**Component Specification**

* **Data Preprocessing Module:**
  + Input: raw data in XML form.
  + Output: organized data set in pandas data frame format.
  + What it does: This function primarily verifies the integrity of the data and the data type. If the data is not complete/damaged or in a format that unaccepted, this function will print an error message to the user before leading to a software crash.
* **Data Management Module:**
  + Input: corresponding dataset in the pandas data frame and user input.
  + Output: Data section/Data subset queried by the user in the pandas data frame. The output is a python list with all data frames generated.
  + What it does: This function asks the user for query conditions and takes the entire data frame. It slices and converts the original data frame into small data subframes per user requests. Each data subframe contains two columns of non-nan value entry. If multiple conditions required, the function will return multiple data frames to meet each token of the user query. All data subframe is passed in the form of a python list.
* **Data Visualization Module**
  + Input: data subframes in python list, user choice of visualization methods.
  + Output: The function has no returns but displays plots and graphs.
  + What it does: Plot data as user wish by using matplotlib and folium.
* **Interactions to accomplish use cases**

First of all, the raw data set should be processed by the Data Preprocessing Module, which checks the integrity and format of raw data. If the raw data is not “healthy”, an error message will be returned to the user. Otherwise, a processed data frame is generated. Next, the Data Management Module will come up and wait for user interaction. The Data Management Module is the first module that a user will face. It takes data selection from the user and returns desired data subsets in the form of a python list. Then, this python list is passed to the Data Visualization Module internally and waits for the user to select a visualization method. Thus, the Data Visualization Module would be the second interface user would use. The visualization could be a pie chart, bar chart or map with data layers. Finally, the Data Visualization Module takes input from the user and data subset from the previous module and generates plots desired by the user. This module returns nothing to the system but displays charts.

* **Preliminary plan**

Our tentative plan would be the following:

* The Data Processing Module (Week 8)
* The Data management Module (Week 8)
* The Data Visualization Module (Week 9): This module can be separated and be done parallelly to each group member. This module can contain many functions which are designed by the case of the plotting method. Once all functions are completed and pass the test, then this module would be done shortly by the integration of all functions.
* Complete module testing and final reviewing. (Week 9)
* Some sample use cases and basic analysis can be performed as part of the final deliverable. (Week 10 to Week 11)