

**Name:** Siddarth Raju Brahmandaberi  
**Email Id:** [sidhusiddarth0@gmail.com](mailto: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;`

Output:

Rank_Calculation	Skill
1	Statistical Analysis and Data Mining
1	Middleware and Integration Software
1	SAP ERP Systems
1	Storage Systems and Management
1	SEO/SEM Marketing
1	Business Intelligence
1	Mobile Development
1	Perl/Python/Ruby
1	Network and Information Security
1	Recruiting
1	Statistical Analysis and Data Mining
1	Recruiting
1	Web Architecture and Development Framework
1	Middleware and Integration Software

**Q2.** What are the unique skills in the dataset?

Query used:

- `SELECT DISTINCT(skill) FROM LinkedIn_Skills;`

Output:

Skill
Statistical Analysis and Data Mining
Middleware and Integration Software
SAP ERP Systems
Storage Systems and Management
SEO/SEM Marketing
Business Intelligence
Mobile Development
Perl/Python/Ruby
Network and Information Security
Recruiting
Java Development
Non-Profit, Fundraising, and Grantmaking
Computer Graphics and Animation

**Q3.** How many unique number of skills are present?

Query used:

- `SELECT COUNT(DISTINCT(skill)) FROM LinkedIn_Skills;`

Output:

COUNT(DISTINCT(skill))
70

**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:

```
COUNT(skill)
800
```

```
COUNT(skill)
410
```

```
count(skill)
390
```

**Q5.** Set of data whose sorting\_calculation is 10?

Query used:

- SELECT \* FROM Linkedin\_Skills WHERE sorting\_calculation = 10;

Output:

#	Sortin...	UniqueC...	Top_5/10	Rank_...	Blank_C...	Country	Number_...	Rank	Skill	Year
10		Australia2...	6-10 Skill	1		Australia	1	10	Recruiting	2014
10		Brazil2014	6-10 Skill	1		Brazil	1	10	Data Engi...	2014
10		Canada2...	6-10 Skill	1		Canada	1	10	Perl/Pyth...	2014
10		France2014	6-10 Skill	1		France	1	10	Computer...	2014
10		Global2014	6-10 Skill	1		Global	1	10	Perl/Pyth...	2014
10		India2014	6-10 Skill	1		India	1	10	Mac, Linu...	2014
10		Netherlan...	6-10 Skill	1		Netherlands	1	10	Network a...	2014
10		South Afri...	6-10 Skill	1		South Africa	1	10	User Inter...	2014
10		United Ar...	6-10 Skill	1		United Ar...	1	10	Web Archi...	2014
10		United Ki...	6-10 Skill	1		United Ki...	1	10	Java Dev...	2014
10		United St...	6-10 Skill	1		United St...	1	10	Perl/Pyth...	2014
10		Australia2...	6-10 Skill	1		Australia	1	10	Perl/Pyth...	2015

## 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:

year	stadium	home	team	goals
1974	Parkstadion (Gelsenkirch...	West Germany	Yugoslavia	9
1954	Hardturm (Zürich)	Switzerland	Hungary	9

**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);

Output:

count(*)
736

team	attendance
Italy	60000
Austria	60000
Italy	50000
Czechoslovakia	50000
England	49348
Soviet Union	49348
Brazil	50928
Soviet Union	50928

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

Queries used:

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

Output:

```
1 COUNT(DISTINCT(home))
16
```

home
Italy
Sweden
Spain
England
Brazil
Uruguay
Mexico
France
USA
Argentina
Germany
South Africa
Chile
West Germany
South Korea/Japan

**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);`

Output:

team	stage
Germany	PLACES 3-4
Austria	PLACES 3-4
France	PLACES 3-4
FRG	PLACES 3-4
Poland	PLACES 3-4
France	PLACES 3-4
Portugal	PLACES 3-4
Soviet Union	PLACES 3-4
Brazil	PLACES 3-4

**Q10.** How many teams played?

Query used:

- `SELECT COUNT(team) FROM WorldCup;`

Output:

```
⋮ COUNT(team)
1672
```

**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:

⋮ Name	Speed
Iron Man	58
Hulk	47
Thor	92
Spider-Man	60
Abomination	53
Loki	47

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

Query used:

- `SELECT sum(strength) FROM superhero;`

Output:

```
⋮ sum(strength)
713
```

**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:

```
SQL> COUNT(*)
6
```

Name	Alignment	Intelligence	Strength	Speed	Durability	Power	Combat
Iron Man	good	100	85	58	85	100	64
Captain America	good	63	65	35	56	46	100
Hulk	good	88	100	47	100	41	85
Hawkeye	good	62	12	23	14	26	80
Thor	good	69	100	92	100	100	85
Spider-Man	good	88	55	60	74	58	85

**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:

Name	Combat
Iron Man	64
Ultron	64
Loki	56
Green Goblin	28

**Q15.** How many super heroes are there in dataset?

Query used:

- `SELECT COUNT(name) FROM superhero;`

Output:

```
SQL> COUNT(name)
12
```

## Data used: Chinook-Data

**Q16.** List the songs done by artists?

Query used:

- SELECT title,name FROM albums INNER JOIN artists;

Output:

Title	Name
For Those About To Rock We Salute You	AC/DC
For Those About To Rock We Salute You	Accept
For Those About To Rock We Salute You	Aerosmith
For Those About To Rock We Salute You	Alanis Morissette
For Those About To Rock We Salute You	Alice In Chains

**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;

Output:

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	Breaking The Rules



**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:

FirstName	LastName	Company	Total
Luis	Gonçalves	Embraer - Empresa Brasileira de ...	3.98
Luis	Gonçalves	Embraer - Empresa Brasileira de ...	3.96
Luis	Gonçalves	Embraer - Empresa Brasileira de ...	5.94
Luis	Gonçalves	Embraer - Empresa Brasileira de ...	0.99
Luis	Gonçalves	Embraer - Empresa Brasileira de ...	1.98

**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.InvoiceId;

Output:

FirstName	LastName	TrackId
Luis	Gonçalves	3247
Luis	Gonçalves	3248
Luis	Gonçalves	447
Luis	Gonçalves	449

**Q20.** What is the count of different customers?

Querie used:

- SELECT COUNT(DISTINCT(firstname)) FROM customers;

Output:

COUNT(DISTINCT(firstname))
57

**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:

```
⋮ COUNT(*)  
268
```

```
⋮ COUNT(*)  
500
```

```
⋮ COUNT(*)  
768
```

**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);

Output:

```
⋮ COUNT(*)  
419
```

**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:

Pregna...	Glucose	BloodPre...	SkinThick...	Insulin	BMI	DiabetesP...	Age	Outcome
1	96	122	0	0	22.4	0.207	27	0

**Q24.** Women with no pregnancy?

Query used:

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

Output:

COUNT(*)
111

**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);

Output:

Pregna...	Glucose	BloodPre...	SkinThick...	Insulin	BMI	DiabetesPedigre...	Age	Outcome
9	134	74	33	60	25.9	0.46	81	0

**Data used:** Covid19\_India

**Data URL:** [https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid\\_19\\_india.csv](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:

State_or_UnionTerritory	Deaths
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0
Kerala	0

max(deaths)
70284

Sno	Date	Time	State_or_...	Confirme...	Confirme...	Cured	Deaths	Confirmed
14495	5/3/2021	8:00 AM	Maharashtra	-	-	3981658	70284	4722401

**Q27.** What is the total deaths and confirmed cases?

Querie used:

- SELECT sum(deaths), sum(confirmedindiannational), sum(confirmedforeignnational) FROM Covid19\_India;

Output:

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

**Q28.** What are the cured cases for each state?

Query used:

- SELECT state\_or\_unionterritory, cured FROM Covid19\_India WHERE cured > 0;

Output:

State_or_UnionTerritory	Cured
Kerala	3
Kerala	3
Kerala	3
Kerala	3
Kerala	3
Kerala	3

**Q29.** What is the date and time when maximum deaths occurred?

Queries used:

- SELECT max(deaths) FROM Covid19\_India;
- SELECT date,time,deaths FROM Covid19\_India WHERE deaths = (SELECT max(deaths) FROM Covid19\_India);

Output:

Date	Time	Deaths
5/3/2021	8:00 AM	70284

**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);

Output:

Date	Time	Deaths
5/3/2021	8:00 AM	70284

### **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 counted 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\_items' 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.