

Data Collection and Preprocessing Phase

| | |
|---------------|--|
| Date | 11 March 2025 |
| Team ID | LTVIP2025TMID25127 |
| Project Title | Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau |
| Maximum Marks | 2 Marks |

Data Collection Plan & Raw Data Sources Identification Template

I collected the data from Kaggle and ensured its quality by handling missing values, removing duplicates, and standardizing formats. Additionally, I validated data consistency, integrated relevant external datasets, and created derived variables like house age to enhance analysis accuracy.

Data Collection Plan Template

| Section | Description |
|-----------------------------|---|
| Project Overview | The project uses tableau to visualize housing market trends, analyzing sale prices and key factors to simplify real estate data for better decision-making. |
| Data Collection Plan | I have collected the data from Kaggle |
| Raw Data Sources Identified | The data source is from Kaggle which is of 24 fields and 21609 rows |

Raw Data Sources Template

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|-------------|---|---|--------|---------|--------------------|
| Kaggle | The Data is about Housing Sales Price , renovations etc | https://www.kaggle.com/datasets/rituparnaghosh18/transformed-housing-data-2 | CSV | 2.41 MB | Public |
| Kaggle | The dataset provides housing market data in California, including location, housing features, demographics, and median house values | https://www.kaggle.com/datasets/shibumohapatra/house-price | CSV | 1.18 MB | Public |