



Proyecto Final

DS4A

DS
4A

Análisis para lograr determinar las variables influyentes en la mortalidad de los pacientes hospitalizados por COVID -19



Con las siguientes tablas y gráficos se quiere dar a conocer un reporte general de las personas ingresadas por COVID-19 a las diferentes unidades de servicio médico, a su vez, dar a conocer su tasa de mortalidad de cada género.

Sebastian Bustamante Tobón
David Santiago Figueroa Forero
Elizabeth Ramírez Patiño
Samuel Arturo Torres Orozco

Resumen

Con el presente proyecto se realiza un análisis exploratorio de cada una de las variables presentes, logrando una familiarización de los datos para a partir de acá, encontrar correlaciones entre estas para dar respuesta a diferentes preguntas que surgen al momento que un paciente entra a un hospital

Objetivo General

Determinar la condición de un paciente a la hora de ingresar a alguna área del hospital.

Objetivos Particulares

1. Seleccionar variables de interés y características claves.
2. Explorar columnas de manera individual.
3. Realizar gráficas a partir de más de una columna.

Base de Datos

https://1drv.ms/x/s!Ak50SpPZ1oSUgRf0Gpm6AXOA_a7J?e=fyvaCu

Esta base de datos está conformada por ocho variables cuantitativas:

Identificación, Frecuencia cardiaca, Presion arterial, Saturacion oxigeno, Temperatura, Edad, Sexo, Moratalidad.

Y una variable tipo cadena denominada Servicio.

Reporte COVID

Identificación

Todas

Identificación	Edad	Mortalidad	Género
53	78	Vivo	Femenino
607	78	Muerto	Femenino
726	33	Vivo	Masculino
745	39	Vivo	Femenino
1075	83	Muerto	Femenino
1396	83	Muerto	Masculino
1423	82	Muerto	Masculino
1757	86	Vivo	Femenino
2108	26	Muerto	Masculino

Promedio frecuencia cardiaca

126.68

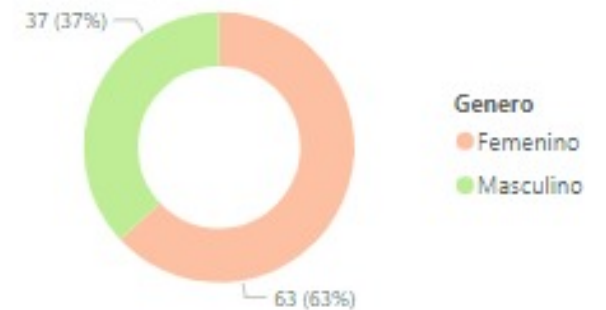
Promedio presión arterial

217.31

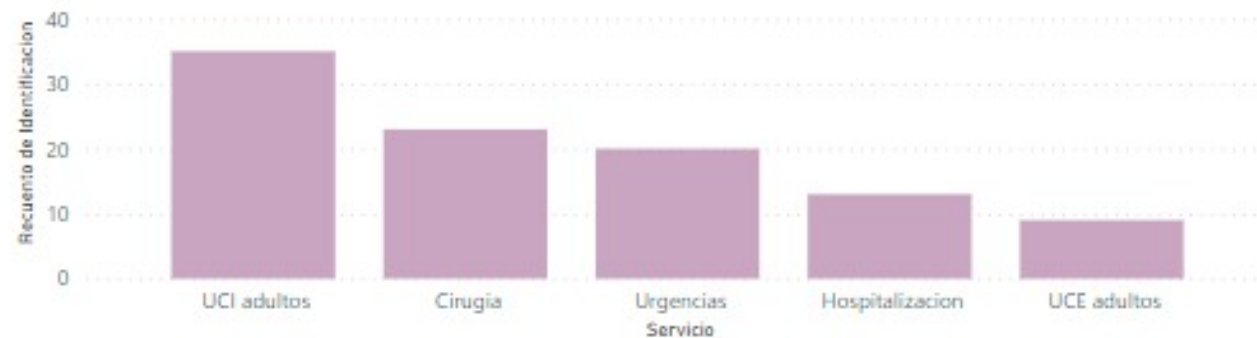
Promedio saturación oxígeno

106.27

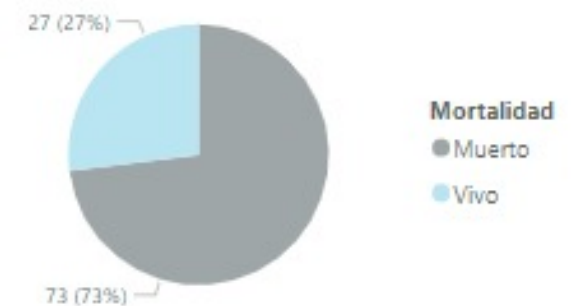
Número de personas por género



Número de personas por servicio



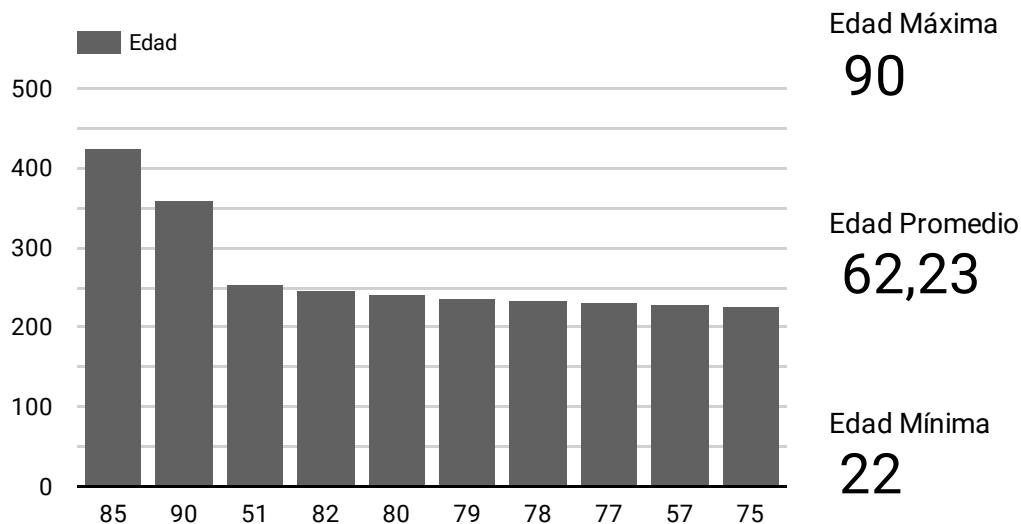
Tasa de mortalidad



Tomado de:

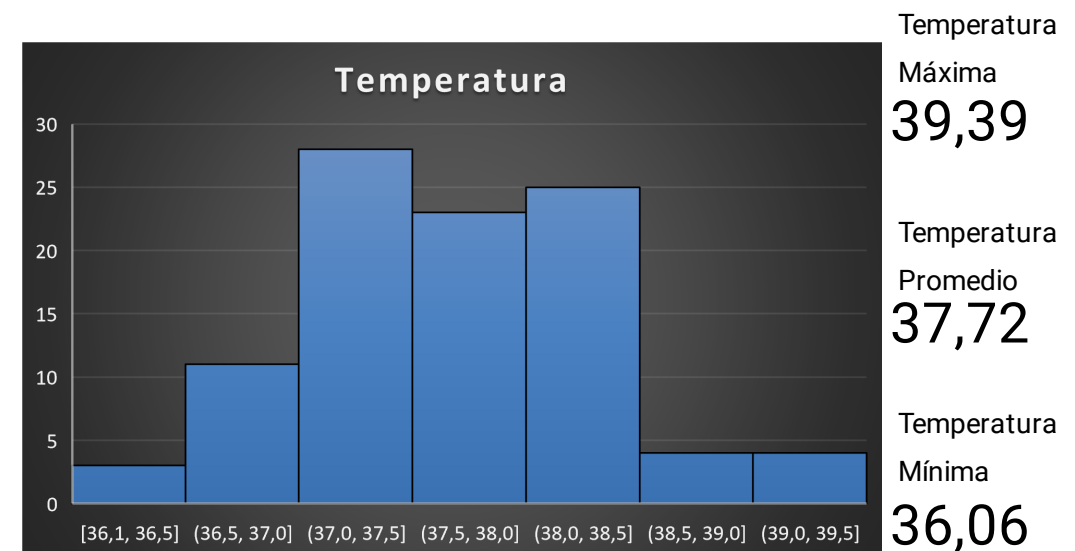
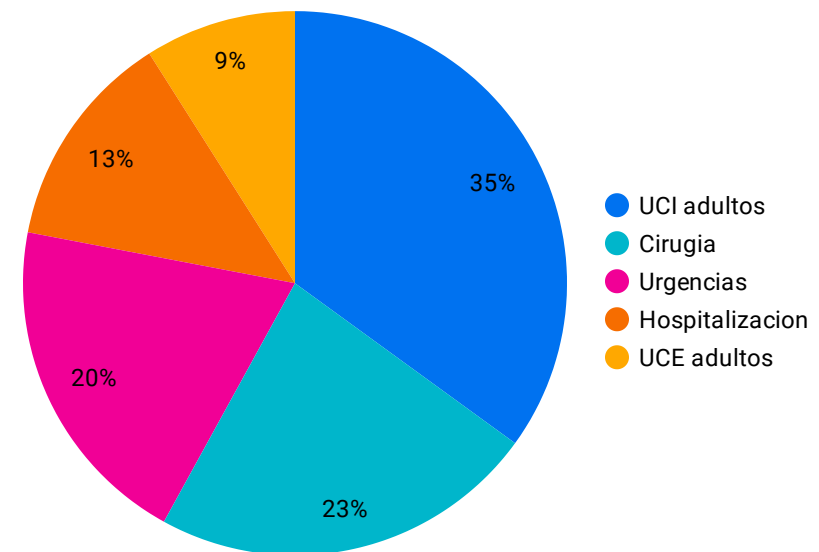
[https://app.powerbi.com/view?
r=eyJrljoiYzdkYTU2OTUtMWRlNS00MWE5LWJhZWQtNGY4YzJjN2E3ZWNIiwidCI6IjFkMmZmNTczLWUzYmUtNGViOS05OTRkLTZhODRjMWE5ZGU2ZSJ9&pageName=Re
portSectionb43cfee63ac9a1136009](https://app.powerbi.com/view?r=eyJrljoiYzdkYTU2OTUtMWRlNS00MWE5LWJhZWQtNGY4YzJjN2E3ZWNIiwidCI6IjFkMmZmNTczLWUzYmUtNGViOS05OTRkLTZhODRjMWE5ZGU2ZSJ9&pageName=ReportSectionb43cfee63ac9a1136009)

Para comenzar con el análisis, es primordial saber cuál es la población que mayormente se va a estudiar, o dicho de otra forma, la edad predominante en la base de datos



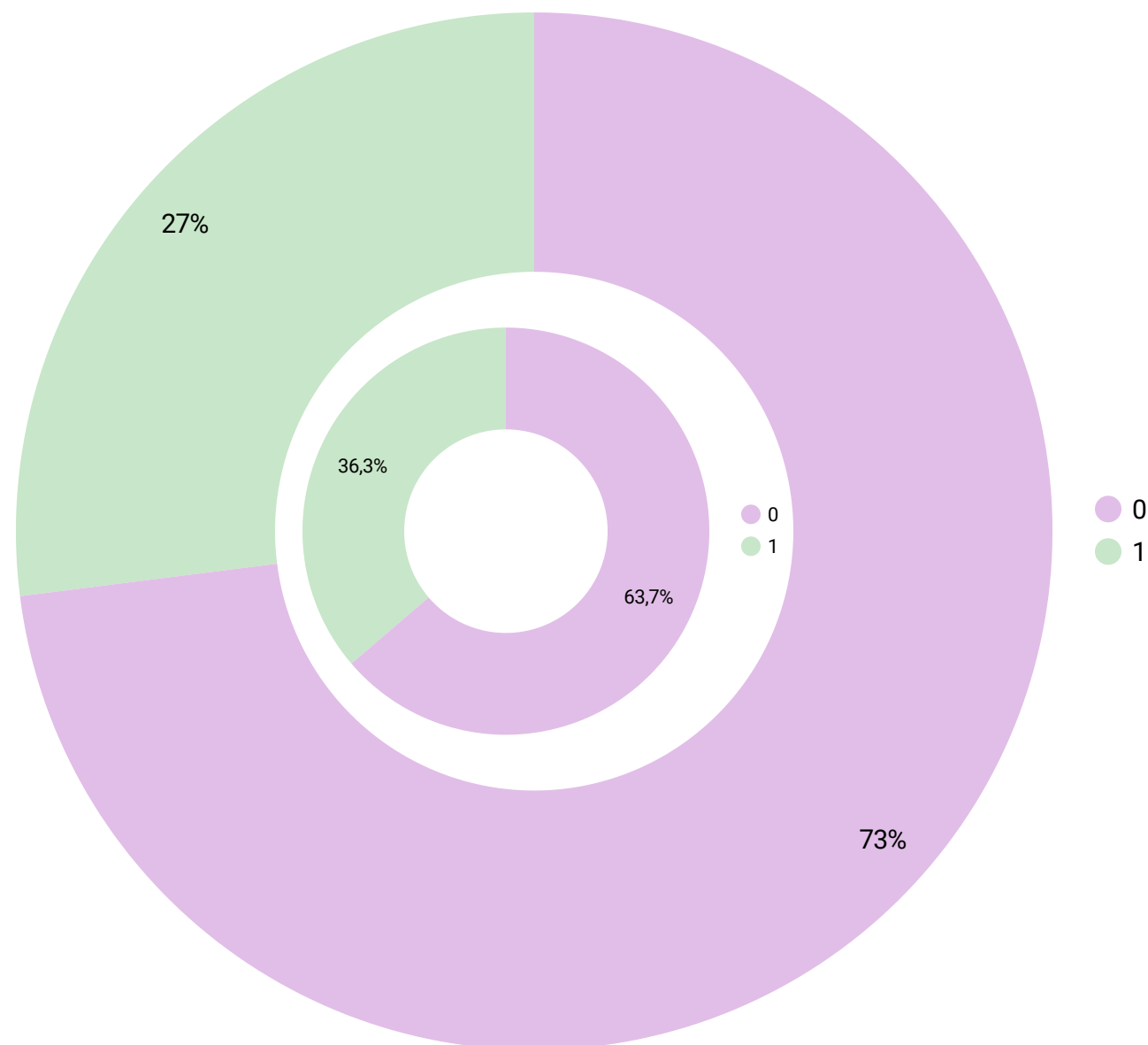
Y también, la temperatura es un valor primordial a la hora de mirar la salud de alguien, pues a pesar de que esta no da mucha información de la enfermedad posible del paciente, se puede decir, que esta da información veraz del estado del paciente. Es bien conocido que si una persona tiene una variación de temperatura por debajo de 36.1°C y por encima de 37.2°C se encuentra una anomalía en el estado de este.

Además, el lugar del hospital del cual se tiene una mayor información



La variable más importante, por simple intuición para definir el estado de una persona a la hora de ser internada en el hospital evidentemente es la mortalidad.

Una variable interesante, aunque no diga mucho en este análisis es el sexo, sin embargo, se debe tener en cuenta que las enfermedades relacionadas con problemas químicos y ambientales tienen más incidencia entre la población femenina porque ellas acumulan un 15% más de grasa en el organismo, menos agua y su sistema metabólico es diferente [*La Salud Según el Género - La Vanguardia*].

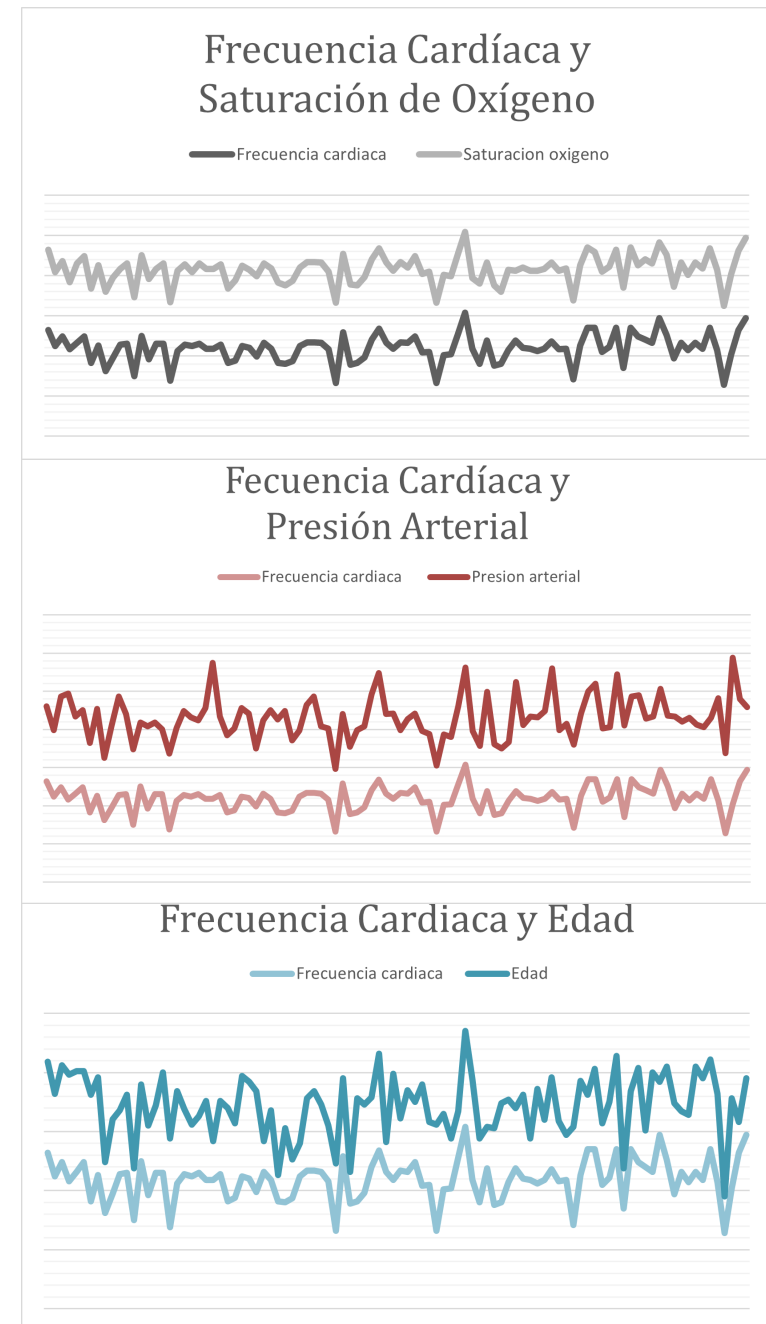


Círculo interno, proporción del sexo, donde 0 es femenino y 1 masculino.

Círculo externo, proporción de la mortalidad, donde 0 es muerto y 1 es vivo.

A medida que una persona va envejeciendo, su vitalidad va empeorando en la mayoría de los casos, trayendo consigo un deterioro en distintos órganos del cuerpo humano.

Con las gráficas siguientes, se encuentran las relaciones más comunes que se dan con respecto al deterioro de su frecuencia cardíaca, presión arterial y saturación de oxígeno



Nota 1: Se omitieron los valores de los ejes ya que el objetivo principal era comparar el comportamiento de la variable más importante (la frecuencia cardíaca) con las distintas variables que presentan deterioros o cambio a medida que una persona aumenta su edad

Nota: Se debe tener en cuenta, que al realizar las gráficas, se encontraron dos datos que impedían una buena escalación de los datos, en la presión arterial había un valor de 628 y en la saturación de oxígeno se encontró uno de 927. Por lo tanto, dichos datos se eliminaron

Conclusiones

La mayoría de los pacientes se encuentran en alguno de los siguientes sectores del hospital: UCI para adultos, cirugía o en urgencias. Por lo tanto, es una base de datos en la cual las personas tienen un estado delicado de salud.

La mayoría de pacientes tienen edades que rondan entre 85 y 75, con edad máxima en 90, mínima en 22 y una edad promedio más o menos en 62 años.

Se encontró que un número considerable de pacientes tienen temperaturas en las cuales si es meritorio estar en el hospital ya que esto afecta directamente la salud de cualquier persona.

Una variable que verifica lo ya dicho respecto a la delicada situación de salud de los pacientes presentes en la base de datos es la mortalidad, dando como resultado que del total de los pacientes hay un 73% que murieron. La mayoría de estos fueron mujeres (26 defunciones contra 19 de hombres). No obstante, la mayoría de pacientes fueron del género femenino con un 63.7% frente a un 36.3% de hombres.

Conclusiones

Finalizando, la variable que predominó en complicaciones de salud en los pacientes fue la frecuencia cardíaca, pues dado el comportamiento de esta, el paciente presenta problemas en la presión arterial o en la saturación del oxígeno y mucho más en este último, exhibiendo máximos y mínimos justo en los mismos puntos.

Por último, la muestra de la base de datos es pequeña, y para tener una mayor precisión o profundidad o incluso, un criterio más fuerte en el momento de determinar la condición de un paciente a la hora de ingresar a alguna área del hospital, se debe tomar una muestra mucho más grande e implementar otro tipo de análisis con mayor valor estadístico y causal.

Python

#¿Cuántos pacientes tuvieron la saturación entre 90 y 95% en el momento de ingreso?

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv("ingresos_abril_06_2022.csv")
sat = df["Saturacion oxigeno"]
sat = sat.to_numpy()
```

```
ind1 = sat <= 95
ind2 = sat >= 90
ind = ind1 & ind2
total = np.sum(ind)
print(f"Total pacientes con saturación en rango: {total}")
```

#¿Cuál es la frecuencia cardiaca mínima, máxima y promedio para las mujeres del archivo?

```
import pandas as pd
import numpy as np
df = pd.read_csv("ingresos_abril_06_2022.csv")
fc = df["Frecuencia cardiaca"]
fc = fc.to_numpy()
sexo = df["Sexo"]
sexo = sexo.to_numpy()
ind = sexo == 1
fc_mujeres = fc[ind]
minimo = np.min(fc_mujeres)
maximo = np.max(fc_mujeres)
promedio = np.mean(fc_mujeres)
print(f"FC mínima mujeres: {minimo} lpm")
print(f"FC máxima mujeres: {maximo} lpm")
print(f"FC promedio mujeres: {np.round(promedio,2)} lpm")
```

Python

```
# En los hospitales se usan escalas cuantitativas para
evaluar la gravedad que tiene
# un paciente en el momento de ingreso. Vamos a calcular
el puntaje para cada paciente
# tomando solo la saturación de oxígeno de la siguiente
manera:
# • 0 puntos: Saturacion oxigeno >= 96%
# • 1 puntos: Saturacion oxigeno entre 94 y 95%
# • 2 puntos: Saturacion oxigeno entre 92 y 93%
# • 3 puntos: Saturacion oxigeno <= 91%
# Entre más grande el valor, peor es la condición de salud
del paciente.
# Calculamos el puntaje de cada paciente y muéstrelo en
pantalla.
import pandas as pd
import numpy as np
df = pd.read_csv("ingresos_abril_06_2022.csv")
sat = df["Saturacion oxigeno"]
sat = sat.to_numpy()
satShape = sat.shape
satNRows = satShape[0]
for i in range(satNRows):
    if sat[i] >=96:
        print(f"Paciente {i} puntaje: 0")
    if (sat[i] ==95) or (sat[i] ==94):
        print(f"Paciente {i} puntaje: 1")
    if (sat[i] ==93) or (sat[i] ==92):
        print(f"Paciente {i} puntaje: 2")
    if sat[i] <=91:
        print(f"Paciente {i} puntaje: 3")
```

```
# El archivo "ingresos_abril_06_2022.txt" contiene la
misma información del archivo
# numérico, pero ahora cada paciente está separa por un
guion (" - ").
# ¿Cuántas personas ingresaron al servicio de UCI adultos?
file = open("ingresos_abril_06_2022.txt")
data = file.read()
total = data.count("UCI adultos")
print(f'Total pacientes en UCI adultos: {total}')
```

Python

#¿Cuántos pacientes tienen entre 50 y 60 años en el momento de ingreso?

```
file = open("ingresos_abril_06_2022.txt")
```

```
data = file.read()
```

```
registros = data.split(" - ")
```

```
cont = 0
```

```
for i in registros:
```

```
    aux = i.split("Edad: ")
```

```
    aux1 = aux[1]
```

```
    edad = int(aux1[0:2])
```

```
    if (edad<=60) and (edad>=50):
```

```
        cont = cont + 1
```

```
print(f"Total de pacientes entre 50 y 60 años: {cont}")
```

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.read_csv("ingresos_abril_06_2022.csv")
```

```
ind1 = df["Temperatura"] >= 38.0
```

```
ind2 = df["Sexo"] == 0
```

```
ind3 = df["Servicio"] == "UCI adultos"
```

```
ind4 = df["Servicio"] == "Hospitalizacion"
```

```
ind = ind1 & ind2 & (ind3 | ind4)
```

```
df2 = df[ind]
```

```
print(f"Pacientes que cumplen las condiciones:  
{len(df2)}")
```

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.read_csv("ingresos_abril_06_2022.csv")
```

```
ind1 = df["Temperatura"] >= 38.0
```

```
ind2 = df["Sexo"] == 0
```

```
ind3 = df["Servicio"] == "UCI adultos"
```

```
ind4 = df["Servicio"] == "Hospitalizacion"
```

```
ind = ind1 & ind2 & (ind3 | ind4)
```

```
df2 = df[ind]
```

```
presion_promedio = df2["Presion arterial"].mean()
```

```
print(f"Presion promedio: {presion_promedio}")
```