Rohan Maharjan

Course Code: CPS151

Course Name: Data Analytics

Date Of Submission: 01/13/2024

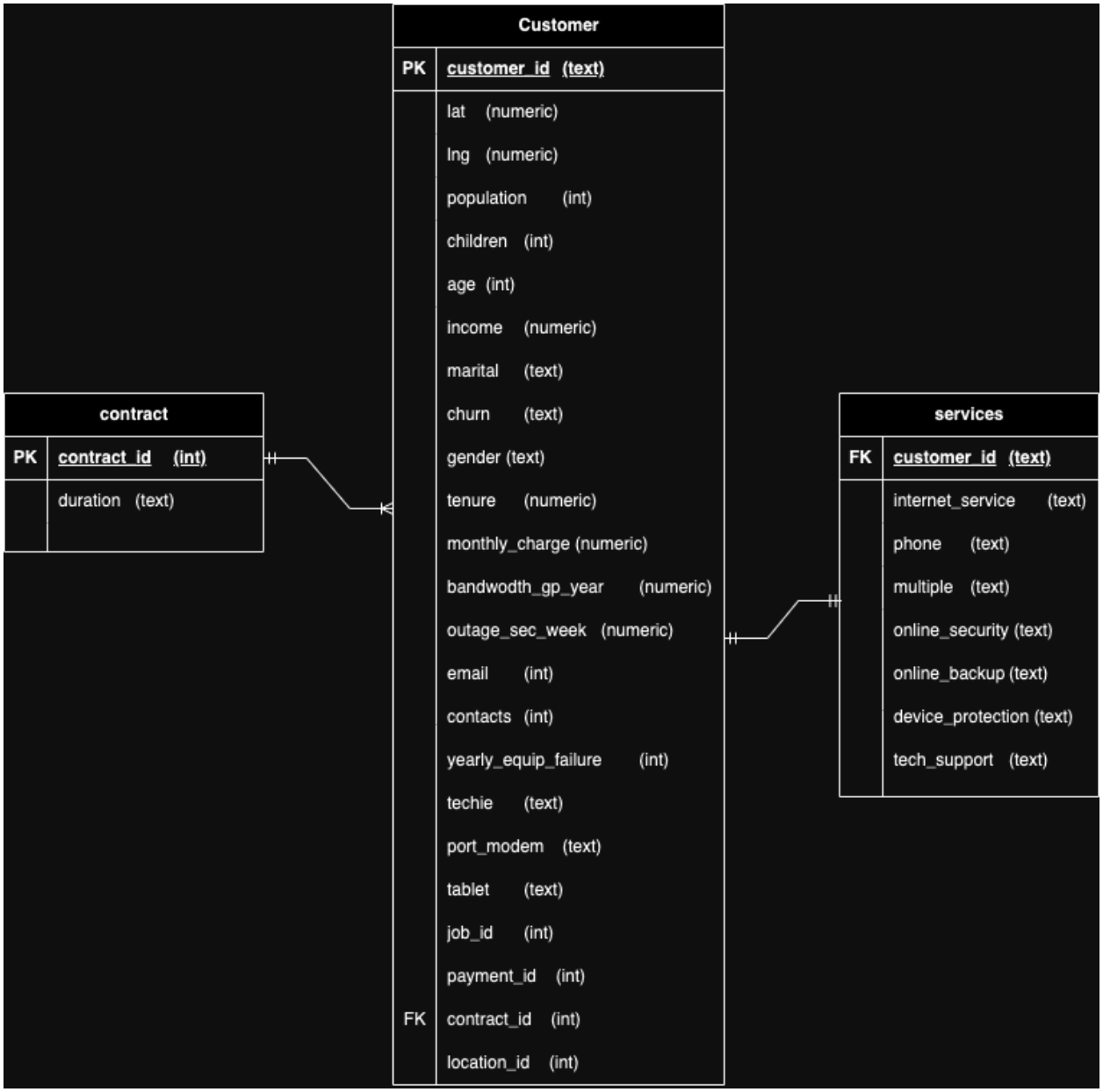
School: Millersville University

A: Research Question

The research question is "What is the total average bandwidth usage of customers using DSL and Fiber Optic internet services on a 12-month contract?" The primary goal of this research is to determine the average bandwidth usage among customers with a 12-month contract, specifically focusing on those using DSL and Fiber Optic internet services. This analysis is crucial for understanding how customer segments utilize Internet services, which is instrumental in shaping network management and service optimization strategies. By pinpointing the bandwidth demands of these two key customer groups, the company can better tailor its infrastructure and service offerings to meet the actual needs of its customer base.

Data will be sourced from the 'Churn' database and an external 'Services' CSV file to conduct this analysis. Necessary data includes the customer_id (text) and bandwidth_gp_year (num) from the 'Customer' table to measure bandwidth usage, alongside the contract_id (text) and duration from the 'Contract' table to identify customers on a 12-month contract. The 'Services' table will provide the internet_service (text) type, distinguishing between DSL and Fiber Optic users. The insights gained from this study will be useful for strategic planning in network capacity, marketing initiatives, and infrastructure development, ensuring that the company's resources are aligned with its customers' usage patterns and preferences.

B. Entity Relationship Diagram



In the ERD diagram, there is a many-to-one relationship between the 'Customer' and 'Contract' tables, meaning many customers can be linked to a single contract. Additionally, there is a one-to-one relationship between the 'Customer' and 'Services' tables, indicating that each customer has a unique set of services.

B1. Creating a table for the CSV file

```
CREATE TABLE public.services
(
    customer_id text COLLATE pg_catalog."default" NOT NULL,
    internet_service text COLLATE pg_catalog."default" NOT NULL,
    phone text COLLATE pg_catalog."default" NOT NULL,
    multiple text COLLATE pg_catalog."default" NOT NULL,
    online_security text COLLATE pg_catalog."default" NOT NULL,
    online_backup text COLLATE pg_catalog."default" NOT NULL,
    device_protection text COLLATE pg_catalog."default" NOT NULL,
    tech_support text COLLATE pg_catalog."default" NOT NULL,
   PRIMARY KEY (customer_id),
  CONSTRAINT customer_id_fkey FOREIGN KEY (customer_id)
REFERENCES public.customer (customer_id) MATCH SIMPLE
ON UPDATE NO ACTION
ON DELETE NO ACTION
NOT VALID
);
ALTER TABLE public.services
OWNER to postgres;
```

The SQL code for creating a table is designed to establish a public.services table in the database, defining columns like customer_id as foreign key, internet_service, phone, and others related to customer services, all set as text data types.

B2. Loading CSV file

```
CSV import

COPY services

FROM 'C:\LabFiles\Services.csv'

DELIMITER ','

CSV HEADER;
```

The second SQL code snippet is for importing data into the services table from a specified CSV file located at 'C:\LabFiles\Services.csv'. This import process uses PostgreSQL's COPY command, with a comma delimiter and assuming the first row of the CSV contains column headers.

C. SQL Code

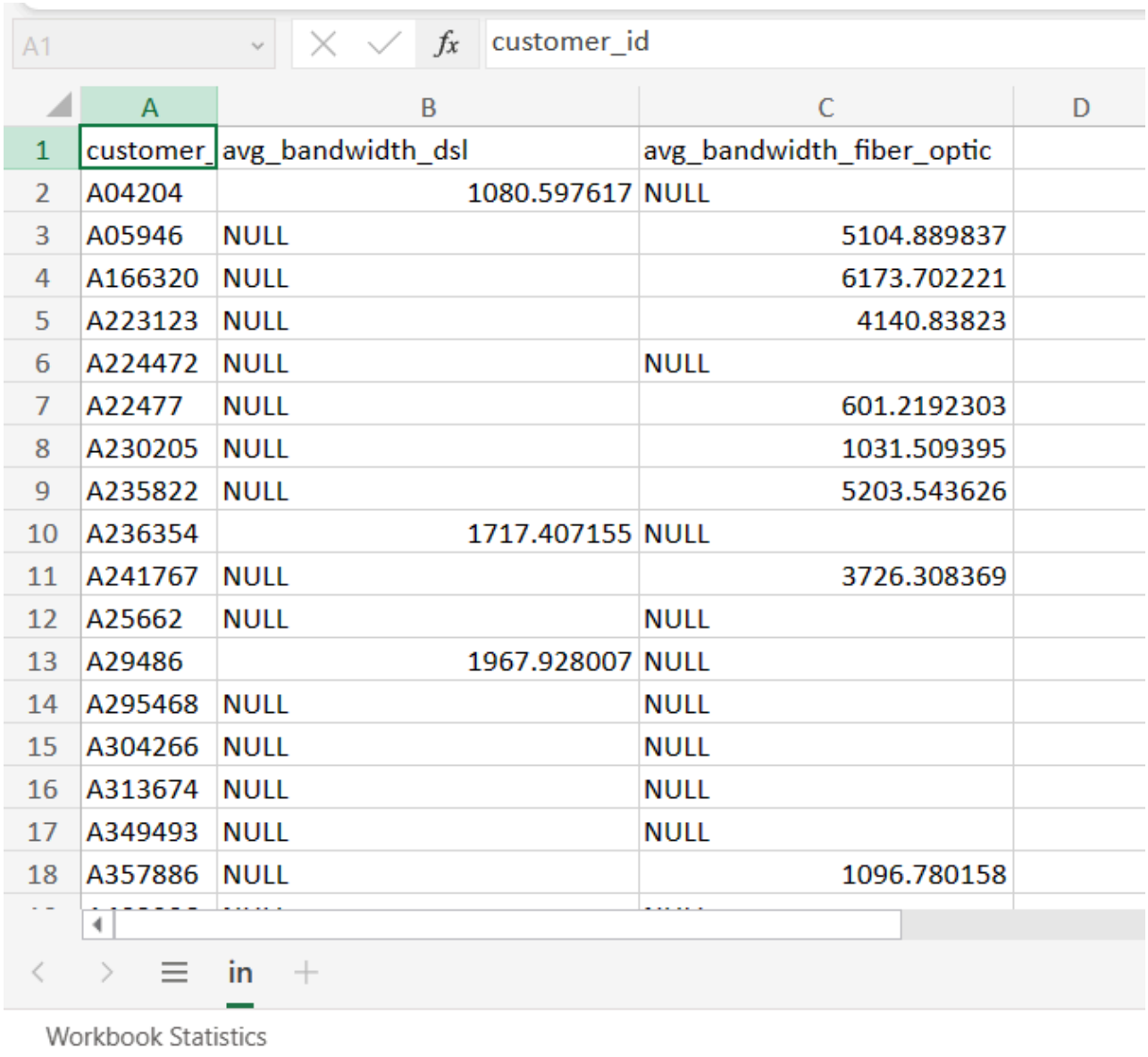
```
SELECT

  customer.customer_id,
 (CASE WHEN services.internet_service = 'DSL' THEN
            customer.bandwidth_gp_year END) AS avg_bandwidth_dsl,
 (CASE WHEN services.internet_service = 'Fiber Optic' THEN
            customer.bandwidth_gp_year END) AS avg_bandwidth_fiber_optic
FROM customer
JOIN contract ON customer.contract_id = contract.contract_id
JOIN services ON customer.customer_id = services.customer_id
WHERE contract.duration = 'One year';
```

This SQL query is designed to list individual customers on a one-year contract and show their yearly bandwidth usage based on the internet service they use - either DSL or Fiber Optic. The query works by joining three tables: 'customer', 'contract', and 'services'. It selects the customer_id from the 'customer' table and uses a CASE statement to determine the type of internet service each customer has. If a customer uses DSL, it shows their yearly bandwidth (bandwidth_gp_year) and does the same for Fiber Optic users. However, this information is displayed for each customer separately. The query specifically looks at customers on a 12-month ('One year') contract, as indicated in the 'contract' table.

```
SELECT
   AVG(CASE WHEN services.internet_service = 'DSL' THEN
            customer.bandwidth_gp_year END) AS avg_bandwidth_dsl,
   AVG(CASE WHEN services.internet_service = 'Fiber Optic' THEN
            customer.bandwidth_gp_year END) AS avg_bandwidth_fiber_optic
FROM customer
JOIN contract ON customer.contract_id = contract.contract_id
JOIN services ON customer.customer_id = services.customer_id
WHERE contract.duration = 'One year';
```

This SQL query calculates the average yearly bandwidth usage for two groups of customers: those using DSL and those using Fiber Optic internet services, both under a one-year contract. This query also joins the 'customer', 'contract', and 'services' tables. However, instead of listing individual customers, it uses the AVG function combined with a CASE statement to compute the average bandwidth usage. It separately calculates the average for customers with DSL and those with Fiber Optic, based on their yearly bandwidth usage. Like the first query, it filters the data to include only those customers whose contract duration is 'One year'. The result of this query is two average figures – one for DSL and one for Fiber Optic users – giving a general overview of bandwidth usage for each service type.

C1. Results



A1 | customer_id

	A	B	C	D
1	customer_	avg_bandwidth_dsl	avg_bandwidth_fiber_optic	
2	A04204	1080.597617	NULL	
3	A05946	NULL	5104.889837	
4	A166320	NULL	6173.702221	
5	A223123	NULL	4140.83823	
6	A224472	NULL	NULL	
7	A22477	NULL	601.2192303	
8	A230205	NULL	1031.509395	
9	A235822	NULL	5203.543626	
10	A236354	1717.407155	NULL	
11	A241767	NULL	3726.308369	
12	A25662	NULL	NULL	
13	A29486	1967.928007	NULL	
14	A295468	NULL	NULL	
15	A304266	NULL	NULL	
16	A313674	NULL	NULL	
17	A349493	NULL	NULL	
18	A357886	NULL	1096.780158	

in

Workbook Statistics

	A	B	C
1	avg_bandwidth_dsl	avg_bandwidth_fiber_optic	
2	3656.300284	3210.448305	
3			
4			
5			
6			
7			
8			
9			
10			
11			

From the result, it is evident that the total average bandwidth usage of customers on 12 months contract using DSL 3656.3 gb per year is more than those customers using fiber optics 3210.45 gb per year.

D. File Refreshment/Update Interval

The recommended refresh interval for the data to be acquired and refreshed in the database for it to remain relevant to the business is every month.

D1. This monthly update is crucial because bandwidth usage and customer preferences for internet services can change quickly due to technological advancements, shifts in customer requirements, and evolving market competition. Regular monthly updates are also important to accurately reflect any changes in contract specifics and service offerings, like new promotions or upgrades. Such a refresh schedule ensures that the analysis remains current, enabling the company to make well-informed strategic decisions and promptly adjust to changing customer trends and market conditions.

E. Panopto Video

The Panopto video for this project is located in the Student Creators folder.

F. Sources

This project and its report were produced independently, without the use of any external sources or web-based materials.