

HW 4: SQL

Syed Nisar Hussain

sh07126@st.habib.edu.pk

Part 1 (RDBMS used: Microsoft SQL Server)

- Create table queries:

```
CREATE TABLE countries (  
    fact_id INT IDENTITY(1,1) PRIMARY KEY,  
    country_code VARCHAR(3) NOT NULL,  
    country_name VARCHAR(100) NOT NULL,  
    country_region VARCHAR(50),  
    country_sub_region VARCHAR(50),  
    country_area BIGINT,  
    country_area_land BIGINT,  
    country_area_water BIGINT,  
    country_population BIGINT,  
    country_population_growth DECIMAL(5, 2),  
    country_birth_rate DECIMAL(5, 2),  
    country_death_rate DECIMAL(5, 2),  
    country_migration_rate DECIMAL(5, 2)  
);  
  
CREATE TABLE cities (  
    city_id INT IDENTITY(1,1) PRIMARY KEY,  
    city_name VARCHAR(100) NOT NULL,  
    city_population BIGINT,  
    capital BIT,  
    fact_id INT,  
    FOREIGN KEY (fact_id) REFERENCES countries(fact_id)  
);
```

- Count of total number of records in each table

SQLQuery2.sql - T...4\NISAR SYED (54))*

```
SELECT COUNT(*) FROM countries;
SELECT COUNT(*) FROM cities;
```

1 (No column name) 261

1 (No column name) 397

100 %

Results Messages

Query executed successfully.

Interpretation: Matches the total number of records in the original csv files which shows that the data was successfully loaded

- Count of number of cities for each country in descending order of count

SQLQuery2.sql - T...4\NISAR SYED (54))*

```
SELECT c.country_name, COUNT(ci.city_id) AS city_count
FROM countries c
JOIN cities ci ON c.fact_id = ci.fact_id
GROUP BY c.country_name
ORDER BY city_count DESC;
```

100 %

Results Messages

	country_name	city_count
1	South Africa	7
2	Turkey	6
3	United Kingdom	6
4	United States	6
5	Venezuela	6
6	Ukraine	6
7	Argentina	6
8	Australia	6
9	Brazil	6
10	Canada	6
11	China	6
12	Colombia	6
13	France	6
14	India	6

Query executed successfully.

Interpretation: South Africa has the most number of cities in this dataset (with 7 cities), however in second place there are multiple countries which have 6 cities each.

- Count of regions and sub-regions in each country. Sort them by ascending order of country name

SQLQuery2.sql - T...4\NISAR SYED (54))*

```
SELECT country_name,
COUNT(DISTINCT country_region) AS region_count,
COUNT(DISTINCT country_sub_region) AS sub_region_count
FROM countries
GROUP BY country_name
ORDER BY country_name ASC;
```

100 %

Results Messages

	country_name	region_count	sub_region_count
1	Afghanistan	1	1
2	Akrotiri	0	0
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	0	0
10	Antigua and Barbuda	1	1
11	Arctic Ocean	0	0
12	Argentina	1	1
13	Armenia	1	1
14	Aruba	1	1

Query executed successfully.

Interpretation: We can observe that most countries are usually associated to one region and one particular sub region within that region. This makes sense because it wouldn't be

logical for a country to belong to different regions of the world. There are some countries which don't have any particular region defined in this dataset (NA in the dataset), they are listed with 0 as their count here.

- Top 10 most populous capital cities. Display country, city and population in descending order

SQLQuery2.sql - T...4\NISAR SYED (54))*

```
SELECT TOP 10 co.country_name, ci.city_name, ci.city_population
FROM cities ci
JOIN countries co ON ci.fact_id = co.fact_id
WHERE ci.capital = 1
ORDER BY ci.city_population DESC;
```

100 %

Results Messages

	country_name	city_name	city_population
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	11169000
10	Indonesia	Jakarta	9769000

Query executed successfully.

Interpretation: Tokyo is the capital city with the highest population and by a huge margin. Second and third place is covered by New Delhi and New Mexico respectively which aren't separated as much as Tokyo is from the rest. It can be observed that most of these capital cities are from Asia.

- Average city population of capital and non-capital cities

SQLQuery2.sql - T...4\NISAR SYED (54))*

```
SELECT capital, AVG(city_population) AS avg_city_population
FROM cities
GROUP BY capital;
```

100 %

Results Messages

	capital	avg_city_population
1	0	2891392
2	1	2140442

Query executed successfully.

Interpretation: Capital cities are more populated on average and the data here supports that. Capital cities tend to be more developed and attract more people from all over the country and can explain why they are more populated than non-capital cities.

- Average country birth rate for each region and sub-region

SQLQuery2.sql - T...4\NISAR SYED (54))*		
<pre> SELECT country_region, country_sub_region, AVG(country_birth_rate) AS avg_birth_rate FROM countries GROUP BY country_region, country_sub_region ORDER BY avg_birth_rate DESC; </pre>		
100 %		
Results Messages		
country_region	country_sub_region	avg_birth_rate
1 Africa	Sub-Saharan Africa	32.4055814521257
2 Africa	Northern Africa	22.6957144056048
3 Oceania	Melanesia	21.9900001525879
4 Asia	Southern Asia	21.0944442749023
5 Asia	Central Asia	20.5819995880127
6 Asia	South-eastern Asia	19.817777633667
7 Oceania	Micronesia	19.6999999682109
8 Asia	Western Asia	18.503529436448
9 Americas	Latin America and the Caribbean	15.8915383999164
10 Oceania	Polynesia	14.8333333333333
11 Asia	Eastern Asia	11.6739998817444
12 Europe	Northern Europe	11.6621430260795
13 Americas	Northern America	11.1999998092651
14 Europe	Western Europe	10.1633333100213
Query executed successfully.		

Interpretation: Sub-Saharan Africa has an extremely high birth rate, much more than Northern Africa on the second place. A lot of Asian regions are also in this list such as Southern Asia, Central Asia etc. These continents are already affiliated with having very high populations and this data backs that since high birth rates lead to higher populations. In comparison, birth rate in Western Europe is twice as less as the Asian regions listed above.