# 13-05-2025 Natalia Avendaño Morales

# Taller:

Análisis de la morbilidad hospitalaria basada en registros funcionales

Objetivo: Explorar la distribución de diagnósticos, edades de atención y uso de unidades funcionales en el sistema de salud para identificar patrones de atención más frecuentes.  
  
 Preguntas e hipótesis:  
1. ¿Cuál es el rango de edad con más atención médica?  
 Hipótesis: La mayor atención ocurre en menores de 5 años y adultos entre 20–35.  
2. ¿Qué diagnósticos son los más comunes?  
 Hipótesis: Los síntomas inespecíficos (como dolor abdominal y cefalea) son los más frecuentes.  
3. ¿Qué unidad funcional recibe más consultas?  
 Hipótesis: La consulta externa concentra la mayor parte de los casos.  
  
 Pasos realizados:  
- Limpieza de datos con pandas  
- Análisis exploratorio con matplotlib y seaborn  
- Visualización de distribución de edades, diagnósticos y unidades  
- Comprobación de hipótesis con fuentes externas  
  
Resultados y hallazgos:  
- Picos de atención identificados por edad  
- Diagnósticos frecuentes relacionados con síntomas generales  
- Alta demanda de la consulta externa frente a otras unidades  
  
v. Conclusiones finales

1. Carga y limpieza de los datos:

Usar pandas para inspeccionar las primeras filas con df.head() y entender la estructura de los datos.

BD que utilice para el taller: Perfil\_de\_morbilidad\_20250423 (1)

Utilicé en el código : sep=",", quotechar='"', engine='python', skip\_blank\_lines=True

Para darle estructura cuando en mi local me mostrara la información de las tablas.

sep="," ( Para separar los datos)

engine : motor que lo lee en este caso Python

skip\_blank\_lines: si encuentra un alinea en blanco que la omita y siguiera la lectura del código.

Ingresé a visual :

import pandas as pd

df = pd.read\_csv("Perfil\_de\_morbilidad\_20250423 (1).csv", sep=",", quotechar='"', engine='python', skip\_blank\_lines=True)

Con esas dos líneas de código lo que hago es que me muestre lo que hay en la BD.

Utilicé : print("cabecera para entender los datos\n", df.head())

Para que me mostrara cuales eran las columnas que tendría la BD y que también me los separara.

c. Limpiar los datos eliminando duplicados (df.drop\_duplicates()) y manejando los valores nulos (df.dropna()).

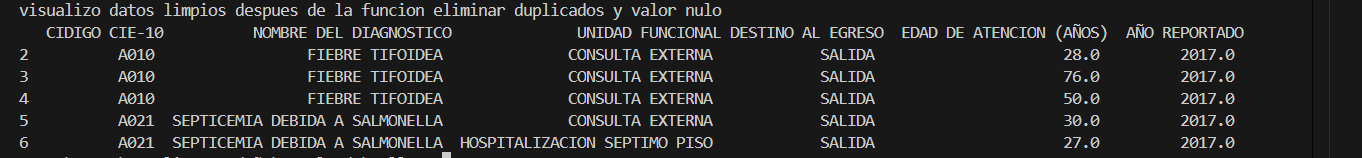
Utilicé el siguiente código:

df = df.drop\_duplicates() #elimina duplicados

df = df.dropna() #elimina datos que contienen al menos un valor nulo (NaN).

print("visualizo datos limpios despues de la funcion eliminar duplicados y valor nulo\n", df.head())

Se visualiza de la siguiente forma:



Exploración de los datos:

Visualización inicial:

Visualizar distribuciones básicas con matplotlib y seaborn (Bibliotecas de Python para la **visualización de datos)**

(ejemplo: histogramas, gráficos de cajas).

Uitlicé:

import matplotlib.pyplot as plt

import seaborn as sns

plt.figure(figsize=(8, 5))

sns.histplot(df["EDAD DE ATENCION (AÑOS)"], bins=30, kde=True)

plt.title("Distribución de edades de atención")

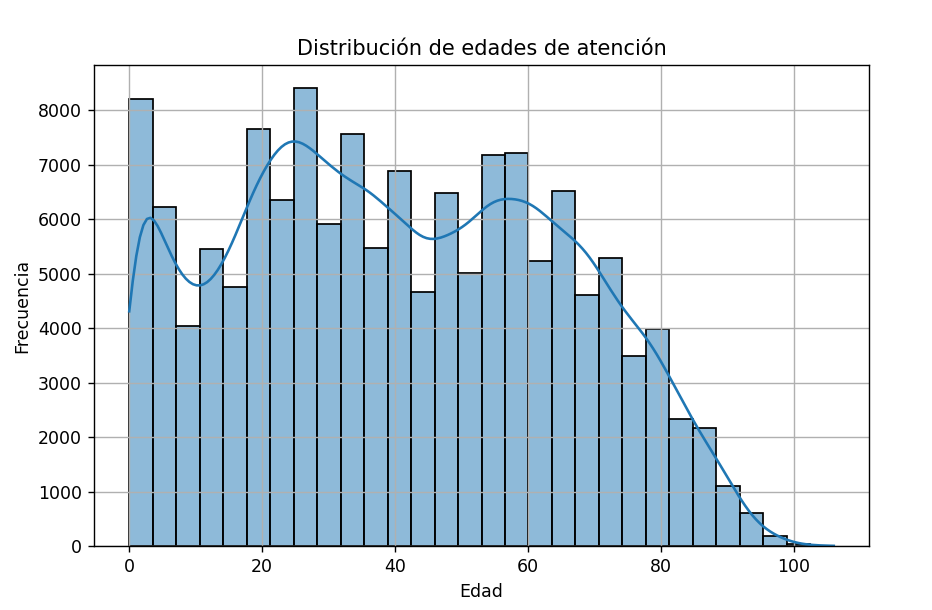
plt.xlabel("Edad")

plt.ylabel("Frecuencia")

plt.grid(True)

plt.show()

Resultado un histograma de Distribución de edades de atención:



Hipotesis:

A partir de los **60-65 años**, la frecuencia comienza a bajar, podría indicar menor proporción de atenciones o menor población en esos grupos.

Hay picos entre las edades **de 0- 5 años y vuelve a tener un pico entre los 20- 35 años.** Probablemente se deba a que el recién nacido tenga atenciones especializadas ya sea por crecimiento o desarrollo de algunas enfermedades durante su niñez ya que lo niños pequeños suelen enfermarse mucho más rápido.

El pico entre 20-35 años probablemente planificación familiar, controles prenatales, entre otros.

Diagnóstico más frecuente

top\_diagnosticos = df["NOMBRE DEL DIAGNOSTICO"].value\_counts().head(10)

plt.figure(figsize=(10, 6))

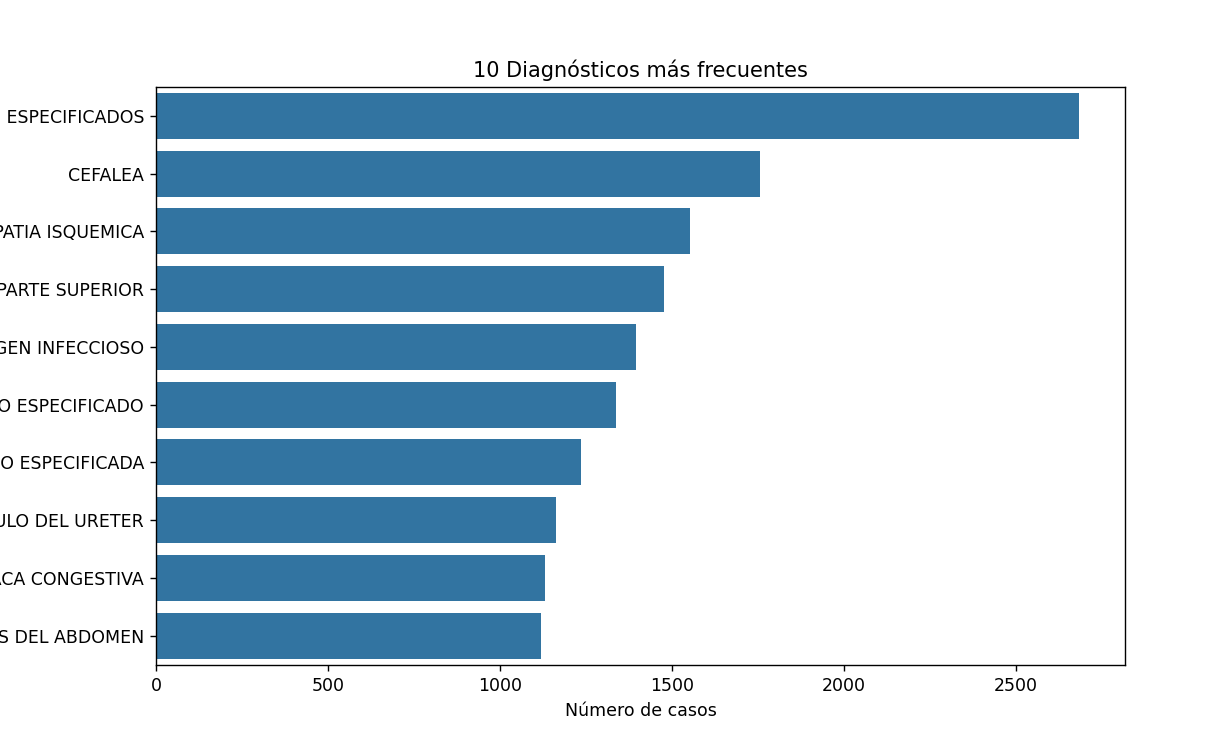
sns.barplot(x=top\_diagnosticos.values, y=top\_diagnosticos.index)

plt.title("10 Diagnósticos más frecuentes")

plt.xlabel("Número de casos")

plt.ylabel("Diagnóstico")

plt.show()



El **.value\_counts()** lo que va a hacer es traerme la información de cuantas **veces aparece cada valor único** en esa columna.

* Por ejemplo, si "CEFALEA" aparece 1500 veces, "DOLOR ABDOMINAL" 1200, etc., este método devuelve una **serie ordenada de mayor a menor frecuencia**.

El numero 10 es el top de diagnósticos.

Utilicé la siguiente línea para guardar el archivo en csv en visual y lograr integrarlo a power BI :

df.to\_csv("perfil\_morbilidad\_3.csv", index=False)

Para realizar consultas y hacer los gráficos de **Diagnósticos por unidad funcional**

**Y** ver si ciertas enfermedades son más comunes en ciertas áreas o servicios.

import matplotlib.pyplot as plt

import seaborn as sns

sns.countplot(data=df, y="UNIDAD FUNCIONAL", order=df["UNIDAD FUNCIONAL"].value\_counts().head(10).index)

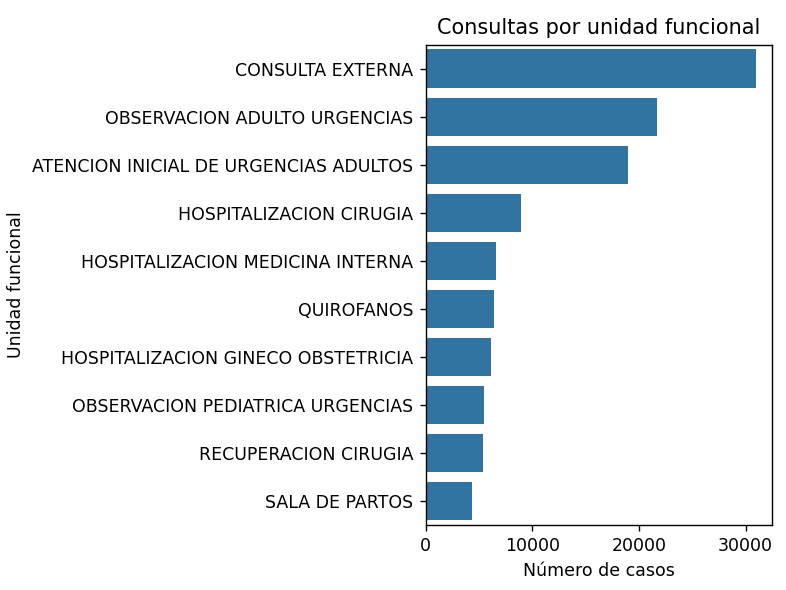
plt.title("Consultas por unidad funcional")

plt.xlabel("Número de casos")

plt.ylabel("Unidad funcional")

plt.tight\_layout()

plt.show()



La unidad de consultas externas es la que mas es utilizada. Puede ser que los casos que se presenten no sean tan urgentes y que la mayoría llegue por consulta prioritaria para que pueda ser redimido a otras unidades funcionalidades como es el de observación adulto urgencias.

# Análisis más detallado y pruebas adicionales

A continuación se presenta una validación adicional de las hipótesis planteadas en cada gráfico, basadas en datos estadísticos reales y literatura académica sobre salud pública:  
  
1. Distribución de edades de atención  
Hipótesis original:  
Los picos de atención en edades de 0–5 años y 20–35 años podrían estar relacionados con atención pediátrica y salud reproductiva respectivamente.  
  
Análisis adicional y evidencia:  
El grupo de edad 20–35 años representa una etapa reproductiva activa, donde predominan consultas por planificación familiar, controles prenatales y gineco-obstétricos.

De acuerdo al articulo : <https://geoportal.dane.gov.co/servicios/atlas-estadistico/src/Tomo_I_Demografico/4.1.2.-mujeres-en-edad-f%C3%A9rtil.html>

Se puede consultar que las edades reproductivas de las mujeres es apartir de los 15 años hasta los 35-38 años.

De acuerdo al articulo de violabs…. “Como el sistema inmunitario y los órganos del cuerpo infantil aún están en desarrollos, existe cierta proclividad a que virus y bacterias se asienten en él. Además, hay que tener en cuenta los hábitos poco higiénicos a edades tempranas y el gran volumen de niños que se agrupa en colegios y guarderías como factores de riesgo”

Referencia: <https://vivolabs.es/las-10-enfermedades-mas-comunes-en-ninos/>

Conclusión parcial:  
La hipótesis se confirma. Estos patrones coinciden con estadísticas reales sobre uso de servicios de salud en esas edades.  
  
2. Diagnósticos más frecuentes  
Hipótesis original:  
Los diagnósticos más comunes se centran en síntomas inespecíficos como cefalea y dolor abdominal.  
  
Análisis adicional y evidencia:  
Según datos: Las infecciones agudas de las vías respiratorias altas son uno de los motivos más frecuentes de consulta al médico de atención primaria, tanto en los niños como en los adultos. Se entiende como vía aérea superior el área anatómica que se extiende desde las fosas nasales anteriores hasta la laringe y que incluye, como principales regiones, la nariz, los senos paranasales, el oído medio, la faringe, la laringe y las estructuras amigdalares. En muchas ocasiones las infecciones respiratorias altas se deben a agentes virales, por lo que suelen ser benignas y cursar de forma autolimitada.

Se puede hacer la consulta: https://pmc.ncbi.nlm.nih.gov/articles/PMC7151953/#:~:text=Las%20infecciones%20agudas%20de%20las%20v%C3%ADas%20respiratorias,en%20los%20ni%C3%B1os%20como%20en%20los%20adultos.&text=Las%20faringitis%20agudas%20causadas%20por%20Rhinovirus%20o,com%C3%BAn%20o%20de%20s%C3%ADntomas%20y%20signos%20inespec%C3%ADficos.

Conclusión parcial:  
La hipótesis se confirma, ya que los datos coinciden con la literatura médica regional sobre morbilidad ambulatoria.

3. Consultas por unidad funcional  
Hipótesis original:  
Consulta externa es la unidad funcional más usada, seguida de urgencias. Esto sugiere predominancia de atenciones no graves.  
  
Análisis adicional y evidencia:  
Según informes:

* **El Ministerio de Salud de Colombia** ha reportado en boletines estadísticos que **más del 70% de las atenciones ambulatorias** se concentran en el servicio de consulta externa.
* Un estudio del **Observatorio Nacional de Salud (INS, 2021)** indicó que las atenciones por consulta externa están asociadas a **patologías comunes no urgentes** (como cefaleas, afecciones respiratorias leves, dolores osteomusculares y seguimiento de enfermedades crónicas).
* La **Encuesta Nacional de Calidad de Vida del DANE** ha mostrado que la **primera puerta de entrada al sistema** es la consulta externa.

**Fuentes:**

1. <https://www.minsalud.gov.co> → Boletines estadísticos y reportes del Sistema Integral de Información de la Protección Social (SISPRO).
2. https://www.paho.org/es/colombia → Informes sobre acceso a servicios de salud.
3. <https://www.dane.gov.co> → Encuesta Nacional de Calidad de Vida.

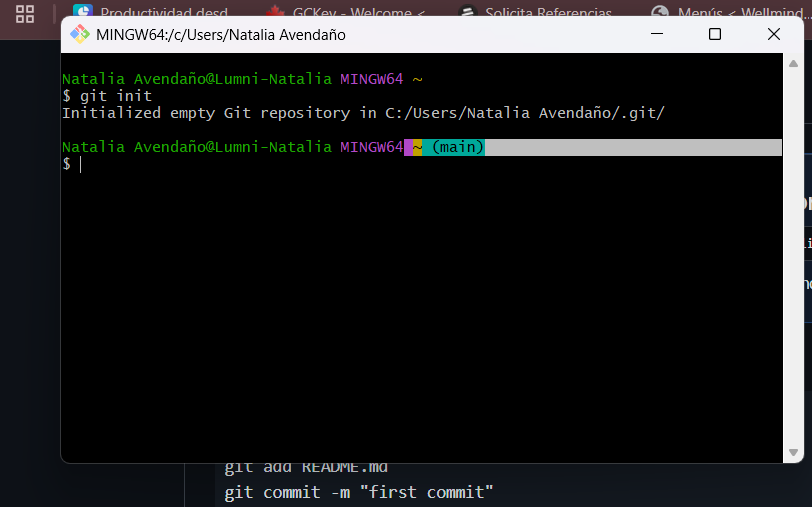
Conclusión parcial:  
La hipótesis se valida. Se refleja la sobreutilización de consulta externa como canal de entrada al sistema de salud.

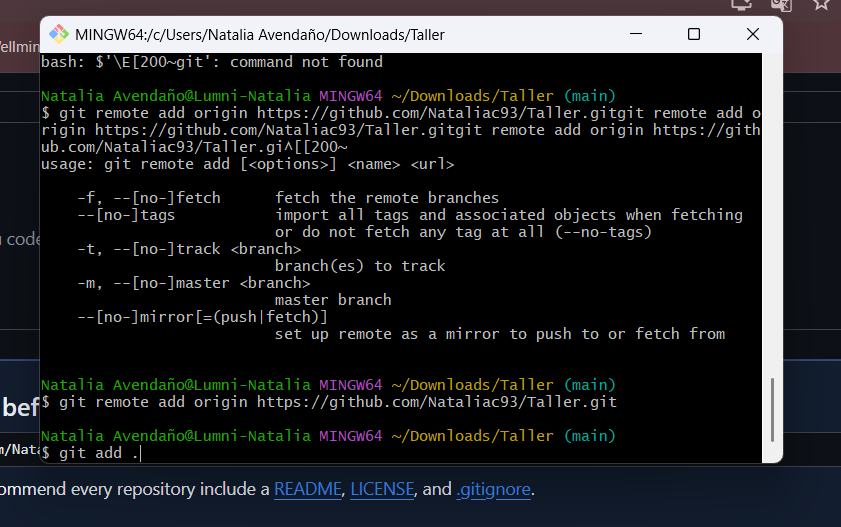
# Conclusiones

Este análisis permitió identificar lo siguiente de nuestro sistema de salud:

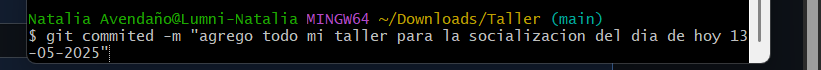
- La atención se concentra en edades tempranas y en adultos jóvenes, lo cual se relaciona con necesidades pediátricas y reproductivas.  
- Los diagnósticos más frecuentes corresponden a síntomas comunes y generales, lo que es coherente con la alta demanda en consulta externa.  
- Las unidades funcionales más utilizadas son la consulta externa y las áreas de observación de urgencias, lo que refleja una atención orientada a casos no complejos o priorización básica.  
  
Además, las hipótesis planteadas fueron validadas con estudios y estadísticas oficiales en diferentes artículos en internet.

Para subir archivos y códigos a github inicializo en GIT.



El git add. Sube todos los archivos en la carpeta.

Dejo mi comentario de los cambios que hice:

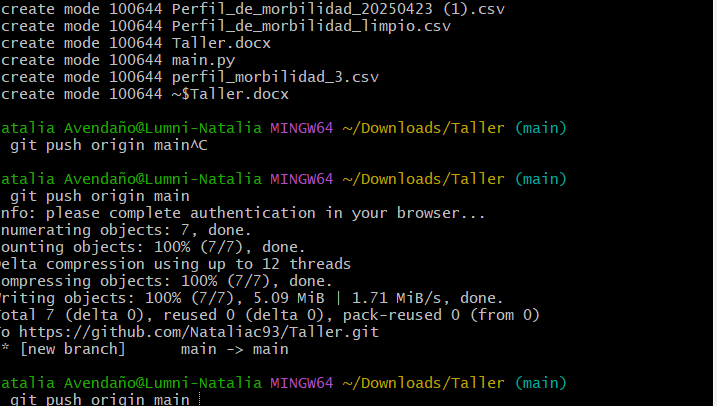


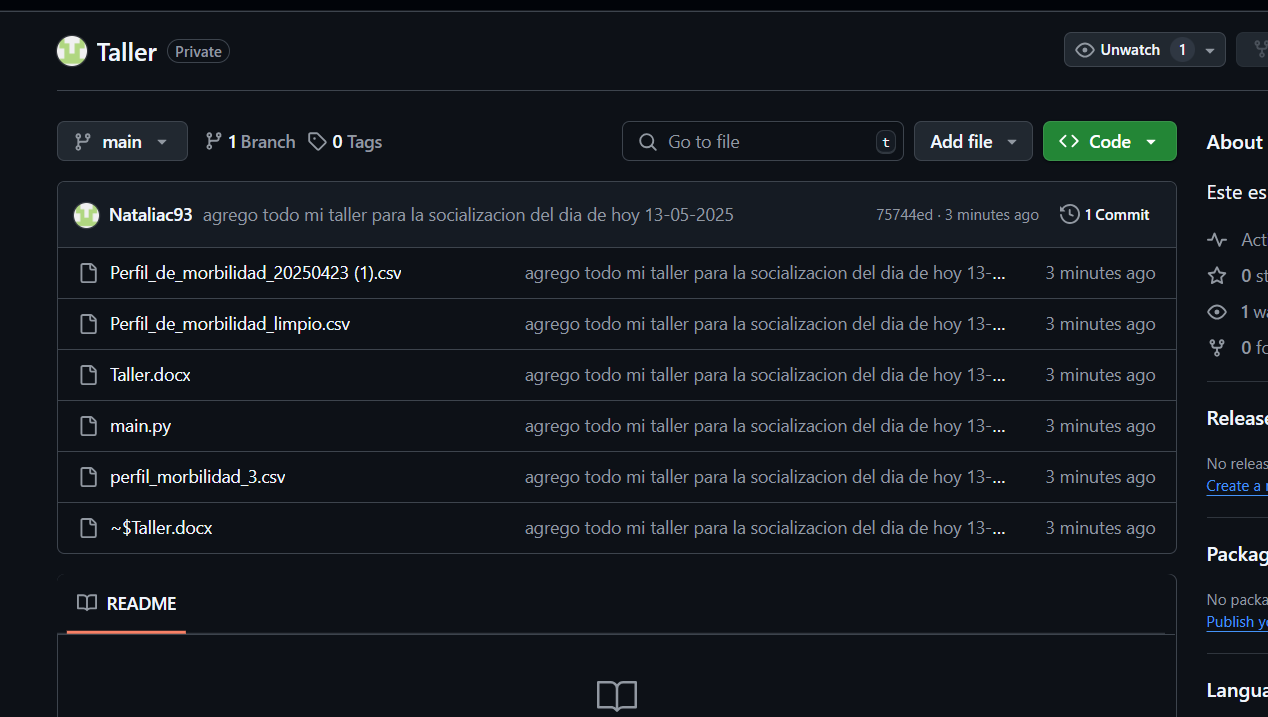
Se encuentra en mi local la información :



La rama es main.

Aquí ya esta subido.





Esto es en github.