# American University *of* Armenia

**Title:** Predictive Analysis of Apartment Conditions Using Image Data

**Submitted by:** Nemrout Safarian and Petros Tepoyan
**Supervisor:** Arman Asryan
**Program Chair:** Habet Madoyan
**Department:** Data Science, American University of Armenia

---

## Executive Summary:

This report outlines the progress made on the capstone project under the guidance of our supervisor, Arman Asryan. Our project aims to develop a sophisticated machine learning model for real estate price prediction, leveraging scraped data from various real estate websites. We detail the tasks completed, including data acquisition, image processing, and preliminary literature review, as well as our next steps towards feature engineering and model training.

---

## Introduction:

The real estate market is a dynamic and complex field, where accurate price predictions can significantly impact investors, buyers, and sellers. This capstone project seeks to enhance prediction accuracy by incorporating a wide array of data points, including images stripped of watermarks, to provide a clearer analysis. Our approach is grounded in a meticulous literature review and is being developed with adherence to the best practices in software development.

---

# Work Completed:

- **Data Acquisition:** We have successfully scraped various real estate websites to collect extensive data sets that include property prices, utilities, amenities, and geolocations. The scraping process also yielded a substantial number of property images, essential for our image analysis algorithm.

- **Image Processing:** An algorithm was developed to remove watermarks from the scraped images. This step is crucial for ensuring the accuracy of the subsequent image-based price predictions, as watermarks can interfere with the model's performance.

- **Literature Review**: We conducted a thorough literature review to understand the existing methodologies and to guide the formation of our report. This review has laid the foundation for our approach and will be instrumental in writing our final paper.

---

# Repository and Code Development

A dedicated repository has been created for this project, which is closely monitored by our supervisor, Arman Asryan. Continuous integration and regular feedback have been established through this platform. The codebase has been refined to follow object-oriented programming principles, protocols, and pipeline creation for efficient data handling.

---

# Immediate Next Steps:

- Feature Engineering: We will begin processing the scraped data to identify and construct meaningful features that can be used for machine learning models.

- Model Training: Concurrently, we will start training our predictive models with the processed data.

---

## Ongoing Considerations:

- Data Quality: A significant focus will remain on the integrity and quality of the data, which underpins the entire project.
- Project Timeline: While we will ensure thorough data analysis, we also recognize the need to adhere to our timeline. Data scraping, particularly the image collection, has taken additional time due to the complexity of handling multimedia content. We are adjusting our schedule accordingly to accommodate this without compromising other project components.
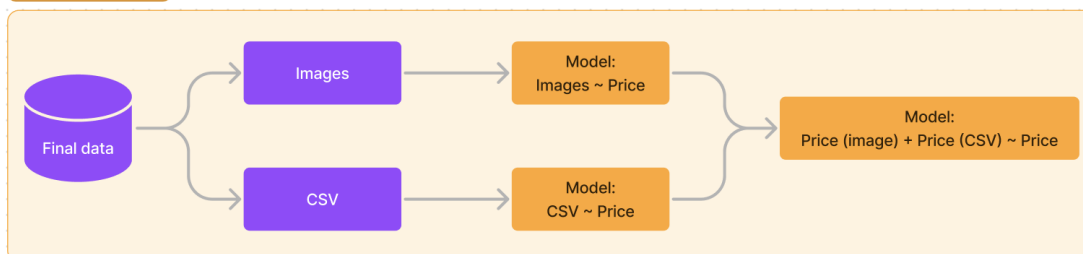
## Project Roadmap:

Attached below is the project roadmap that outlines the major milestones and the projected timeline for the completion of the capstone project. This visual representation serves as a guide for the project's progression and helps ensure that all team members and stakeholders are aligned with the project's objectives and deadlines.
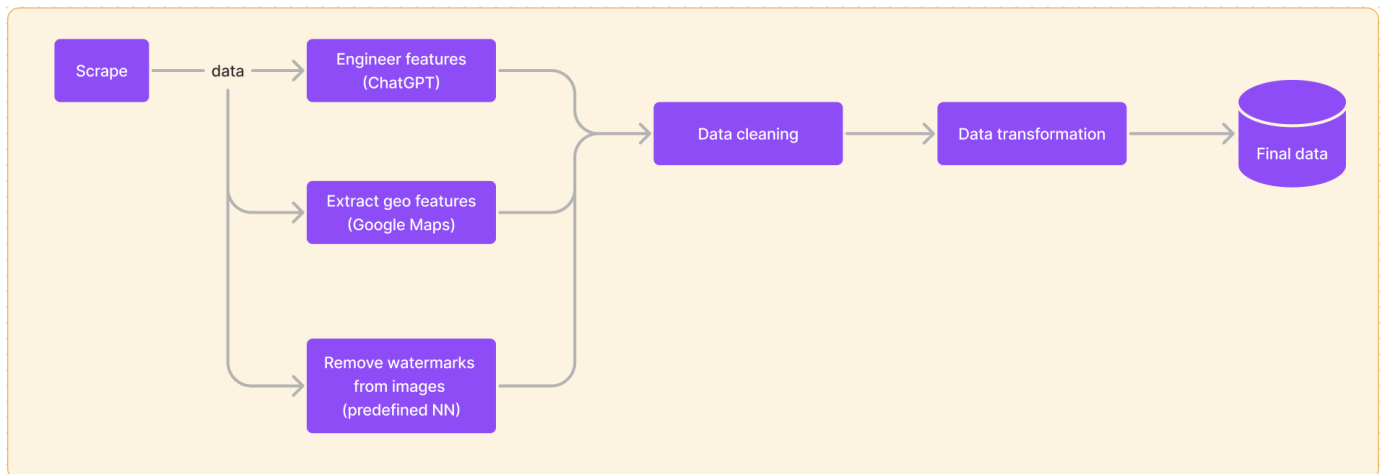
## Long-term Goals:

- Completion of the Feature Engineering Phase
- Model Optimization and Validation: Following feature engineering, we will optimize and validate our predictive models to ensure accuracy and reliability.
- Final Report and Paper Preparation: The insights and findings will be compiled into a comprehensive report and potentially a paper suitable for publication.

---

**Signed,**

Nemrout Safarian and Petros Tepoyan