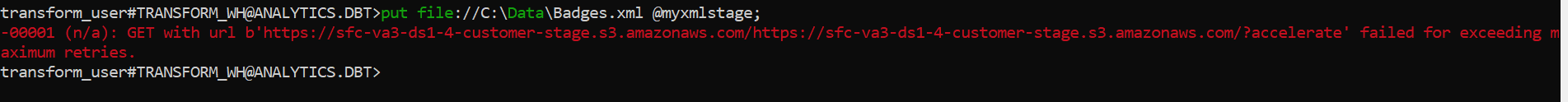
**WEEK 5**

**MINI ASSIGNMENT**

ARUNIMA RAJ B

**Data Upload**

1. I have tried to upload data using snowsql , but that gave me an error.



1. I tried to upload the entire file using Snowflake web interface. But that too gave an error due to large file size.
2. **I wrote a Python Script to chunk the large files into small ones** of affordable size.

Text

Description automatically generated

1. I tried to implement a script to upload these small files to snowflake, but I could only find documentations and clues for csv file upload using **snowflake-python connector with pandas**.
2. Since there was a time constraint for the submission of assignment, I decided to upload manually the small files I got by chunking using the script using Web Interface. Example:

CREATE TABLE Badge\_Value(

badge VARIANT

);

Graphical user interface

Description automatically generated

1. Thereby, **I was able to upload the entire data provided for the assignment** to the database without missing any.
2. Then , I wrote queries to create a table.

CREATE TABLE BADGE\_TABLE(

id number primary key,

userid number,

name varchar(50),

date datetime,

class number,

tagbased varchar(15)

);

1. Then, I wrote queries to parse xml data to table format and insert it into database. An example is given below.

INSERT INTO BADGE\_TABLE

SELECT

badge:"@Id" as "Id",

badge:"@UserId" as "UserId",

badge:"@Name"::String as "Name",

badge:"@Date"::DateTime as "Date",

badge:"@Class" as "Class",

badge:"@TagBased"::String as "TagBased"

FROM Badges\_Value;

1. Queries for all the 5 given questions are written as separate dbt models with names under qstn1, qstn2, qstn3, qstn4, qstn5.
2. Associated tests for not null and unique are also given for id columns