### Capstone Mini-Project: Data Wrangling

The dataset is acquired from kaggle, originally acquired using [Weather API](https://openweathermap.org/api) on the [OpenWeatherMap website](https://openweathermap.org/)

### The dataset being used for the capstone project consists of a single file, temperature.csv. It is checked for missing values and outliers. The code is shown in the jupyter notebook.

### The dataset consists of 45,253 rows and 37 columns. Each row represents hourly temperature for the 36 cities. First column is the Datetime and is being set as the index.

### First row consisted of NaNs and was discarded with no impact to the remaining data.

### There were around 700 NaNs each for the cities, Vancouver, Miami, New York, Beersheba, Tel Aviv District, Eilat, Haifa, Nahariyya and Jerusalem. As these accounted for very small percentage of the data, these NaNs were dropped.

### Few NaNs found in the dataset were forward filled.

### The data for each city consists of outliers but these are retained since these are not errors and are extreme temperature values.

### The cleaned data is saved into Temperature\_cleaned.csv .