Project Report

Mentor: Pooja Bassin   Bhavil Sharma (IMT2021041)

1 Objective

This initiative entails the development of a system dedicated to the generation of case files and the creation of stability modeling visualizations on Tableau.

2 My Responsibility

In this role, my responsibilities encompassed stability modeling, the generation of case-files, and the visualization of results using Tableau. Throughout the project's duration, we conducted experimentation with diverse methodologies to enhance our analytical approach.

3 Sprint Report

Sprint 1: Week-1

- **Sprint Number:** 1
- **Start Date:** 01-08-2023
- **End Date:** 31-08-2023
- **Tasks Planned:**
    - Read about Sustainable Development Goals. Go through the Karnataka Data Lake(KDL) website
    - Go through SDG 3 - Good Health and Well-being data story on the KDL website.
    - Understand the stability modelling code for 1 dimension capability vector [IMR]
    - Understand distance metrics (L1 norm, L2 norm)
- **Tasks Completed:** Completed All the tasks
- **Comments:** I had mis-understood the tasks, and did them Incorrectly.

Sprint 2: Correcting

- **Sprint Number:** 2
- **Start Date:** 01-09-2023
- **End Date:** 05-09-2023
- **Tasks Planned:**
    - Correcting the Sprint-1
- **Tasks Completed:**
    - Corrected the Sprint-1, however, there were some talukas, that were not reading the data, I later corrected this bug, but it took a long time

Sprint 3: After $1^{s}t$ project review

- **Sprint Number:** 3
- **Start Date:** 06-09-2023
- **End Date:** 26-09-2023
- **Tasks Planned:**
    - Implement SM for 2D and a common intervention (ANC)
    - Visualize Stability vs Impact using Python libraries and Tableau Public.
- **Tasks Completed:** Completed both the tasks, however, the error mentioned before was still there

Sprint 4: Post Mid-Sem exams

- **Sprint Number:** 4
- **Start Date:** 04-10-2023
- **End Date:** 10-10-2023
- **Tasks Planned:**
    - Implement SM for 3D- [IMR,MMR, PAW] and a common intervention (HM)
    - Visualize Stability vs Impact using Python libraries and Tableau Public.
    - Prepare the casefiles. Casefile 1 had the column names mapped to an id. Casefile 2 and 3 have the metadata for Single-Variate and Multi-Variate Linear regression for the PIA files.
- **Tasks Completed:** Completed the tasks. The bug found and troubleshooted.

Sprint 5: PSS Talk

- **Sprint Number:** 5
- **Start Date:** 12-10-2023
- **End Date:** 25-10-2023
- **Tasks Planned:**
    - Add dissonance to the SM
    - Talk preparation
- **Tasks Completed:** Gave the talk with Ma'am on $25^{t}h$

Sprint 6: Final Project review prep

- **Sprint Number:** 6
- **Start Date:** 26-10-2023
- **End Date:** 29-11-2023
- **Tasks Planned:**
    - Scaled the impact for 3D files.
    - Various experiments with the SI Score.
    - Preparation for the final Project review.
- **Tasks Completed:** Successfully conducted the talk.
- **Comments:** Some score values were getting too big, so, changed the formula that day, and corrected it the same day.

Sprint 7: Finalizing

- **Sprint Number:** 7
- **Start Date:** 7-12-2023
- **End Date:** 12-12-2023
- **Tasks Planned:**
  - Wrapping up with GitHub and readme
  - Peer review for Nimish. **My Review:** It was okay, his work didn't involve code, so I only checked if his visuallization were working. And they were working.
  - Made this report for final submission.
- **Tasks Completed:** Wrapped up with the work and made the final submission.

4 Final Summary of the sprint

Sprint Summary			
Title	Start Date	End Date	Tasks
Week-1	01-08-2023	31-08-2023	Getting Started-KDL, Previous Code
Corrections	01-09-2023	05-09-2023	Corrected the Sprint-1 Work
Post review-1	06-09-2023	26-09-2023	Visualize Stability vs Impact for 2D
Post Mid-Sem	04-10-2023	10-10-2023	Implement 3D SM, Visualize Stability vs Impact, Prepare casefiles
PSS Talk	12-10-2023	25-10-2023	Gave talk on 25th
Review-3 Prep	26-10-2023	29-11-2023	Scaled impact for 3D, SI$_S$score
Wrapping Up	7-12-2023	12-12-2023	GitHub, Readme and Peer-Review

5 Source Code details

1. **Link:** Policy Support System

2. The code generated some excel files, which were visualized using Tableau. The tableau workbooks are in the github repository.

3. The code is a simple notebook in python. We will just have to press 'Run All' at the top, and the files will get generated.
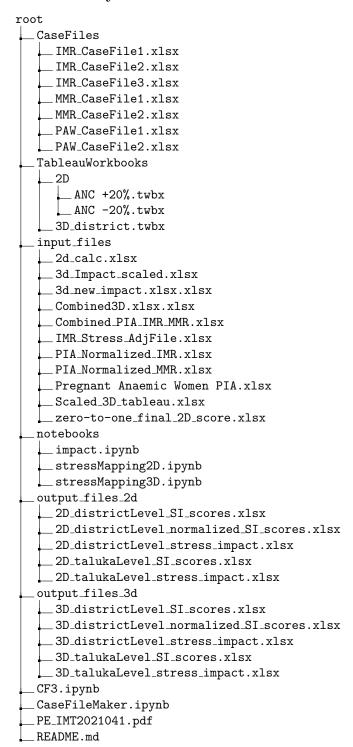
6 Code Structure:

Purpose of Folders:

1. CaseFiles: Contain the final output of casefiles generated. They have the metadata of the given files.

2. Tableau Workbooks: Have the final visualizations for the 2D and 3D models. They contain Stability Modelling and Scores.

3. input_files: Have all the files needed by the notebooks as input.

4. notebooks: Contain the main source code for the project. The take input PIA files from input_files and process them. Then, they store the output in the output_files. There is a file to scale the impact.

5. output_files_2D: Here, we have the 2D output files for both district and taluka level.

6. output_files_3D: Here, we have the 3D output files for both district and taluka level.

6.1 Directory Tree:

```
root
├── CaseFiles
│   ├── IMR_CaseFile1.xlsx
│   ├── IMR_CaseFile2.xlsx
│   ├── IMR_CaseFile3.xlsx
│   ├── MMR_CaseFile1.xlsx
│   ├── MMR_CaseFile2.xlsx
│   ├── PAW_CaseFile1.xlsx
│   ├── PAW_CaseFile2.xlsx
├── TableauWorkbooks
│   ├── 2D
│   │   ├── ANC +20%.twbx
│   │   ├── ANC -20%.twbx
│   ├── 3D_district.twbx
├── input_files
│   ├── 2d_calc.xlsx
│   ├── 3d_Impact_scaled.xlsx
│   ├── 3d_new_impact.xlsx.xlsx
│   ├── Combined3D.xlsx.xlsx
│   ├── Combined_PIA_IMR_MMR.xlsx
│   ├── IMR_Stress_AdjFile.xlsx
│   ├── PIA_Normalized_IMR.xlsx
│   ├── PIA_Normalized_MMR.xlsx
│   ├── Pregnant Anaemic Women PIA.xlsx
│   ├── Scaled_3D_tableau.xlsx
│   ├── zero-to-one_final_2D_score.xlsx
├── notebooks
│   ├── impact.ipynb
│   ├── stressMapping2D.ipynb
│   ├── stressMapping3D.ipynb
├── output_files_2d
│   ├── 2D_districtLevel_SI_scores.xlsx
│   ├── 2D_districtLevel_normalized_SI_scores.xlsx
│   ├── 2D_districtLevel_stress_impact.xlsx
│   ├── 2D_talukaLevel_SI_scores.xlsx
│   ├── 2D_talukaLevel_stress_impact.xlsx
├── output_files_3d
│   ├── 3D_districtLevel_SI_scores.xlsx
│   ├── 3D_districtLevel_normalized_SI_scores.xlsx
│   ├── 3D_districtLevel_stress_impact.xlsx
│   ├── 3D_talukaLevel_SI_scores.xlsx
│   ├── 3D_talukaLevel_stress_impact.xlsx
├── CF3.ipynb
├── CaseFileMaker.ipynb
├── PE_IMT2021041.pdf
├── README.md
```

7 Libraries

1. Pandas - 2.0.3

2. NetworkX - 3.1

3. sklearn

4. statsmodels

8 System Requirements

This code simply runs a python code, so we just need python3. If we don't have anything, we can simply run the code on a google collab.

9 Challenges faces(Bugs and corrections)

1. **Graph Generation:** In the initial version of the Stability Modeling code, only the 1D model was implemented. An issue arose in the Adjacency file where a taluka was missing at the 81st index, resulting in the creation of a graph with numerous double edges beyond the 80th index. This issue posed a significant challenge, as it initially led me to suspect that pandas might not be accurately reading information from the provided PIA files.

   Upon manual inspection of numerous talukas, I identified the error, which originated from certain talukas lacking values in my code. Detecting this discrepancy was particularly challenging in the 1D code since all values fell within the [0, 1] range. However, as the dimensions increased, the error became more conspicuous.

2. **Large Values of Scores:**Previously, the calculation of taluka scores involved aggregating scores that were assigned disproportionately large values for some districts. Subsequently, we opted for a more refined approach by directly calculating scores from the District-level data.

10 Talks given

On October 25, 2023, I gave a talk with Pooja Ma'am, my mentor, pertaining to the presentation on PSS (Policy Support System). The talk encompassed an in-depth analysis of the citations incorporated in the paper, as well as a comprehensive review of the code I had written for the project.