Forecasting the Future of Crop Yields By Data Analysis

Predicting the next crop yield success

SAURABH JAMNARE

Crop Yield Forecast

Analyzing historical data and weather patterns to predict crop yields accurately this season.

Weather Forecast Analysis

Analyze past data to predict future crop yields accurately.

Pesticide Impact

Understanding pesticides usage crucial for determining crop yield potential.

Yield analysis

Utilize data for precise crop monitoring and management.

Predicting Crop Yields

Crop yield prediction is a crucial element in agricultural planning and decision-making processes.

Yield Change Over Years:

The average yield of wheat in the specified area has fluctuated over the years. It's seen both increases and decreases, with notable spikes in certain years like 2002 and 2003.

```sql
-- How has the yield of a specific crop changed over the years in a particular area?
SELECT Item,Year, ROUND(AVG(Value),2) AS Average_Yield
FROM yield
WHERE Area = 'Afghanistan' AND Item = 'Wheat'
GROUP BY Item,Year
ORDER BY Year;
```

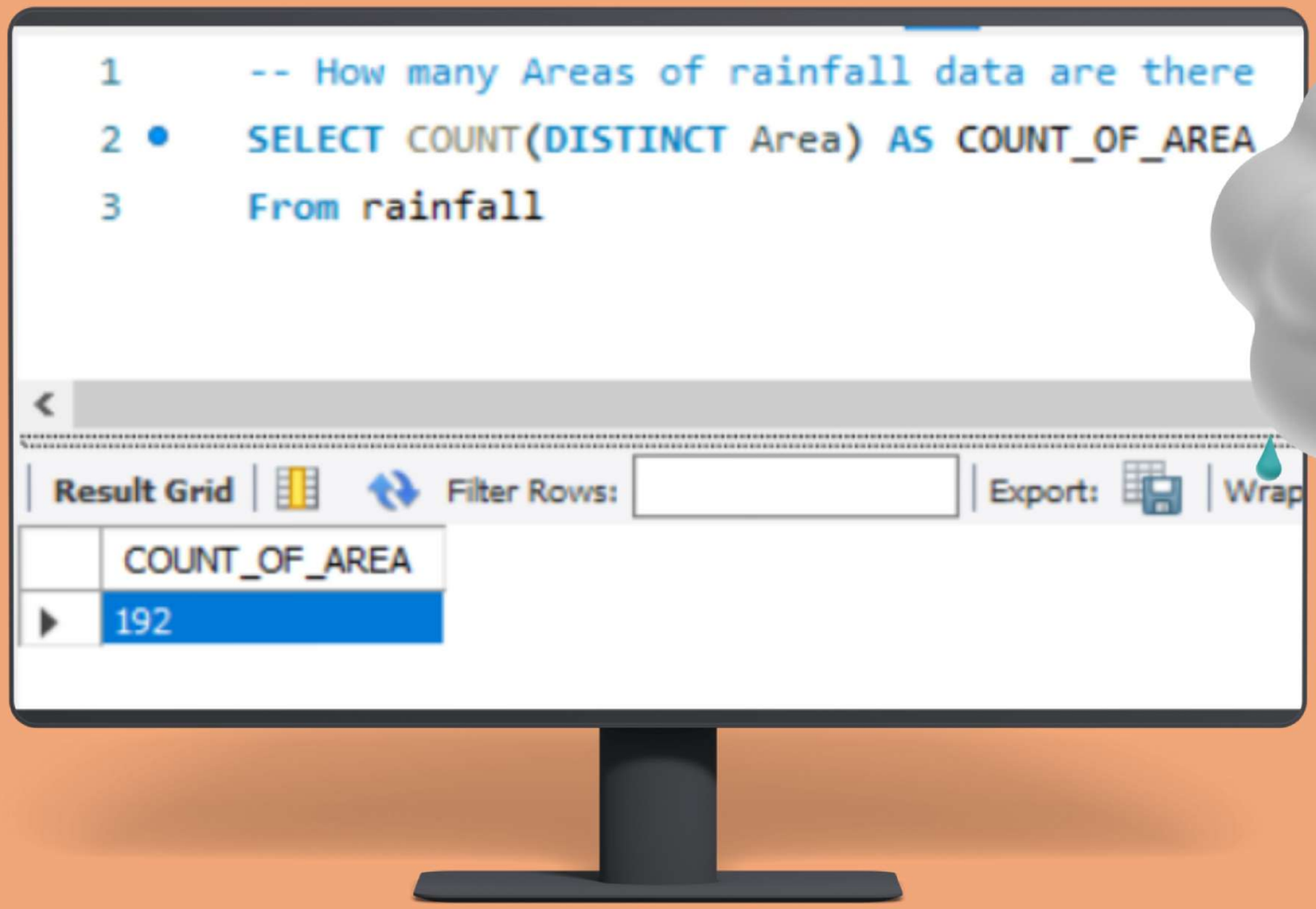
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

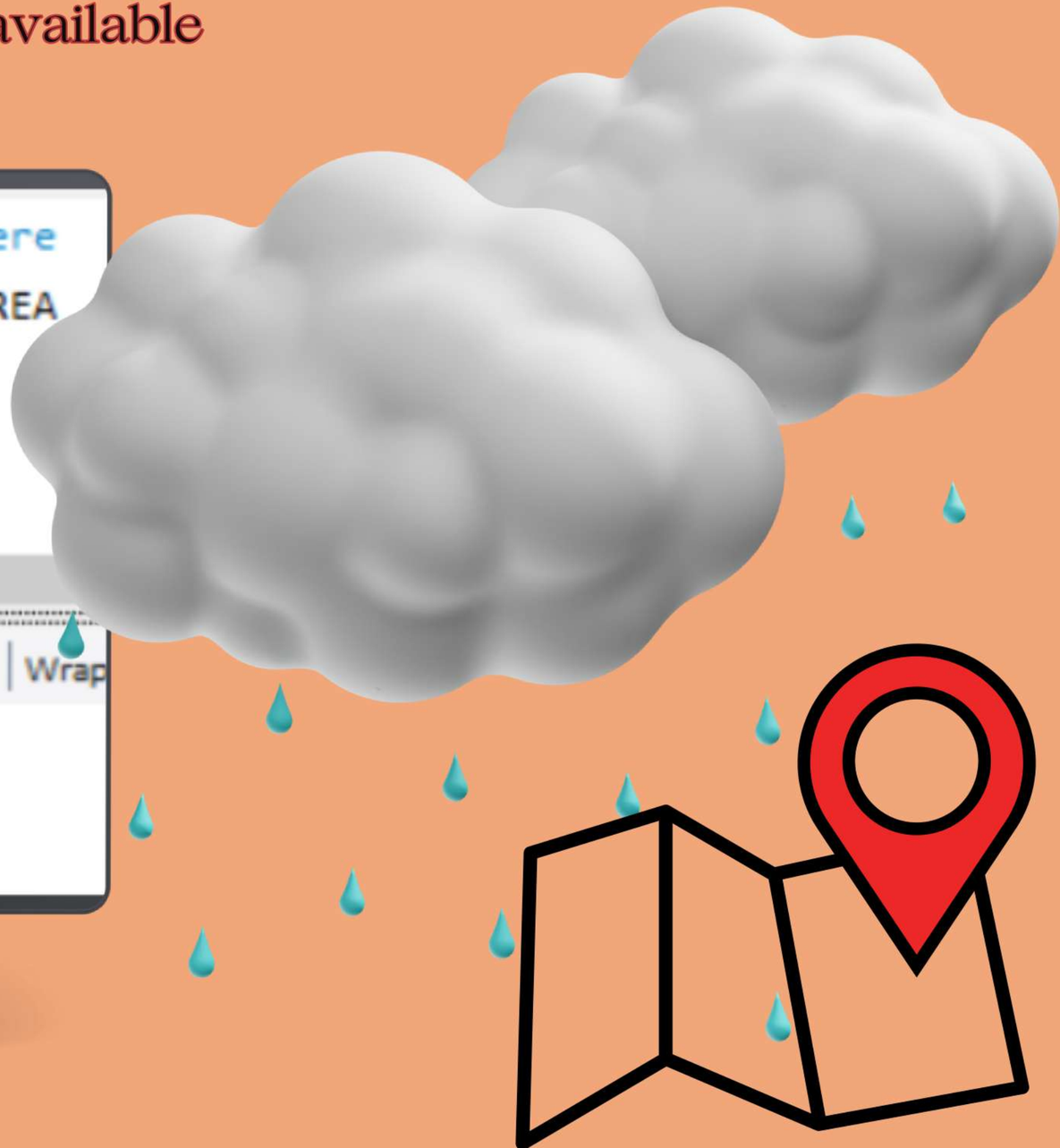
Item	Year	Average_Yield
Wheat	1997	12764.00
Wheat	1998	12964.00
Wheat	1999	12329.00
Wheat	2000	7240.00
Wheat	2001	8977.00
Wheat	2002	15419.00
Wheat	2003	15000.00
Wheat	2004	12659.00

- Predict that wheat yield will continue to fluctuate based on historical patterns, influenced by various factors such as weather conditions, technological advancements, and agricultural practices.

Areas of Rainfall Data:

- There are 192 areas for which rainfall data is available





- Trends in Pesticide Usage:

Pesticide usage has generally increased over the years, with occasional fluctuations but a noticeable upward trend from 1990 to 2016.

```
1    -- Are there any trends in pesticide usage over the years?
2 •  SELECT Year, SUM(Value) AS Total_Pesticides
3    FROM pesticides
4    GROUP BY Year
5    ORDER BY Year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Total_Pesticides
1990	2302485
1991	2276734
1992	2342271
1993	2402381
1994	2562421
1995	2712357
1996	2828472
1997	2938375
1998	2989981
1999	3093702
2000	3059521
2001	3019371

- Predict a continued upward trend in pesticide usage due to increasing demand for food production, although there may be periodic fluctuations based on regulatory changes and advancements in pest management techniques

- Top 3 Areas with Highest Average Yield and Correlation:

United Kingdom, Belgium, and Denmark are the top 3 areas with the highest average yield. The correlation between yield and factors like rainfall, temperature, and pesticide usage varies across these areas.


```sql
-- Which top 3 area has the highest average yield, and how does it correlate with rainfall, temperature, and pesticide usage?
SELECT y.Area, ROUND(AVG(y.Value),2) AS Avg_Yield, ROUND(AVG(r.average_rain_fall_mm_per_year),2) AS Avg_Rainfall,
ROUND(AVG(t.avg_temp),2) AS Avg_Temp, ROUND(AVG(p.Value),2) AS Avg_Pesticides
FROM yield y
JOIN rainfall r ON y.Area = r.Area AND y.Year = r.Year
JOIN temp t ON y.Area = t.country AND y.Year = t.year
JOIN pesticides p ON y.Area = p.Area AND y.Year = p.Year
GROUP BY y.Area
ORDER BY Avg_Yield DESC
LIMIT 3;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Area	Avg_Yield	Avg_Rainfall	Avg_Temp	Avg_Pesticides
United Kingdom	240956.48	1220.00	9.6	28159.43
Belgium	216468.46	847.00	11.02	7357.92
Denmark	214033.02	703.00	8.98	3909.82

- The top areas with the highest average yield may change over time as agricultural practices evolve, and the correlation between yield and environmental factors will continue to be studied for optimization.

- Comparison of Yield Before and After Pesticide Policy Change:

The average yield has generally increased from 2010 to 2016, possibly indicating a positive impact from a change in pesticide policy or introduction of new pesticides.

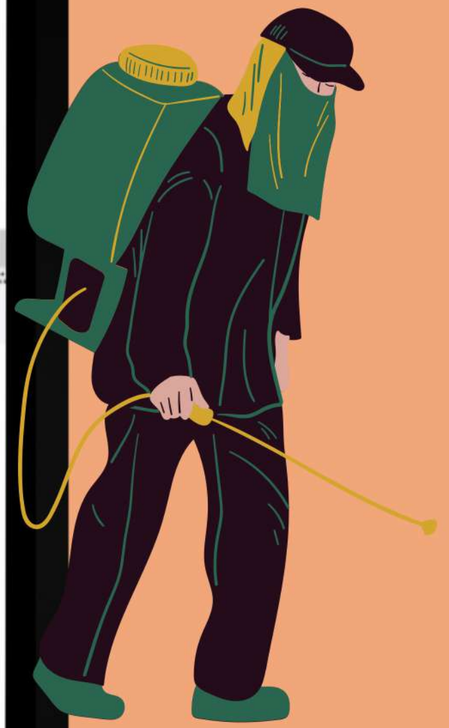
before

after

```
1   -- Comparison of yield before and after a specific change in pesticide
2   -- policy or introduction of a new type of pesticide.
3   SELECT Year, ROUND(AVG(Value),2) AS Average_Yield
4   FROM yield
5   WHERE Area = 'South Africa' AND Item = 'Potatoes' AND Year >= '2010'
6   GROUP BY Year;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Average_Yield
2010	342609.00
2011	349826.00
2012	351066.00
2013	353546.00
2014	356027.00
2015	358507.00
2016	360988.00



- Yield improvements seen after a pesticide policy change will continue if the new policies promote sustainable and effective pest management practices.

- **Rainfall vs. Crop Yields in Top Producing Areas:**

United Kindom, Belgium, Denmark, Netherland, Ireland having average yields across top producing areas, and their correlation with avergae rainfall.

```sql
-- How does rainfall compare to crop yields in the top producing areas?
SELECT y.Area, ROUND(AVG(y.Value),2) AS Average_Yield, ROUND(AVG(r.average_rain_fall_mm_per_year),2) AS Average_Rainfall
FROM yield y
JOIN rainfall r ON y.Area = r.Area AND y.Year = r.Year
GROUP BY y.Area
ORDER BY Average_Yield DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Area	Average_Yield	Average_Rainfall
United Kingdom	238186.25	1220.00
Belgium	216559.54	847.00
Denmark	210341.51	703.00
Netherlands	204268.57	778.00
Ireland	195911.18	1118.00

- The relationship between rainfall and crop yields will continue to be studied and optimized, with technology playing a crucial role in precision farming and water management.

- The trends indicate a complex relationship between average yield and average temperature, influenced by various factors including technological advancement, agricultural practices, and geographical conditions.

```sql
-- What are the trends in yield and average temperature for the yielding areas?
WITH TopYieldingAreas AS (
  SELECT Area
  FROM yield_df
  GROUP BY Area
  ORDER BY AVG(hgha_yield) DESC
)
SELECT y.Area,
       ROUND(AVG(hgha_yield),2) AS Average_Yield,
       ROUND(AVG(y.avg_temp),2) AS Average_Temperature
FROM yield_df y
JOIN TopYieldingAreas t ON y.Area = t.Area
GROUP BY y.Area
ORDER BY y.Area ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Area	Average_Yield	Average_Temperature
Argentina	89304.43	17.71
Armenia	71811.11	9.58
Australia	112951.41	16.59
Austria	113044.35	9.12
Azerbaijan	39727.41	12.44
Bahamas	65443.54	25.46
Bahrain	153237.55	26.7
Bangladesh	52518.09	26.1
Belarus	74679.56	6.74
Belgium	216468.46	11.02
Botswana	7353.92	19.61
Brazil	73583.80	22.67
Bulgaria	45384.60	9.39
Burkina F...	33061.61	28.74

- Regions with temperate climates tend to have higher yields, while those with extreme temperatures (either very high or very low) tend to have lower yields.

- Several countries have shown notable improvements in agricultural yields while simultaneously reducing pesticide usage for Example Albania, Columbia Etc.

```sql
-- Which areas have shown improvement in yield despite a decrease in pesticide usage?
SELECT DISTINCT Area, Year, Last_Year_Yield, hgha_yield, Last_Year_Pesticides, pesticides_tonnes
FROM (
    SELECT Area, Year,
        LAG(hgha_yield, 1) OVER (PARTITION BY Area ORDER BY Year) AS Last_Year_Yield,
        hgha_yield,
        LAG(pesticides_tonnes, 1) OVER (PARTITION BY Area ORDER BY Year) AS Last_Year_Pesticides,
        pesticides_tonnes
    FROM yield_df
) AS subquery
WHERE hgha_yield > Last_Year_Yield
  AND pesticides_tonnes < Last_Year_Pesticides;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Area	Year	Last_Year_Yield	hgha_yield	Last_Year_Pesticides	pesticides_tonnes
Albania	2011	39905	59869	1311	1303
Albania	2012	42301	67290	1303	766
Algeria	1992	10809	22828	2462	1215
Algeria	1993	9939	12500	1215	1022
Algeria	1996	8924	17154	2697	480
Algeria	1997	13090	21417	480	401
Algeria	1998	8016	17222	401	323
Algeria	2000	10711	36186	654	604
Algeria	2005	13582	33864	3748	3663
Algeria	2009	11038	30914	11556	3868
Algeria	2010	15975	25827	3868	1296
Algeria	2013	17639	33649	17380	17279
Angola	1992	8000	42295	79	23
Angola	1994	8876	58596	169	26
Angola	1996	11855	48077	336	40

Result 4

- Albania:

From 2011 to 2012, the yield increased from 39,905 to 42,301 hgha, while pesticide usage dropped from 1,311 to 766 tonnes.

- Colombia:

Between 2007 and 2010, Colombia's yield increased from 108,337 to 106,094 hgha, while pesticide usage significantly dropped from 98,329 to 48,618 tonnes.

The data reflects global agricultural practices with high pesticide usage concentrated in major agricultural economies, including countries in North and South America, Europe, and Asia.

```sql
-- For each area, what is the maximum value of pesticides used
-- in a single year from the "pesticides" dataset? out of which
-- give top 10 Areas
SELECT Area, MAX(Value) AS Max_Pesticides_Used
FROM pesticides
GROUP BY Area
ORDER BY Max_Pesticides_Used DESC
Limit 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Area	Max_Pesticides_Used
China, mainland	1807000
United States of America	434541
Brazil	395646
Argentina	207706
Colombia	117881
France	114695
Italy	100596
USSR	92200
Ukraine	90815
Thailand	88548

Developed vs. Developing Countries: Both developed (e.g., United States, France, Italy) and developing countries (e.g., China, Brazil, Colombia) exhibit high pesticide usage, indicating a widespread reliance on pesticides in agriculture irrespective of the country's development status.

Environmental and Health Implications: The high levels of pesticide use in these regions may have significant implications for environmental health and safety, necessitating ongoing monitoring and potential regulation to mitigate adverse effects.

- **Highest Average Yield Across All Areas and Years (Top 10):**

The top crops in terms of average yield across all areas and years are potatoes, cassava, sweet potatoes, yams, and others.

```sql
-- Which items have the highest average yield (hgha_yield) across all areas and years
SELECT Item, ROUND(AVG(hgha_yield),2)AS Avg_Yield
FROM yield_df
GROUP BY Item
ORDER BY Avg_Yield DESC;
```

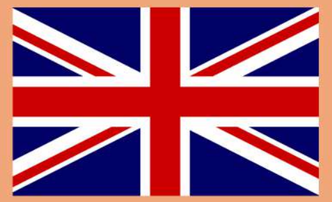
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Item	Avg_Yield
Potatoes	199801.55
Cassava	150479.47
Sweet potatoes	119057.79
Yams	114140.35
Plantains and others	106041.32
Rice, paddy	40730.43
Maize	36310.07
Wheat	30116.27
Sorghum	18635.78
Soybeans	16731.09

crops with consistently high average yields will continue to be prioritized in agricultural production, with ongoing research and development to further enhance their productivity.

- **Top 5 Areas with Highest Average Yield in Last 15 Years:**

United Kingdom, Belgium, Netherlands, Ireland, and New Zealand are the top areas with high average yields over the last 15 years.

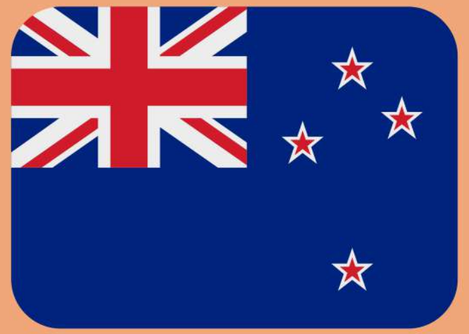


```sql
-- Identify the top 5 areas with the highest average yield (hgha_yield) in the last 15 years
SELECT Area, ROUND(AVG(hgha_yield),2) AS Avg_Yield_Last_15_Years
FROM yield_df
WHERE Year >= YEAR(CURDATE()) - 15 -- Last 15 years
GROUP BY Area
ORDER BY Avg_Yield_Last_15_Years DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

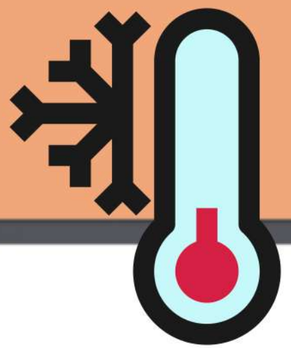
Area	Avg_Yield_Last_15_Years
United Kingdom	240367.30
Belgium	219642.80
Netherlands	218711.20
Ireland	203142.20
New Zealand	198744.05

areas with historically high average yields will continue to innovate and invest in agricultural technologies to sustain and improve productivity.



- **Correlation Between Average Temperature and Yield:**

The correlation between average temperature and yield fluctuates over the years, with some years showing positive correlations and others negative.



```sql
-- Identify if there is any corelation between avg_temp and hgha_yield
SELECT
    t.Year,
    (SUM(t.avg_temp * y.hgha_yield) - SUM(t.avg_temp) * SUM(y.hgha_yield) / COUNT(*))
    /
    (SQRT((SUM(POW(t.avg_temp, 2)) - POW(SUM(t.avg_temp), 2) / COUNT(*)) * (SUM(POW(y.hgha_yield, 2)) - POW(SUM(y.hgha_yield), 2) / COUNT(*))))
    AS Temperature_Yield_Correlation
FROM
    temp t
JOIN
    yield_df y ON t.Year = y.Year
GROUP BY
    t.Year;
-- 1 means 1st temp increase as well as yield increase, -1 means one variable decrese then other increase
-- 0 means no co realtion
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Year	Temperature_Yield_Correlation
1990	-0.00000000000086865450977436 29
1991	-0.00000000000103053603705070 25
1992	-0.00000000000252103371286801 13
1993	-0.00000000000122542596639298 36
1994	-0.00000000000120044002442331 4
1995	0.00000000000203318091574934 65
1996	0.00000000000088208945541515 47
1997	0.00000000000023678549819432957
1998	-0.00000000000254129584011285 77
1999	-0.00000000000052704379634590 84
2000	0.00000000000151636478392437 64

The data indicates an extremely weak and insignificant correlation between average temperature and hgha_yield over the years. This suggests that average temperature alone is not a reliable predictor of crop yield, and other factors should be given more attention in efforts to optimize agricultural output.

Thank you!