

Question_1:

Ans:

project:props_AMC data automation

Explanation:

It is a Data Automation Project developed in Python Django Framework. Design for internal use of the company for the Data Crawling ,visualization & Verification from Various sources related to Real Estate transactions from various sites .

Technology stack used for development:

- Used python as programming language for backend and django as web framework for fronted end we used html,css,js in django templates.
- For the web scraping part we have used some libraries like requests and selenium and also integrated Django celery and rabbitMQ for task queuing and scheduling tasks.
- For data cleaning and data formatting we have used regex and pandas df also and also used some libraries for data extraction from pdf files and image captcha extraction like camelot and easyocr.
- Integrated webhook for otp handling and some third party proxy services also.
- For data storage we have used Postgres as structured database and amazon s3 buckets for file storage also created ETL pipe line for data transfer using Apache kafka from postgres to NOSQL database Cassandra.