Sasikumar Jayapal – x21153272

Group 14 - Project Journal

Database Analytics Programming

National College of Ireland

* Managed entire team to complete this project Natural Disaster Analysis.
* Discussed about the project domain, datasets, research questions and driving methodology.
* Datasets: This analysis uses three datasets, two of which are web data and one of which is a csv file.
* Dataset 1: https://en.wikipedia.org/wiki/List\_of\_countries\_by\_natural\_disaster\_risk
* Dataset 2: https://en.wikipedia.org/wiki/List\_of\_natural\_disasters\_by\_death\_toll
* Dataset 3: [https://www.kaggle.com/datasets/](https://www.kaggle.com/datasets/andradaolteanu/country-mapping-iso-continent-region) [andradaolteanu/country-mapping-iso-continent-region](https://www.kaggle.com/datasets/andradaolteanu/country-mapping-iso-continent-region)
* I have created a Data architectural plan and Database schema model for the entire project.
* I have scrapped web data using beautiful soup python library. Identifying tags and fetching relevant data from the web was a challenging one.
* I had trouble configuring MongoDB for staging semi-structured data, so I browsed online for solutions and read more of our MongoDB tutorials. Almost it took a whole day to configure MongoDB.
* The data I have obtained on the web is raw, it is not clean. In order to load them into a database, it must comply with database normalization. I spent most of our time on cleaning and preparing the entire web data, which was an extremely challenging part of this project.
* Our final version of the entire data will be stored in Microsoft Azure SQL database since it's a cloud-based database, which allows the entire team to access it at the same time to load content.
* I have managed to setup Azure SQL database in cloud for the entire team and I had to do the azure setup multiple times as the free credits are running out quickly.
* I collaborated with the entire team and ensured the Azure SQL database contained relevant data for the integration. In addition to integrating multiple tables into Azure SQL Server, I also worked with the team to analyze and visualize the results of our research.
* I have automated all the data processing flows into one single pipeline with one python jupyter notebook.
* I worked with the team to gather the necessary content, and collaborated with the overleaf that I am maintaining to complete the report.
* Continually keeping everyone in the team informed about progress, I helped to solve any blockers if there were any.
* Finally, we have merged all our code into github: https://github.com/MscDATSasi/Data-Analytics-Programming

**Artifacts Created and Worked:**

* DAP\_PRJ.ipynb
* Master\_Disaster\_Integration\_Datasets.ipynb
* Master\_Disaster\_Event\_Visuals.ipynb
* Master\_Disaster\_Events\_Automated.ipynb

**Tables Created and Loaded in Azure:**

1. [dbo].[country]
2. [dbo].[RISK\_COUNTRY]
3. [dbo].[Death\_Toll\_by\_Disaster]
4. [dbo].[Master\_Disaster\_Event]