

**Nanodegree:** Data Analyst

**Project Submission:** Exploring Weather Trends

**Prepared by:** Rustem Yeleussinov

**Outline:**

1. Workflow

- a. Define objective / project goal: Come up with observations based on analysis of local and global temperature data and compare the temperature trends where I live (San Jose, CA) to overall global temperature trends with application of moving average calculation.
- b. Business understanding: reviewed all the materials in Workplace available, understood project goals, tools and expected outcomes
- c. Data understanding: reviewed dataset using SQL queries including schema, table names, table columns, number of rows and etc. Here is the example of simple SQL query for data discovery:

*SELECT \**

*FROM 'table\_name'*

- d. Data preparation: Created a dataset based on available tables using SQL query then downloaded the output to csv file. The file was ready to use without further data cleansing. Transformation was not required as dataset structure is simple and optimal for next steps. Here is SQL query to pull the data into one dataset:

*SELECT city\_data.year, city\_data.city, city\_data.country, city\_data.city, city\_data.avg\_temp as*

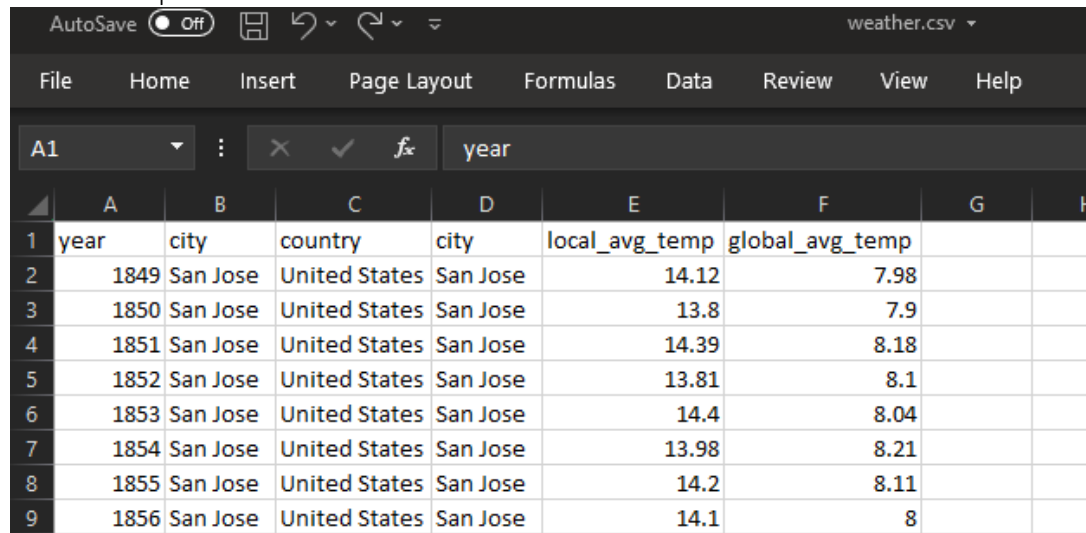
*local\_avg\_temp, global\_data.avg\_temp as global\_avg\_temp*

*FROM city\_data*

*LEFT JOIN global\_data ON city\_data.year = global\_data.year*

*WHERE country='United States' and city='San Jose'*

The final output in csv would look like this:



	A	B	C	D	E	F	G	H
1	year	city	country	city	local_avg_temp	global_avg_temp		
2	1849	San Jose	United States	San Jose	14.12	7.98		
3	1850	San Jose	United States	San Jose	13.8	7.9		
4	1851	San Jose	United States	San Jose	14.39	8.18		
5	1852	San Jose	United States	San Jose	13.81	8.1		
6	1853	San Jose	United States	San Jose	14.4	8.04		
7	1854	San Jose	United States	San Jose	13.98	8.21		
8	1855	San Jose	United States	San Jose	14.2	8.11		
9	1856	San Jose	United States	San Jose	14.1	8		

- e. Modeling: Python code created using pandas and plotly libraries. iloc and rolling functions were applied to calculate moving average in 10 years window. The output of such code would look like this:

```
In [4]: df_pivot['moving_avg_10_local'] = df_pivot.iloc[:,0].rolling(window=10).mean()
df_pivot.head(15)
```

```
Out[4]:
```

	global_avg_temp	local_avg_temp	moving_avg_10_global	moving_avg_10_local
year				
1849	7.98	14.12	NaN	NaN
1850	7.90	13.80	NaN	NaN
1851	8.18	14.39	NaN	NaN
1852	8.10	13.81	NaN	NaN
1853	8.04	14.40	NaN	NaN
1854	8.21	13.98	NaN	NaN
1855	8.11	14.20	NaN	NaN
1856	8.00	14.10	NaN	NaN
1857	7.76	14.78	NaN	NaN
1858	8.10	14.19	14.177	8.038
1859	8.25	13.71	14.136	8.065
1860	7.96	13.81	14.137	8.071
1861	7.85	14.88	14.186	8.038
1862	7.56	14.43	14.248	7.984
1863	8.11	14.43	14.251	7.991

Jupyter Notebook was chosen as working environment for Python code as simple and effective tool. Plotly Python library was used as visualization tool and create the output in png file (please refer to page 3 of current document). The choice of visualization tool is determined by the authors decision to learn new tool. MS Excel of any other tabular editor would be more than sufficient to create simple line chart for this use case.

- f. Evaluation and deployment: Reviewed the output of python visualization and wrote observations. All the outputs were uploaded to dedicated folder in GitHub repository for easy access and sharing.
2. GitHub folder [https://github.com/rustemyeleussinov/Udacity/tree/master/Data\\_Analyst/Weather](https://github.com/rustemyeleussinov/Udacity/tree/master/Data_Analyst/Weather) containing the following files:
  - a. Jupyter Notebook
  - b. SQL query
  - c. Line chart picture saved as a png file
  - d. SQL query output as a csv file
3. Tools used:
  - a. SQL
  - b. Python
  - c. Jupyter
  - d. GitHub Desktop and Web

#### Observations:

1. Average temperature in closest city is considerably lower than global average.
2. The difference is not consistent over the observed period of time in short term
3. Global temperature is more volatile than local temperature
4. However, there is a similarity in overall trend: both temperature measures are rising in long term with considerable spike last 20-40 years.

#### Line Chart:

## 10 years moving average Global vs San Jose

