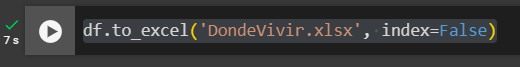
**Primero: Descarga el archivo DondeVivir.csv**

**https://github.com/tuxitobb/examenfeliz-md**

**EVALUACION:**

**2 Puntos:**

* Utiliza Google Colab para agregarle una columna ID al archivo. 1) Agrega <https://colab.research.google.com/drive/1w1SPOgIc4QIIvwIKcfE7UFmEH-vMaGd4?usp=sharing> el enlace público al archivo colab que utilizaste. Exporta tu archivo y ponle DondeVivir-TUNOMBRE.xls. 2) Añade el código utilizado y las imágenes de ejecución:
* df.to\_excel('DondeVivir.xlsx', index=False)
* 

**3 Puntos:**

* Utiliza el mismo archivo de Google Colab e identifica **2 de los 8 problemas de calidad de datos que vimos en clase.** Identificalos en el conjunto de datos **DondeVivir**. Añade **AQUÍ**

1) el código utilizado

def verificar\_valores\_nulos(data\_frame):

    columnas\_con\_nulos = data\_frame.columns[data\_frame.isnull().any()].tolist()

    if not columnas\_con\_nulos:

        print("No hay valores nulos en el conjunto de datos.")

    else:

        print("Columnas con valores nulos:")

        for columna in columnas\_con\_nulos:

            print(f"- {columna}")

matriz\_correlacion = df.corr()

print(matriz\_correlacion)

plt.figure(figsize=(10, 8))

sns.heatmap(matriz\_correlacion, annot=True, cmap='coolwarm', fmt=".2f", linewidths=.5)

plt.title('Matriz de Correlación')a

plt.show()

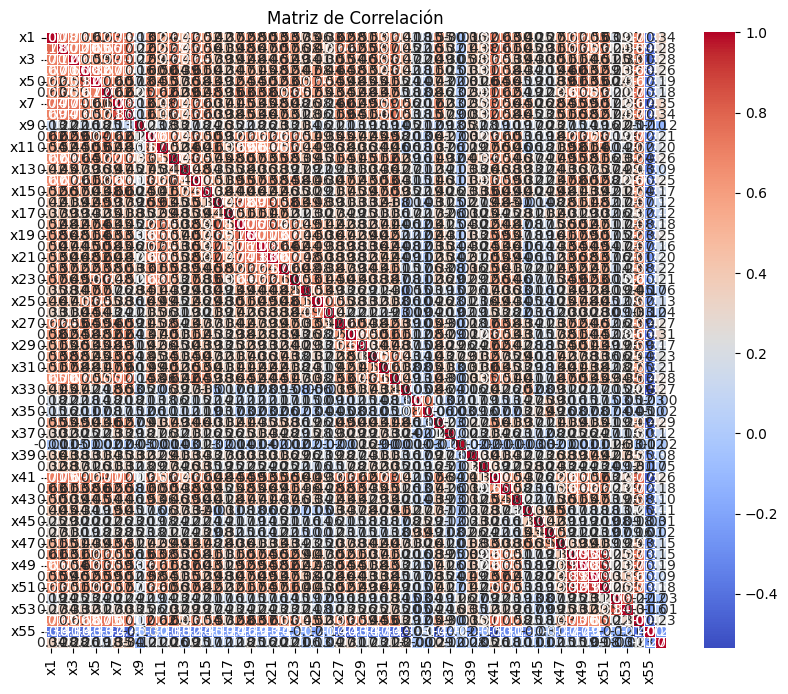
2) el texto en google colab que justifica por qué es un problema de calidad y 3) las imágenes de la ejecución, a este documento.

**1. Completitud:** Es la relación entre los campos que se tienen y los que se ocupan por el negocio. Es decir, si tenemos 100 campos peros solo 80 se requieren para el hacer cuncionar el negocio, la evaluación será dividir el(número de campos completos que se requieren para el negocio) / 80.

**2. Consistencia:** Es la relación entre los datos que deberían estar correctos en términos de formato, tiempo o lugar y los que lo están. En particular se puede hablar de:\*

**Consistencia en los valores:** Los valores son los mismos a través de los sistemas de información. Es decir, el total de depósitos en el CRM es igual al totoal de depósitos en el ERP.\*

**Consistencia en la trazabilidad:** Los valores son consistentes de acuerdo a las reglas de negocio. Por ejemplo, si el estatus de un trabajador es 'baja' el acceso debe ser 'denegado'.



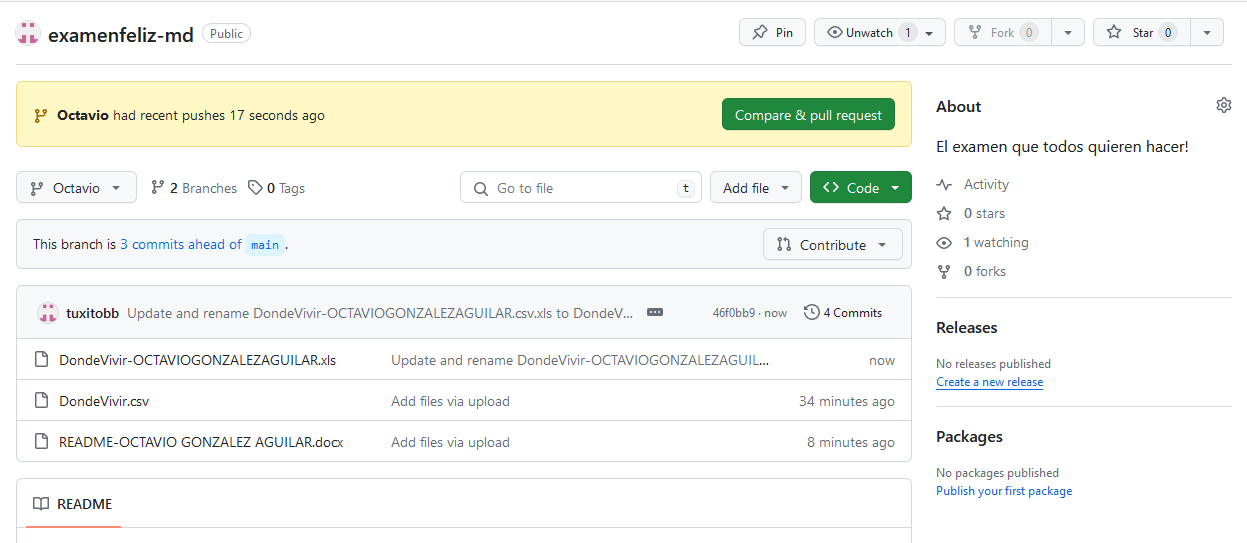
}

**3 Puntos:**

* Utiliza el mismo archivo de Google Colab y adapta el código visto en clase de **aprendizaje supervisado** o **aprendizaje no supervisado** al conjunto de datos **DondeVivir**. Añade el código utilizado y las imágenes de la ejecución **AQUÍ**.

**2 Puntos:**

* Sube un archivo a una rama con tu nombre y un README.docx con tu nombre. El documento no tiene que tener contenido. Solo tu nombre. También sube tu archivo .xls que realizaste en el 1 ejercicio. Si tienes éxito, debes obtener algo así. Un punto por crear la rama y 1 punto por subir los archivos:



**Calificación entre equipos**

De los 4 equipos en los que no participaste, indica como primero al mejor, segundo al segundo mejor, tercero al menos peor y cuarto al peor.

**Participación**

Explica en 10 líneas, ¿cómo ayudaste a tus compañeros en el proyecto? Redacta en cuántas reuniones participaste, qué parte del proyecto hiciste, qué problemas resolviste, qué ideas son tuyas, cómo tomaban las decisiones, a través de qué medio se juntaban, quién tomaba el liderazgo, cómo eligieron el tema, quién hizo la parte técnica, quién hizo la presentación, quién quería exponer. **Recuerda que los relatos deben coincidir.** Si no participaste, sé un caballero o una dama y simplemente dilo, siempre es mejor la honestidad.