**Content of row data**

The dataset contains ~5 years of **high temporal resolution** (hourly measurements) data of various weather attributes, such as temperature, humidity, air pressure, etc. This data is available for 30 US and Canadian Cities, as well as 6 Israeli cities. I've organized the data according to a common time axis for easy use. Each attribute has it's own file and is organized such that the rows are the time axis (it's the same time axis for all files), and the columns are the different cities (it's the same city ordering for all files as well). Additionally, for each city we also have the country, latitude and longitude information in a separate file.

Taxi dataset is based on the [2016 NYC Yellow Cab trip record data](https://cloud.google.com/bigquery/public-data/nyc-tlc-trips) made available in Big Query on Google Cloud Platform. The data was originally published by the [NYC Taxi and Limousine Commission (TLC)](http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml). The data was sampled and cleaned for the purposes of this project . Based on individual trip.

### Acknowledgements

The weathr dataset was aquired using [Weather API](https://openweathermap.org/api) on the [OpenWeatherMap website](https://openweathermap.org/), and is available under the [ODbL License](https://opendatacommons.org/licenses/odbl/).

### Inspiration

Weather data is both intrinsically interesting, and also potentially useful when correlated with other types of data. For example, demand for taxi ride is famously known to be correlated with weather conditions.

**Processed data :**

Using all data above and using data cleaning method, processed data is available.

## **File descriptions**

* X\_train (26676, 7)
* y\_train (26676,)
* X\_test (8892, 7)
* y\_test (8892,)

## **Data fields :**

|  | **datetime** | **temperature** | **humidity** | **pressure** | **wind\_speed** | **wind\_direction** | **rides** | **date** | **hour** | **month** | **day** | **hour\_sin** | **hour\_cos** | **day\_sin** | **day\_cos** | **month\_sin** | **month\_cos** |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 2012-10-01 12:00:00 | 15.070000 | 58.0 | 1012.0 | 7.0 | 260.0 | 31 | 2012-10-01 | 12 | 10 | 1 | -0.136167 | -0.990686 | 0.017213 | 0.999852 | -0.866025 | 0.5 |  |  |
| **1** | 2012-10-01 13:00:00 | 15.070000 | 58.0 | 1012.0 | 7.0 | 260.0 | 31 | 2012-10-01 | 13 | 10 | 1 | -0.398401 | -0.917211 | 0.017213 | 0.999852 | -0.866025 | 0.5 |  |  |
| **2** | 2012-10-01 14:00:00 | 15.097676 | 57.0 | 1012.0 | 7.0 | 260.0 | 30 | 2012-10-01 | 14 | 10 | 1 | -0.631088 | -0.775711 | 0.017213 | 0.999852 | -0.866025 | 0.5 |  |  |
| **3** | 2012-10-01 15:00:00 | 15.176940 | 57.0 | 1012.0 | 7.0 | 260.0 | 30 | 2012-10-01 | 15 | 10 | 1 | -0.816970 | -0.576680 | 0.017213 | 0.999852 | -0.866025 | 0.5 |  |  |
| **4** | 2012-10-01 16:00:00 | 15.256203 | 57.0 | 1012.0 | 7.0 | 260.0 | 30 | 2012-10-01 | 16 | 10 | 1 | -0.942261 | -0.334880 | 0.017213 | 0.999852 | -0.866025 | 0.5 |  |  |

* **Datatime**- a unique identifier for each trip
* **temperature** - a code indicating the temperature
* **Humidity** - a code indicating the humidity
* **Wind speed** - a code indicating the wind speed
* **Wind direction** - a code indicating the wind direction
* **Rides** – It indicates the number of the rides
* **Date** – extracted from date and time
* **Hour** - extracted from date and time
* **month** - extracted from date and time
* **Hour Sin** - sine transformation
* **Hour Cos** - consine transformation
* **Month Sin** - sine transformation
* **month Cos** - consine transformation
* **Day Sin** - sine transformation
* **Day Cos**  - consine transformation