# 3XA3 L03 Group 17
# Project Approval Document

Karuka Khurana, Utsharga Rozario, Samarth Kumar

January 24, 2022

**Team Name:** Group 17

**Group members/Mac ID:**
Karuka Khurana (khurak1)
Utsharga Rozario (rozariou)
Samarth Kumar (kumars38)

**Original Project Name:** Notion PDF Scrapper

**Software Purpose:** The purpose of this project is to implement a PDF scraping feature to an existing open source React Notion renderer.

**Software Scope:** The scope of this project is to design, develop, and implement a Python based PDF scraper that would scrape a document and extract important dates/deadlines. We will create a dashboard that lists all the important dates and categorize them based on courses a student is taking.

**URL for original project:** https://github.com/NotionX/react-notion-x

**Any specialized hardware requirements?** No

**Any software license required that McMaster does not own?** No

**Programming language:** React, TypeScript, Python

**Is programming language feasible for your team?** Yes

**Is the domain knowledge understandable within one term?** Yes

**Number of lines of code:** Approximately 10,000

**License:** https://github.com/NotionX/react-notion-x/blob/master/license (MIT License)

**License allows public redevelopment?** Yes

**Can you compile the original projects source code?** Yes

**What would be some test cases for the existing software?**

– Create sample Notion web pages and render them using the existing software.
– Create a table/list of dates in a sample Notion web page

**What changes you intend to make to the project?**
Our group intends to integrate some minor features such as YouTube integration and on-hover preview to the pre-existing Notion React renderer. The YouTube integration feature would allow a user to paste a YouTube URL within a Notion page and the video could be viewed in a player. The on-hover feature would allow a user to paste any URL in a Notion page and once a user hovers over the link, a preview of the website will be displayed. A major feature we hope to implement is a PDF scraper. A user can upload a PDF document (e.g. course outline). The software will scan the document for important due dates and deadlines by identifying pre-defined key words. It will then take this information and create a dashboard for the user that will visualize all the dates in an organized, easy-to-read manner.