Project Report: Melbourne Development Activity Monitoring

1. Introduction

This document presents a comprehensive analysis of the development activity in Melbourne, focusing on commercial and residential projects and their status from the 2013 dataset.

2. Data Overview

The dataset consists of 1,424 records spanning 42 features, encompassing development type, location, status, and other relevant attributes.

3. Data Preprocessing

The data underwent several preprocessing steps including cleaning, normalization, and transformation to ensure high quality and consistency for the analysis.

4. Analysis Results

4.1 Median Values

Median values for key numeric features are listed below, providing insight into the typical property attributes within the dataset:

- **Clue Block:** 432.0
- **Property ID:** 109,147.0
- **Floors Above Ground:** 8.0
- **Residential Dwellings:** 4.0
- All other residential and commercial attributes like studio apartments and office spaces showed medians of 0, reflecting a sparsity in those specific developments.

4.2 Status Distribution

The development status of properties indicated:

- **Applied:** 6.67%
- **Approved:** 17.21%
- **Completed:** 69.73%
- **Under Construction:** 6.39% These ratios underscore the progression and conclusion of most development activities by 2013.

4.3 Clustering Outcomes

The K-Means and Hierarchical Clustering techniques provided the following insights:

- **Optimal Number of Clusters (K-Means):** 10
- **Best Silhouette Score (K-Means):** 0.9331
- **Purity Score (K-Means):** 24.7%

- **Purity Score (Hierarchical Clustering):** 24.7% These results reflect moderate effectiveness in the clustering, indicating overlapping features across different types of developments.

4.4 DBSCAN Results

DBSCAN clustering was applied with varying parameters, the best results being:

- **Optimal EPS:** 30
- **Optimal Min Samples:** 6
- **Number of Clusters Identified:** 2
- **Max Silhouette Score:** 0.579
- **Purity Score:** 23.9% These metrics suggest that DBSCAN was less effective compared to K-Means and Hierarchical clustering, likely due to the high density and overlap in the dataset's features.

5. Conclusions

The detailed analysis offers a thorough understanding of Melbourne's development landscape, revealing a predominant completion of projects and a complex overlap in the features of development activities. The insights from clustering highlight the diversity in development types and can guide future urban planning initiatives.