Afeefah Manzoor

Course Project

ITN213

**PART 1 (200 points):**

AWS Lake Formation

1. For this service, describe the typical way to interact with the service (that is, a list of actions or steps users' takes to implement this service. Similar to  instructions provided to create a MySQL DB, Image Rekognition etc.)

To create an AWS data Lake with Lake Formation, the user will need to know where the data is located and what data it needs to access as well as the secuiry policies that the user will want to apply. Then the Lake Formation will collect and catalog all the data from the databases and object storages and move them into the new AWS S3 data lake. After that the Lake Formation will clean, classify the data using machine learning and secure access to sensitive data.

1. How does AWS charge for this service?

There is no additional charge for Lake formations but the user is charged if using AWS Glue, S3, EMR, Athena, or Redshift.

(3) Is it a part of the free tier?

Yes

(4) Provide URLs to documentation for this service.  
 <https://docs.aws.amazon.com/lake-formation/index.html>

(5) URLs to AWS videos about this service.

<https://youtu.be/SeZ78yX8SC8>

<https://youtu.be/nsiLMqg654s>

<https://youtu.be/hJ8Hdt7y0qw>

<https://youtu.be/hXXytLly7tw>

<https://youtu.be/g34xUaJ4WI4>

(6) Watch the videos you located.

(7) For one of the videos you listed, answer the following questions: <https://youtu.be/SeZ78yX8SC8>

* Summarize a few key points made in the video.
  + Growth of data is insane; 10x every 5 years
  + Data lakes provides flexibility to store any type of data
* Identify two interesting quotes that were made.
  + “Customers use glue to build their data lake and catalog it right….”
  + “You need to configure your ETL jobs to load data on a continuous basis to cleanse this data…”
  + “it helps you build a secure data lake in days…”
* What new facts did you learn from watching this video.
  + Creating data lakes are still quite manual
  + That you need to create a bucket for intermediate staging when creating a data lake.
  + The AWS data lake formation is basically a wizard to help make a secure data lake.
  + Can use Athena SQL with lake formation
  + You can train to recognize data, foreample a data that says the episode number verses the episode and can train to recognize that it is the same episode.
* What was the best part of the video?  Why?
  + When she demonstrated how to make a data lake in AWS because I am a visual learner and gave me a better understanding.
* What questions remain in your mind after watching the video?  Why?
  + I do not have any questions since the video was very thorough.

**PART 2 (200 points):**

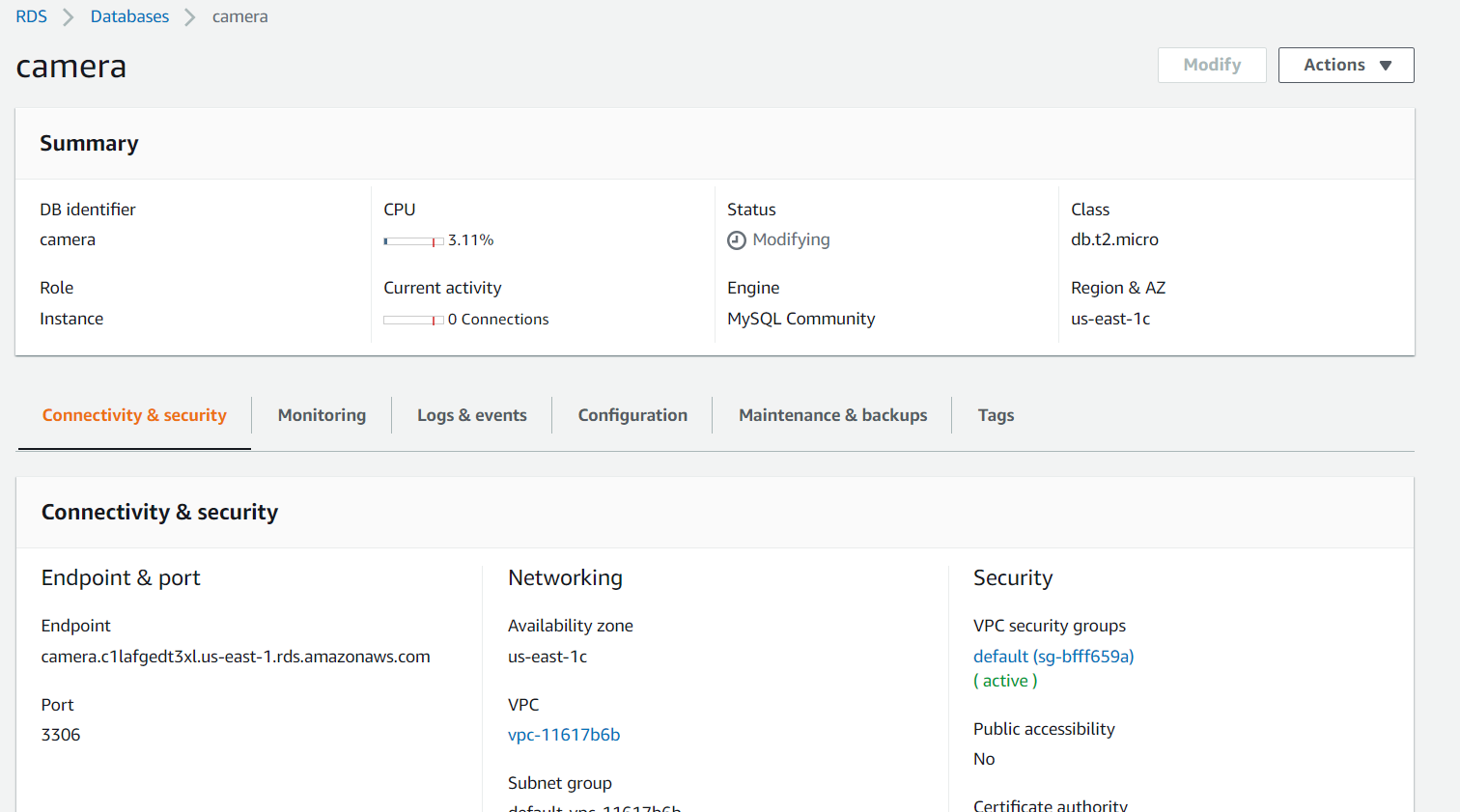
Create a word  document and mark the six answers clearly in it . (You can upload multiple documents to the same assignment, as long as it is before the due date).

Download**one** of the datasets from Unit8/Assignment 8 (Use something different from what you use in assignment 8):

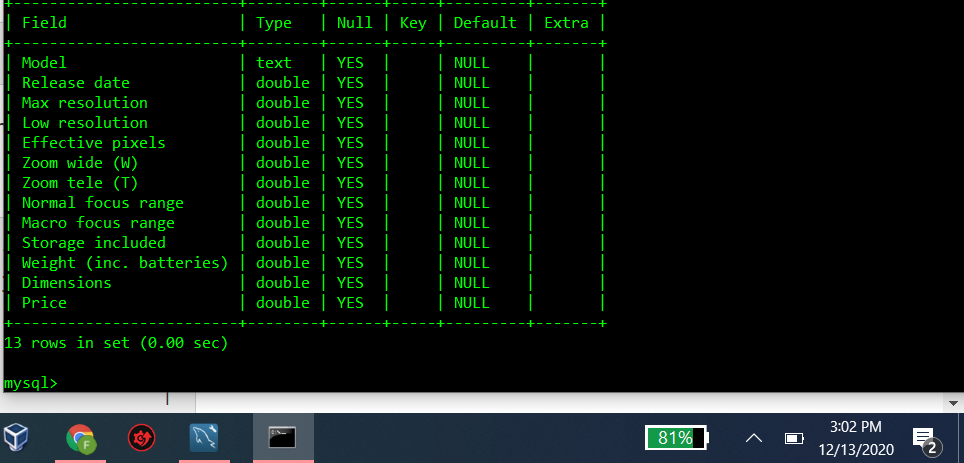
1. Document and describe the dataset you selected.

I decided to use AWS RDS which is a relational database web service which is all stored on the cloud.

1. Create an RDS or DynamoDB solution to store the records from this dataset.

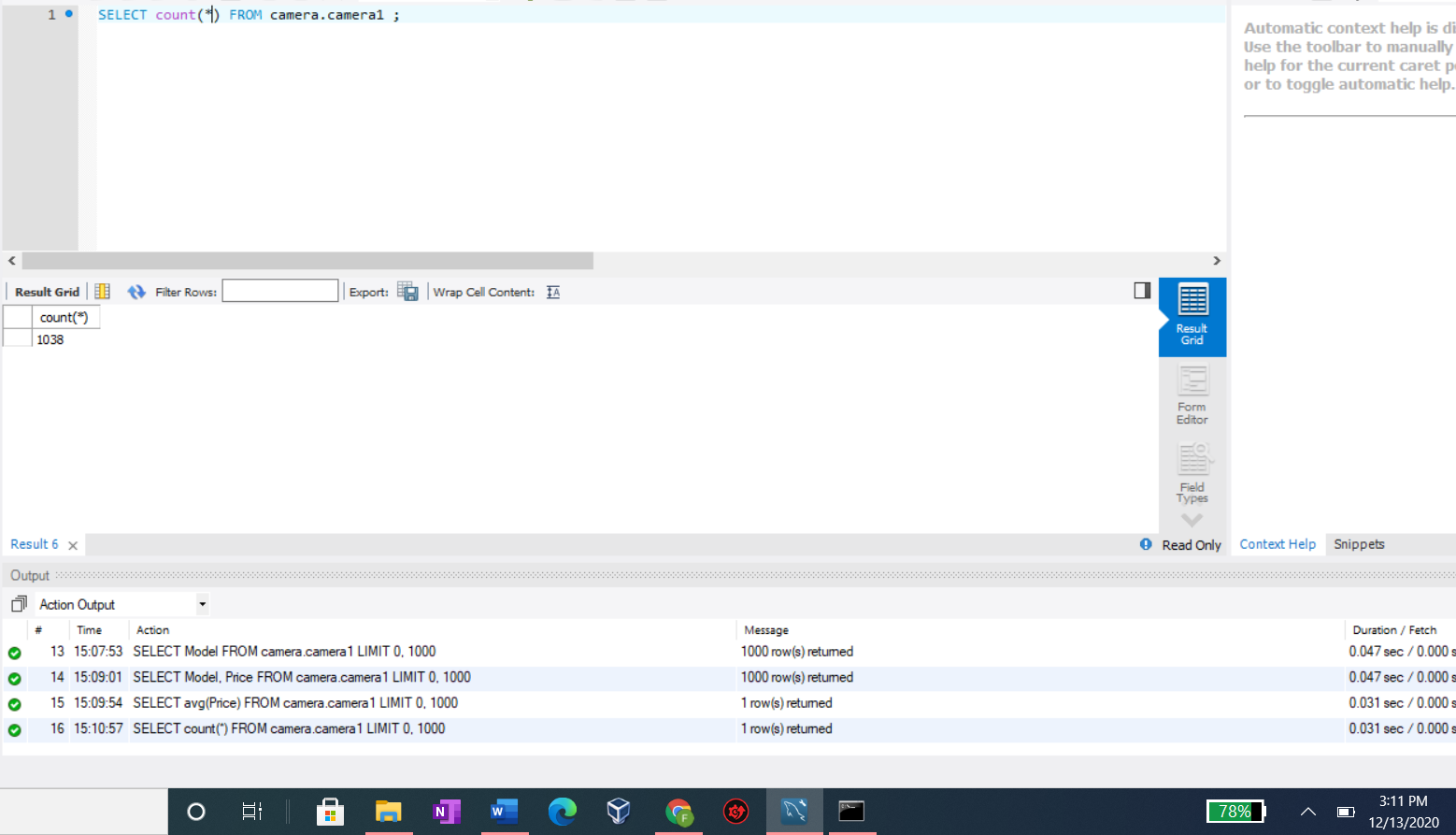
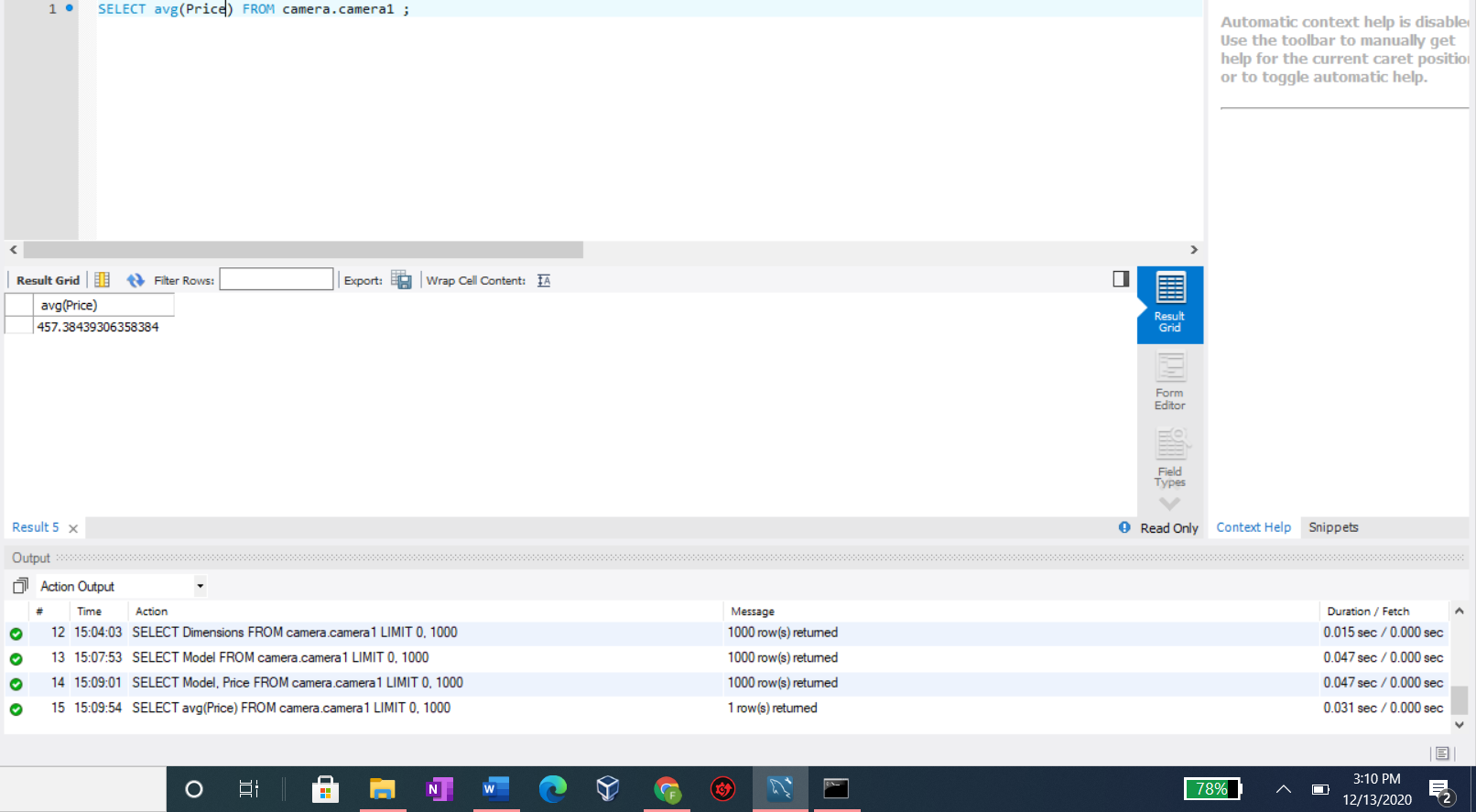
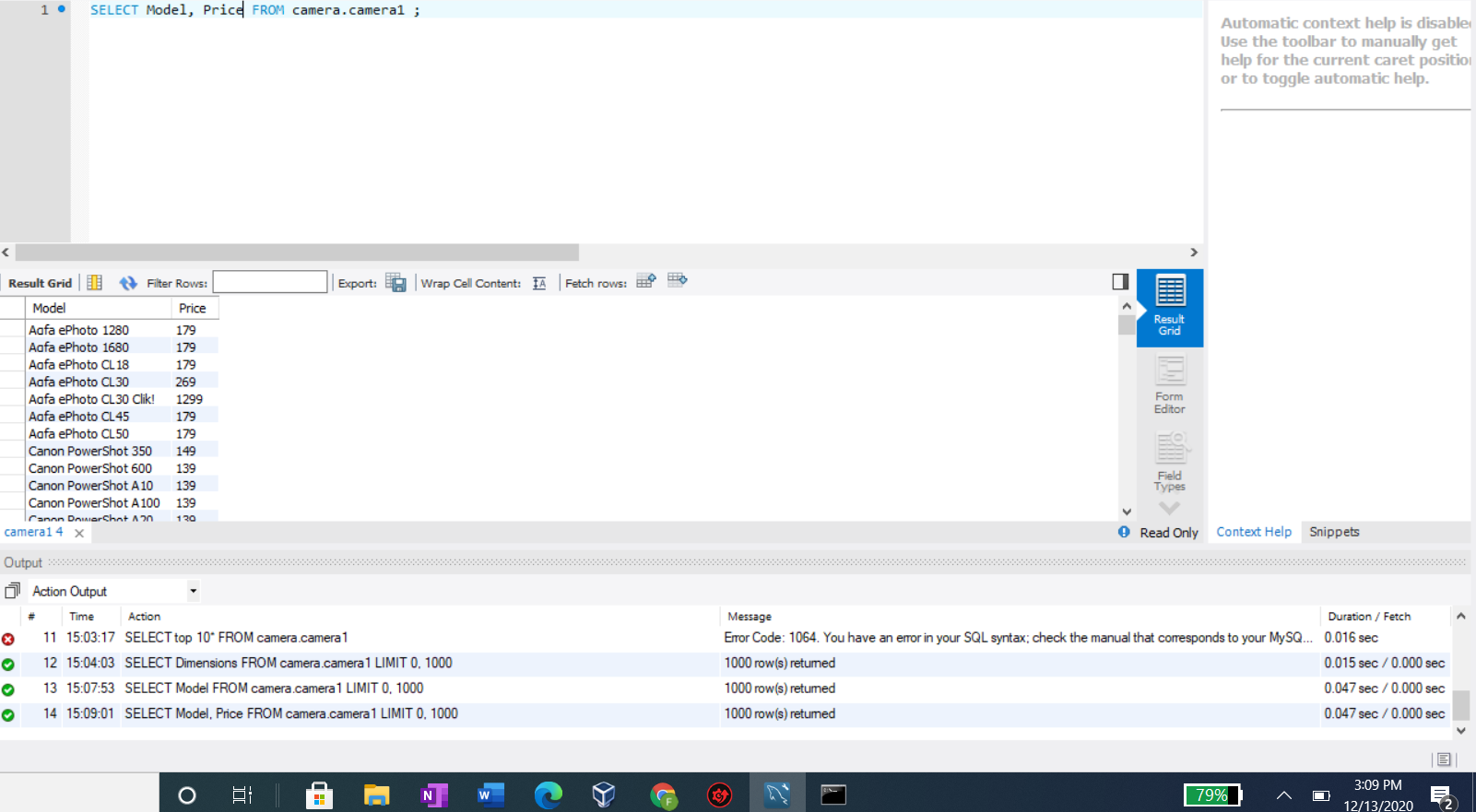
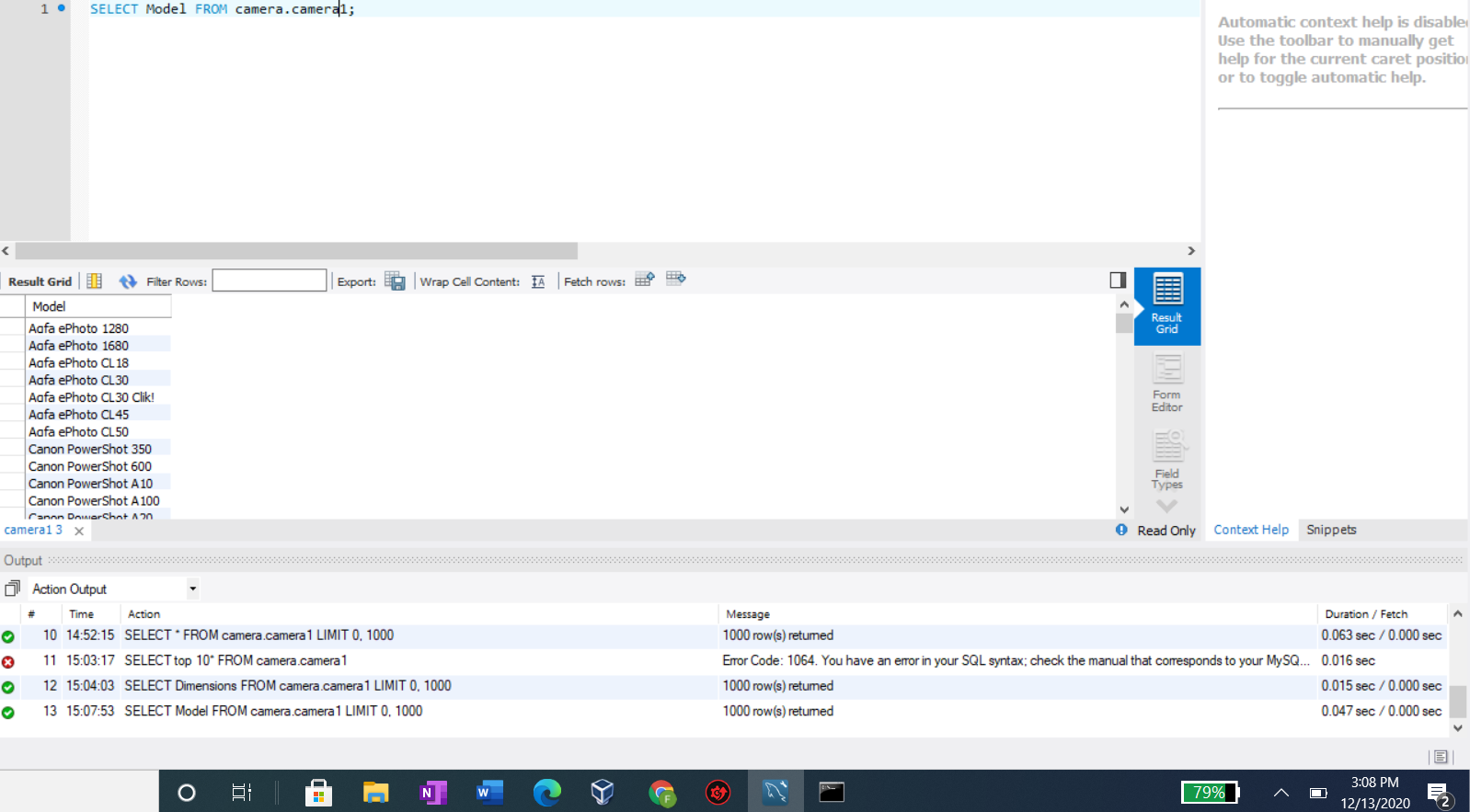
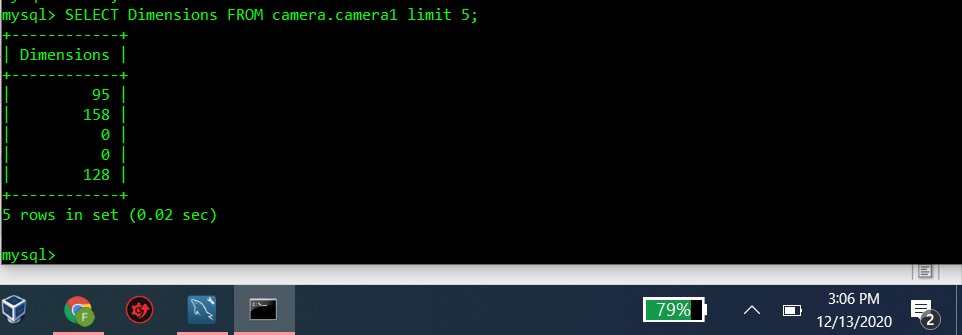


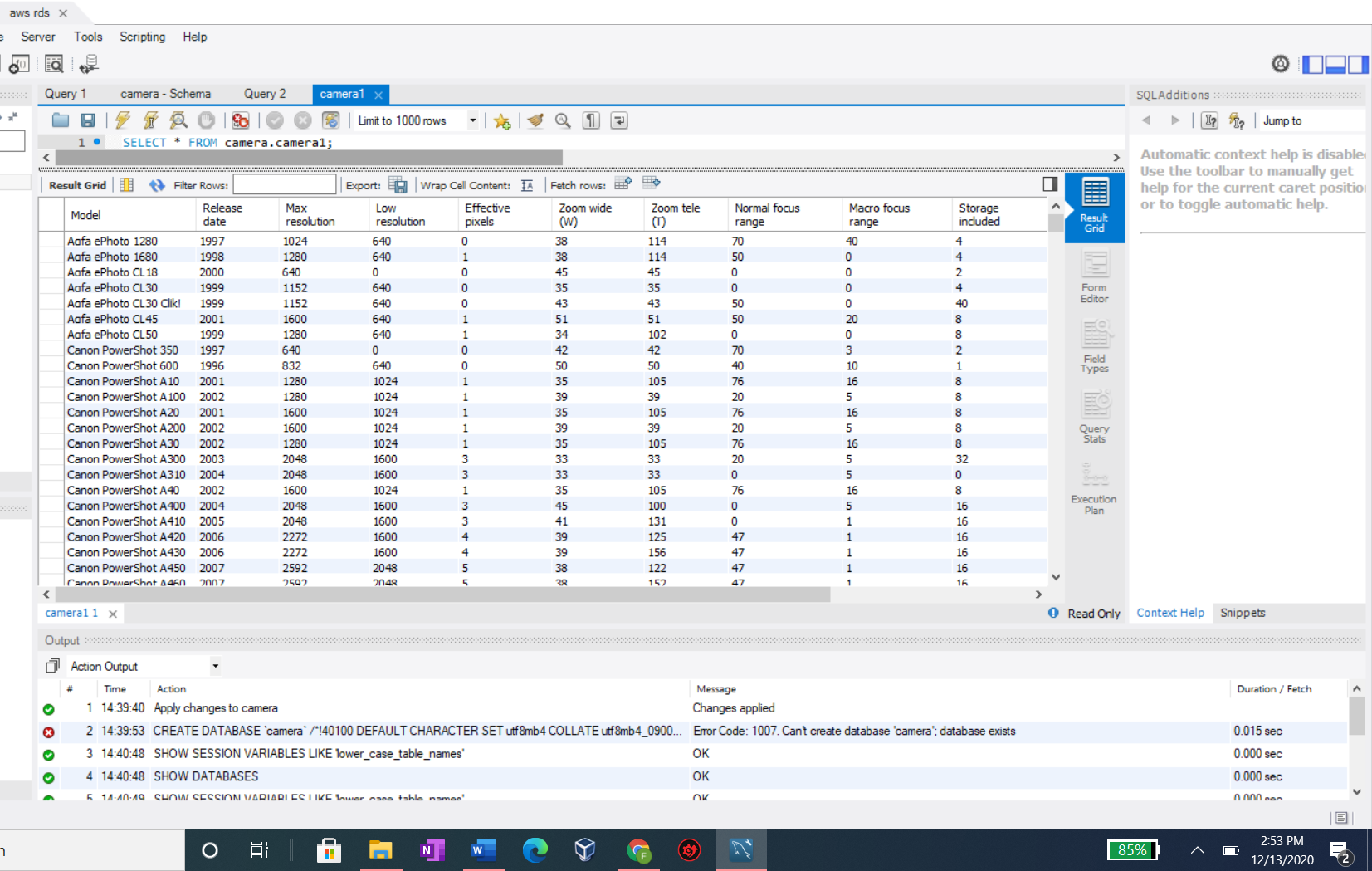
1. Document the structure of the storage solution you defined in AWS.



1. Upload various records into your storage solution that you downloaded from the dataset.

5. **Perform about 4 to 5 different queries of your choice.**



6. Please take a screenshot showing me either your mySQLWorkbench or the table scan from the Items tab of the RDS/DynamoDB table you created to store this dataset. 

7. Describe why you select the AWS service you used to store the dataset records.

I chose RDS because it was a SQL database and DynamoDB is a noSQL database, so I went with RDS.