

# Capstone Project Data Platform

Data Process:

1. Activate Data Pipeline
  - a. Activate Zillow Scraper – Databrick
    - i. If the last data in the database is more than 1 month old, scrape for new data. Otherwise deactivate scraper.
  - b. Activate New Construction (Building Permits) Scraper – Databrick
    - i. If there are any new files available on the US Census website, scrape those files. Otherwise deactivate scraper.
  - c. Activate US Census API call for Median Income – Databrick
    - i. If there is a new year available for the Income Data then scrape that year. Otherwise deactivate scraper.
  - d. Activate St. Louis Fed Scraper – Databrick
    - i. Once a month, scrape the page and check if any new information is available.
2. If there are data updates, clean the data and send it to the data lake.
3. From the data lake, stream data to a new blob for the Zillow data to simulate data streaming, and pass all data to the database.
4. Gather data from the database for the dashboard and machine learning models.

