**BÁO CÁO TUẦN**

Thời gian: 22/11 – 29/11/2020

**Danh sách nhóm:**

1920058 Lê Tất Thiện

1927038 Trần Phương Tĩnh

1920068 Dương Quang Tuấn

1927040 Nguyễn Đức Tuấn

**Các công việc đã thực hiện**

* Thiện

Setup backend

Hiện thực crawl dữ liệu bằng selenium và cheerio

Viết báo cáo và document

Crawl dữ liệu trên trang batdongsan.com.vn và muabannhadat.vn

Làm module xuất dữ liệu ra file csv

Viết report, hướng dẫn sử dụng

Tạo sample dữ liệu 400 record từ batdongsan.com

* Tuấn (Dương): crawl dữ liệu trên trang batdongsan321.com
* Tuấn (Nguyễn): crawl dữ liệu trên trang 123nhadat.vn
* Tĩnh: crawl dữ liệu trên trang homedy.com, viết hàm convert thời gian sử dụng cho các trang không ghi rõ định dạng ngày

**Các khó khăn gặp phải**

* Ban đầu nhóm sử dụng Selenium để tiến hành crawl dữ liệu, nhưng khi sử dụng vòng lặp với selenium thì có vấn đề với bất đồng hộ, kết quả chạy không được như mong muốn
* Nhóm chuyển qua sử dụng Cheerio và nhận thấy thời gian crawl tăng đáng kể và khắc phục được vấn đề trên.

**Dự kiến công việc tuần tới**

* Triển khai Frontend để thuận tiện cho việc crawl các dữ liệu cần thiết
* Sử dụng MongoDB để lưu trữ và graphQL để thao tác và truy vấn API
* Hoàn thiện module xuất file csv