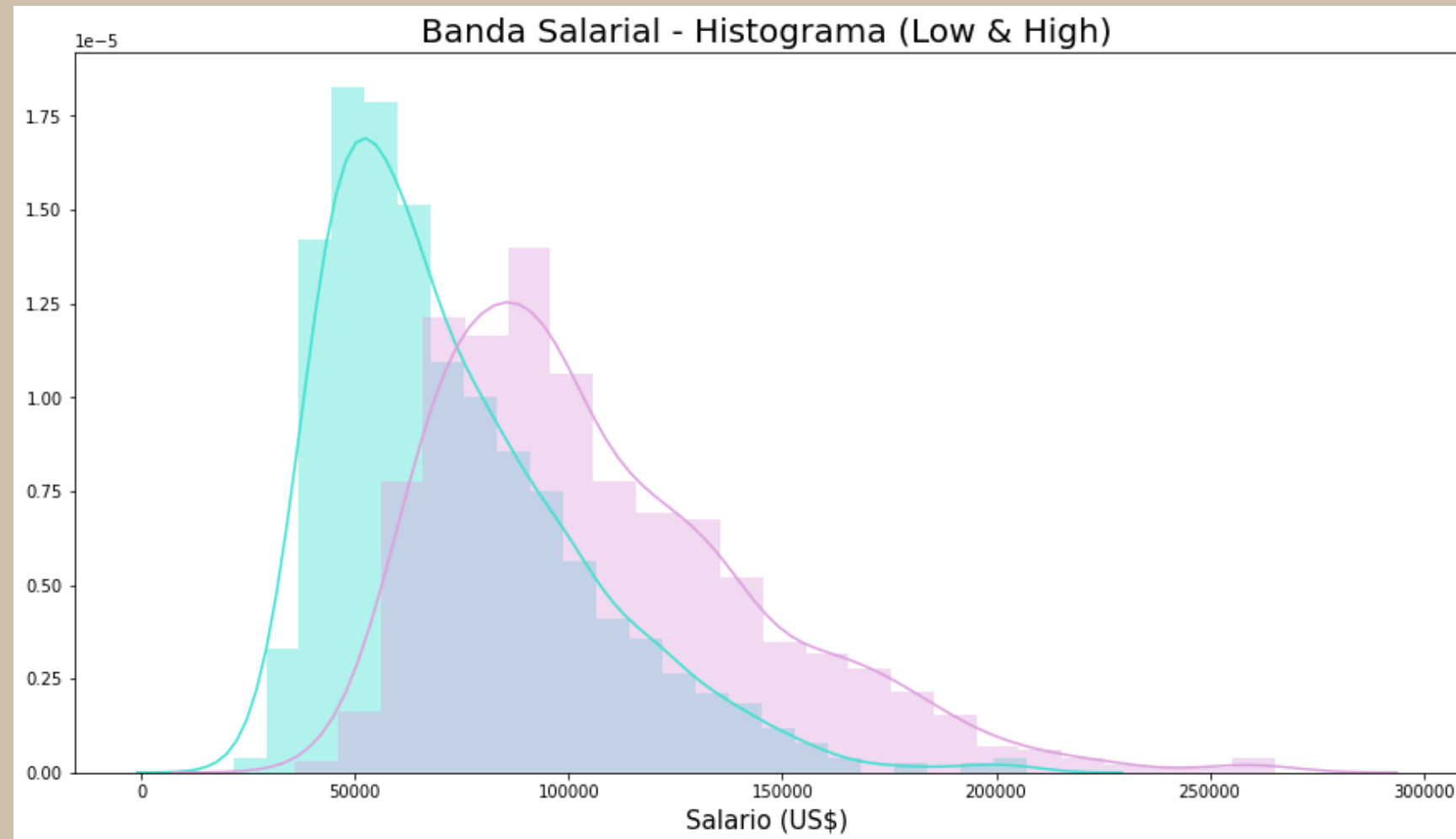
- El 50% de la oferta proviene de empresas de naturaleza privada y el 32% de caracter público.
- Entre los primeros lugares se encuentran las empresas de selección especializada de personal (Hays, myltjobs y Gotfriends)
- Amazon y las empresas de consultoría estratégica (como BCG, McKinsey) son, según esto, quienes más se están desarrollando en esta área a nivel especializado (DS y ML)

Salario medio por antigüedad de la empresa y número de empleados



Antigüedad y número de empleados

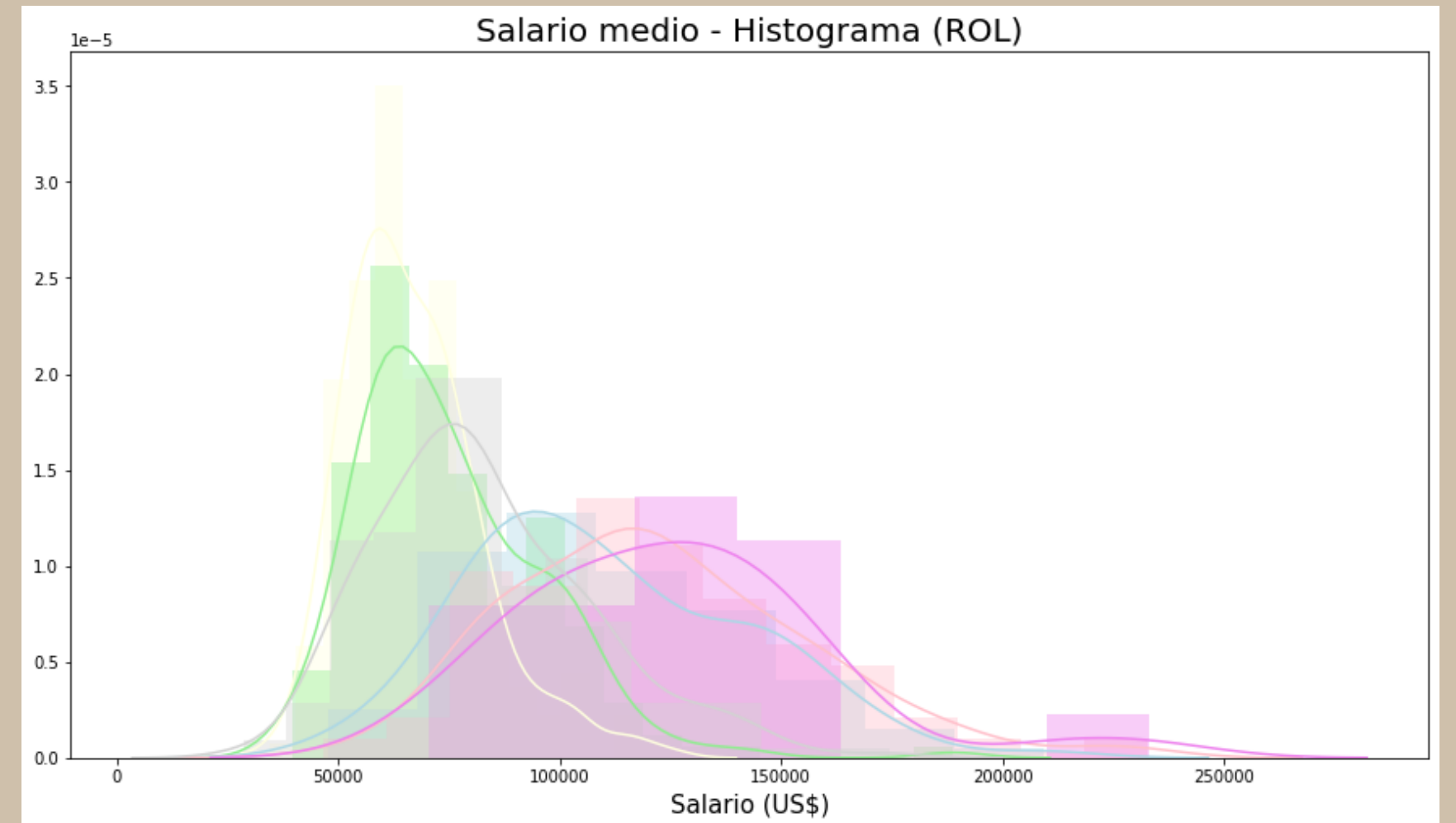
- Alrededor de 1/3 de las ofertas está en empresas pequeñas (menos de 200 empleados), 1/3 en ofertas de empresas entre 200 y 10.000 empleados y el otro tercio restante en empresas grandes (más de 10.000 empleados).
- Los salarios más altos están en empresas de antigüedad menor (más innovadoras y menos tradicionales) y de menor tamaño (desde el punto de vista de número de empleados)



Bandas - mínimos y máximos

La banda salarial se mueve entre los 50.000 dólares anuales a 90.000 dólares anuales.

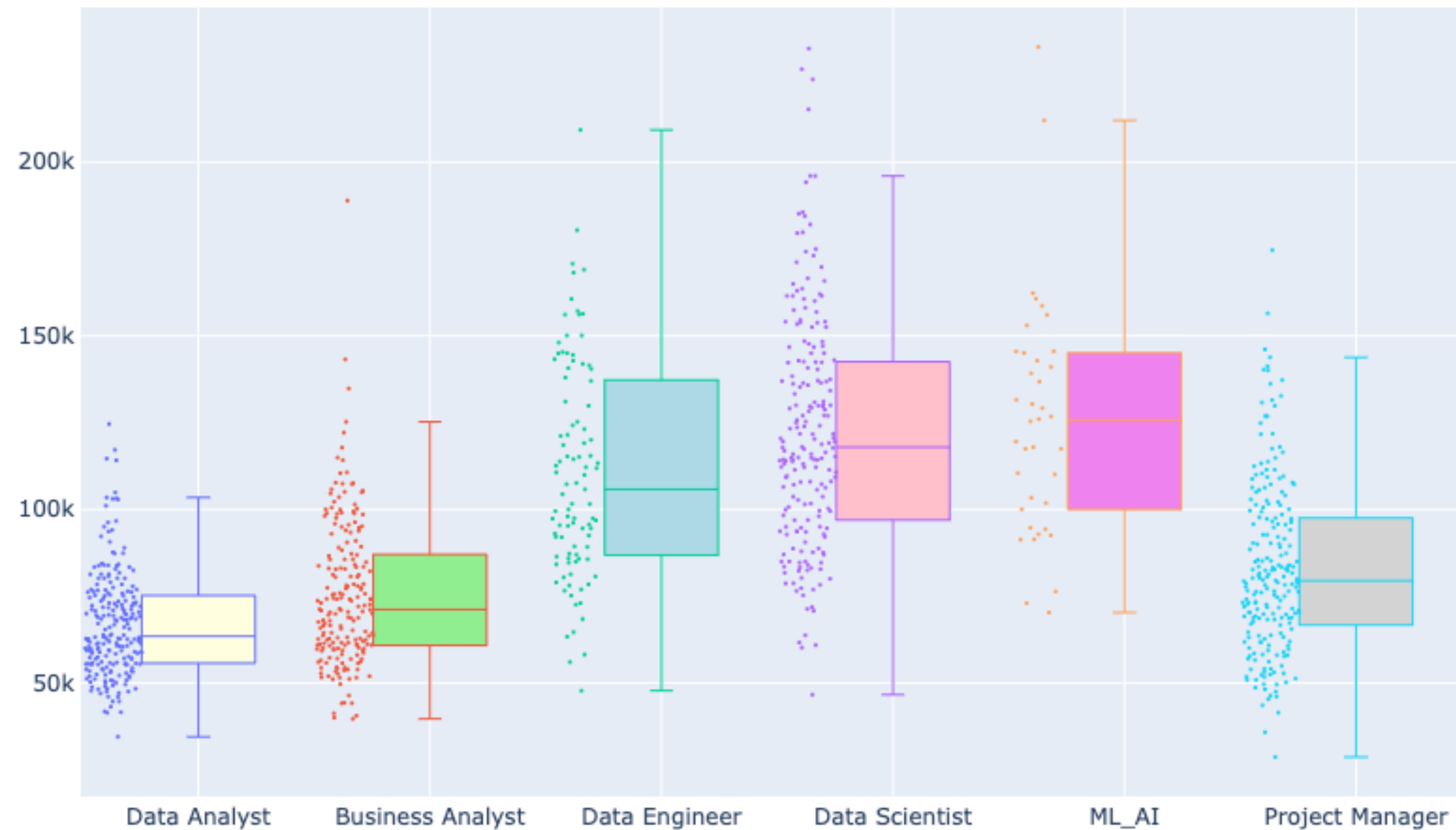
Ambas curvas se asemejan a una campana de gauss y están dispersos en torno a los valores de los picos de la curva (con mayor dispersión para el caso del tope superior de la banda salarial donde hay mayor desviación con respecto a la media]



Salarios por rol

Se observa con claridad la variación del salario en función al rol, siendo los menos retribuidos los roles asociados a Analistas de Datos seguidos por los Analistas de Negocios . El salario aumenta a medida que el rol va requiriendo más experticia técnica o más nivel de especialización (Ingenieros de Datos, Científicos de Datos y Machine Learning)

Box Plot - Salario (US\$) por ROL (Data Science & Project Management)



Distribución / Dispersión del salario por rol

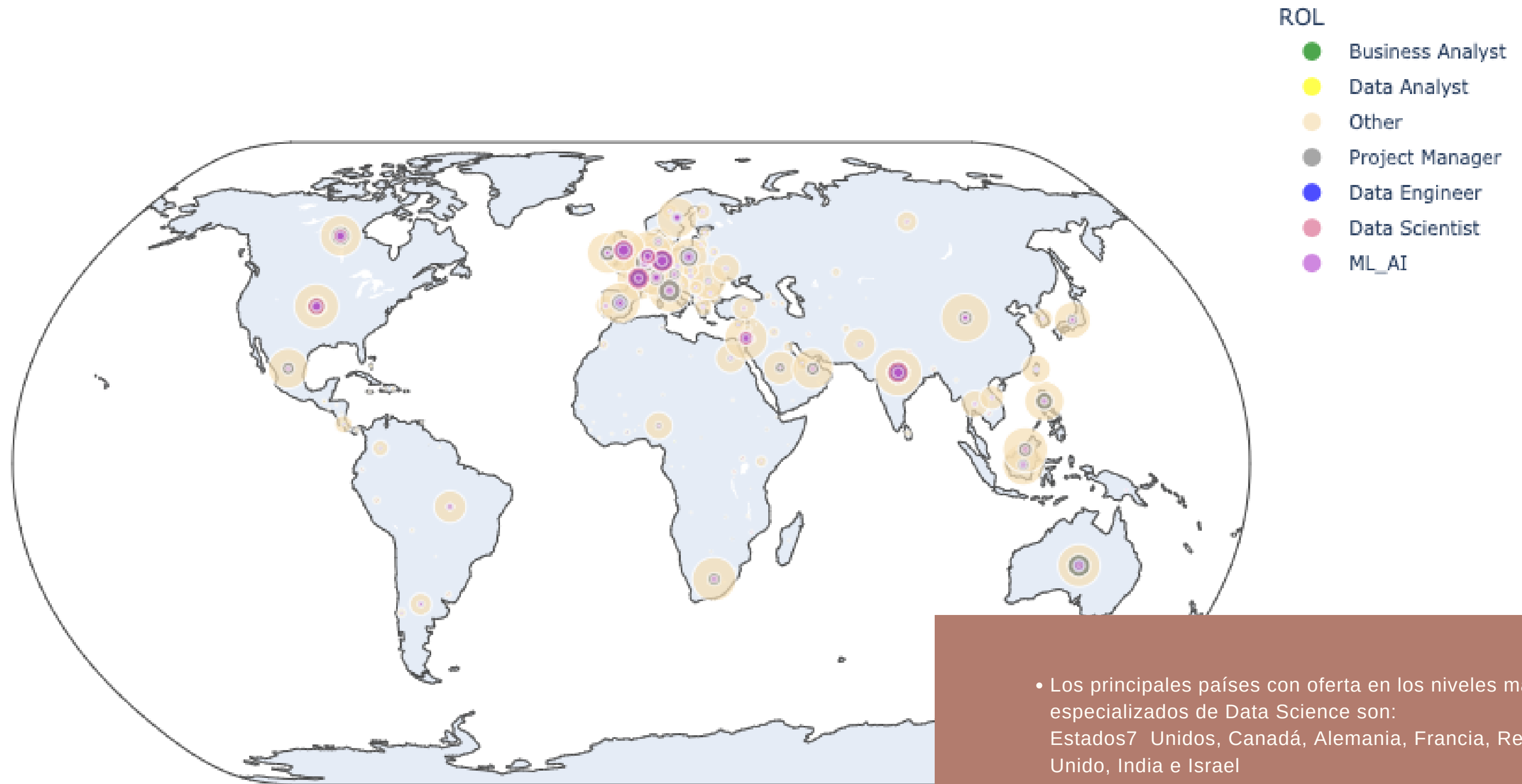
Outliers (Salario) Empresas

Las empresas outliers son en su mayoría empresas tecnológicas pequeñas (menos de 500 empleados) o del sector salud (biotecnología y farmacéuticas) (más de 5000 empleados)-

En términos de salarios pertenecen principalmente a los sectores de Tecnología de Información, de investigación (farmacéutica, salud, biotecnología) y Aeroespacio y Defensa



OFERTA LABORAL EN DATA SCIENCE POR ROL - MAPA MUNDIAL

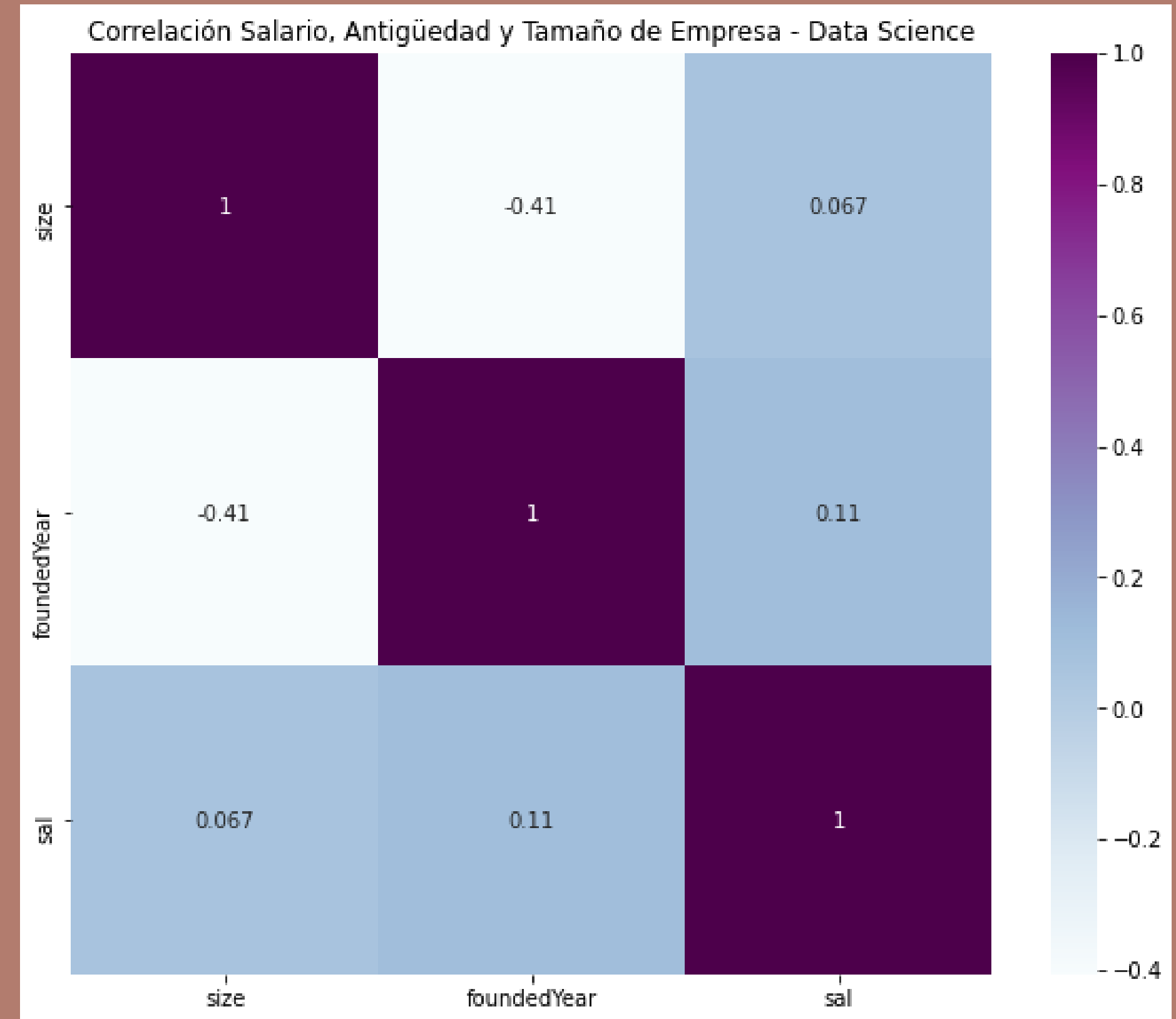


- Los principales países con oferta en los niveles más especializados de Data Science son:
Estados Unidos, Canadá, Alemania, Francia, Reino Unido, India e Israel
Para los niveles más básicos (Analistas), se agregan a esta lista Suiza, Bélgica, Holanda y Australia.-
Europa es el continente donde se concentra la mayor cantidad de ofertas en esta área

CORRELACIÓN

No se observa correlación alguna entre las variables,

NOTA: Si analizamos la correlación por rol, para el caso de Ingenieros de datos o Ciencias de datos se puede ver una correlación negativa baja (un poco mayor a 0.5) entre el año de fundación de la empresa y el tamaño de la misma.



¿Qué aprendí? ¿Qué haría diferente?

DATA ACTUALIZADA
(RECIENTE Y DE ESPAÑA)

A TRAVES DE UN API U OTRAS FUENTES

MEJOR ESTIMACIÓN DE TIEMPOS
(DATA WRANGLING)

BALANCEAR MEJOR LOS TIEMPOS Y CARGA

MAYOR APROVECHAMIENTO DE
GRÁFICOS

MEJOR COMUNICACIÓN DE RESULTADOS

ANALIZAR HERRAMIENTAS
REQUERIDAS POR ROL

PARA TENER UNA VISIÓN MÁS COMPLETA

¡GRACIAS!

MONICA VILLASUSO

DATA SCIENCE BOOT CAMP

THE BRIDGE
Septiembre 2020