



CS/CE 457/464 Data Science Fall 2024

Homework 4

Submitted by:
Muhammad Abdullah

Student ID:
ma07534

September 2024

Contents

1	Problem 1
---	-----------

2

1 Problem 1

Part 1 (70 points):

Deliverables: You can either include screenshots of your pgAdmin screen showing query and output table or copy-paste your queries and output table for your answers and submit a PDF version.

Instructions

1. Create SQL database tables for both datasets `countries.csv` and `cities.csv` separately using an RDBMS (PostgreSQL preferred). Submit the `CREATE TABLE` query in the final document.

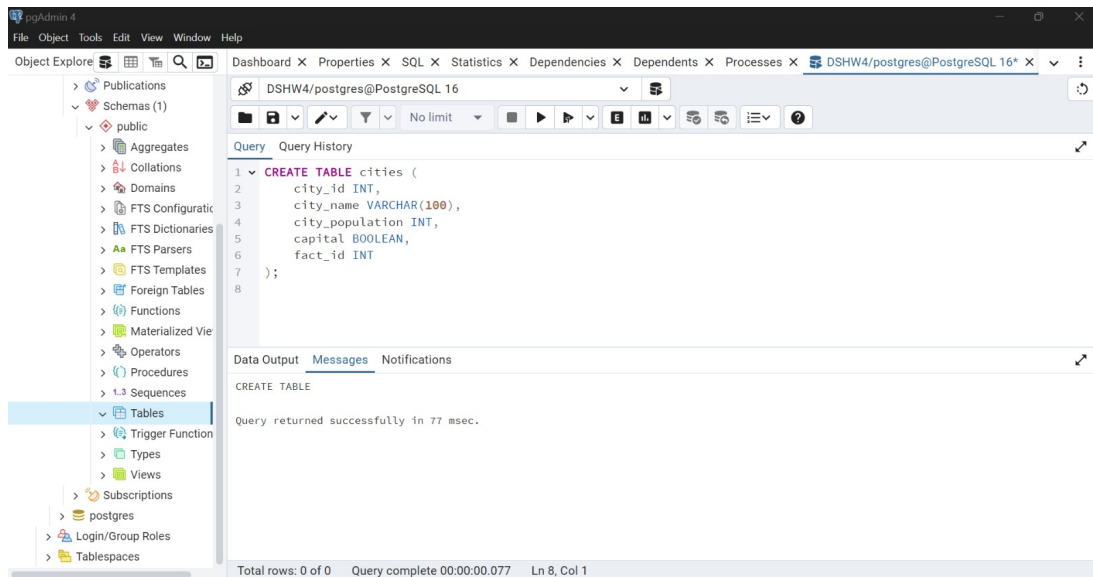


Figure 1: Create City Table Query

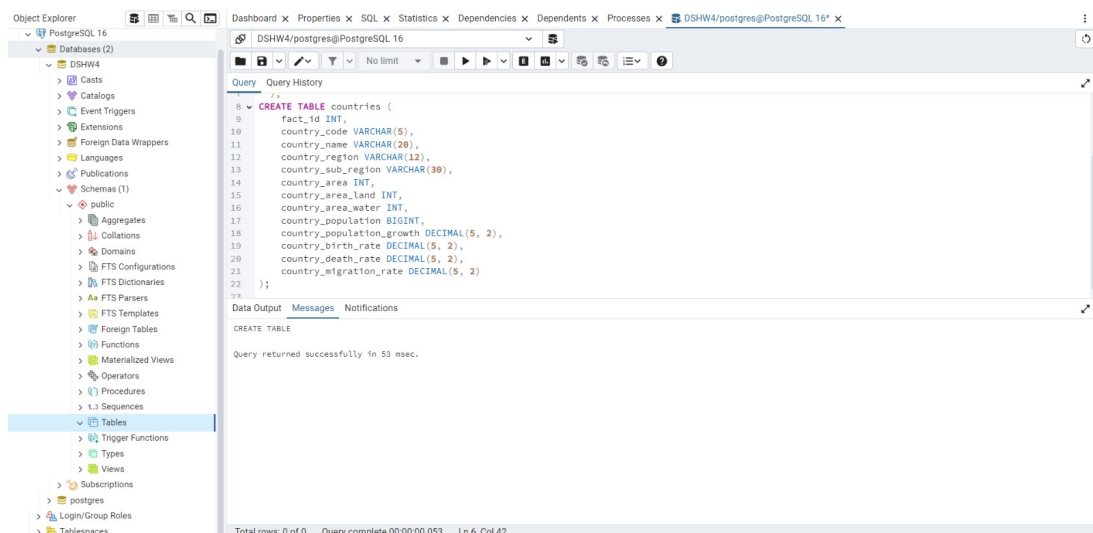


Figure 2: Create Country Table Query

2. Load/Import the dataset into the table.
3. Query the database tables and interpret the results:

A note worth mentioning here is that I initially created the tables using these queries, however I failed to import the CSV files into the tables. In the process of debugging I changed the size of VARCHARS, increasing them in order to clear out any possible errors due to size limits. Even though this did not help fix my error, I had to import the data using CMD.

TLDR; some of my VARCHAR sizes are greater than the ones in the screenshot attached.

1. Count of Total Number of Records in Each Table



Figure 3: City Table Count Records Query

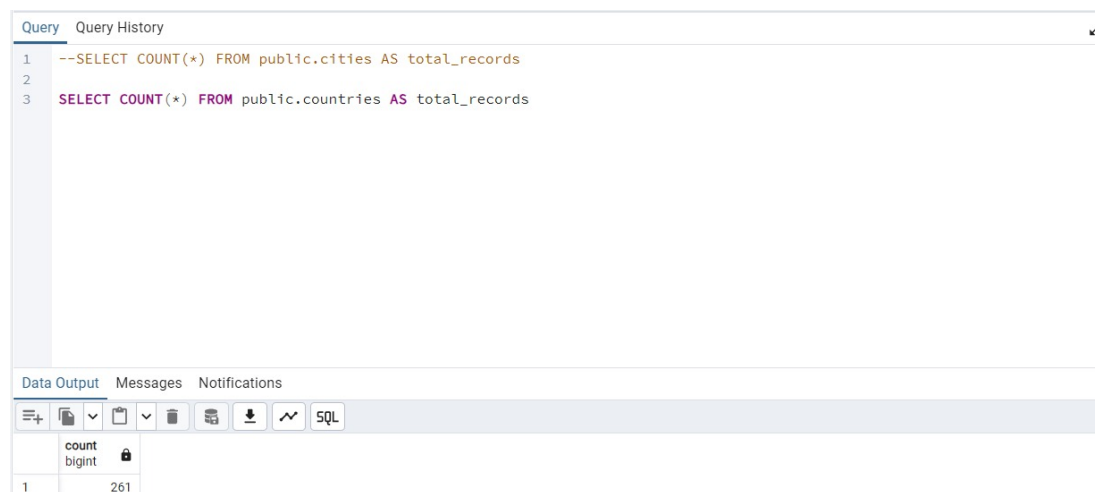


Figure 4: Country Table Count Records Query

2. Count of Number of Cities for Each Country in Descending Order (use GROUP BY)

Query		Query History
		2
		3 --SELECT COUNT(*) FROM public.countries AS total_records
		4
		5 SELECT
		6 countries.country_name AS Country,
		7 COUNT(cities.city_name) AS City_Count
		8 FROM
		9 countries
		10 JOIN
		11 cities ON countries.fact_id = cities.fact_id
		12 GROUP BY
		13 countries.fact_id, countries.country_name
		14 ORDER BY
		15 City_Count DESC;
		16
		17

Data Output		Messages	Graph Visualiser	×	Notifications
		SQL			
	country	city_count			
	character varying (100)	bigint			
1	South Africa	7			
2	Indonesia	6			
3	Turkey	6			
4	Iran	6			
5	Japan	6			
6	United States	6			
7	United Kingdom	6			
8	Brazil	6			
9	Australia	6			
10	India	6			

Figure 5: Number of Cities for Each Country in Descending Order Query

Data Output		Messages	Graph Visualiser	×	Notifications
		SQL			
	country	city_count			
	character varying (100)	bigint			
1	South Africa	7			
2	Indonesia	6			
3	Turkey	6			
4	Iran	6			
5	Japan	6			
6	United States	6			
7	United Kingdom	6			
8	Brazil	6			
9	Australia	6			
10	India	6			
11	Colombia	6			
12	Canada	6			
13	China	6			
14	Venezuela	6			
15	Ukraine	6			
16	Russia	6			
17	Iraq	6			
18	Morocco	6			
19	Korea, South	6			
20	Nigeria	6			
21	Pakistan	6			
22	Mexico	6			
23	France	6			
24	Italy	6			
Total rows: 210 of 210		Query complete 00:00:00.126	Ln 16, Col 1		

Figure 6: Number of Cities for Each Country in Descending Order Query Result

3. Count of Regions and Sub-Regions in Each Country (Sorted by Country Name)

Query Query History

```
14 --ORDER BY
15 --    City_Count DESC;
16
17 SELECT
18     countries.country_name AS Country,
19     COUNT(countries.country_region) AS Region_Count,
20     COUNT(countries.country_sub_region) AS Sub_Region_Count
21 FROM
22     countries
23 GROUP BY
24     countries.fact_id, countries.country_name
25 ORDER BY
26     countries.country_name ASC;
27
28
29
```

Data Output Messages Graph Visualiser X Notifications

	country	region_count	sub_region_count
	character varying (100)	bigint	bigint
1	Afghanistan	1	1
2	Akrotiri	1	1
3	Albania	1	1
4	Algeria	1	1
5	American Samoa	1	1
6	Andorra	1	1
7	Angola	1	1
8	Anguilla	1	1
9	Antarctica	1	1
10	Antigua and Barbuda	1	1

Total rows: 261 of 261 Query complete 00:00:00.103 Ln 29, Col 1

Figure 7: Count of Regions and Sub-Regions in Each Country (Sorted by Country Name) Query

4. Top 10 Most Populous Capital Cities

Dashboard x Properties x SQL x Statistics x Dependencies x Dependents x Processes x DSHW4/postgres@PostgreSQL 16 x

DSHW4/postgres@PostgreSQL 16

Query Query History

```
1 SELECT
2     countries.country_name AS Country,
3     cities.city_name AS City,
4     cities.city_population AS Population
5 FROM
6     countries
7 JOIN
8     cities ON countries.fact_id = cities.fact_id
9 WHERE
10     cities.capital = TRUE
11 ORDER BY
12     cities.city_population DESC
13 LIMIT 10;
14
```

Data Output Messages Notifications

country	city	population
character varying (100)	character varying (255)	bigint
1 Japan	Tokyo	37217000
2 India	New Delhi	22654000
3 Mexico	Mexico City	20446000
4 China	Beijing	15594000
5 Bangladesh	Dhaka	15391000
6 Argentina	Buenos Aires	13528000
7 Philippines	Manila	11862000
8 Russia	Moscow	11621000
9 Egypt	Cairo	11149000
10 Indonesia	Jakarta	9769000

Figure 8: Top 10 Most Populous Capital Cities Query & Result

5. Average City Population of Capital and Non-Capital Cities (use GROUP BY)

Query QueryHistory	
1	SELECT
2	cities.capital AS capital,
3	AVG(cities.city_population) AS avg_population
4	FROM
5	cities
6	GROUP BY
7	cities.capital
8	
9	

Data Output Messages Notifications	
capital boolean	avg_population numeric
1 false	2891392.063492063492
2 true	2140442.307692307692

Figure 9: Average City Population of Capital and Non-Capital Cities Query & Result

6. Average Country Birth Rate for Each Region and Sub-Region (use GROUP BY)

Query Query History

1 SELECT

2 countries.country_region AS region,

3 countries.country_sub_region AS subregion,

4 AVG(countries.country_birth_rate) AS avg_birthrate

5 FROM

6 countries

7 GROUP BY

8 countries.country_region, countries.country_sub_region

9

10

Data Output Messages Notifications

SQL

region character varying (50)	subregion character varying (100)	avg_birthrate numeric
1 Europe	Northern Europe	11.6621428571428571
2 Asia	Southern Asia	21.0944444444444444
3 Oceania	Micronesia	19.7000000000000000
4 Africa	Sub-Saharan Africa	32.4055813953488872
5 Americas	Northern America	11.2000000000000000
6 Europe	Western Europe	10.1633333333333333
7 NA	NA	9.2364583333333333
8 Americas	Latin America and the Caribbean	15.8915384615384615
9 Europe	Southern Europe	9.8693333333333333
10 Asia	South-eastern Asia	19.8177777777777778
11 Europe	Eastern Europe	10.1520000000000000
12 Asia	Central Asia	20.5820000000000000
13 Oceania	Melanesia	21.9900000000000000
14 Oceania	Australia and New Zealand	4.2466666666666667
15 Asia	Western Asia	18.503204117647059
16 Africa	Northern Africa	22.4957142857142857
17 Oceania	Polynesia	14.8333333333333333
18 Asia	Eastern Asia	11.6740000000000000

Figure 10: Average Country Birth Rate for Each Region and Sub-Region Query & Result

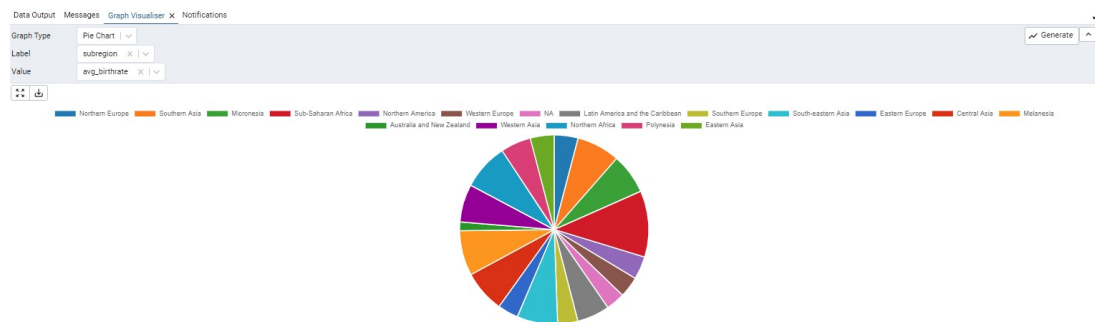


Figure 11: Average Country Birth Rate for Each Region and Sub-Region Pie Graph