

## **IRisk Lab Data Discovery and Consolidation - Task 2**

### **Week 12 Report**

*Nov. 14*

*Rohit Valmeekam*

#### **What was accomplished**

- ❖ Conducted discussions with Group 1 to understand their specific dataset merging needs.
- ❖ Clarified the key columns or criteria for merging the datasets effectively.
- ❖ Demonstrated how to use Pandas to read specific sheets from Excel files, showcasing its flexibility in handling various data formats.
- ❖ Guided Group 1 through the process of loading datasets into Pandas DataFrames.
- ❖ Merging Techniques:
  - Introduced different merging techniques such as inner, outer, left, and right merges based on the desired outcome.
  - Explained the importance of choosing the appropriate merging method to retain relevant information.
- ❖ Handling Missing Data:
  - Addressed strategies for handling missing or mismatched data during the merging process
- ❖ Optimizing Performance:
  - Shared tips on optimizing performance, especially for large datasets, to enhance the efficiency of the merging process.
- ❖ Teaming up with Zifan, worked on creating an API interface for mapping functionalities
  - Created tech stack for product : React.js for frontend connected to an API created on the backend using Flask and Python. The web scraping tool would be made in Python using the BeautifulSoup module

#### **Next Steps**

- ❖ Now that we have the full plan, we build out the rest of the interface