# DSCI551 - Team 67

## Team Members and Background

**Armand Patel**
- Skilled in Data Analysis and Probability and Statistics techniques. Domain expertise in healthcare and technology. I have experience with python, HTML/CSS, R studio, and general project communication.

**Joshua Payapulli**
- Academic background in computer science. Comfortable with Python, Java, C, SQL, JavaScript. Professional experience in regulatory compliance and finance as a Business Analyst and Consultant. Also have close to a year's experience as a Full-Stack Software Engineer.

**Paul Yoo**
- Project management experience in manufacturing industry digitization projects with domain knowledge in Business, Finance, and Strategic decision-making. Skills in project management, business insights, python programming, MS Office, and IT project communication.

---

# Tweet and Sentiment Analysis with Databases

**Jan 24, 2024 - May 3, 2024**

## Project Introduction

### Project Requirements:

This project entails the development of two interconnected components. The first is a user interface tailored for database managers, providing functionalities for inserting, deleting, and modifying data within the database. The insert function should intelligently direct data to the appropriate database, potentially based on the hash value of the partition key for optimized data organization. The second objective is to develop a web-based application designed for end users, which leverages the database to manage its data. The specific functionality and type of data managed by the application are left open-ended.

### Project Objectives:

Sentiment Analysis of tweets related to the presidential election using the Twitter API. The first database would be stacked with raw data from tweets collected directly from Twitter

with keywords related to the upcoming presidential election. The second database will classify tweets into two different categories, positive and negative, and will be later used to find out the possibility of being elected for each candidate.

## Team Responsibilities

Armand Patel
- Leader of the team, responsible for making overall decisions on the direction the project should take and what the overall goals are. Responsible for establishing connection to our main sources of data (most likely Twitter API) and putting these into our databases.
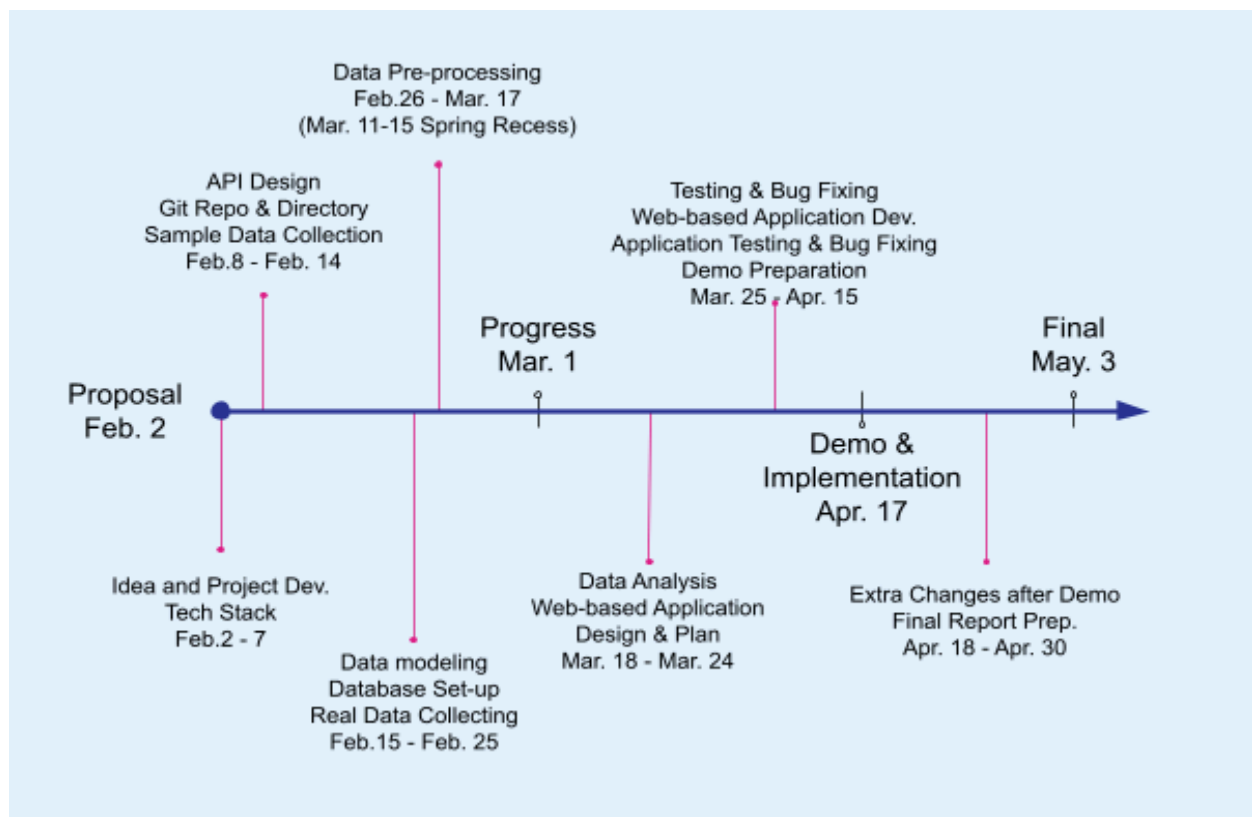
Paul Yoo
- Project manager, responsible for time management, scheduling, and set-up meetings. Responsible for data cleaning, after the raw data is inserted into the databases, preprocessing these tweets so that we can perform sentiment analysis on them.

Joshua Payapulli
- Technical Chief, will work on the end-to-end design of the web application and will define the logic for how to collect relevant tweets and perform sentiment analysis on them.

## Project Timeline

# Milestones

1. Project environment set-up

   1) Idea clarification

   2) Git repository set up.

   3) Establish data extraction pipeline, most likely through Twitter API.

   4) Define logic for which tweets to be considered as relevant.

2. Real data collection and preprocessing

   1) Data collection from Twitter

   2) Database/tech stack environment set-up

   3) Pre-processing/cleaning collected data

3. Analysis & Web-based Application Design

   1) Selecting appropriate ML Model and overall Algorithm.

   2) Sentiment analysis of tweets implementation.

   3) Analysis feedback & Bug Fix.

   4) Web-based application design and development planning.

4. Web-based application development

   1) Web-based application development.

   2) Application testing.

   3) Application Bug Fix & Feedback.

   4) Demo preparation.

5. Final Report Preparation

   1) Web-based application feedback from Demo.

   2) Final Report Writing.