**REPORT**

To begin setting up the project's tasks, the team meeting was held. The targeted scope of the project and its development phases were settled upon at the meeting's conclusion. In the end, we can claim that these project-specific goals were accomplished through productive collaboration.

**Rishikesh Waghela** – Finding legal website for Web Scraping, Web crawling and storing data into Mongo dB atlas (cloud)

**Shreya Padte** – Data Cleaning

**Saurabh Gupta** – Frame Question and finding Insights

**REFLECTION OF LEARNING**

We have gathered the data from the site for resale properties, makaan.com. First , we tried using Scrapy library for web crawling. In Scrapy, a "spider" is basically a computer program that specifies how to crawl a website, which pages to visit, and how to gather data from those pages. To handle complicated circumstances like pagination, login pages, and more, scrapy spiders can be tailored. Additionally, Scrapy has functionality for handling concurrency, proxies, and request and response data. While 'Beautiful Soup' can assist in extracting particular elements from a target web page. So, first we try using spider that scrap around web and we were unsuccessful as the site block us for scraping [makaan.com].

Therefore, we use beautiful soup to parse html and find all the required tags. We have fetch three cities data like Mumbai, Delhi and Chennai were the properties are for sale. In which, we have scrap tags like status, Apartment name, prices, Number of bhk, area , per sqft area , longitude and latitude. We scrape 5000 pages for Mumbai, 2000 pages for Delhi and Chennai. Secondly, save all this data in tabular format making three different csv file for our data.

Thirdly, we concatenate the three csv into one and started the data cleaning process. After that, we uploaded our cleaned data to the MongoDB cloud so that other group members could access it. We connected directly to mongo dB Atlas(CLOUD) from our Visual studio code. Also, we can make API of this data which can be further use by anyone for prescriptive , descriptive analysis.

Finally, we answered framed questions through finding the insights of the data while manipulating the data. Three question we answered like most expensive areas in Mumbai, upcoming Projects in three cities which are in sale, and all properties of Lodha group.

In conclusion, this project can be robust and deploy for many aspects like for predictive analysis like the prices for the area while gather more information. Also, fetching the details like number hospitals in area ,crime rate, public sectors which affects this prices. With such, information we can conclude why this areas have such prices.