Data set Name: Baby\_Names.CSV

Question 1: How to load a csv file?

path="/databricks-datasets/online\_retail/data-001/ Baby\_Names.csv"

df=spark.read.load(path, format='com.databricks.spark.csv', header='true', inferSchema='true')

Question 2: Show 20 rows

display(df)

Question 3: Show me number of rows

select COUNT(\*) from `baby\_names\_csv`

Question 4: Show me first 10 rows

select \* from `baby\_names\_csv` LIMIT 10

Question 5: Show DataFrame Schema

df.printSchema()

Question 6: Handle Missing values for count and country

Question 7: Show “Country” column value

df.select("Country").show()

Question 8: Remove Duplicates Country Name.

select distinct `Country` from `baby\_names\_csv`

Question 9: show total number of child in country= NEW YORK

Select Count(Children) from ` Baby\_Names.CSV` where country=’ NEW YORK ‘

Question 10: show total number of child in year= 2013

select sum(count) from ` Baby\_Names.CSV `

` where Year = 2013

Question 11:Show total number of children of JACKSON

display(pathLoad.select(pathLoad["First Name"],pathLoad["Count"]).groupBy(pathLoad["First Name"]).sum().filter(pathLoad["First Name"]==" JACKSON "))

Question 12: show total number of Male and Female living in country= NEW YORK

select sum(`Count`) as `total number of Male and Female` from `baby\_names\_csv` where `Sex`="M" between "F" and `Country`= "NEW YORK"

Question 13: save result as table

r1.write.saveAsTable(" ")