

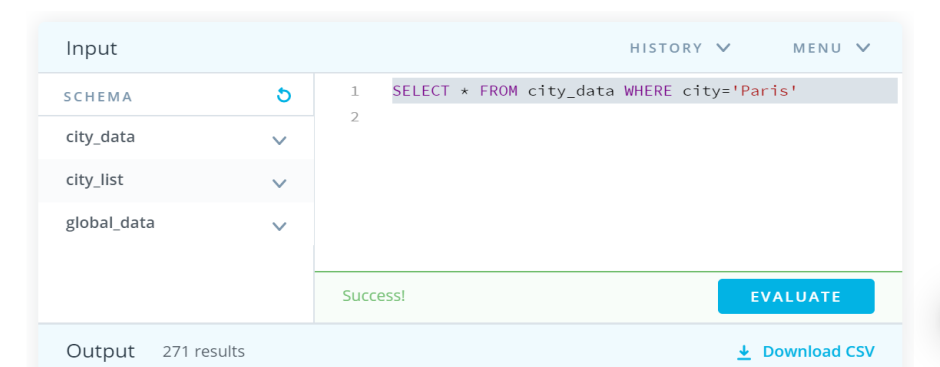
Data Analyst Nanodegree : Udacity

Project 1: Explore Weather trends

Extract the data from the database.

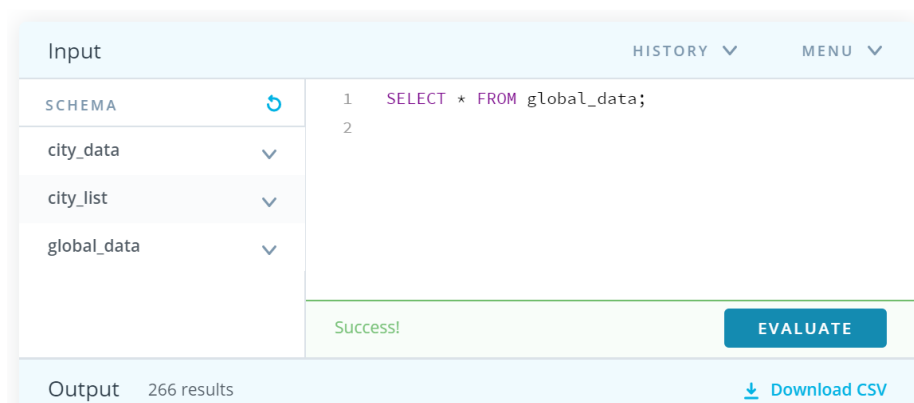
For completing this task I need to have two datasets from SQL for global temperature and for my local city (Paris).

The code for getting the local dataset I apply the following SQL code as I need all the dataset for only the city of Paris



The screenshot shows a SQL query editor interface. On the left, under the 'Input' tab, there is a 'SCHEMA' section with a refresh icon and a list of tables: 'city_data', 'city_list', and 'global_data', each with a dropdown arrow. The main area contains a SQL query: `SELECT * FROM city_data WHERE city='Paris'`. Below the query, there is a green 'Success!' message and a blue 'EVALUATE' button. At the bottom, the 'Output' section shows '271 results' and a 'Download CSV' link.

And for the global temperature dataset my SQL code will be the following as I need all the global dataset



The screenshot shows a SQL query editor interface. On the left, under the 'Input' tab, there is a 'SCHEMA' section with a refresh icon and a list of tables: 'city_data', 'city_list', and 'global_data', each with a dropdown arrow. The main area contains a SQL query: `SELECT * FROM global_data;`. Below the query, there is a green 'Success!' message and a blue 'EVALUATE' button. At the bottom, the 'Output' section shows '266 results' and a 'Download CSV' link.

After having both the datasets in CSV format I will start working in Python.

As each project starts with importing the necessary packages, I did import them. After I open them in Python with the help of pandas. After I tried to have some general overview of the data looking at first, last 5 rows of both datasets, also I check the summaries of these datasets.

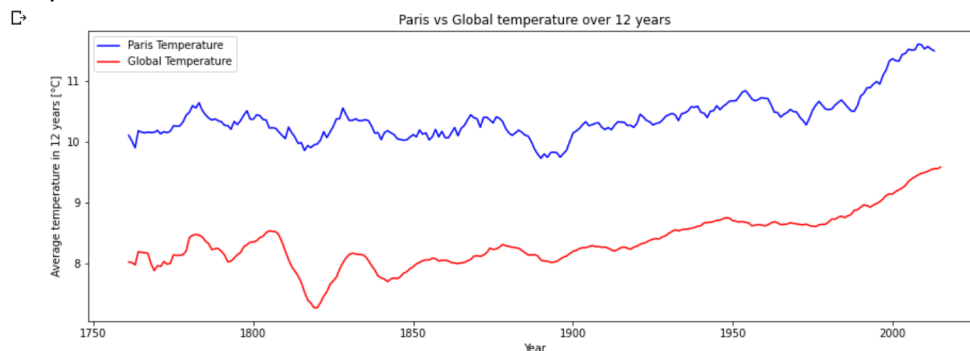
```
global_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 266 entries, 0 to 265  
Data columns (total 2 columns):  
#   Column      Non-Null Count  Dtype  
--  --  
0   year        266 non-null    int64  
1   avg_temp    266 non-null    float64  
dtypes: float64(1), int64(1)  
memory usage: 4.3 KB
```

```
[20] paris_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 271 entries, 0 to 270  
Data columns (total 4 columns):  
#   Column      Non-Null Count  Dtype  
--  --  
0   year        271 non-null    int64  
1   city        271 non-null    object  
2   country     271 non-null    object  
3   avg_temp    267 non-null    float64  
dtypes: float64(1), int64(1), object(2)  
memory usage: 8.6+ KB
```

To compare Paris's temperatures with the global temperatures I will use rolling means. Here window=12 means that I am taking 12 years and calculate mean temperature over 12 years. Just with simple .describe function I could answer the questions but the visualization is an easier and better way.



Conclusions

1. Paris has had significantly hotter temperature over the global temperature during the past centuries. It was always hotter.
2. Though the global temperature did not show huge fluctuations
3. Global temperature decreased only during the period of 1800-1850
4. Based on the outcomes, the world is getting hotter as the trend is increasing(a clear global warming).
5. Paris also has increasing trend (quite logical it makes part of the world) and gets only hotter since the last decades.