Questions to answer:
What percentage of the total electricity generated in Germany comes from solar energy and how has this ratio evolved since the years?
Which countries in Europe generated a higher percentage of solar energy relative to their total generated energy between 2017 and 2022?

Exploration of the table (SQL BigQuery):
The electricity production values are mixed in the VALUE column, where data from solar energy production, total production, and distribution losses are included. To determine which parameter the VALUE column refers to, we need to look at the associated values in the PRODUCT column. The only data of interest for the study, which will help us answer the questions, will be found in the VALUE column when PRODUCT = "Net electricity production" and when PRODUCT = "Solar". We will exclude "OECD Total", "IEA Total", "OECD Americas", "OECD Europe", "OECD Asia Oceania" as they are sets of countries rather than specific countries.

Data Cleaning and Verification (SQL BigQuery)

- In the COUNTRY column, there are 52 distinct values corresponding to countries or groups of countries, and only 2 of them have incomplete data for "Solar" and "Net electricity production" in the last 5 years (Costa Rica and Iceland) (Figure 1).
- Some averages might be biased, as missing data for certain years for certain countries could also lead to missing data for certain months. This was not the case, as all countries had data for each month during the last 5 years for both "Solar" and "Net electricity production". Additionally, none of these monthly data points are 0 (Query result: 6000 = 5 years * 50 countries * 12 months * 2 values Solar and Total) (Figure 2).
- Germany has data from 2010 to 2022 for all months, and in none of the cases, the value is 0 (Figure 3).

Analysis and creation of necessary tables for visualization (Python Pandas)

- Create a table showing how much electricity each country produced between 2018 and 2022 (inclusive) in total and from solar energy, along with a column reflecting the percentage of the total that solar energy represented (Figure 4).
- Create a table showing, for Germany, for all years from 2010 to 2022, the total electricity production, total electricity production from solar energy, the ratio between the two, and the annual increase in this ratio. Additionally, create a table comparing the % of solar energy in Germany with that of OECD Europe and OECD Total (Figure 5).

```sql
SELECT
  COUNTRY,
  COUNT(DISTINCT IF(PRODUCT = 'Solar', YEAR, NULL)) AS solar_years,
  COUNT(DISTINCT IF(PRODUCT = 'Net electricity production', YEAR, NULL)) AS
total_years
FROM
  `future-silicon-416914.Project1.electricity_production`
WHERE
  YEAR > 2017
GROUP BY
  COUNTRY
ORDER BY solar_years, total_years
```

(Figure 1)

```sql
SELECT
  COUNT(*) AS count_not_null_values
FROM
  `future-silicon-416914.Project1.electricity_production`
WHERE
  YEAR > 2017
  AND PRODUCT IN ('Net electricity production', 'Solar')
  AND VALUE IS NOT NULL
  AND COUNTRY NOT IN ('Iceland', 'Costa Rica')
```

(Figure 2)

```sql
SELECT
  YEAR,
  COUNT (DISTINCT MONTH) AS number_of_months
FROM future-silicon-416914.Project1.electricity_production
WHERE COUNTRY = "Germany" AND VALUE IS NOT NULL
GROUP BY YEAR
```

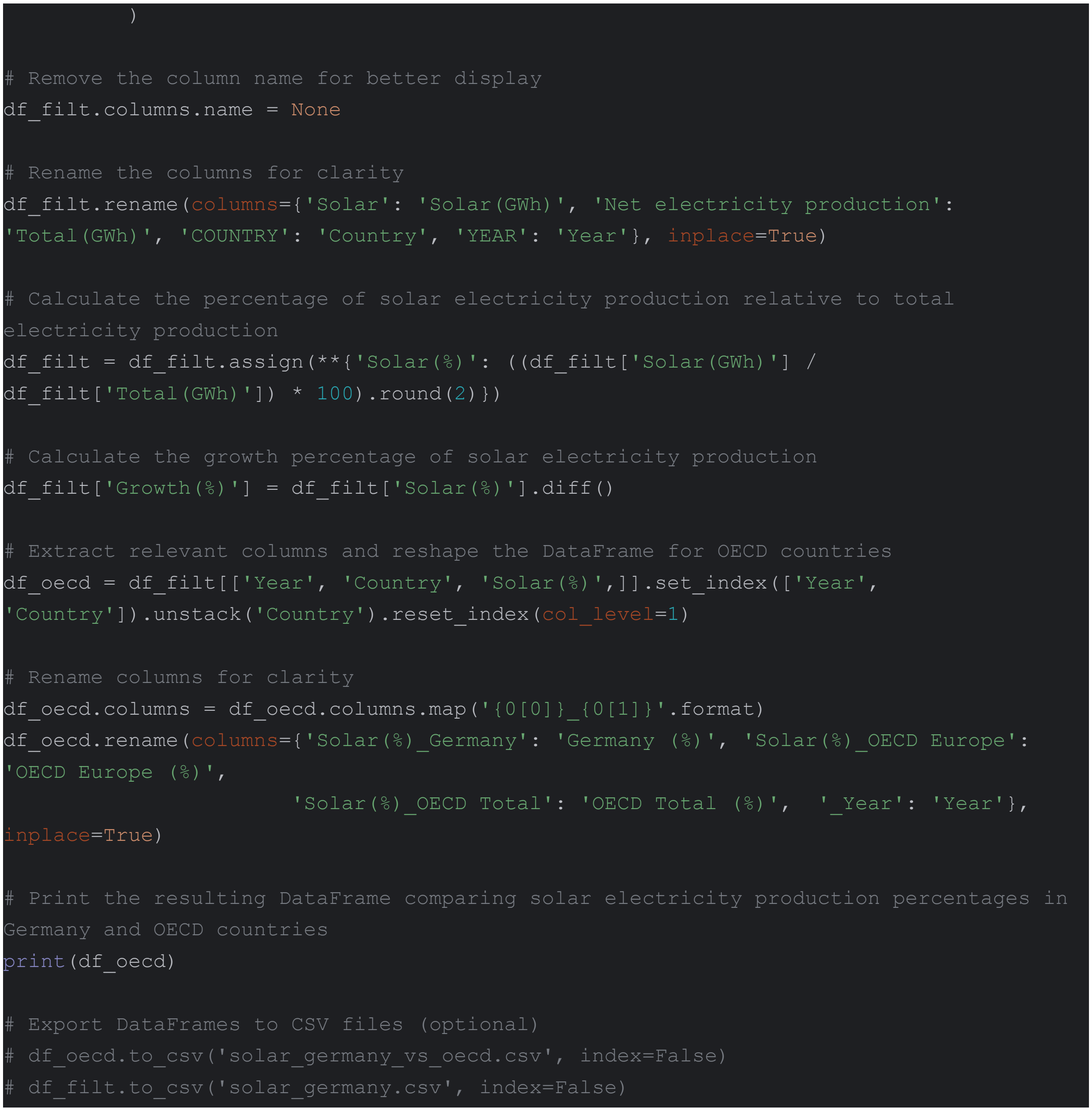
(Figure 3)

```python
import pandas as pd

# Reading the CSV file into a DataFrame
df = pd.read_csv('C:/Users/Usuario/Desktop/Project 1/electricty_production.csv')
```

```
# Filtering the DataFrame based on specific conditions, calculating the total, and
reshaping the data
df_filt = (df.query("PRODUCT in ['Solar', 'Net electricity production'] and YEAR >
2017 and COUNTRY not in ['Costa "
                   "Rica', 'Iceland','OECD Total','IEA Total', 'OECD Americas','OECD
Europe','OECD Asia Oceania']")
          .groupby(['COUNTRY', 'PRODUCT'])['VALUE']
          .sum()
          .unstack('PRODUCT')
          .reset_index()
          .round(2)
          .sort_values('Net electricity production', ascending=False)
          )
# Removing the name of the columns' index
df_filt.columns.name = None

# Renaming the columns 'Solar' and 'Net electricity production'
df_filt.rename(columns={'Solar': 'Solar(GWh)', 'Net electricity production':
'Total(GWh)', 'COUNTRY': 'Country'}, inplace=True)

# Calculating the percentage of 'mean_solar' relative to 'mean_total' and assigning it
to a new column 'solar_pct'
df_filt = df_filt.assign(**{'Solar(%)': ((df_filt['Solar(GWh)'] /
df_filt['Total(GWh)']) * 100).round(2)})

# Sorting the DataFrame by the column 'solar_pct' in descending order and printing the
result
print(df_filt.sort_values('Solar(%)', ascending=False))
```

(Figure 4)

```
# Import the pandas library
import pandas as pd

# Read the CSV file into a DataFrame
df = pd.read_csv(r'C:/Users/Usuario/Desktop/Project 1/electricty_production.csv')

# Filter the DataFrame based on specific conditions, then group and aggregate the data
df_filt = (df.query("COUNTRY in ['Germany', 'OECD Europe', 'OECD Total'] and PRODUCT
in ['Solar', 'Net electricity production']")
          .groupby(['YEAR', 'PRODUCT', 'COUNTRY'])['VALUE']
          .sum()  # Sum the values
          .unstack('PRODUCT')  # Unstack the 'PRODUCT' level of the index
          .reset_index()  # Reset the index of the DataFrame
          .round(2)  # Round the values to two decimal places
          .sort_values('YEAR', ascending=True)  # Sort the DataFrame by the 'YEAR'
column in ascending order
```



```
          )

# Remove the column name for better display
df_filt.columns.name = None

# Rename the columns for clarity
df_filt.rename(columns={'Solar': 'Solar(GWh)', 'Net electricity production': 'Total(GWh)', 'COUNTRY': 'Country', 'YEAR': 'Year'}, inplace=True)

# Calculate the percentage of solar electricity production relative to total electricity production
df_filt = df_filt.assign(**{'Solar(%)': ((df_filt['Solar(GWh)'] / df_filt['Total(GWh)']) * 100).round(2)})

# Calculate the growth percentage of solar electricity production
df_filt['Growth(%)'] = df_filt['Solar(%)'].diff()

# Extract relevant columns and reshape the DataFrame for OECD countries
df_oecd = df_filt[['Year', 'Country', 'Solar(%)',]].set_index(['Year', 'Country']).unstack('Country').reset_index(col_level=1)

# Rename columns for clarity
df_oecd.columns = df_oecd.columns.map('{0[0]}_{0[1]}'.format)
df_oecd.rename(columns={'Solar(%)_Germany': 'Germany (%)', 'Solar(%)_OECD Europe': 'OECD Europe (%)',
                       'Solar(%)_OECD Total': 'OECD Total (%)',  '_Year': 'Year'}, inplace=True)

# Print the resulting DataFrame comparing solar electricity production percentages in Germany and OECD countries
print(df_oecd)

# Export DataFrames to CSV files (optional)
# df_oecd.to_csv('solar_germany_vs_oecd.csv', index=False)
# df_filt.to_csv('solar_germany.csv', index=False)
```

(Figure 5)