Name: Siddarth Raju Brahmandaberi Email Id: sidhusiddarth0@gmail.com

Batch: Apr 15, 2021 09:30 PM

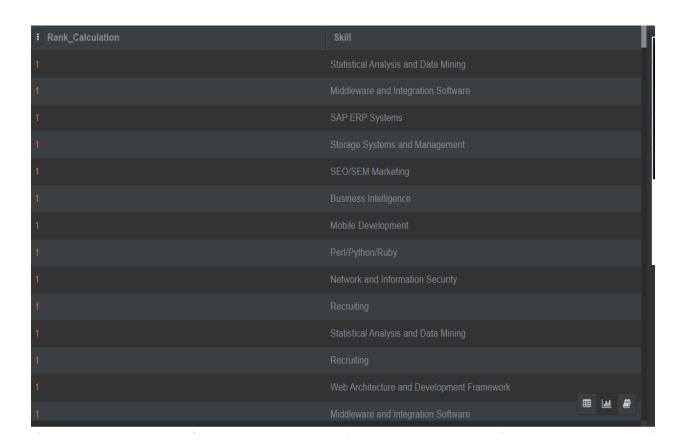
SQL(Major project)

Data used: Linkedin-skills

Q1. What is the skill which is ranked 1St?

Querie used:

• SELECT rank_calculation, skill FROM Linkedin_Skills WHERE rank_calculation = 1;

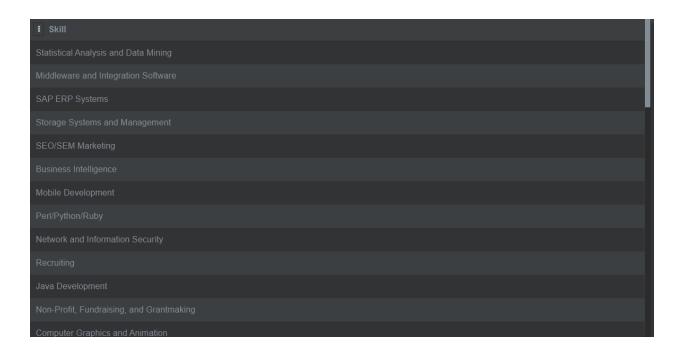


Q2. What are the unique skills in the dataset?

Querie used:

• SELECT DISTINCT(skill) FROM Linkedin_Skills;

Output:



Q3. How many unique number of skills are present?

Querie used:

• SELECT COUNT(DISTINCT(skill)) FROM Linkedin_Skills;



Q4. Find total no of skills and with no of 0 & 1 rank calculated skills?

Queries used:

- SELECT COUNT(skill) FROM Linkedin_Skills;
- SELECT COUNT(skill) FROM Linkedin_Skills WHERE rank_calculation = 1;
- SELECT count(skill) FROM Linkedin_Skills WHERE rank_calculation == 0;

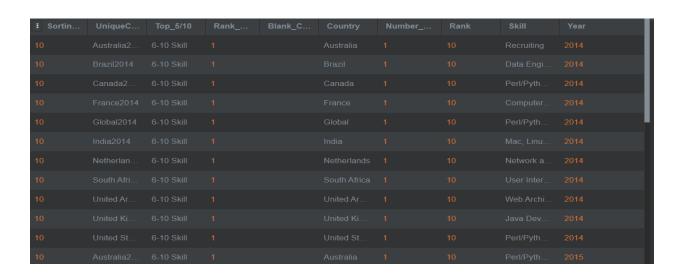
Output:



Q5. Set of data whose sorting_calculation is 10?

Querie used:

SELECT * FROM Linkedin_Skills WHERE sorting_calculation = 10;



Data used: World cup

Q6. Where and in which year, team highest goal was scored?

Queries used:

- SELECT max(goals) FROM WorldCup;
- SELECT year, stadium, home, team, goals FROM WorldCup WHERE goals = (SELECT max(goals) FROM WorldCup);

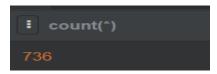
Output:

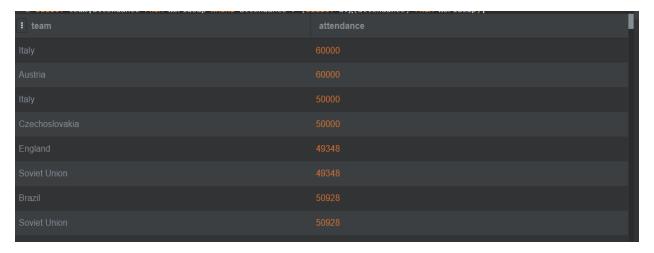


Q7. How many teams have more than average attendance, what are they?

Queries used:

- SELECT avg(attendance) FROM WorldCup;
- SELECT team, attendance FROM WorldCup WHERE attendance > (SELECT avg(attendance) FROM WorldCup);
- SELECT count(*) FROM WorldCup WHERE attendance > (SELECT avg(attendance) FROM WorldCup);



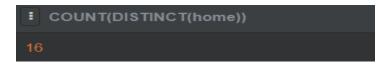


Q8. How many unique home countries are their and what are they?

Queries used:

- SELECT DISTINCT(home) FROM WorldCup;
- SELECT COUNT(DISTINCT(home)) FROM WorldCup;

Output:





Q9. Which teams reached max stage?

Queries used:

- SELECT max(stage) FROM WorldCup;
- SELECT team,stage FROM WorldCup WHERE stage = (SELECT max(stage) FROM WorldCup);



Q10. How many teams played?

Querie used:

SELECT COUNT(team) FROM WorldCup;

Output:



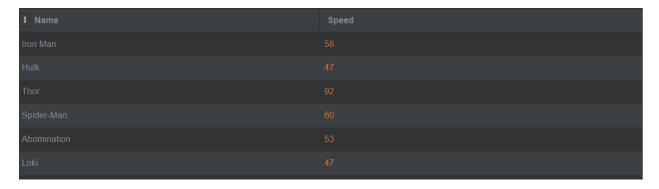
Data used: Super Hero

Q11. Name the Super heroes who have more than average speed?

Queries used:

- SELECT avg(speed) FROM superhero;
- SELECT name, speed from superhero where speed > (SELECT avg(speed) FROM superhero);

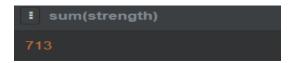
Output:



Q12. How much is the combined strength of all super heroes?

Querie used:

SELECT sum(strength) FROM superhero;



Q13. How many factors of dataset consists alignment as good and what are they?

Queries used:

- SELECT COUNT(*) FROM superhero WHERE alignment = 'good';
- SELECT * FROM superhero WHERE alignment = 'good';

Output:



■ Name	Alignment	Intelligence	Strength	Speed	Durability	Power	Combat
Iron Man							
Captain Ameri							
Hulk							
Hawkeye							
Thor							
Spider-Man							

Q14. Whose combat is more than average durability?

Queries used:

- SELECT avg(durability) FROM superhero;
- SELECT name,combat FROM superhero WHERE combat < (SELECT avg(durability) FROM superhero);

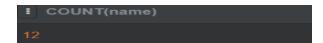
Output:



Q15. How many super heroes are there in dataset?

Querie used:

SELECT COUNT(name) FROM superhero;



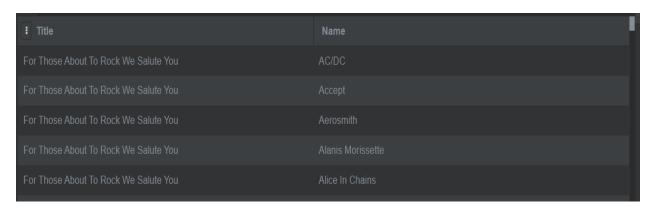
Data used: Chinook-Data

Q16. List the songs done by artists?

Querie used:

• SELECT title,name FROM albums INNER JOIN artists;

Output:



Q17. List the album, artist and his song?

Queries used:

- CREATE TABLE sss as SELECT title,name,albumid FROM albums INNER JOIN artists;
- SELECT sss.title as 'Album', sss.name as 'Artist', tracks.name as 'Songs' FROM sss INNER JOIN tracks on sss.albumid = tracks.albumid;

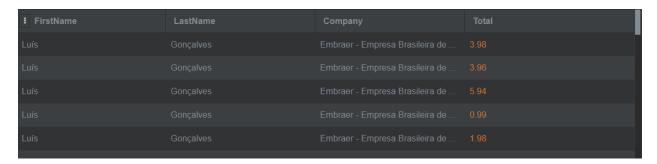
Album	Artist	Songs
For Those About To Rock We Salute You	AC/DC	For Those About To Rock (We Salute You)
For Those About To Rock We Salute You	AC/DC	Put The Finger On You
For Those About To Rock We Salute You	AC/DC	Let's Get It Up
For Those About To Rock We Salute You	AC/DC	Inject The Venom
For Those About To Rock We Salute You	AC/DC	Snowballed
For Those About To Rock We Salute You	AC/DC	Evil Walks
For Those About To Rock We Salute You	AC/DC	C.O.D.
For Those About To Rock We Salute You	AC/DC	

Q18. List the customers and their total bill from company?

Querie used:

 SELECT firstname,lastname,company,total FROM customers INNER JOIN invoices ON customers.CustomerId = invoices.CustomerId;

Output:



Q19. List the track id of each customer?

Queries used:

- CREATE TABLE business as SELECT * FROM customers INNER JOIN invoices ON customers.CustomerId = invoices.CustomerId;
- select firstname,lastname,company,trackid FROM business INNER JOIN invoice_items ON business.invoiceid = invoice_items.Invoiceld;

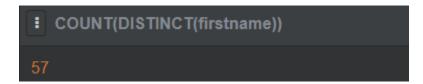
Output:



Q20. What is the count of different customers?

Querie used:

SELECT COUNT(DISTINCT(firstname)) FROM customers;



Data used: Diabetes

Data URL: https://www.kaggle.com/mathchi/diabetes-data-set

Q21. What is the total members in dataset, specify no of diabetic and non-diabetic?

Queries used:

- SELECT COUNT(*) FROM Diabetes WHERE outcome = 1;
- SELECT COUNT(*) FROM Diabetes WHERE outcome = 0;
- SELECT COUNT(*) FROM Diabetes;

Output:



Q22. How many members have less than average glucose?

Queries used:

- SELECT avg(glucose) FROM Diabetes;
- SELECT COUNT(*) FROM Diabetes WHERE glucose < (SELECT avg(glucose) FROM Diabetes);



Q23. List the person's report who has maximum BP?

Queries used:

- SELECT max(bloodpressure) FROM Diabetes;
- SELECT * FROM Diabetes WHERE bloodpressure = (SELECT max(bloodpressure) FROM Diabetes);

Output:

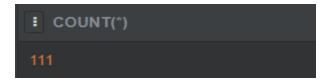


Q24. Women with no pregnancy?

Querie used:

SELECT COUNT(*) FROM Diabetes WHERE pregnancies = 0;

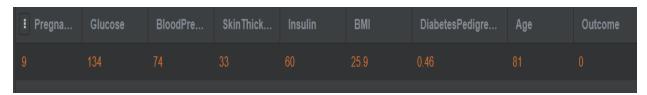
Output:



Q25. What is the maximum age in the dataset, How many members have it along with their report?

Queries used:

- SELECT max(age) FROM Diabetes;
- SELECT COUNT(*) FROM Diabetes WHERE age = (SELECT max(age) FROM Diabetes);
- SELECT * FROM Diabetes WHERE age = (SELECT max(age) FROM Diabetes);



Data used: Covid19 India

Data URL: https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_19_india.csv

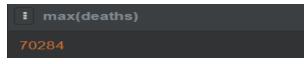
Q26. What are deaths in each state and maximum death?

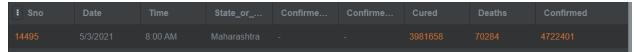
Queries used:

- SELECT state_or_unionterritory, deaths FROM Covid19_India;
- SELECT max(deaths) FROM Covid19_India;
- SELECT * FROM Covid19_India WHERE deaths = (SELECT max(deaths) FROM Covid19_India);

Output:







Q27. What is the total deaths and confirmed cases?

Querie used:

SELECT sum(deaths), sum(confirmedindiannational), sum(confirmedforeignnational)
FROM Covid19_India;

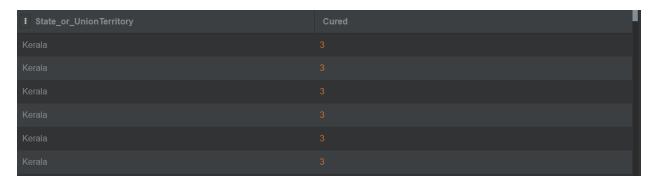
: sum(deaths)	sum(confirmedindiannational)	sum(confirmedforeignnational)
37013014	5436	667

Q28. What are the cured cases for each state?

Querie used:

SELECT state_or_unionterritory, cured FROM Covid19_India WHERE cured > 0;

Output:

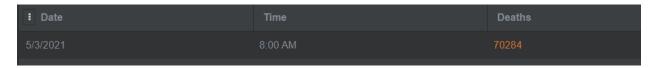


Q29. What is the date and time when maxium deaths occured?

Queries used:

- SELECT max(deaths) FROM Covid19_India;
- SELECT date, time, deaths FROM Covid19_India WHERE deaths = (SELECT max(deaths) FROM Covid19_India);

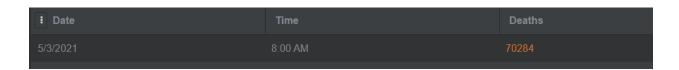
Output:



Q30. What is maximum cured no of cases and which state done it?

Queries used:

- SELECT max(cured) FROM Covid19_India;
- SELECT state_or_unionterritory,cured FROM Covid19_India WHERE cured = (SELECT max(cured) FROM Covid19_India);



Observation:

The dataset used in Q1-Q5 is Linkedin_skills.

It contains 800 rows and 10 columns.

- Q1. Skills with rank 1
- Q2. Unique skills without repetition by using 'DISTINCT'
- Q3. The count of number of unique skills by using 'COUNT' with Q2
- Q4. Count of total skills in dataset = 800 Count of skills with rank 1 = 410 Count of skills with rank 0 = 300
- Q5. Data with sorting_calculation = 10

The dataset used in Q6-Q10 is World-cup.

It contains 1672 rows and 19 columns.

- Q6. Teams which scored highest maximum goals, first the highest goals were found by using 'max' = 9, then the teams along with max score.
- Q7. Teams with more than avg attendance, First avg attendance is found using 'avg' = 44651.4019138756 then the teams were couted using 'COUNT'
- Q8. Unique home countries
- Q9. Teams which reached max stages.
- Q10. Total number of teams took part = 1672.

The dataset used in Q11-Q15 is Super-Hero.

It contains 13 rows and 8 columns.

Q11. Heroes with more than avg speed (44.0833333)

- Q12. Combined strength of all heroes using 'sum'
- Q13. Count and the data that is alignment good
- Q14. Heroes with more than avg durability(69.66666)
- Q15. Total number of super heroes.

The dataset used in Q16-Q20 is Chinook-Data.

It contains various tables.

- Q16. Joined the tables and got the Title and Artist name out of it.
- Q17. From the above join, created a new table 'sss' and joined Tracks table with sss and got Title from sss as 'album', Name as 'Artist' and Name from track 'Song'
- Q18. Joined Customers and Invoice and got cust_name, company and total bill.
- Q19. Created a table 'business' from Q18 and again joined 'business' with 'Invoice_iteams' and got customers with their trackid.
- Q20. Count of unique customers by using 'DISTINCT'

The dataset used in Q21-Q25 is Diabetes taken from Kaggle.

It contains 769 rows and 9 columns.

- 1 Diabetic
- 0 Non-diabetic
- Q21. Count of diabetic, non-diabetic and total
- Q22. Members with less than avg glucose (120.89453125)
- Q23. Found max BP = 122, then data of people with max BP.
- Q24. Woman without pregnancy
- Q25. First found persons with max age then count of them with their data

The dataset used in Q26-Q30 is Covid19 in india taken from Kaggle.

It contains 14511 rows and 9 columns.

Q26. Found deaths in each state also the count of max deaths.

The state with max deaths is Maharasta with 70284

- Q27. Total deaths and Confirmed cases of each state
- Q28. Cured cases of each state
- Q29. Found date and time of max deaths in states
- Q30. States with maximum number of cases cured.