

Earthquakes

Data Visualization Website



April 1, 2022

Saad Ahmad

**Table of Contents**

1. **Description**………..……………………………………..…..2
2. **Screenshots**………………………………………………....3

Front End………………………………………….........4

Synthetic Data………………………………………….5

1. **Architecture Diagrams**..……………………………….…..6

Earthquakes…….……………………………….……..6

Frequency Predictions………………………….……..6

Magnitude Predictions………………………….……..7

Sentiment Analysis…….…………………...….….…..7

1. **Description**

This project's numerical data comes from the Kaggle API and is saved in Dynamo DB. It contains information on global seismic activity from 1965 to 2016. Although each earthquake has its own unique timestamp, the time period for each earth is classed into either the early or last 6-months of the year. In this approach, we get 5 separate sets of data indicating seismic values for each of the five regions for each 6-month period between 1965 and 2016.The data includes the timestamp of the earthquake's occurrence, such as the first or second six months of any year, the frequency of earthquakes in that region, and the average magnitude of all earthquakes in that region. In short, each data item in the database has four fields: timestamp, region, frequency, and magnitude. The ultimate desired result is to use a choropleth map and a slider to modify the value of year to illustrate the relative seismic activities in these five regions since 1965.

The textual data set is made up of all newspaper stories published between 1999 and 2022 that mention the phrases "earthquake" and "magnitude." This information is kept in a database that has two fields for each data item: timestamp and text. The purpose of this project is to do text sentiment analysis on the textual data and create a graph depicting the change in sentiment for earthquake-related articles in the Guardian newspapers since 1999.

Finally, users will be able to compare, contrast, and analyse various types of historical seismic activity information from throughout the world.

Textual data stored in the database triggers a lambda function that performs sentiment analysis on each entered data item in the database as it is entered and transfers the data to another database table that stores the timestamp and sentiment value of textual data.

There is no need to handle the numerical data because all of the necessary processing, such as updating the timestamp and categorising the earthquake into its geographical region, is already done on the client side before the data is uploaded to the cloud.

Finally, this database data is accessed via a WebSocket API, which invokes lambda functions that query the database and communicate the results to the client. Plotly is used to display the textual data as a filled-line plot, with the area under the line representing the percentage of positive/negative/neutral/mixed data between 1999 and 2022. The numerical data is displayed using a choropleth map, in which the world map appears to be divided into 5 equal parts based on longitude, and the transparency of each region changes as the year value is adjusted using the slider, highlighting the magnitude and frequency intensity compared to the maximum frequency and magnitude observed over the entire time period.

Since the numerical data only went up to 2016, this project used AWS SageMaker to try to forecast future seismic activity values for the next ten years. The SageMaker requires three datasets: a test dataset that contains all of the data, a training dataset that excludes the last 20 values from the test dataset for the reach region, and a validation dataset that is used to query the endpoint and generate predictions.

The data was divided into 5 json lines for each of the five regions, with the start value being 1965, except for the validation set, which had a start value of 1987 because it consisted of the 60 data values in the data set for each region, with all values within a json line being spaced by a 6-month period.

These files were built with JavaScript code and uploaded to AWS S3 to build a training job and, eventually, an endpoint that could be queried for forecasts for the next ten years (2017-2026). A Lambda function was used to call the endpoint that generated the predictions and stored them in DynamoDB, and another lambda function was created to route the predictions data to the front-end code via the WebSocket API.

1. **Screenshots**

This section contains screenshots from the website's front end, as well as visual representations of the synthetic data and the predictions it generates.

* 1. **Front End**

S3 URL: https://cst3130-cw2-frontend.s3.amazonaws.com/earthquake.html

**A picture containing graphical user interface

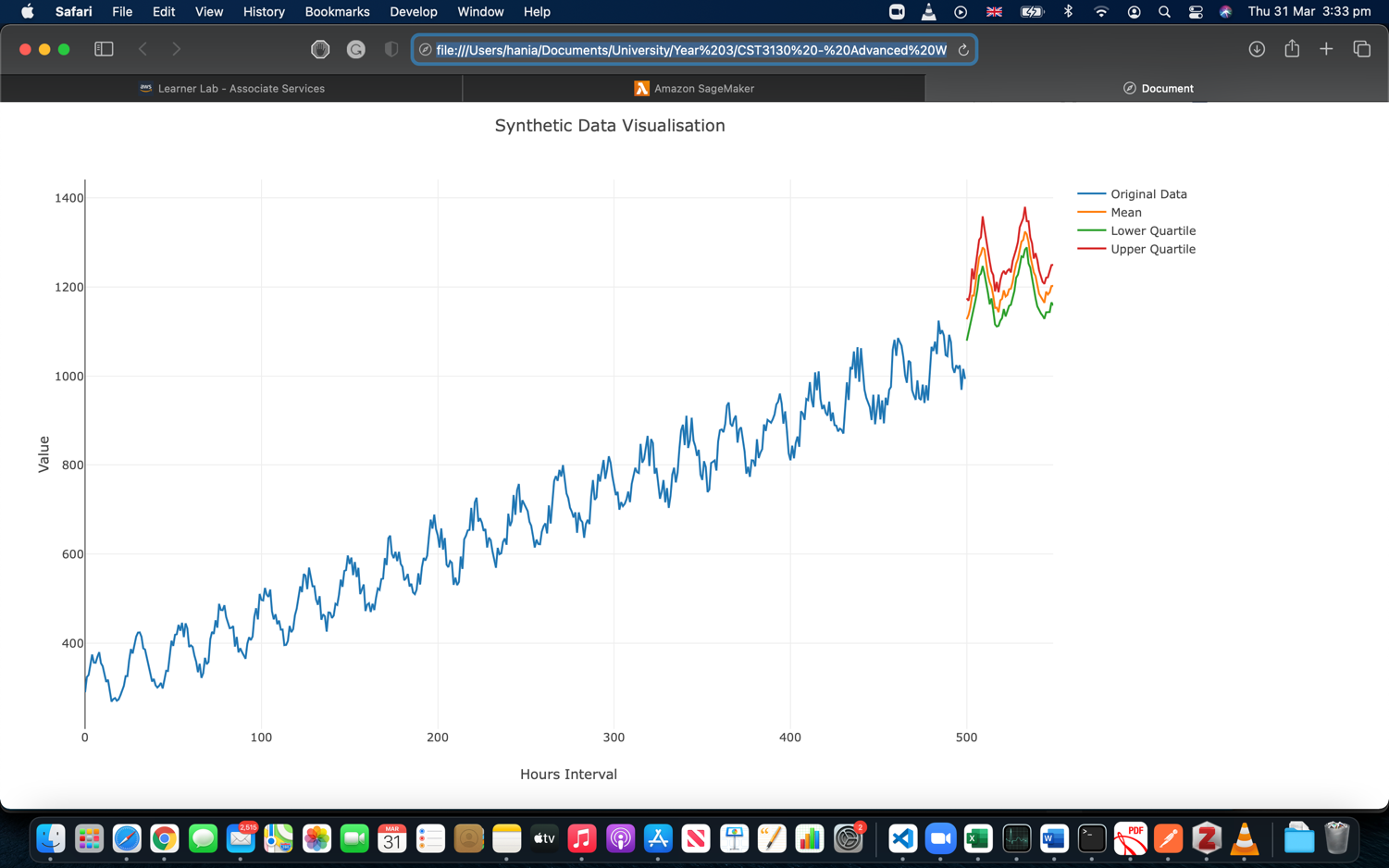
Description automatically generatedA picture containing timeline

Description automatically generated**

**A picture containing graphical user interface

Description automatically generated**

**2.2 Synthetic Data**

****

1. **Architecture diagrams** 
   1. **Earthquakes**

Kaggle

Earthquakes data

Earthquakes data

Type Script Code

Earthquakes

Trigger

Client

Earthquakes data

Connection IDs

wsEarthquakesData

WebSocketClients

Earthquakes data

Earthquakes

* 1. **Frequency Predictions**

Predictions data

Predictions\_EarthquakesFrequency

Predictions data

getPredictionsFrequency

Validation data

Trigger

SageMaker

Predictions data

Client

wsSendPredictionsFrequency

WebSocketClients

Connection IDs

Predictions data

Predictions\_EarthquakesFrequency

* 1. **Magnitude Predictions**

Predictions data

Predictions\_EarthquakesMagnitude

getPredictionsMagnitude

Predictions data

Validation data

Trigger

SageMaker

Predictions data

Client

wsSendPredictionsMagnitude

WebSocketClients

Predictions data

Connection IDs

Predictions\_EarthquakesMagnitude

* 1. **Sentiment Analysis**

Guardian API

Articles headlines

Guardian\_newspaper

Articles headlines

Type Script Code

Articles headlines

GenerateSentiment

Sentiment data

SentimentData

Trigger

Client

WebSocketClients

wsSentiment

Sentiment data

Connection IDs

Sentiment data

SentimentData