**PROJECT DOCUMENTATION**

**Table of Contents**

1. Background
2. Overview
3. Problem Statement
4. Project Objectives
5. Project Approach
6. Project Deliverables
7. Methodology
8. Data Gathering
9. Data Cleaning
10. Data Merging
11. Conclusion
12. Result
13. Limitations and Challenges.
14. **BACKGROUND**

**Overview**

Mappa is a company that aims to redefine the experience of finding a new home by using the power of data and technology. Their goal with this project is to create a master list of listed properties across the Greater London area (United Kingdom) in one dataset. The purpose of this dataset is to provide house hunters with a comprehensive and up-to-date list of available properties, enabling them to make informed decisions and secure the best value for their money.

**Project Objectives**

The main objective of this project is to extract and blend property datasets across the Greater London area in one dataset. This involves scraping data from various sources, cleaning and processing the data, and then combining it into a single, comprehensive dataset. The specific objectives of the project are:

1. To identify and scrape property data from various sources across the Greater London area.
2. To clean and pre-process the scraped data to ensure consistency and accuracy.
3. To combine the scraped data into a single, comprehensive dataset.
4. To make the final dataset available for use by house hunters.

**Project Approach**

To achieve the project objectives, we will be using web scraping techniques to extract property data from various sources, including property listing websites and real estate agents' websites. We will then use data cleaning and pre-processing techniques to ensure that the data is consistent and accurate. Once the data has been cleaned and pre-processed, we will merge the data into a single dataset using data blending techniques. Finally, we will make the dataset available for use by house hunters.

**Project Deliverables**

The main deliverable of this project is the comprehensive dataset of listed properties across the Greater London area. This dataset will be made available in a format that can be easily accessed and used by house hunters. In addition, we will also provide documentation on the data sources used, the data cleaning and pre-processing techniques used, and the data blending techniques used to create the final dataset.

1. **METHODOLOGY**

For this project, we followed a three-stage approach, which included:

**Data Gathering**

1. Data cleaning, and
2. Data merging.