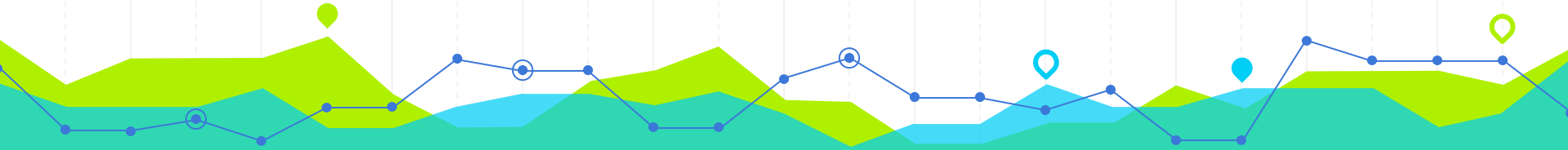**Using Employee Logs and Retention Data to predict Industry Correlation and Corporate Norms**
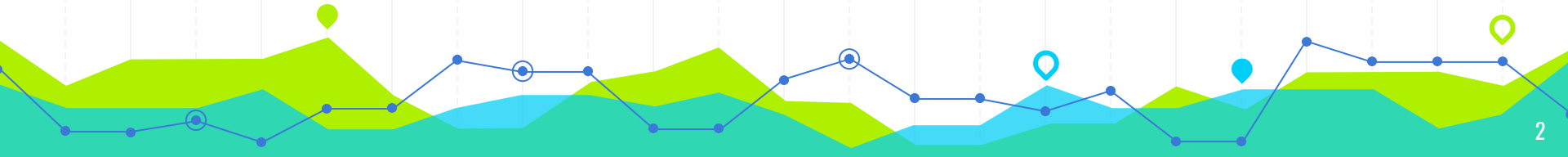
Aryaman Mishra
19BCE1027
Guide: Dr. Suguna M

OUTLINE

Area of the Project

This project would make use of the following area:

- Sentiment Analysis
- Recommendation Techniques
- Function Interactivity
- Data Visualization
- Open Source API Usage
- Machine Learning Techniques
- Data Science Monitoring
- Optimization Techniques
- Pivot Tables
- Data Filtering

Background of the Problem

As the aspects of digitization has reached into the education and job industry, it has become vulnerable to hidden discrepancies regarding the candidate's abilities to perform upto the standards expected and employer expectations with said candidates. This includes criterias like employee engagement, difficulty of tasks being handed out, how any grievances are dealt with, information security and confidentiality. These are some issues which are not generally documented digitally-leading to stagnation of progress in the company and failure of vital operations in their respective fields.

Using sentiment analysis containing employer-employee review, we will cross reference them by their unique IDs to measure their chances in companies of either lower tier or are required to shift their objectives to another industry.This would also help us in providing an insight of their educational backgrounds and find the correlation between what Industries require and the employee's educational background.

By using the various technique mentioned prior to the project,we will analyze the skills required for job,employee behaviour,experience required and met for an individual,find necessary metrics for how salary is given out in each industry and integrate geo-location techniques to cater to an individual's need to find an appropriate job and whether or not he or she will be paid substantially and for how much time will he/she be able to persist in the work environment of that industry.

Problem Statement

In this project, we will pull Jobs related datasets and Employee Review Datasets to pivot and merge them and use them to reflect presence of corporate norms which needs to be eradicated from industries ranging from Software Engineering, Media, Information Technology, Medical, Textile and Pharmacies.

We will then clean the data, as it is highly possible that such datasets may contain dirty values which means that data-cleaning will be a prioritized objective in the beginning phases of this project.

Thus,our Problem Statement states:

"

Using the availability of employee logs,job openings and employer-employee cross referenced data, the possibility of deducing retainability of an employee, gender pay-gap and skill-set requirements can be deduced by understanding the important metrics in order to aid people to assess or find their quality-oriented jobs from various industries,all with the use of job postings and employee data."