

WORKSHOP 1



JUAN PABLO GRANADOS

JAVIER ALEJANDRO VERGARA
ZORRILLA

ETL

UNIVERSIDAD AUTÓNOMA DE
OCCIDENTE

FACULTAD DE INGENIERÍA

SANTIAGO DE CALI

Documentación del Proyecto: Análisis de Datos de Candidatos

Descripción del Proyecto

El proyecto tiene como objetivo analizar y visualizar datos de candidatos que han participado en procesos de selección. Los datos están en un archivo CSV y se migrarán a una base de datos relacional. Luego, se crearán visualizaciones para mostrar información relevante sobre los candidatos contratados, como su distribución por tecnología, año, seniority y país.

Proceso del Proyecto

1. Preparación del Entorno

Primero, es necesario configurar el entorno de trabajo. Esto incluye:

- Instalar las herramientas necesarias: Asegúrate de tener instaladas las bibliotecas de Python como pandas para manipulación de datos, sqlalchemy para operaciones con bases de datos, y matplotlib y seaborn para la visualización de datos.
- Configurar el entorno de desarrollo: Utiliza un entorno como Jupyter Notebook para documentar y ejecutar el código de manera interactiva.

2. Carga de Datos

El primer paso es cargar los datos desde el archivo CSV a un DataFrame. Esto incluye:

- Leer el archivo CSV: Utiliza una herramienta adecuada para importar los datos del archivo CSV a un DataFrame.
- Explorar los datos: Revisa las primeras filas del DataFrame para entender la estructura y el contenido de los datos.

EJEMPLO:

	first_name	last_name	email	application_date	country	yoe	seniority	technology	code_challenge_score	technical_interview_score
0	Bernadette	Langworth	leonard91@yahoo.com	2021-02-26	Norway	2	Intern	Data Engineer	3.0	3.0
1	Camryn	Reynolds	zelda56@hotmail.com	2021-09-09	Panama	10	Intern	Data Engineer	2.0	10.0
2	Larue	Spinka	okey_schultz41@gmail.com	2020-04-14	Belarus	4	Mid-Level	Client Success	10.0	9.0
3	Arch	Spinka	elvera_kulas@yahoo.com	2020-10-01	Eritrea	25	Trainee	QA Manual	7.0	1.0
4	Larue	Altenwerth	minnie.gislason@gmail.com	2020-05-20	Myanmar	13	Mid-Level	Social Media Community Management	9.0	7.0

3. Limpieza y Preparación de los Datos

Una vez que los datos están cargados, el siguiente paso es limpiarlos y prepararlos para su migración a la base de datos. Este proceso incluye:

- Verificación de valores faltantes: Identifica y maneja los valores nulos o faltantes.
- Conversión de tipos de datos: Asegúrate de que los datos están en el formato correcto, como fechas y números enteros.
- Determinación de candidatos contratados: Aplica la lógica para identificar a los candidatos que fueron contratados, basándote en las puntuaciones de las pruebas.

4. Migración de Datos a la Base de Datos

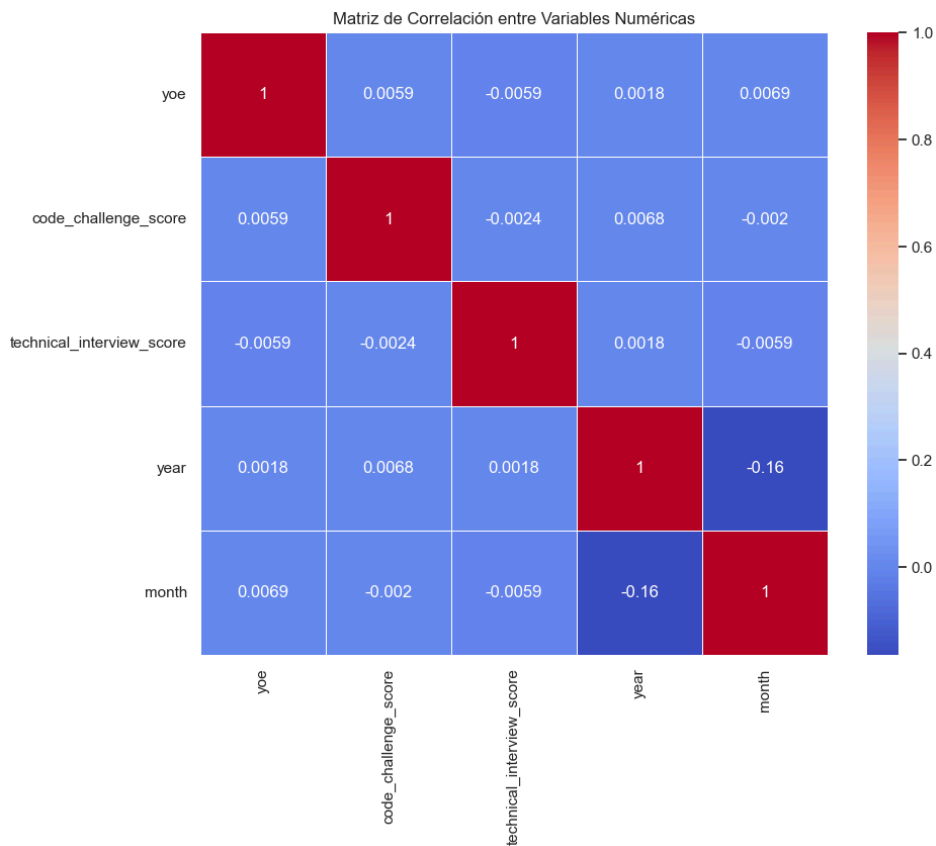
Con los datos limpios, el siguiente paso es migrarlos a una base de datos relacional:

- Configurar la base de datos: Configura un sistema de gestión de base de datos relacional (PostgreSQL).
- Crear la base de datos y las tablas: Diseña la estructura de la base de datos y crea las tablas necesarias para almacenar los datos.
- Insertar los datos en la base de datos: Carga el DataFrame en las tablas de la base de datos.

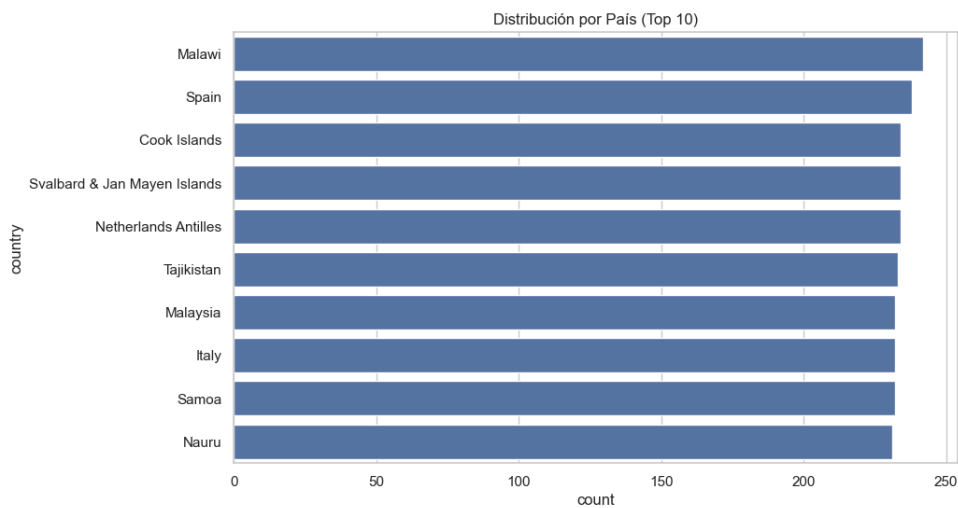
5. Análisis Exploratorio de Datos (EDA)

Realiza un análisis exploratorio de los datos para obtener insights iniciales. Esto incluye:

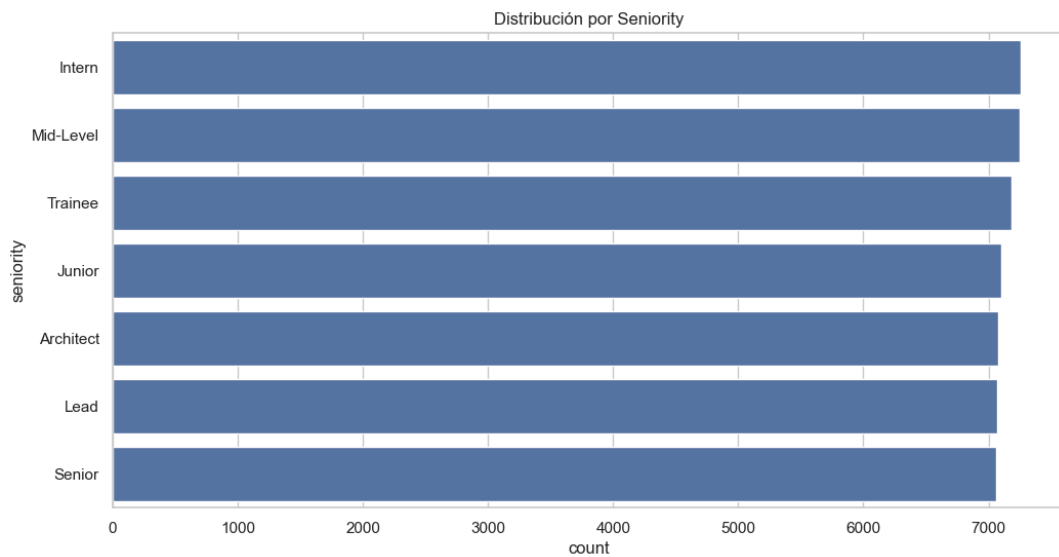
- **Correlación entre variables numéricas:** Examina qué variables numéricas tienen relación entre sí



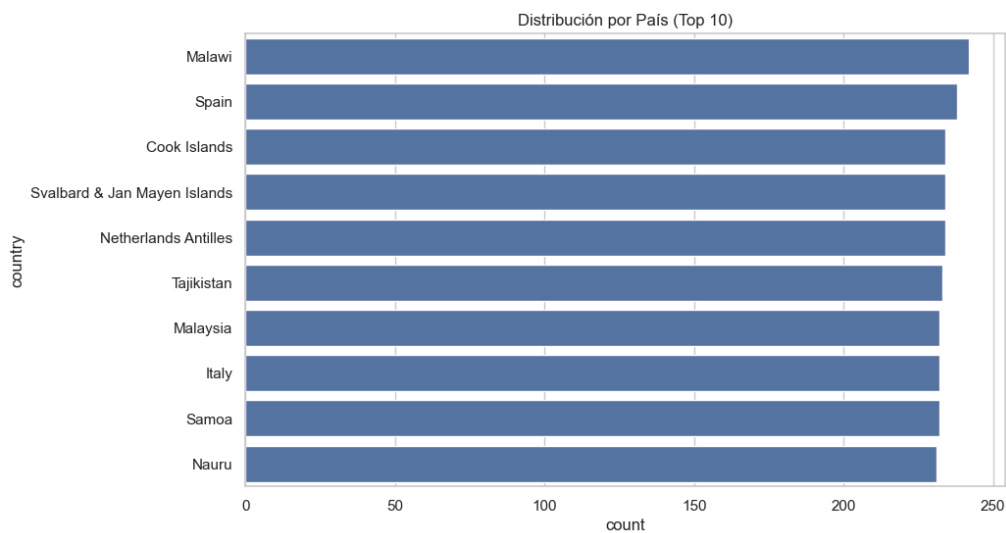
- **Aplicaciones por año:** Analiza cómo varía el número de aplicaciones a lo largo de los años.



- **Contrataciones por nivel de seniority:** Investiga la distribución de contrataciones según el nivel de seniority.



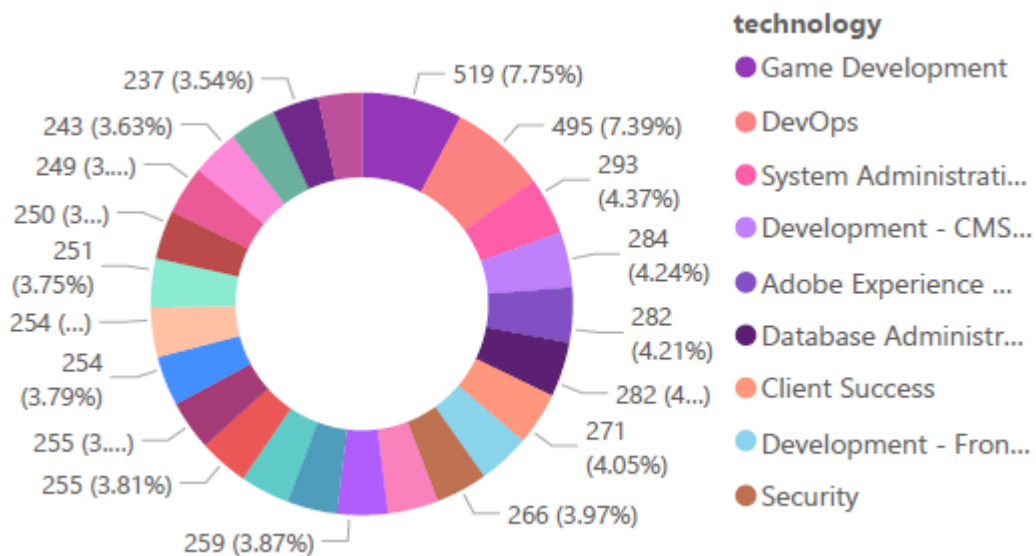
- **Contrataciones por país:** Evalúa cómo se distribuyen las contrataciones por país,



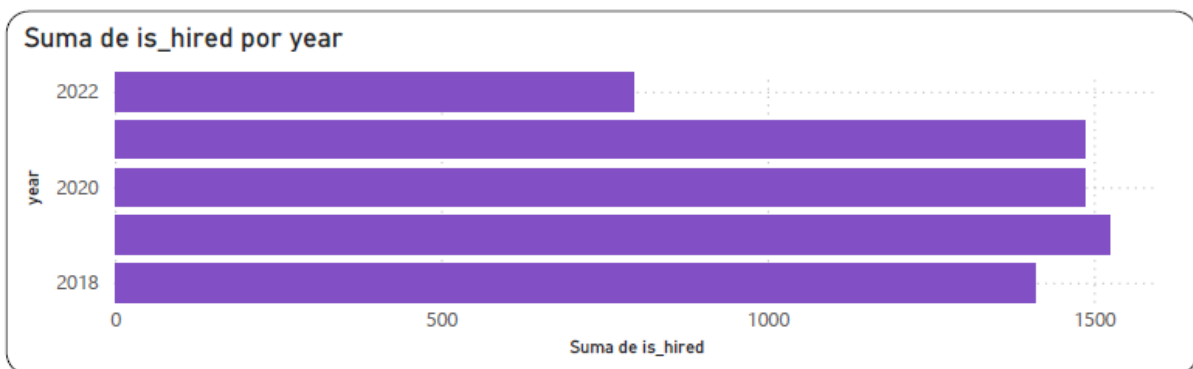
6. Creación de Visualizaciones

Con los datos consultados, el siguiente paso es crear visualizaciones para representar la información de manera gráfica:

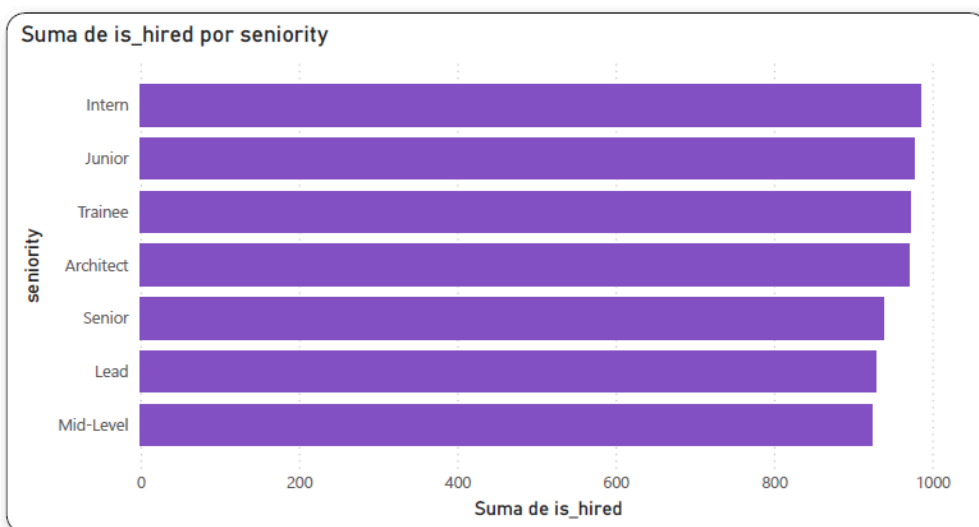
- **Gráfico de pastel (Pie chart):** Muestra la distribución de contrataciones por tecnología.



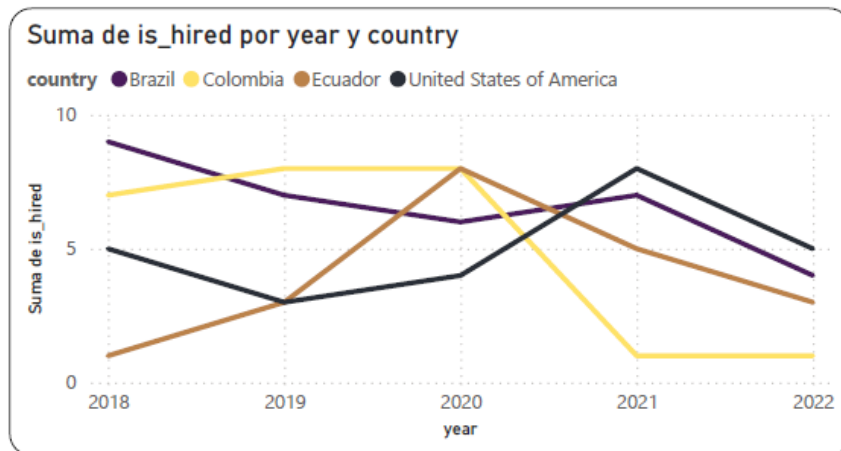
- Gráfico de barras horizontal: Visualiza el número de contrataciones por año.



- Gráfico de barras: Representa el número de contrataciones por nivel de seniority.



- Gráfico de líneas múltiples (Multiline chart): Muestra la evolución de contrataciones por país a lo largo de los años.



7. Documentación y Presentación

Finalmente, documenta el proceso en Jupyter Notebook:

- Explica cada paso: Incluye descripciones claras de cada etapa del proceso, desde la carga de datos hasta la visualización.
- Resultados y análisis: Presenta los resultados de las consultas y las visualizaciones, proporcionando un análisis interpretativo de los datos.

Conclusión

Este proyecto proporciona una visión integral de cómo manejar datos de candidatos desde su carga inicial hasta la visualización final. Asegúrate de seguir cada uno de los pasos descritos para garantizar que el análisis sea preciso y completo. La documentación clara y detallada ayudará a comunicar eficazmente los resultados y las conclusiones del análisis.