Workshop 1

Federico Gonzalez Cardenas

Javier Alejandro Vergara Zorrilla

ETL



Ingeniería de datos e inteligencia artificial, Facultad de ingeniería

Universidad Autónoma de Occidente

Santiago de Cali

2024

Aspectos a tener en cuenta:

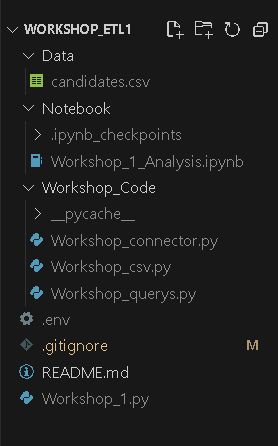
* A lo largo de los archivos de python y el notebook se usaron consultas sql, estas se usaron principalmente para la creación y modificación de la nueva tabla con el campo hired y la visualización de esta, todo desde python (Ninguna consulta SQL se ejecutó en un intérprete o motor SQL)
* Absolutamente todo el código que trae consigo variables con lenguaje de consultas SQL a sido hechas desde python conectada a la base de datos, ninguno se realizó fuera de python (Se verán consultas realizadas en MySQL Workbench en este documento solo como prueba de que todo lo que se hizo en python se ha guardado y se ha podido visualizar correctamente desde la misma terminal de la base de datos).
* El notebook trae consigo todo el código que fue el proceso de visualizar,migrar y tratar los datos del csv mientras que en el github se mantuvo la estructura de la manera en que se estuvo trabajando el workshop a lo largo del corte (diferentes archivos, cada uno con su proceso individual) - Aspecto importante para los archivos .py
* En este documento se mostrara las graficas y analisis individual que se realizo en el notebook, para resaltar la parte que divide el codigo del github con el notebook, de igual manera se resalta ya que varias de las graficas que realizaron con el fin de tratar las conclusiones que se pedían en el dashboard de una manera anticipada al desarrollo de esta, como una “preparación” de lo que iba y se tenia que ver en el tratamiento de los datos en el dashboard

Pruebas y descripción de todo el proceso:

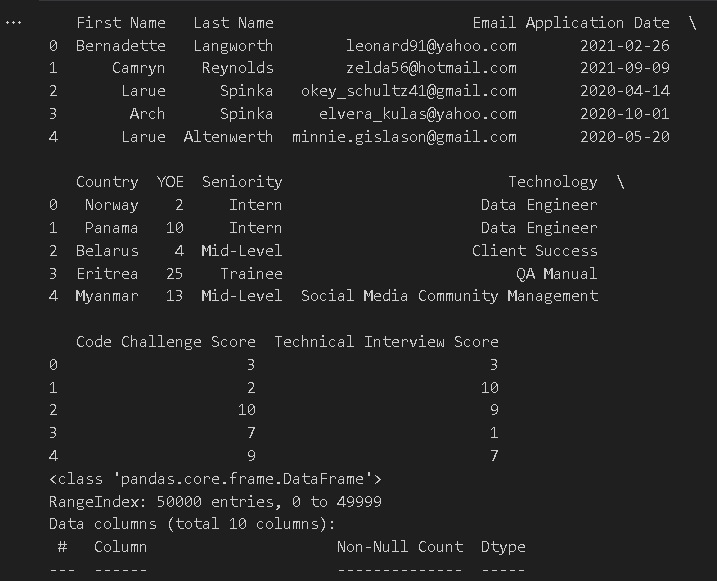
1. Herramientas utilizadas para el desarrollo del trabajo

* Python ( Pandas, Matplotlib, Sqlalchemy dotenv)
* MySQL
* Jupyter Notebook
* Csv (Candidates)
* Encriptación de las credenciales usando un archivo .env (Variables de entorno)

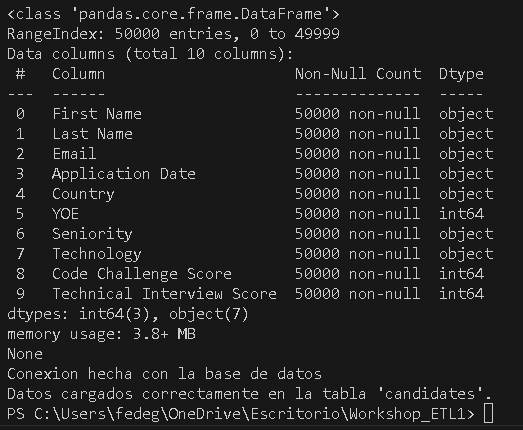
1. Visualización del manejo de los archivos y carpetas en vs



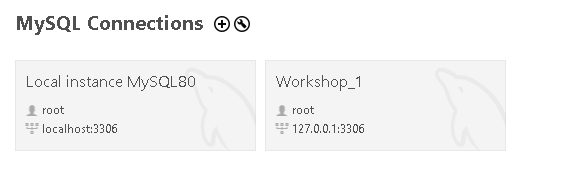
1. Lectura del csv candidates (Workshop\_csv.py)

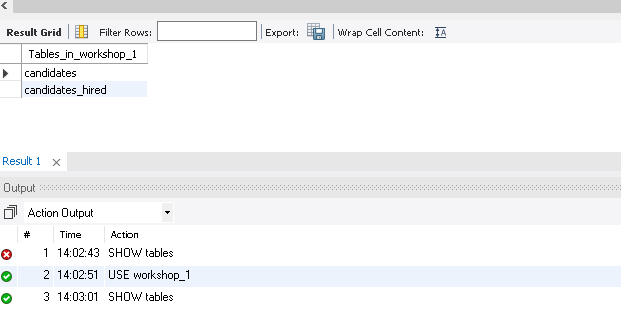


1. Conexión y subida del dataset a la base de datos (Workshop\_connector.py)

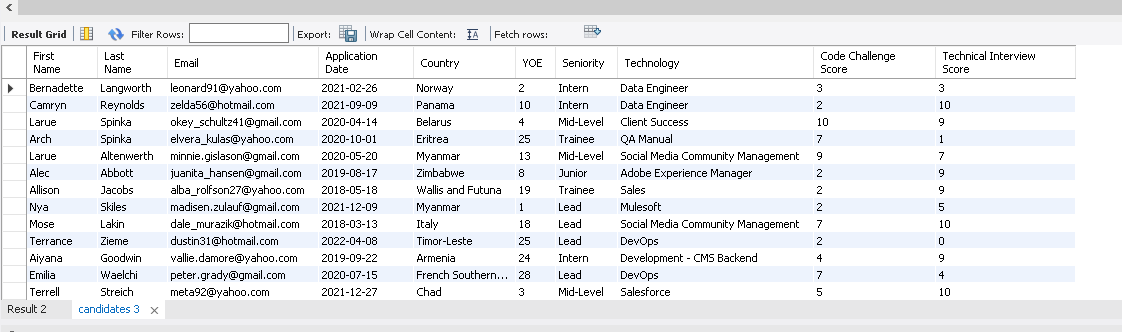


- Verificación y visualización de la conexión y subida de datos desde MySQL workbench



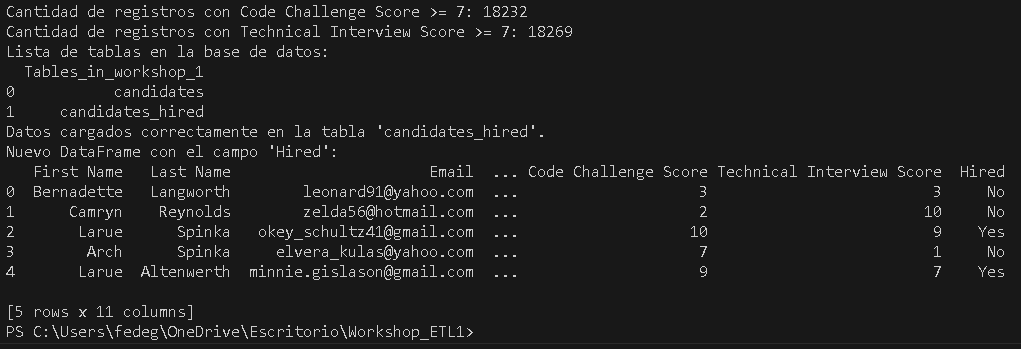




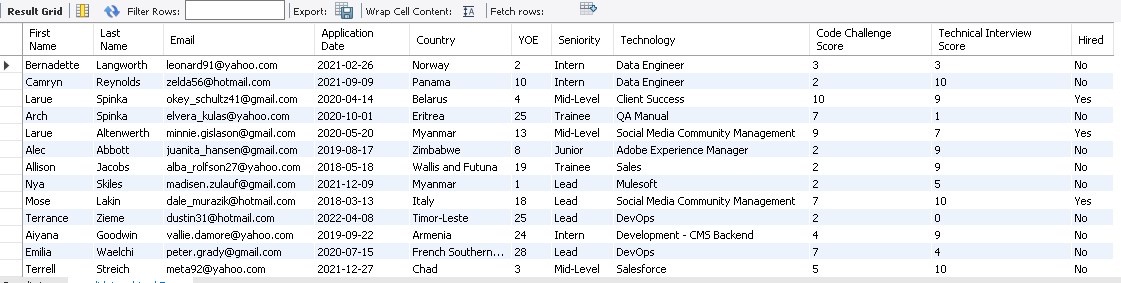


1. Se clona la tabla candidates agregando a esta nueva tabla el campo “Hired”, se guarda esta nueva tabla en la base de datos con el nombre de candidates\_hired

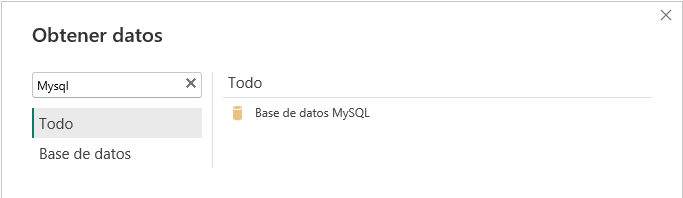
(Workshop\_queys.py)

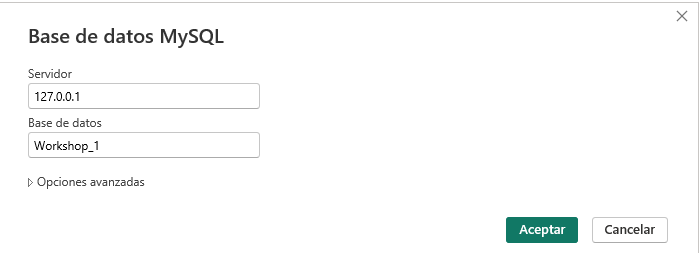


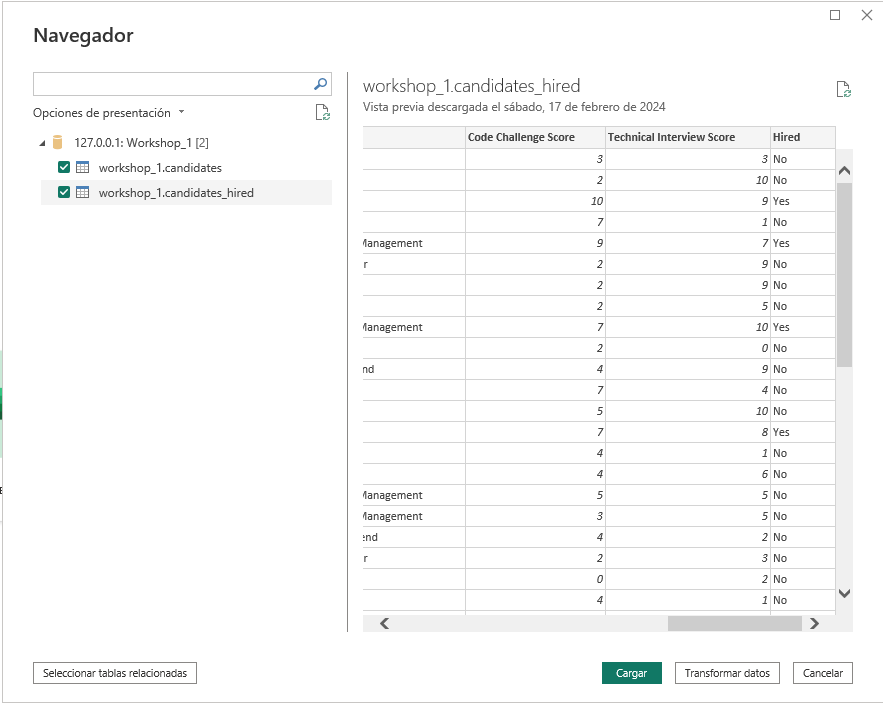


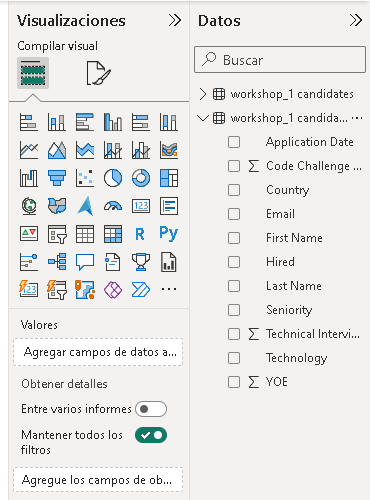


1. Conexión de Power BI con la base de datos

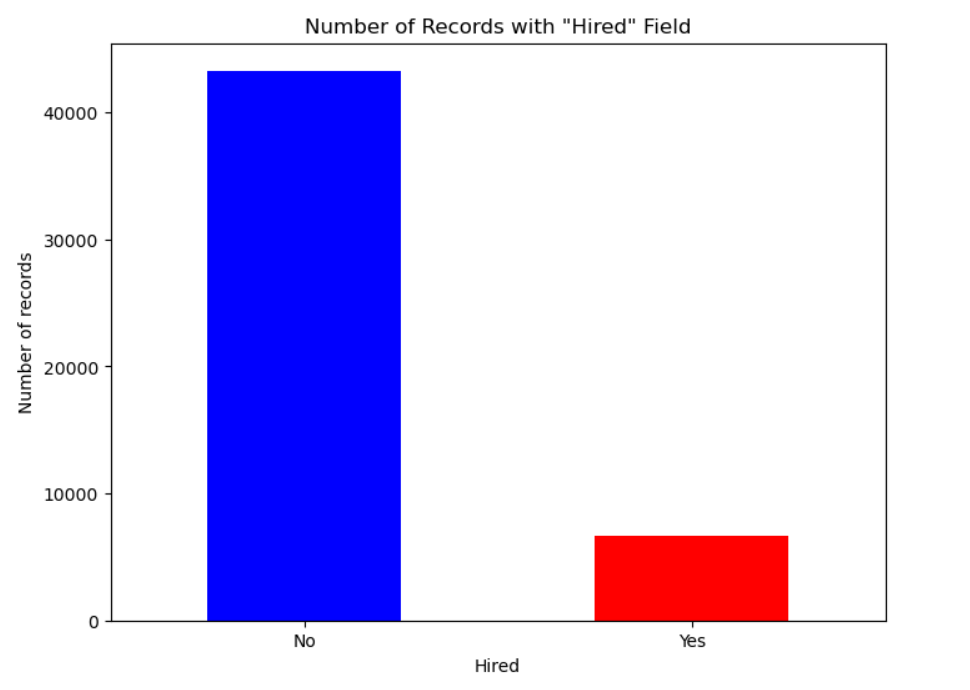




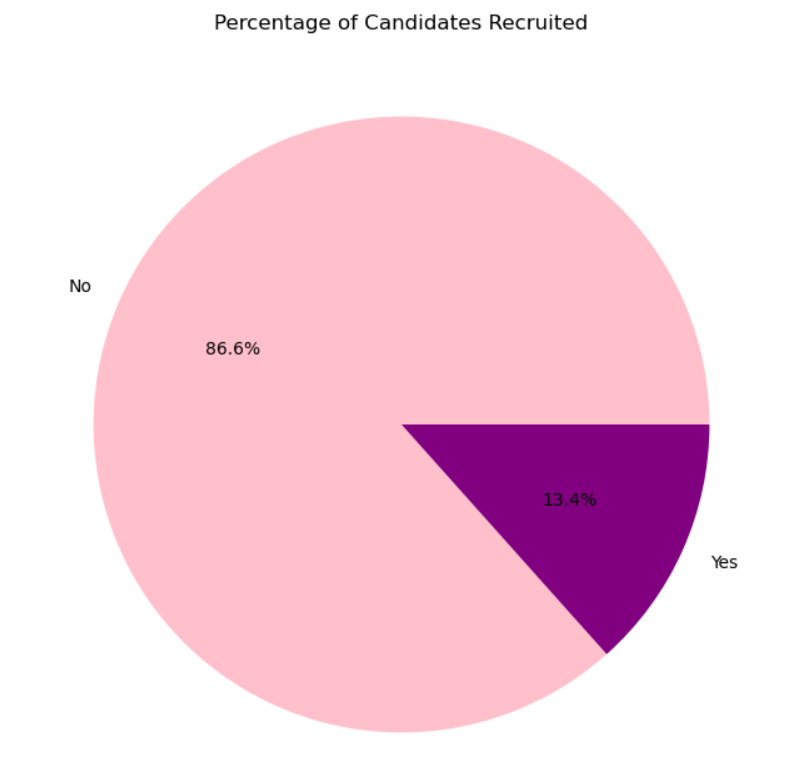


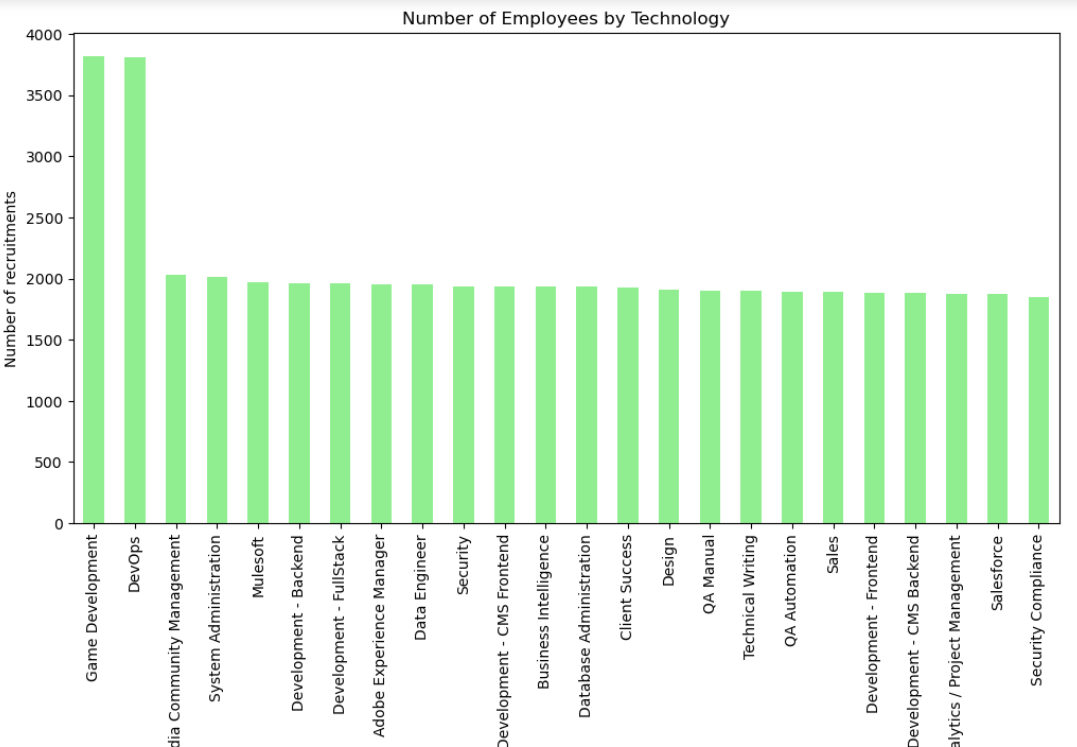


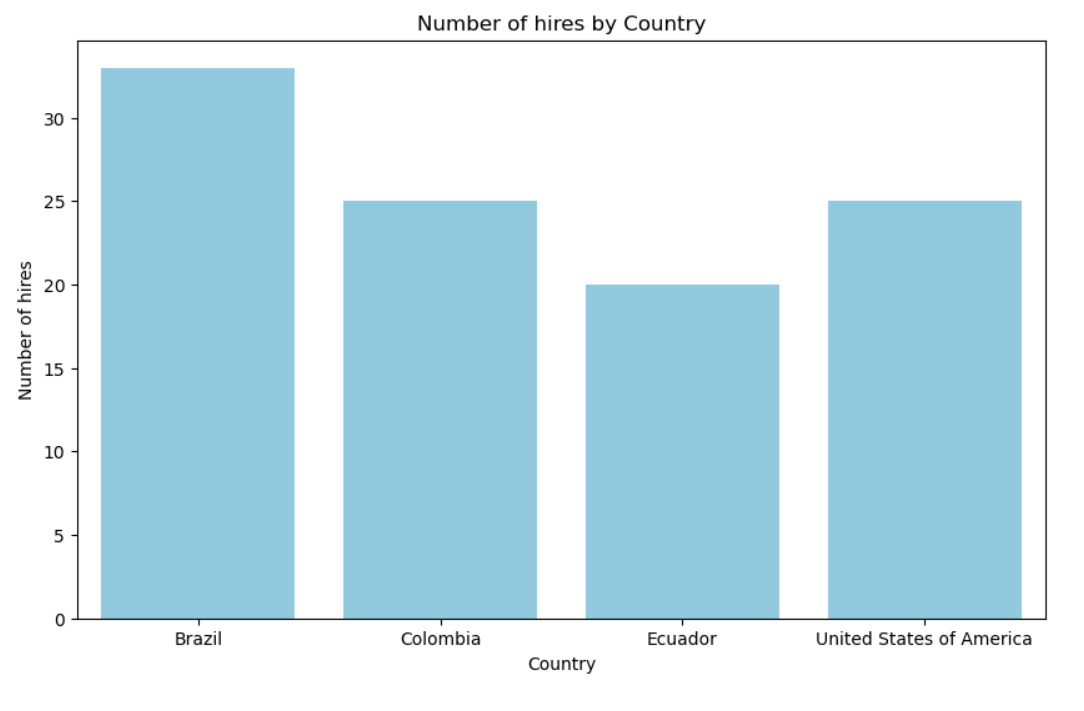
1. Jupyter notebook (Análisis individual fuera de los mostrados anteriormente)

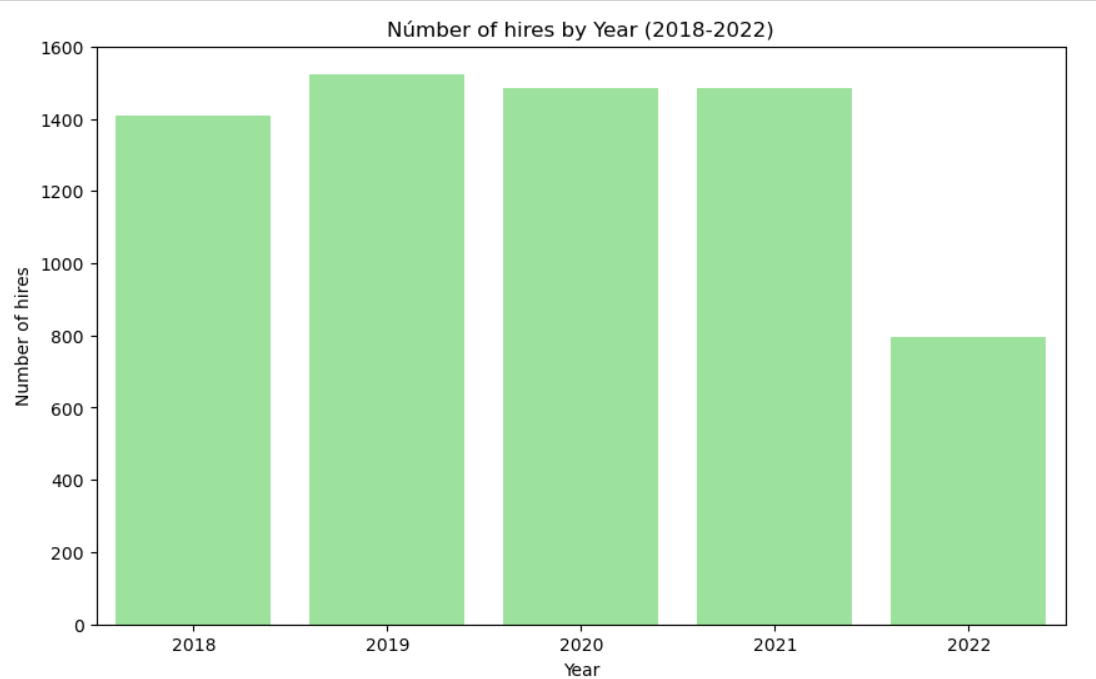


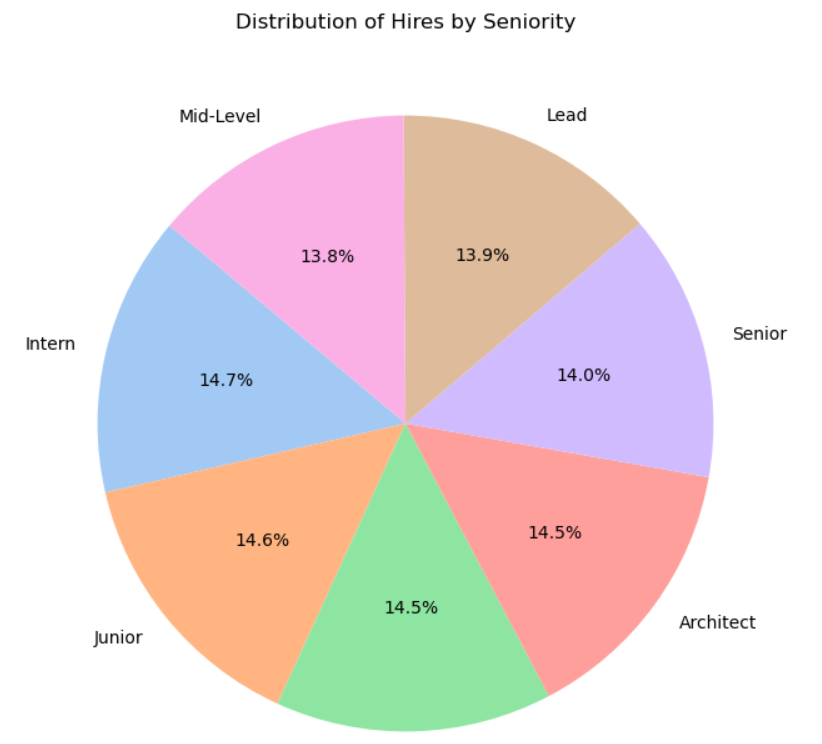












1. Conclusiones

* Hubo una tendencia de dos tecnologías entre las contrataciones y estas fueron la de Game development con 519 contratados (7,75%) y Dev Ops con 495 (7,39%), teniendo estas dos tecnologías una diferencia de alrededor de 200 registros contratados en comparación con la tercera tecnología con mayor contrataciones System Administration con 293 contratados (4,37%). (Los porcentajes hacen referencia a su cantidad porcentual en comparación con todas las tecnologías que tuvieron contrataciones).
* El país con más contrataciones a lo largo de los años fue brazil con un total de 33 contrataciones, seguido de colombia y estados unidos teniendo los dos un total de 25 contrataciones, dejando esto a ecuador como el país con menos contrataciones de estos 4 con un total de 20 contrataciones.
* El año con más contrataciones fue el 2019 con 1524 contrataciones, seguido por 2020 y 2021 con 1485 contrataciones y por último 2022 con 795 contrataciones teniendo una diferencia de más de 600 registros con respecto a los otros años ya que este año y sus registros en el dataset están incompletos, en el sentido de que todos los años cumplen con tener registros hasta un mes específico, mientras que los registros de 2022 son hasta un mes anterior a los del tope que tienen los de los otros años (Supongamos que cada año abarca en su totalidad registros de los 12 meses del año, el 2022 no cumple con esa cobertura de todos los meses en el dataset)
* A lo largo de varios años, se observó que hubo tres niveles de seniority que destacaron al ser más contratados en comparación con los demás. En el primer lugar se encuentra el seniority de arquitecto, liderando las contrataciones en los años 2018 y 2020. En segundo lugar, el nivel de Trainee se destacó liderando las contrataciones en 2019 y 2021. Por último, en tercer lugar, el nivel de Intern lideró las contrataciones en 2022.