TGM1 — TGM1 Task 1: Data Acquisition

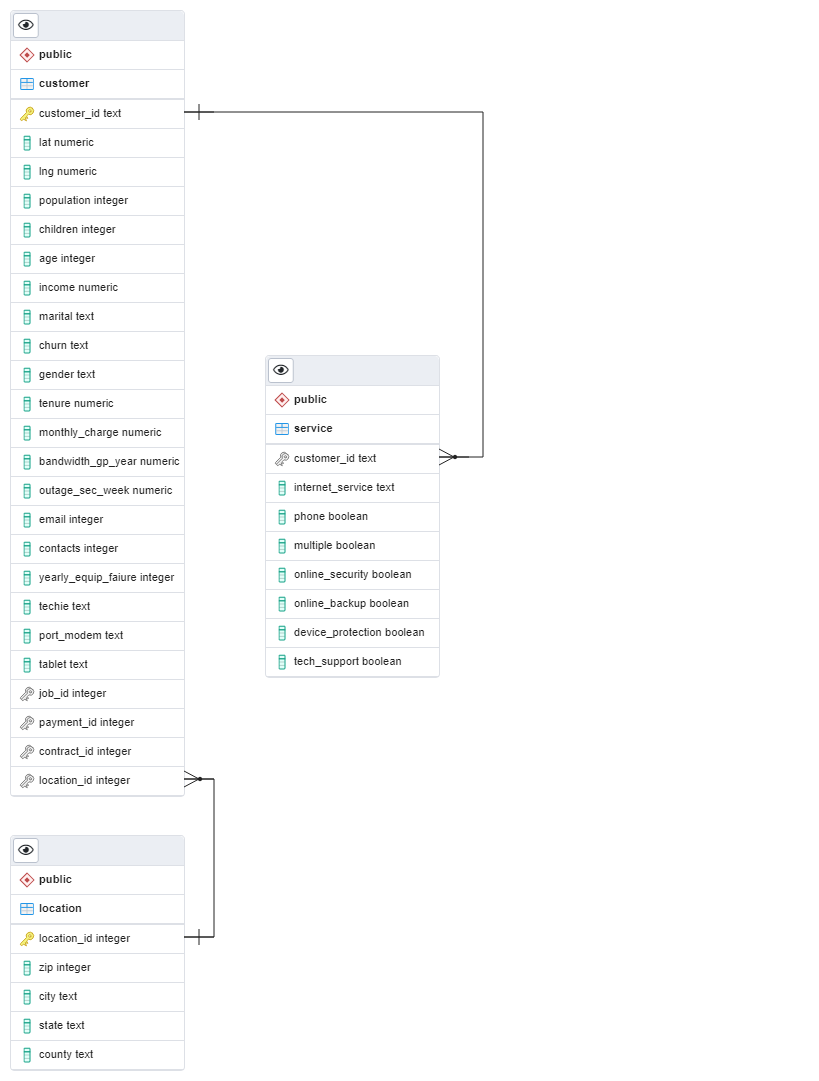
# Data Acquistion — D205

PRFA — TGM1

1. *Summarize a research question that can be answered using both the original database and the add-on CSV data. The question should require data from both these data sources.*  
     
   **Research Question:** For each state in the United States based on the data does a state use more Fiber or more DSL as an internet service?  
   1. Identify which data from the original data set and the add-on CSV file are needed to answer the research question.

|  |  |
| --- | --- |
| **Data Set** | **Tables/Columns** |
| Original Data Set | * customer   + customer\_id (join the data) * location   + location\_id (join the data)   + state (for grouping) |
| Add-on CSV Data Set | * service   + customer\_id (join the data)   + internet\_service (getting the type of internet) |

1. *Create a logical data model for the add-on CSV file by evaluating the data contained in the file and emphasizing the relational constraints.*



* 1. *Write SQL code that creates a table that accommodates the extension of the logical data model to a physical data model by specifying the field types and relevant keys.*

|  |
| --- |
| CREATE TABLE service(  customer\_id TEXT NOT NULL  ,internet\_service TEXT NOT NULL  ,phone BOOLEAN NOT NULL DEFAULT FALSE  ,multiple BOOLEAN NOT NULL DEFAULT FALSE  ,online\_security BOOLEAN NOT NULL DEFAULT FALSE  ,online\_backup BOOLEAN NOT NULL DEFAULT FALSE  ,device\_protection BOOLEAN NOT NULL DEFAULT FALSE  ,tech\_support BOOLEAN NOT NULL DEFAULT FALSE  ,FOREIGN KEY(customer\_id) REFERENCES customer(customer\_id)  ); |

* 1. *Write SQL code that loads the data from the add-on CSV file into the table created in part B1.*

|  |
| --- |
| COPY public.service (  customer\_id  ,internet\_service  ,phone  ,multiple  ,online\_security  ,online\_backup  ,device\_protection  ,tech\_support)  FROM 'C:\LabFiles\Services.csv'  DELIMITER ','  CSV  HEADER  QUOTE '"'  ESCAPE ''''  FORCE NOT NULL customer\_id  ,internet\_service  ,phone  ,multiple  ,online\_security  ,online\_backup  ,device\_protection  ,tech\_support; |

1. *Write SQL statement(s) for a query or queries that inform the research question summarized in part A.*

|  |
| --- |
| SELECT l.state  ,CASE WHEN(  COUNT(CASE s.internet\_service WHEN 'Fiber Optic' THEN 1 ELSE NULL END)   > COUNT(CASE s.internet\_service WHEN 'DSL' THEN 1 ELSE NULL END)  )   THEN 'More Fiber' ELSE 'More DSL' END AS MoreServiceOf  FROM customer AS c  INNER JOIN location AS l ON l.location\_id = c.location\_id  INNER JOIN service AS s ON s.customer\_id = c.customer\_id WHERE s.internet\_service IN ('Fiber Optic', 'DSL') GROUP BY l.state ORDER BY l.state; |

* 1. Provide a CSV file or files that capture the results from the query or queries.  
       
     File: [question-query-result.csv](https://westerngovernorsuniversity-my.sharepoint.com/:x:/g/personal/ada1962_wgu_edu/EXKvYVqx7cBLjBopNE7tJ4YB9Ez2GsYuvtLnRYfCAYmDIQ?e=gAXJbn) (File will also be uploaded with assignment.)

1. Determine how often the add-on file should be acquired and refreshed in the database for the data to remain relevant to the business and the research question.  
     
   The add-on file should be acquired and refreshed monthly. The data resulting from the question that determines if Fiber Optic or DSL is the dominant service being provided is simply used to gauge which type technicians (Fiber Optic or DSL expert) to have more of employed.
2. *Create an SQL script that performs the process of loading the add-on data.*  
     
   COPY public.service (  
    customer\_id  
    ,internet\_service  
    ,phone  
    ,multiple  
    ,online\_security  
    ,online\_backup  
    ,device\_protection  
    ,tech\_support)   
   FROM 'C:\LabFiles\Services.csv'   
   DELIMITER ','   
   CSV   
   HEADER   
   QUOTE '"'   
   ESCAPE ''''   
   FORCE NOT NULL customer\_id  
    ,internet\_service  
    ,phone  
    ,multiple  
    ,online\_security  
    ,online\_backup  
    ,device\_protection  
    ,tech\_support;
3. *Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.*  
     
   ***File Name:*** *D205 - PA - André Davis(ada1962@wgu.edu)*  
     
   ***File Location:***  
   [*https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c755ffa5-1bc3-41fb-bdee-af3c001f3b1f#*](https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c755ffa5-1bc3-41fb-bdee-af3c001f3b1f#)
4. *Record the web sources used to acquire data or segments of third-party code to support the application. Be sure the web sources are reliable.*
   1. The data sets required to be aquired for this project were present on the LAB machine in the LabFiles folder.
5. *Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.*  
   1. PostgreSQL Tutorial . (n.d.). *Import CSV file into postgresql table*. PostgreSQL Tutorial. Retrieved October 24, 2022, from <https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-posgresql-table/>
6. Demonstrate professional communication in the content and presentation of your submission.