Cleaning data with python

import pandas as pd

#uploaded file

from google.colab import files

uploaded = files.upload()

filepath="hospital\_patient\_records.csv"

df=pd.read\_csv(filepath)

df.head()

A screenshot of a computer

AI-generated content may be incorrect.

##date time

df['DischargeDate']=pd.to\_datetime(df['DischargeDate'] , errors='coerce')

df['DischargeDate'].dtype

A computer code with text

AI-generated content may be incorrect.

#converte NAT to 0

df['DischargeDate'].fillna(0 , inplace=True)

df['DischargeDate']

A screenshot of a phone

AI-generated content may be incorrect.

#date time

df['AdmissionDate']=pd.to\_datetime(df['AdmissionDate'] , errors='coerce')

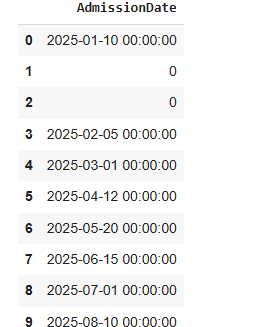
df['AdmissionDate'].dtype

dtype('<M8[ns]')

#converte NAT to 0

df['AdmissionDate'].fillna(0 , inplace=True)

df['AdmissionDate']



#to lowercase

df['Status']=df['Status'].str.lower()

df['Status'].head()

df['Status'].dtype

A screenshot of a computer

AI-generated content may be incorrect.

#convert age to integer

df['Age']=df['Age'].replace('forty-five' , 45)

df['Age'].fillna(0 , inplace=True)

df['Age']=pd.to\_numeric(df['Age'] , errors = 'coerce')

df['Age'].dtypes

df['Age']

A white rectangular box with black numbers

AI-generated content may be incorrect.

#missing value for doctor coulumn

df['Doctor'].fillna('Unknown' , inplace=True)

df['Doctor']

A screenshot of a phone

AI-generated content may be incorrect.

#missing value for name coulumn

df['Name'].fillna('Unknown' , inplace=True)

df['Name']

A group of names on a white background

AI-generated content may be incorrect.

#ststus to lowercase

df['Status']=df['Status'].str.lower()

df['Status']

A screenshot of a phone

AI-generated content may be incorrect.

#remove dublicate

df.drop\_duplicates(inplace=True)

#there is no dublicat data

#checking missing value

#validate cleaning data

missing\_values = df.isnull().sum()

print("Missing values after cleaning:")

print(missing\_values)

A screenshot of a computer

AI-generated content may be incorrect.

# check coulum datatype

df['Age'] = df['Age'].astype(int)

df.dtypes

A screenshot of a computer

AI-generated content may be incorrect.

print("Cleaned dataset:")

print(df)

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

# Count the number of patients per department

patient\_counts = df['Department'].value\_counts()

print("Number of patients per department:")

print(patient\_counts)

A white background with black text

AI-generated content may be incorrect.

# Create a bar chart

plt.figure(figsize=(8, 5))  # Set the figure size

patient\_counts.plot(kind='bar', color='skyblue')

A screenshot of a computer

AI-generated content may be incorrect.

#pie chart

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.