import findspark

findspark.init()

import pyspark

from pyspark.sql import SparkSession

from pyspark.sql.functions import \*

import os

def etl\_function()

.appName("ETL")\.config("spark.jars", “Users/shiva/Desktop/assignment task/postgresql- 42.2.6.jar") \

.getOrCreate()

df = spark.read.format("csv").option("header","true").load('file:///Users/shauvik/Desktop/file1.csv')

df = df.filter((df.Orderno != '')

df = df.filter(df.Quantityorder != '')

df = df.filter(df.Priceofone >=600 )

df = df.filter((df.date >01/01/2022 '')

df = df.withColumn("date", to\_date(df["date"], "dd/mm/yyyy"))

df = df.withColumn("Id",df.Id.cast('bigint'))

df = df.withColumn("Orderno",df.Orderno.cast('int'))

df = df.withColumn("Quantityorder",df.Quantityorder.cast('int'))

df = df.withColumn("Priceofone",df.Priceofone.cast('int'))

result\_df.show()

result\_df.write\

.format("jdbc")\

.mode("append")\

.option("url", "jdbc:postgresql://localhost:5434/)\

.option("dbtable", "demo")\

.option("user", "postgres")\

.option("password", "password")\

.option("driver", "org.postgresql.Driver")\

.save()

result\_df.show()

spark.stop()

if \_\_name\_\_ == "\_\_main\_\_":

etl\_function()