**JOURNAL**

Name – BABITA CHAINI

GROUP – 14

Student ID – [x21139211@student.ncirl.ie](mailto:x21139211@student.ncirl.ie)

* Dataset 3 – US Disaster data is extracted using API call in JSON format from URI : https://www.fema.gov/api/open/v1/FemaWebDisasterDeclarations .
* As the data is unstructured data so the data is inserted in Mongo DB by connecting to Mongo Client and using the pymongo library. JSON is converted in BSON format by default in MongoDB.
* I fetched challenge in getting data in proper format from MongoDB as the JSON data stored was in format of List of dictionary. So while extracting data it was coming as 2 lists with all dictionary items inside one of the list. But I required only the dictionary items because that is where my data lies.
* After exploring more about this issue I got to know something about explode a list. Using this technique we can separate data inside the list into different columns and thus can be converted into a data frame.
* Then I did pre-processing and transformation on the data frame. Like I did date-time formatting using to\_datatime function in pandas and extracted only the year from the fields having both date and time. As I required only year data to do further analysis. I also changed the column name to some meaningful name. I also converted all data into a proper format to have a good visualization. I also deleted unwanted columns that are not required for analysis.
* Finally I checked for any NULL values in the data frame but there were none.
* After that I performed some meaningful visualizations using matplotlib, seaborn, pair plots and plotly.
* Finally I identified some columns to put in Azure SQL for further merging with datasets of other team members. For that, I identified 7 columns that can be used for further analysis and created a table US\_Disaster\_Events in Azure SQL and inserted the data of selected column with a column as a Primary key and another as a Foreign Key.
* While using Azure SQL we identified that its subscription gets expired once our azure credit gets exhausted. So for this issue, we created another account after the first account we created got expired.
* Finally after importing datasets in Azure SQL we finally created a final table with the merged data for further analysis. We extracted the merged data into a CSV file.
* I along with other group members then worked on creating visualization for the final dataset.
* Finally after all these processes are done we created our reports and merged it, also we have a Git repository with each group member’s name where we have been committing our codes. Below is the URL to access my code from GIT : <https://github.com/MscDATSasi/Data-Analytics-Programming/tree/master/Babita>
* I Coordinated with all the team members to complete the project.