

Project gồm 4 phần:

1. vietnambusinessinsider: folder crawler data.
 - file code crawl data: vnbi.py
 - pipeline đưa vào CSDL MySQL: pipelines.py
2. SQL: folder chứa truy vấn và tạo bảng.

Dữ liệu được lưu trữ vào CSDL MySQL version 5.1.1

- post.sql: là file để import database.
 - 2 file còn lại dùng để xem truy vấn và kết quả của 2 câu thống kê theo yêu cầu đề bài dưới dạng json:
 - Thống kê số lượng bài viết theo tác giả:
_select_author_count_from_post_group_by_author_order_by_count_de_202305271050.json
 - Thống kê số lượng bài viết theo tháng trong vòng 1 năm:
_select_month_date_post_as_month_year_date_post_as_year_count_t_i_202305271050.json
3. vnbi_analysis.ipynb: file phân tích theo 2 câu thống kê sử dụng thư viện pandas, matplotlib.
 4. post.json: file chứa data được crawl dưới dạng json.