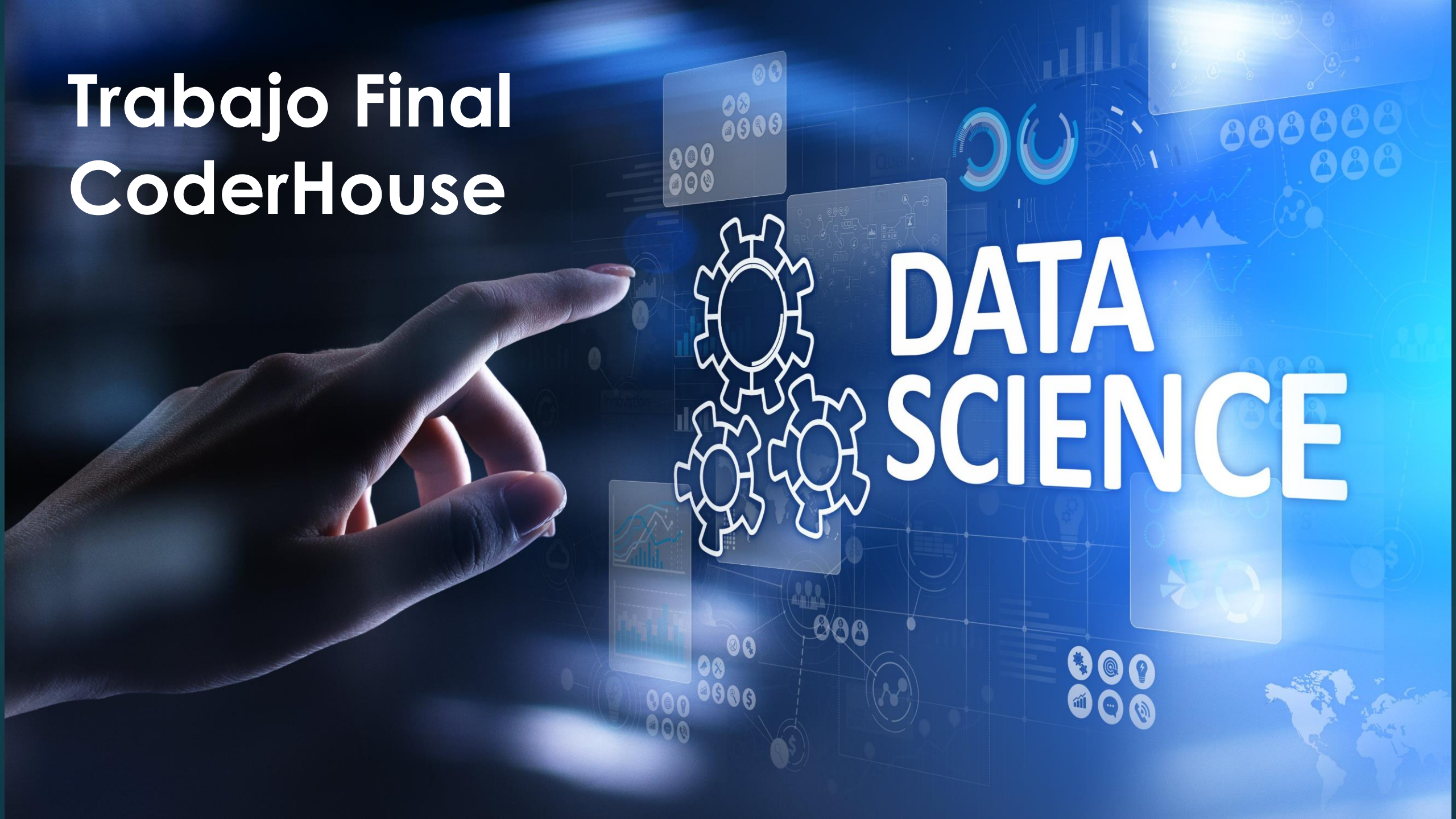


Trabajo Final CoderHouse



DATA SCIENCE



INDICE



Integrantes



Introduccion



Preguntas y
Objetivos

Integrantes

- Lucas San Cristobal



Introduccion

Para este análisis, se han utilizado datos del dataset "Stroke Prediction Dataset" de Kaggle (<https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>). Este dataset contiene información sobre pacientes y si han sufrido un accidente cerebrovascular (ACV) o no. Los datos fueron recolectados en hospitales de la India y constan de 12 variables, que incluyen información demográfica, de estilo de vida y de salud del paciente.





Introduccion

El dataset contiene 5109 observaciones con 12 variables.
Las variables son las siguientes:

id: el numero único de paciente

gender: genero del paciente (femenino = 0, masculino = 1, otro = 2)

age: edad del paciente

hypertension: si el paciente tiene o no hipertensión (1 = si, 0 = no)

heart_disease: si el paciente tiene o no DVC o enfermedad cardiovascular (1 = si, 0 = no)

ever_married: si el paciente estuvo o no casado

Residence_type: tipo de residencia del paciente (urbano = 1 o rural = 0)

avr_glucose_level: nivel de glucosa en sangre del paciente

bmi: masa corporal del paciente

stroke: si el paciente sufrió o no ACV (1 = si, 0 = no)

work_type_Govt_job: Si el paciente trabaja para el gobierno (1 = si, 0 = no)

work_type_Never_worked: Si el paciente nunca trabajo (1 = si, 0 = no)

work_type_Private: Si el paciente trabaja como privado (1 = si, 0 = no)

work_type_Self-employed: Si el paciente es empleador (1 = si, 0 = no)

work_type_children: Si es un niño (1 = si, 0 = no)

smoking_status_Unknown: Si se desconoce si el paciente fuma (1 = si, 0 = no)

smoking_status_formerly smoked: Si el paciente fumaba (1 = si, 0 = no)

smoking_status_never smoked: Si el paciente nunca fumo (1 = si, 0 = no)

smoking_status_smoke: Si el paciente fuma (1 = si, 0 = no)



Introduccion

Los accidentes cerebrovasculares (ACV) son eventos médicos graves que pueden tener consecuencias devastadoras en la vida de las personas.

La detección temprana de los factores de riesgo es fundamental para prevenir estos eventos.

En este análisis, utilizaremos el dataset "Stroke Prediction Dataset" para explorar las variables relacionadas con pacientes y su propensión a sufrir un ACV.

Nuestro objetivo es identificar patrones y desarrollar un modelo predictivo para ayudar en la prevención de accidentes cerebrovasculares.

Preguntas y Objetivos

Contexto empresarial:

Un equipo de investigación médica ha recopilado datos de pacientes con enfermedades cardiovasculares durante los últimos tres años. Quieren identificar patrones y factores de riesgo que puedan estar relacionados con la aparición de enfermedades cardiovasculares en pacientes. El equipo ha recopilado información detallada sobre cada paciente, como su edad, género, nivel de actividad física, presión arterial, nivel de colesterol, tabaquismo, etc. Quieren que construyas visualizaciones que les ayuden a identificar patrones y factores de riesgo en el conjunto de datos, lo que les permitiría desarrollar intervenciones preventivas para reducir la incidencia de enfermedades cardiovasculares.



Preguntas y Objetivos

Problema comercial:

Su tarea es formatear los datos proporcionados y proporcionar visualizaciones que respondan las preguntas específicas que tiene el equipo de investigación médica, que se mencionan a continuación.



Preguntas y Objetivos

Contexto analítico:

Se le proporciona un archivo CSV (almacenado en la carpeta datos ya creada) que contiene detalles sobre cada paciente, como edad, género, nivel de actividad física, presión arterial, nivel de colesterol, tabaquismo, etc. Realizará las siguientes tareas con los datos:

1. Leer, transformar y preparar datos para su visualización
2. Realizar análisis y construir visualizaciones de los datos para identificar patrones en el conjunto de datos.



Preguntas y Objetivos

El equipo de investigación médica tiene un conjunto específico de preguntas a las que le gustaría obtener respuestas. Deberá proporcionar visualizaciones para acompañar estos:

1. ¿Hay alguna relación entre la edad y la presencia de DVC?
2. ¿Afecta el género a la aparición de DVC?
3. ¿Hay más fumadores entre los pacientes con DVC que entre los pacientes sin DVC? ¿Y con la hipertensión?
4. ¿Hay algún patrón de riesgo cardiovascular en diferentes grupos de edad y género?

5. ¿Hay correlación entre la edad y los ACV?

6. ¿Cuál es la distribución de fumadores por edad?

7. Entre los pacientes con ACV ¿Hay relación entre el nivel de glucosa en sangre y la edad con las DVC y la hipertensión?

8. Entre los pacientes con ACV ¿Hay relación entre el BMI y la edad con el estado de fumador?

