

SQL Data Pre-Processing

```
select * from crop_production
```

```
select COUNT(*) from crop_production
```

```
select * from crop_production where Production IS NULL
```

To Trim spaces from front and back of String:

```
UPDATE
  crop_production
SET
  District_Name = LTRIM(RTRIM(District_Name))
```

```
UPDATE
  crop_production
SET
  State_Name = LTRIM(RTRIM(State_Name))
```

```
UPDATE
  crop_production
SET
  Crop = LTRIM(RTRIM(Crop))
```

```
UPDATE
  crop_production
SET
  Crop_Year = LTRIM(RTRIM(Crop_Year))
```

Created CTE to hold the average production values for each crop of specific State and District:

```
WITH avg_production AS (
  SELECT Crop, State_Name, District_Name , AVG(Production) AS avg_production
  FROM crop_production
  WHERE Production IS NOT NULL
  GROUP BY Crop, State_Name, District_Name
)
UPDATE crop_production
SET crop_production.Production = a.avg_production
FROM avg_production AS a
WHERE crop_production.Crop = a.Crop and crop_production.District_Name =
a.District_Name
  and crop_production.State_Name = a.State_Name
AND crop_production.Production IS NULL;
```

Created CTE to hold the average production values for Years for States:

```
WITH avg_production_Year AS (
  SELECT State_Name, Crop_Year, AVG(Production) AS avg_production
  FROM crop_production
  WHERE Production IS NOT NULL
  GROUP BY Crop_Year, State_Name
)
```

```
UPDATE crop_production
SET crop_production.Production = a.avg_production
FROM avg_production_Year AS a
WHERE crop_production.State_Name = a.State_Name
AND crop_production.Production IS NULL;
```

Select Query to Know the states with highest Production:

```
select State_Name, SUM(Production) as Production from crop_production
group by State_Name
order by Production desc
```

Select Query to Know the states and its Districts with highest Production:

```
select District_Name, State_Name, SUM(Production) as Production from crop_production
group by District_Name, State_Name
order by Production desc
```

Select Query to Know the Crop with highest Production:

```
select Crop, SUM(Production) as Production from crop_production
group by Crop
order by Production desc
```

Select Query to Know the Crop and States with highest Production:

```
select Crop, State_Name, SUM(Production) as Production from crop_production
group by Crop, State_Name
order by Production desc
```

Select Query to Know the Year with highest Production:

```
select Crop_Year, SUM(Production) as Production from crop_production
group by Crop_Year
order by Production desc
```

Select Query to Know the Year and the Crop with highest Production:

```
select Crop_Year, Crop, SUM(Production) as Production from crop_production
group by Crop_Year, Crop
order by Production desc
```

Select Query to Know the Season with highest Production:

```
select Season, SUM(Production) as Production from crop_production
group by Season
order by Production desc
```

Select Query to Know the Season and the Crop with highest Production:

```
select Season, Crop, SUM(Production) as Production from crop_production
group by Season, Crop
order by Production desc
```

Select Query to Know the Year, the Season and the Crop with highest Production:

```
select Season, Crop, Crop_Year, SUM(Production) as Production from crop_production
group by Season, Crop, Crop_Year
order by Production desc
```

Select Query to find the highest Average of Production for the States:

```
declare @production float
select @production = SUM(Production) from crop_production
select
    State_Name,
    (SUM(Production) * 100.0 / @production) AS ProductionPercentage
from
    crop_production
group by State_Name
order by ProductionPercentage desc
```

Select Query to find the Average of Production for Tamil Nadu:

```
declare @production1 float
select @production1 = SUM(Production) from crop_production
select
    Crop_Year,
    (SUM(Production) * 100.0 / @production1) AS ProductionPercentage
from
    crop_production
where State_Name = 'Tamil Nadu'
group by Crop_Year
order by ProductionPercentage desc
```

Select Query to find the Average of Production for Tamil Nadu based on the Year and Crop:

```
declare @production2 float
select @production2 = SUM(Production) from crop_production
select
    Crop_Year, Crop,
    (SUM(Production) * 100.0 / @production2) AS ProductionPercentage
from
    crop_production
where State_Name = 'Tamil Nadu'
group by Crop_Year, Crop
order by ProductionPercentage desc
```

Select Query to find the overall Average of Production:

```
select Avg(Production) from crop_production
```

