Name: Ahsan Zaman
Subject: D191 – Advanced Data Management
Student ID # 002765859
Title: Performance Assessment

A. Summarize one real-world business report that can be created from the attached Data Sets and Associated Dictionaries.

Answer: One real-world benefit that can be extracted from this data, is number of payments made from a particular zip code. This allows us to see the most profitable regions and calculate total revenue for the zip codes. Knowing this will let us concentrate advertisements appropriately for those regions versus the ones that have done poorly in sales.

1. Describe the data used for the report.

Answer: The data used for this report will be each payment made by customer and the address information associated with that customer. The address information will be used to get the postal code/zip code for that payment. Additionally, we will use city information to get a clearer picture of where the district is located.

2. Identify two or more specific tables from the given dataset that will provide the data necessary for the detailed and the summary sections of the report.

Answer: Tables used will be payment, customer, address and city.

3. Identify the specific fields that will be included in the detailed and the summary sections of the report.

Answer: The Detailed and Summary sections will contain the following fields.

Detailed Section: payment_id, customer_id, amount, store_id, postal_code, district, city

Summary Section: postal_code, region, payment count, total payments

4. Identify one field in the detailed section that will require a custom transformation and explain why it should be transformed. For example, you might translate a field with a value of 'N' to 'No' and 'Y' to 'Yes'.

Answer: City and district will be appended in the summary section to specify region. Even though the postal codes are there for specific region information, it is easier to read through the report with the district and city specified.

5. Explain the different business uses of the detailed and the summary sections of the report.

Answer: Detailed section will allow us to see all the individual payments made according to districts and postal codes. Store id is also included so that we can see which stores are performing well. Summary section will show the number of payments made by customers in a zip code and region. This will allow us to understand which regions have the most activity and are generating the most revenue. Once the ETL process will be automated with a stored procedure, we can generate print the Detailed and Summary table at the start of each month to see regions by popularity. If there is a program for

rewarding stores for performance in the future, this would allow us to reward the regions that have been doing well.

6. Explain how frequently your report should be refreshed to remain relevant to stakeholders.

Answer: This report should be refreshed each month. Most companies update financial data on a quarterly basis but for individual stores it is better to gather this data monthly because this would let us track payment and revenue trends as they correlate to events in that particular month.

B. Write a SQL code that creates the tables to hold your report sections.

Answer: /* Creating detailed table */

```
CREATE Table Detailed (
        Payment_id INT,
        Customer_id SMALLINT,
        Amount NUMERIC(5,2),
        Store_id SMALLINT,
        District VARCHAR(20),
        Postal_code VARCHAR(10),
        City VARCHAR(50)
);
```

/* Creating summary table*/

```
 CREATE TABLE Summary(
        Postal_code VARCHAR(10),
        Region VARCHAR(80),
        Payments_count INT,
        Total_payments NUMERIC(8, 2)
);
```

C. Write a SQL query that will extract the raw data needed for the Detailed section of your report from the source database and verify the data's accuracy.

Answer:

/* Inserting data into detailed from payment, customer, address and city*/

```
INSERT INTO Detailed (
        Payment_id,
```

```
        Customer_id,
        Amount,
        Store_id,
        District,
        Postal_code,
        city
)
SELECT
        p.payment_id AS “Payment ID”,
        c.customer_id AS “Customer ID”,
        p.amount AS amount,
        c.store_id AS “Store ID”,
        a.district AS district,
        a.postal_code AS postal_code,
        ci.city AS city
FROM payment p
INNER JOIN customer c ON c.customer_id = p.customer_id
INNER JOIN address a ON c.address_id = a.address_id
INNER JOIN city ci ON a.city_id = ci.city_id;
```

D. Write code for function(s) that perform the transformation(s) you identified in part A4.

Answer: /*Creating sales_summary_update() function for concatenating district and city and updates summary table with all the required data from detailed table*/

```
CREATE FUNCTION sales_summary_update()
RETURNS TRIGGER
LANGUAGE plpgsql
AS $$
BEGIN
/* Emptying table */
```

```
DELETE FROM Summary;

/* Inserting new data into summary from detailed with transformation on district and city */

INSERT INTO Summary(

        SELECT

                Postal_code,

                Concat_ws(' ', district, city) AS "Region",

                COUNT(payment_id) AS "payments_count",

                SUM(amount) as total_amount

        FROM Detailed

        GROUP BY postal_code

        ORDER BY total_amount DESC

);

RETURN NEW;

END; $$
```

E. Write a SQL code that creates a trigger on the detailed table of the report that will continually update the summary table as data is added to the detailed table.

Answer: /*Creating Trigger called summary_update which will be triggered when new data is inserted into detailed table. */

```
CREATE TRIGGER summary_update

AFTER INSERT ON Detailed

FOR EACH STATEMENT

EXECUTE PROCEDURE sales_summary_update()
```

F. Create a stored procedure that can be used to refresh the data in both your detailed and summary tables. The procedure should clear the contents of the detailed and summary tables and perform the ETL load process from part C and include comments that identify how often the stored procedure should be executed.

Answer: The stored procedure below can refresh the data in both detailed and summary tables. It will first empty the Detailed table and then insert new data into detailed table. The trigger that we created earlier will refresh the summary table since it activates on the insert into detailed. As shown at the end of the code, this procedure can be executed with the CALL keyword. The tables can then be displayed with a simple select statement.

```
/* creating a store procedure called update_proc() */
CREATE PROCEDURE update_proc()

LANGUAGE plpgsql

AS $$

BEGIN

/* Emptying detailed table */

DELETE FROM Detailed;

/* Inserting new data into the detailed table*/

INSERT INTO Detailed(
        Payment_id,
        Customer_id,
        Amount,
        Store_id,
        District,
        Postal_code,
        city
)
SELECT
        p.payment_id AS “Payment ID”,
        c.customer_id AS “Customer ID”,
        p.amount AS amount,
        c.store_id AS “Store ID”,
        a.district AS district,
        a.postal_code AS postal_code,
        ci.city AS city
FROM payment p
INNER JOIN customer c ON c.customer_id = p.customer_id
INNER JOIN address a ON c.address_id = a.address_id
INNER JOIN city ci ON a.city_id = ci.city_id;
```

```
END;$$

CALL update_proc(); /* calling the stored procedure */
```

1. Explain how the stored procedure can be run on a schedule to ensure data freshness.

Answer: Data can be refreshed automatically at the start or end of each month by scheduling a job within PgAgent. There is also the option of using pg_cron to schedule the job and make the call to the stored procedure on a set time/date but PgAgent's GUI makes it a bit easier to schedule this job.

G. 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.

Answer: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=3e58ceed-cb71-4e9d-8a7c-ae8e0127299d

H. Record the web sources you used to acquire data or segments of third-party code to support the application if applicable. Be sure the web sources are reliable.

Answer: None used.

I. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

Answer: 1. https://protect-us.mimecast.com/s/XOKNCv2xEotEOAojxHQlJRo?domain=severalnines.com

2. https://protect-us.mimecast.com/s/C_kjCxkzJvtLmxK4kHY_mjA?domain=stitchdata.com/

3. https://www.postgresqltutorial.com/postgresql-triggers/