import pandas as pd

# Load data into pandas DataFrame from "/lakehouse/default/Files/HealthcareDataset1.csv"

df = pd.read\_csv("/lakehouse/default/Files/HealthcareDataset1.csv")

display(df)

# Code generated by Data Wrangler for pandas DataFrame

def clean\_data(df):

    # Drop columns: 'Unnamed: 17', 'Unnamed: 18' and 5 other columns

    df = df.drop(columns=['Unnamed: 17', 'Unnamed: 18', 'Unnamed: 19', 'Unnamed: 20', 'Unnamed: 21', 'Unnamed: 22', 'Unnamed: 23'])

    return df

df\_clean = clean\_data(df.copy())

df\_clean.head()

# Define schema

from pyspark.sql.types import \*

healthcareschema = StructType([

    StructField('PatientID', IntegerType(), True),

    StructField('PatientName', StringType(), True),

    StructField('Gender', StringType(), True),

    StructField('BloodType', StringType(), True),

    StructField('Diagnosis', StringType(), True),

    StructField('Treatment', StringType(), True),

    StructField('AdmissionDate', StringType(), True),

    StructField('DischargeDate', StringType(), True),

    StructField('TotalBill', IntegerType(), True),

    StructField('FullPrescriptionDetails', StringType(), True),

    StructField('Ratings', FloatType(), True),

    StructField('Comments', StringType(), True),

    StructField('NPS', StringType(), True),

    StructField('Category', StringType(), True),

    StructField('NodaysStayed', IntegerType(), True),

    StructField('DoctorsExplainedtheTreatmentwell', StringType(), True)

 ])

# write file data in bronze file path as a delta table

df = spark.read.format("csv").schema(healthcareschema).load("abfss://bd10bb14-b060-4e38-9a42-f9b7241ea216@onelake.dfs.fabric.microsoft.com/a124e9df-ee60-4191-af40-1694bc10b149/Files/HealthcareDataset1.csv")

display(df)

#save as delta table

df.write.format("delta").mode("overwrite").saveAsTable("LakehouseTablehealthcare")

# Bronze to Silver NoteBook  
  
from pyspark.sql.functions import when,rand

from pyspark.sql import SparkSession

from pyspark.sql.functions import col

from notebookutils import \*

from pyspark.sql.functions import input\_file\_name

import os

#read data from bronze table

df = spark.read.format("delta").load("abfss://bd10bb14-b060-4e38-9a42-f9b7241ea216@onelake.dfs.fabric.microsoft.com/a124e9df-ee60-4191-af40-1694bc10b149/Tables/dbo/lakehousetablehealthcare")

#display data

display(df)

# Define the path to your Silver Lakehouse

silver\_path = "abfss://bd10bb14-b060-4e38-9a42-f9b7241ea216@onelake.dfs.fabric.microsoft.com/71f943f7-c745-47d7-8435-1fa072d21d16/Tables/Bronze\_Silver\_Table"

# Save the DataFrame as a Delta table

df.write.mode("overwrite").format("delta").option("overwriteSchema", "true").save(silver\_path)

# Register the Delta table as a SQL table

spark.sql(f"""

    CREATE TABLE IF NOT EXISTS Bronze\_Silver\_Table

    USING DELTA

    LOCATION '{silver\_path}'

""")

# Welcome to your new notebook

# Type here in the cell editor to add code!

from pyspark.sql.functions import when,rand

from pyspark.sql import SparkSession

from pyspark.sql.functions import col

from notebookutils import \*

from pyspark.sql.functions import input\_file\_name

import os

#read data from silver table

df = spark.read.format("delta").load("abfss://bd10bb14-b060-4e38-9a42-f9b7241ea216@onelake.dfs.fabric.microsoft.com/71f943f7-c745-47d7-8435-1fa072d21d16/Tables/Bronze\_Silver\_Table")

display(df)

# Define the path to your Gold Lakehouse

gold\_path = "abfss://bd10bb14-b060-4e38-9a42-f9b7241ea216@onelake.dfs.fabric.microsoft.com/521a46ad-c69d-4c72-a545-15d278c1753b/Tables/Gold\_Patient"

# Save the DataFrame as a Delta table

df.write.mode("overwrite").format("delta").option("overwriteSchema", "true").save(gold\_path)

# Register the Delta table as a SQL table

spark.sql(f"""

    CREATE TABLE IF NOT EXISTS Gold\_Patient

    USING DELTA

    LOCATION '{gold\_path}'

""")