

Capstone 3 Project Proposal

Wesley Hall - Springboard Data Science Track

Analyzing the Impact of Subway and Park Proximity on Los Angeles Residential and Commercial Property Values

1. Problem Statement

With urban expansion and evolving transportation infrastructure, proximity to amenities such as subway stations and parks significantly influences property values. The Los Angeles Metro is currently expanding its subway network, and with growing urbanization, the relationship between public transport, green spaces, and property valuation is becoming more crucial.

However, quantifying how proximity to these amenities affects property values, especially across different property types (commercial vs. residential), remains unclear. This project will seek to answer the following questions:

- How does proximity to subway stations and parks affect property values in Los Angeles?
- Is there a significant difference in this impact for residential versus commercial properties?
- Do properties within certain proximity bands appreciate at a different rate compared to those farther away?

2. Approach

This project will take a data-driven approach to analyze the relationship between proximity to subway stations/parks and property values using real estate data. I will employ statistical techniques such as regression analysis and clustering methods (KMeans) to find patterns. The analysis will include:

- *Exploratory Data Analysis (EDA)*: Identifying trends in property values based on proximity to subway stations and parks across multiple years.
- *Regression Modeling*: To quantify the influence of distance on property prices.
- *Clustering*: To identify clusters of properties with similar value trends and proximity metrics, using techniques such as KMeans to group similar properties.
- *Time Series Analysis*: To determine whether properties closer to subways or parks have appreciated at a higher rate over time, compared to those farther away.

3. Dataset

The primary dataset includes the following:

- *Assessor Parcel Data (2006-Present)*: This dataset contains information on property types, land improvements, and assessed property values. It includes geographical coordinates, making it ideal for spatial analysis.
- *Proximity to Subway and Park Data*: The dataset includes calculated distances between properties and the nearest subway station and park.
- *Cleaned Combined Dataset*: This dataset merges property data with proximity metrics, focusing on commercial and residential properties in the Los Angeles Metro area.

4. Process and Methodology

The steps involved in the analysis are aligned with the Data Science Methodology (DSM):

- **Problem Identification:** The relationship between proximity to subway stations/parks and property values.
- **Data Collection and Cleaning:** The combined dataset has been preprocessed to include property information, proximity calculations, and categorized distance bands. This data will undergo additional filtering (e.g., limiting to recent years and relevant property types like commercial and residential).
- **Exploratory Data Analysis (EDA):** Initial visualizations will reveal trends and outliers in property value distributions. I will compare the average property values across proximity bands for both subway stations and parks.
- **Modeling and Clustering:** Regression models will be developed to assess the influence of subway and park proximity on property values, controlling for property type, square footage, and other factors. KMeans clustering will be employed to identify groups of properties with similar value appreciation trends.
- **Time Series Analysis:** By examining property values across multiple years, I will assess if properties closer to subways appreciate faster over time than those farther away.
- **Model Validation:** The models will be validated using standard metrics such as RMSE for regression analysis, and visualization will help interpret model outputs.

5. Expected Outcomes

The key deliverables of this project will include:

- A regression model that quantifies the impact of proximity to subway stations and parks on property values.
- Clusters of properties with similar value trends based on proximity metrics.
- Time-series visualizations showing how property values evolve in relation to subway proximity over multiple years.
- Clear, actionable insights on how real estate developers, urban planners, and investors might use proximity data to forecast property value changes.

6. Conclusion

This project aims to provide valuable insights into how urban development and proximity to amenities such as public transit and parks affect property values. By leveraging statistical analysis, clustering, and time-series methods, I aim to deliver a comprehensive view of how these factors influence both commercial and residential properties in Los Angeles. The findings of this project could help real estate investors and city planners make more informed decisions regarding future property developments and infrastructure investments.