Export as .csv files, then import to SQLite database

**Table**

“GlobalTempPredictors”

**.csv files (from Kaggle)**

Land Temperatures

“GlobalLandTemperaturesByCity.csv”

**Data Frame**

Atmospheric predictors

“df”

**Table**

“GeoData”

**Data Frame**

Location predictors

“geoData”

**.txt file (from NOAA)**

Pollutant Concentrations

"ftp://ftp.ncdc.noaa.gov/pub/data/

paleo/icecore/antarctica/law/law2006.txt"

1.

Filenames

“globalTemp\_predictors.csv”

“geoData.csv”

“delhiTemps.csv”

Data import, cleaning, and preparation

**Table**

“DelhiData”

**PDF file (from UN report)**

Human Population

“tabula-sixbilpart1.csv”

**Data Frame**

Delhi temperatures

“delhi.data”

------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Data Frame**

“global.temp.predictors”

**Linear Regression model**

Predict global temp. based on atmosphere

**Ensemble model**

Predict local temp. based on atmosphere

2.

**Data Frame**

“geoData”

**Linear Regression model**

Predict local temperature difference from global avg.

**Model Evaluation/Results**

3.

**Data Frame**

“delhi.data”

**Linear Regression trend**

Forecast future temperature

**Exponential smoothing**

Forecast future temperature

**Weighted moving average**

Forecast future temperature

There are three R notebooks:

* “ETL” – data preparation (corresponds to **1** above)
* “PredictiveAnalysis” – multiple regression models (corresponds to **2** above)
* “ForecastingAnalysis” – forecasting (corresponds to **3** above)

Raw Datasets (used by “ETL” notebook):

* “GlobalLandTemperaturesByCity.csv”
* "ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/law/law2006.txt"
* “tabula-sixbilpart1.csv”

Cleaned Datasets (produced by running “ETL” notebook):

* “globalTemp\_predictors.csv”
* “geoData.csv”
* “delhiTemps.csv”

SQL script to create tables and import data: “ClimateProject\_make\_tables.sql”

Project overview:

Part 1: a) Atmospheric data used to predict average global temperature (linear regression model).

b) Geographic location used to predict temperature for given city, relative to global average (linear regression model).

c) The above models are combined to predict temperature for a given city based on atmospheric data.

Part 2: a) Three different forecasting models created and evaluated:

1. Linear regression trend line
2. Exponential smoothing
3. Weighted moving average

**Link to rubric:** https://docs.google.com/spreadsheets/d/1hp50AtS-qwy3TSKkpPp22NdWuoihSwzk9T0Ur5Fl-ts/edit#gid=0