LISA VARGHESE
2037041

ASSIGNMENT-1

SQL QUERIES

#1

--Create a query to extract all data from the first 30 rows of the dataset

```
SELECT * FROM `bigquery-public-data.ml_datasets.census_adult_income` LIMIT 30
```

#2

--Create a query to extract occupations of people aged 30 or above but aged less than 45 years from the first 30 rows of the dataset

```
SELECT
occupation
FROM bigquery-public-data.ml_datasets.census_adult_income
WHERE
age>=30 AND age<45
LIMIT 30
```

#3

--Create a query to get the minimum and maximum number of years of education from the entire dataset

```
SELECT
min(education_num) as minimum_education,
MAX(education_num) as maximum_education
FROM bigquery-public-data.ml_datasets.census_adult_income
```

#4

--Create a query to order the first 50 rows of occupations by descending order of hours per week.

```
SELECT
occupation
FROM bigquery-public-data.ml_datasets.census_adult_income
ORDER BY hours_per_week desc
LIMIT 50
```

#5

--Create a query to order the first 20 rows of occupation and race by hours per week.

```
SELECT
occupation,race
FROM bigquery-public-data.ml_datasets.census_adult_income
ORDER BY hours_per_week
LIMIT 20
```

#6

LISA VARGHESE
2037041

--Create a query to find average capital gain for the first 50 rows of the dataset

```
SELECT
AVG(capital_gain)
FROM bigquery-public-data.ml_datasets.census_adult_income
LIMIT 50
```

#7

--Create a query to extract information about workclass, occupation and marital status for people aged 30 or more but aged below 45 years for the first 50 rows of the dataset.

```
SELECT
age>=30 AND age<45,
workclass,occupation,marital_status
FROM bigquery-public-data.ml_datasets.census_adult_income
LIMIT 50
```

JOINS

#8

-- Create a query to get population by state by joining the population and state code from the census table and zip code table by using the unique key, i.e. zip code. Query to pull only rows where minimum age, maximum age and gender are null since we only want total population summary, grouped by state code.

```
SELECT
zip.state_code,
sum(census.population)population
FROM `bigquery-public-data.census_bureau_usa.population_by_zip_2010`census
LEFT JOIN
`bigquery-public-data.geo_us_boundaries.zip_codes`zip
ON
(census.zipcode=zip.zip_code)
WHERE census.minimum_age is null
AND census.maximum_age is null
AND census.gender is null
GROUP BY state_code
```

#9

-- Create a query to fetch data from two tables, namely github timeline and github nested about repository_url, repository_created_at, repository_has_issues and repository_description, by using the common repository url in both tables for the first 20 rows.

```
SELECT repository_url, repository_created_at, repository_has_issues, repository_description,
FROM `bigquery-public-data.samples.github_timeline` as Time
JOIN `bigquery-public-data.samples.github_nested` as Nest
ON repository_url=repository.url
LIMIT 20
```

LISA VARGHESE
2037041

#10

--Create a query to match the first 30 rows of indicator name with indicator code in the international debt table using the country code in the country debt table.

```
SELECT indicator_name,indicator_code
FROM `bigquery-public-data.world_bank_intl_debt.international_debt` as international
JOIN `bigquery-public-data.world_bank_intl_debt.country_summary` as country
ON country.country_code= international.country_code
LIMIT 30
```