

## SQL QUERIES FOR IRIS FLOWER CLASSIFICATION:-

### STATISTICS:-

1.What are the minimum, maximum, average, and median values for sepal length, sepal width, petal length, and petal width?How many records are there in the dataset?What are the distinct species of iris present in the dataset?

Query:-SELECT

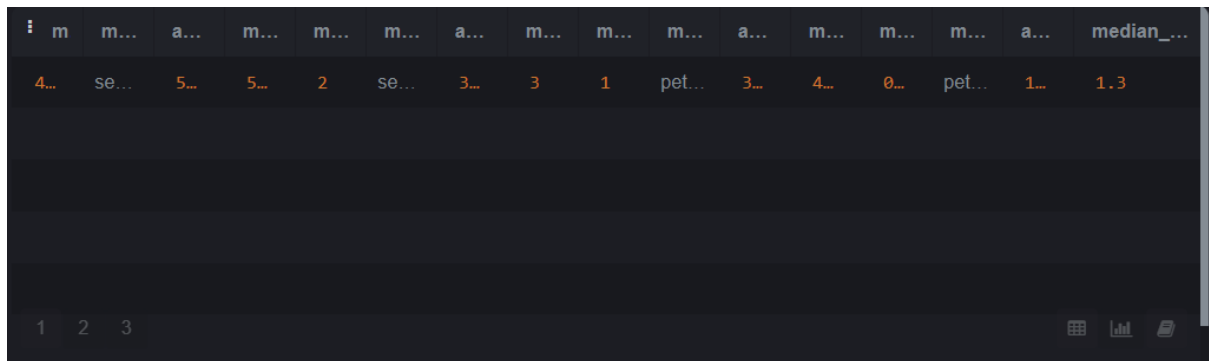
```
MIN(c1) AS min_sepal_length,  
MAX(c1) AS max_sepal_length,  
AVG(c1) AS avg_sepal_length,  
MEDIAN(c1) AS median_sepal_length,  
MIN(c2) AS min_sepal_width,  
MAX(c2) AS max_sepal_width,  
AVG(c2) AS avg_sepal_width,  
MEDIAN(c2) AS median_sepal_width,  
MIN(c3) AS min_petal_length,  
MAX(c3) AS max_petal_length,  
AVG(c3) AS avg_petal_length,  
MEDIAN(c3) AS median_petal_length,  
MIN(c4) AS min_petal_width,  
MAX(c4) AS max_petal_width,  
AVG(c4) AS avg_petal_width,  
MEDIAN(c4) AS median_petal_width
```

FROM IRIS;

```
SELECT COUNT(*) AS total_records FROM IRIS;
```

SELECT DISTINCT c5 FROM IRIS;

Output:-



m	m...	a...	m...	m...	m...	a...	m...	m...	m...	a...	m...	m...	m...	a...	median_...
4...	se...	5...	5...	2	se...	3...	3	1	pet...	3...	4...	0...	pet...	1...	1.3

### SPECIES DISTRIBUTION:-

1.How many flowers belong to each species?What is the percentage distribution of each species?

Query:- SELECT

c5,

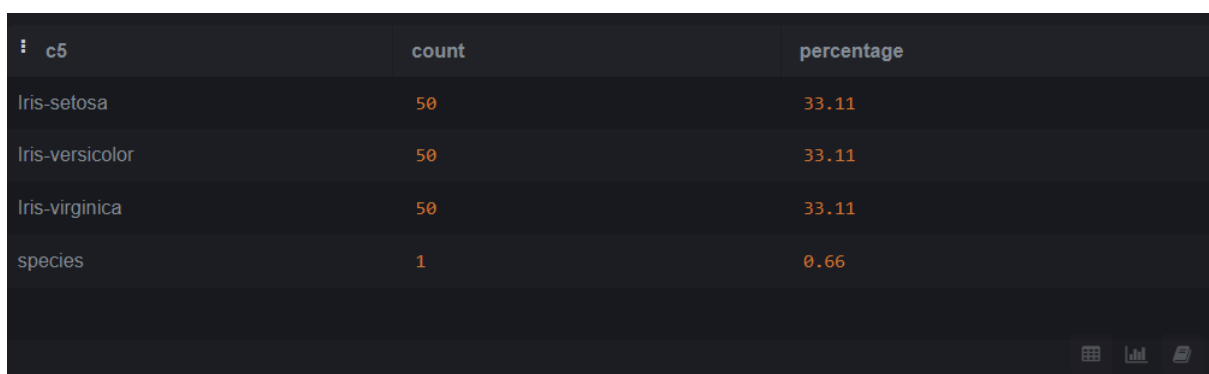
COUNT(\*) AS count,

ROUND((COUNT(\*) \* 100.0) / (SELECT COUNT(\*) FROM IRIS), 2) AS  
percentage

FROM IRIS

GROUP BY c5;

Output:-



c5	count	percentage
Iris-setosa	50	33.11
Iris-versicolor	50	33.11
Iris-virginica	50	33.11
species	1	0.66

## CORRELATION ANALYSIS:-

1. Is there any correlation between sepal length and petal length? Is there any correlation between sepal width and petal width?

Query:- SELECT

```
c5 AS species,  
(COUNT(*) * SUM(c1 * c3) - SUM(c1) * SUM(c3)) /  
(SQRT((COUNT(*) * SUM(c1 * c1) - (SUM(c1) * SUM(c1))) *  
(COUNT(*) * SUM(c3 * c3) - (SUM(c3) * SUM(c3))))) AS  
sepal_petal_length_corr,  
(COUNT(*) * SUM(c2 * c4) - SUM(c2) * SUM(c4)) /  
(SQRT((COUNT(*) * SUM(c2 * c2) - (SUM(c2) * SUM(c2))) *  
(COUNT(*) * SUM(c4 * c4) - (SUM(c4) * SUM(c4))))) AS  
sepal_petal_width_corr  
FROM IRIS  
GROUP BY c5;
```

Output:-

species	sepal_petal_length_corr	sepal_petal_width_corr
Iris-setosa	0.26387409291866826	0.2799728885169049
Iris-versicolor	0.7540489585920166	0.6639987200241148
Iris-virginica	0.8642247329355728	0.5377280262662006
species	NULL	NULL

## Data Distribution:-

1.How are the sepal length and width distributed for each species?How are the petal length and width distributed for each species?

Query:-

```
SELECT
    c5 AS species,
    AVG(c1) AS avg_sepal_length,
    AVG(c2) AS avg_sepal_width,
    AVG(c3) AS avg_petal_length,
    AVG(c4) AS avg_petal_width
FROM IRIS
GROUP BY c5;
```

Output:-

species	avg_sepal_length	avg_sepal_width	avg_petal_length	avg_petal_width
Iris-setosa	5.006	3.418	1.464	0.24400000000000002
Iris-versicolor	5.936	2.77	4.26	1.3259999999999998
Iris-virginica	6.587999999999999	2.973999999999999...	5.552000000000000...	2.026
species	0	0	0	0