**Mapper coding structure**

1. **Ploty:**
   1. Sign up
   2. Generate API Key
   3. Get used to the use of website
2. **Scala:**
   1. **Main App with GUI**
      1. User should have a text field to input username
      2. User should have a text field to input API Key
      3. User should have up to five file fields to choose original data files
      4. A button with operation to ingest files
      5. Test#1: press button, print all input Strings in console to check if it is correct
   2. **Ingest function:**
      1. Take in all files from the given path
      2. Call Spark API to ingest data
      3. Give output in a Saprk\_result\_dataset
      4. Test#2: To be decided
      5. Using machine learning algorithm to make predictions from the Saprk\_result\_dataset, give output in Result\_dataset
      6. Machine learning algorithm: Linear least squares, Lasso, and ridge regression
      7. Test#3: To be decided
      8. Parse Result\_dataset into some structure which could be used by Ploty API to draw the graph, called Ploty\_dataset
      9. Test#4: To be decided
      10. Return Ploty\_dataset
      11. Test#5: To be decided
   3. **Main object:**
      1. Main object use the Ploty\_dataset as parameter to call Ploty API, use Future for result, once finished, redirect user to Ploty website for result looking.
      2. Test#6: To be decided
3. **Source Data:**

Uploaded with 2010 – 2016, daily PM2.5 concentration across America and Air Quality concentration, with latitude, longitude, city and date.