Final Project Report Template

1. Introduction

The real estate market is influenced by various factors such as house age, renovation status, number of bedrooms and bathrooms, and overall size. This project aims to analyze housing market trends and visualize key insights using Tableau to better understand how different features impact sale prices**.**

1.1. Project overviews

The dataset contains Transformed housing data and 21,609 house sale records, including Property features such as Sales price, area, bedrooms, bathrooms, floors and location. There are a total of 31 columns, out of which Sale Price can be supposedly taken as a dependent variable. The other variables are different features, locations and date, etc. regarding the houses. This project, "Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau," aims to explore and analyze housing market trends using the Transformed Housing Data 2 dataset from Kaggle. The objective is to identify key factors influencing house prices, such as location, size, number of bedrooms, bathrooms, floors and basement area.

By leveraging Tableau, the project will create interactive dashboards, story, bar chart, histogram, summary dashboard to visualize patterns, compare regional price variations, and gain insights into how different features impact house sale prices. The analysis will help in making data-driven decisions for buyers, sellers, and real estate professionals.

1.2. Objectives

- Identify **key factors** influencing house prices.
- Analyze the **effect of renovations** on property value.
- Explore the **distribution of house sales** across different price ranges.
- Create **interactive Tableau dashboards** to present findings effectively.

2. Project Initialization and Planning Phase

2.1. Define Problem Statement

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	A first-time homebuyer who wants to make an informed decision	Find a home within my budget that meets my needs	The available market data is difficult to interpret and scattered across multiple sources	There is no centralized, easy-to-use tool that visualizes housing trends based on historical sales data	Confused and overwhelmed, making me hesitant to proceed
PS-2	A real estate investor looking for high-return properties	Identify profitable properties based on price trends and key influencing factors	Existing datasets require extensive manual analysis and lack clear insights	No interactive visualization tool allows me to compare property appreciation trends effectively	Frustrated and uncertain about making investment decisions
PS-3	A real estate agent aiming to assist clients efficiently	Provide accurate and insightful recommend ations based on market data	The data is time-consuming to analyze and spread across various reports	There is no comprehensive tool to aggregate and visualize pricing trends for quick insights	Less efficient, unable to provide quick, data-backed advice to clients

2.2. Project Proposal (Proposed Solution)

Project Overview	
Objective	The primary objective of this project is to analyze and visualize housing market trends by examining sale prices and key property features using Tableau. The project aims to provide insights into market patterns, price fluctuations, and factors influencing home values, helping stakeholders make data-driven decisions.
Scope	• **Geographical Coverage:** Focus on a specific city, region, or nationwide data. • **Data Sources:** Utilize publicly available datasets (e.g., Zillow, Kaggle, government housing data). • **Analysis Areas:** Price trends over time, location-based variations, property features affecting prices. • **Visualization Methods:** Interactive dashboards, Story, Pie chart, Bar charts, and Histogram.
Problem Statement	
Description	The housing market is influenced by multiple factors, including location, property size, number of bedrooms, and economic trends. However, without effective visualization, identifying key trends and insights can be challenging. This project aims to bridge that gap by providing a clear, interactive representation of housing market data.
Impact	• Helps potential buyers and sellers make informed decisions. • Assists real estate professionals in understanding pricing trends. • Provides policymakers with insights into housing affordability and market fluctuations.
Proposed Solution	
Approach	• Collect and preprocess historical housing market data. • Use Tableau to create interactive visualizations. • Apply data analysis techniques to identify patterns and correlations. • Develop dashboards for easy interpretation of housing trends.
Key Features	• **Dynamic Visualizations:** Interactive dashboards showing trends over time. • **Geospatial Analysis:** Heat maps highlighting price variations by location. • **Comparative Analysis:** Charts comparing features like square footage vs. price. • **Predictive Insights:** Basic forecasting of price trends using historical data.

2.3. Initial Project Planning

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	Data Collection & Extraction from Database	VHMTUT-2	Downloading the dataset	2	High	D Krishna Siva Ram	20-6-2025	21-6-2025
Sprint-1	Data Preparation	VHMTUT-4	Explanation video links	1	High	D Krishna Siva Ram	22-6-2025	23-6-2025
Sprint-2	Data Visualization	VHMTUT-6	No. of Unique Visualizations	2	Low	D Krishna Siva Ram	24-6-2025	25-6-2025
Sprint-2	Data Visualization	VHMTUT-7	Visualizations	2	High	P Kumar Swamy	26-6-2025	27-6-2025
Sprint-3	Dashboard	VHMTUT-9	Responsive and Design of Dashboard	1	High	P Kumar Swamy	28-6-2025	29-6-2025
Sprint-3	Story	VHMTUT-11	No of Scenes of Story	2	High	O Kumara Narasimha Swamy	28-6-2025	29-6-2025
Sprint-3	Story	VHMTUT-12	Utilization of Filters	1	Medium	O Kumara Narasimha Swamy	29-6-2025	30-6-2025
Sprint-4	Web integration	VHMTUT-14	Go to Dashboard/story, click on share button on the top ribbon	2	Medium	Lahari Korada	30-6-2025	1-07-2025
Sprint-4	Web integration	VHMTUT-15	Dashboard and Story embed with UI With Flask	2	High	Lahari Korada	30-6-2025	1-7-2025

3. **Data Collection and Preprocessing Phase**

3.1. Data Collection Plan and Raw Data Sources Identified

Section	Description
Project Overview	**Visualizing Housing Market Trends: -** This project, **"Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau,"** aims to explore and analyze housing market trends using the **Transformed Housing Data 2** dataset from Kaggle. The objective is to identify key factors influencing house prices, such as location, size, number of bedrooms, bathrooms, floors and basement area. By leveraging **Tableau**, the project will create interactive dashboards, story, bar chart, histogram, summary dashboard to visualize patterns, compare regional price variations, and gain insights into how different features impact house sale prices. The analysis will help in making data-driven decisions for buyers, sellers, and real estate professionals. **Objectives:** 1. **Understand Sale Price Trends** – Identify price distribution and variations based on property features. 2. **Feature Analysis** – Examine factors influencing house prices, such as **area, bedrooms, and location**. 3. **Market Insights** – Discover trends and patterns to assist **buyers, sellers, and real estate professionals**. 4. **Data-Driven Decision Making** – Use visualizations to interpret market conditions effectively.

Data Collection Plan	The Data is collected from the “KAGGLE” **Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau** Link: https://www.kaggle.com/datasets/rituparnaghosh18/transformed-housing-data-2
Raw Data Sources Identified	Transformed Housing Data 2 from KAGGLE. This dataset contains **21,609** house sale records with various attributes such as **sale price, number of bedrooms, square footage, location, and additional features**. The data has been pre-processed and transformed for analytical purposes. There are a total of 31 columns, out of which Sale Price can be supposedly taken as a dependent variable. The other variables are different features, locations and date, etc. regarding the houses.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Transformed Housing Data 2	This dataset contains 21,609 house sale records with details such as price, area, number of bedrooms, and other property features.	Kaggle Dataset Link	CSV	10 MB	Public

3.2. Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
Transformed Housing Data 2	• Missing values in key attributes. • Duplicate records. • Outliers affecting analysis. • Incorrect data types or formats. • Inconsistent data entry.	Moderate	• Filling missing values using mean/median/mode. • Removing or correcting duplicates. • Normalizing or transforming data types. • Handling outliers using statistical techniques.
Transformed Housing Data 2	Duplicate records found (11 instances).	Moderate	Remove duplicate records using data cleaning techniques.
Transformed Housing Data 2	No of Times Visited' column contains only zero values, which may indicate missing or incorrect data.	High	Verify data source or remove column if it lacks meaningful information.
Transformed Housing Data 2	Zipcode representation is in multiple binary columns instead of a single categorical variable.	Moderate	Convert multiple binary columns into a single categorical 'Zipcode Group' column for better efficiency.
Transformed Housing Data 2	Potential outliers in 'Sale Price' and 'Lot Area' columns.	High	Use statistical methods to identify and handle outliers (e.g., IQR method).

3.3. Data Exploration and Preprocessing

Section	**Description**
Data Overview	The dataset contains Transformed housing data and 21,609 house sale records, including Property features such as Sales price, area, bedrooms, bathrooms, floors and location. There are a total of 31 columns, out of which Sale Price can be supposedly taken as a dependent variable. The other variables are different features, locations and date, etc. regarding the houses.
Data Cleaning	Missing values were handled, duplicates were removed, and data inconsistencies were corrected. Removed 11 duplicate records to ensure data integrity. Verified that no missing values exist.
Data Transformation	Applied filtering, sorting, pivoting, and created calculated fields such as Sales price (Average), price per square foot (SUM) and Age of houses (Average) and filters like Sales price (Top 25), Ages of houses (Top 10) and Measure Names. Converted multiple binary zip code group columns into a single categorical variable. Removed or adjusted potential outliers in 'Sale Price' and 'Lot Area' using statistical techniques.
Data Type Conversion	Corrected data types, ensuring numerical fields (e.g., price, area) are properly formatted. Ensured that numerical fields (e.g., Sale Price, Lot Area) are in the correct format. Converted categorical variables like Zipcode Groups into appropriate data types.
Column Splitting and Merging	Split combined address columns and merged relevant features for better analysis. Merged zip code binary columns into a single categorical field for better data structure.
Data Modeling	Established relationships between features and variables such as house size, number of floors, bathrooms, bedrooms, sale price and Age of houses (in years). Identified key relationships between features such as Sale Price, Lot Area, and House Features for further analysis in Tableau.

Save Processed Data	-

4. Data Visualization

4.1. Framing Business Questions

1. What is the average sale price of houses?
2. Which basement size shows the highest sale price?
3. Which price range has the lowest number of sales?
4. What does the color coding represent?
5. How many different houses age categories are shown?
6. What age groups of houses are included in the chart?
7. What is the count of houses in the dataset for each house age
8. Which price bins have the least number of houses?

4.2. **Developing Visualizations**

Visualizations play a crucial role in analyzing and interpreting housing market trends. They help in uncovering patterns, relationships, and insights from the dataset. Here's how you can approach developing visualizations for your project:

- **Bar Charts** – Compare categorical variables like the number of houses by neighbourhood.
- **Scatter Plots** – Show relationships between sale prices and features like square footage or number of bedrooms.
- **Histograms** – Display distribution of sale prices to understand market trends.
- **Heatmaps** – Represent correlations between different features.
- **Box Plots** – Identify outliers in sale prices and feature distributions.
- **Geospatial Maps** – Show housing price trends across different locations.

5. Dashboard

5.1. Dashboard Design File

1. Link-1
2. Link-2
3. Link-3

6. Report

6.1. Story Design File

1. Link

7. Performance Testing

7.1 Utilization of Data filters

Selected Top (10) Highest Sales Price

Selected Top (10) Age Of House (in years)

Selected Top (10) No Of Floors, Bathrooms, Bedrooms

7.2 No of Calculation Field

- - -

7.3 No of Visualization

- Bar Chart
- Pie Chart
- Bubble Chart
- Donut Chart
- Text Table
- Word Cloud
- Funnel Chart
- Horizontal Chart

8. Conclusion/Observation

Conclusion:

Developing visualizations for housing market trends provides deep insights into the relationships between different property features and sale prices. By leveraging Tableau, interactive and visually appealing dashboards can be created, making it easier to analyze market dynamics. The integration of these visualizations into a Flask-based web application allows users to explore trends effectively. Properly selected charts, such as scatter plots, heatmaps, and geospatial maps, help in identifying patterns, correlations, and anomalies within the dataset.

Observations:

1. **Price Distribution** – Sale prices are not evenly distributed; they tend to cluster around certain price ranges, indicating popular price segments.
2. **Impact of Features** – Factors like square footage, lot size, and the number of bedrooms significantly influence sale prices, while some features have minimal impact.
3. **Neighbourhood Trends** – Certain neighbourhoods consistently show higher property values, possibly due to better infrastructure, schools, or amenities.

4. **Seasonal Variation** – Housing prices fluctuate across different months or years, indicating seasonal demand shifts.
5. **Outliers in Data** – Some properties are priced significantly higher or lower than average, which could be due to unique characteristics or anomalies in data collection.

9. Future Scope

The project on **"Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau"** can be expanded in various ways to enhance its impact and usability. Some potential future directions include:

1. Advanced Predictive Analytics

- Implement **Machine Learning models** (e.g., regression, decision trees, neural networks) to predict future house prices based on historical trends.
- Use **Time Series Forecasting** (e.g., ARIMA, LSTM) to anticipate market fluctuations.

2. Enhanced Interactive Visualizations

- Develop **real-time dashboards** in Tableau by integrating live data sources such as real estate APIs.
- Introduce **user-driven filtering** to allow users to explore data based on their preferences (e.g., price range, neighborhood).

3. Integration with External Data Sources

- Combine **economic indicators** (e.g., interest rates, inflation, job market data) to analyze their impact on housing prices.
- Use **Geospatial Data & GIS Mapping** to show property trends based on location, infrastructure, and amenities.

4. Web Application Enhancements

- Deploy the **Flask-based website** on a cloud platform (e.g., AWS, Heroku) for wider accessibility.
- Add **User Authentication & Personalization**, allowing users to save preferences and compare properties.

10. Appendix

10.1. Source Code (if any)

- - -

10.2. GitHub & Project Demo Link

I. GitHub Link

II. Project Demo Link