

# REAL ESTATE MARKET ANALYSIS & INVESTMENT STRATEGY

---

Section A Group 8

Sector: Residential Real Estate

Faculty: Satyaki Das

Team Members:

- 1) Naveen Kumar (2401020041)
- 2) Satya Yadav (2401010427)
- 3) Satyam Kumar (2401010428)
- 4) Dakarapu Hemamdharnath (2401010137)
- 5) Kapil Karan Mathur (2401010209)
- 6) Saksham Sharma (2401010403)

# Executive Summary

## Problem

A real estate investment firm wants to know:

Which house features increase price the most?

And how should they buy houses to make maximum profit?

## Approach

We studied 21,607 house sale records.

We cleaned the data and created important measures like:

- Average price
- Price per square foot
- Renovation status
- Size groups
- Age groups

## **Key Findings**

- Waterfront houses are almost 88% more expensive.
- Bigger houses sell for higher prices.
- Renovated houses sell for about 30% more.
- Some zipcodes are 5 times more expensive than others.
- Very large houses give less extra value after 4000 sqft.

## **Main Recommendation**

The company should:

- Buy some waterfront houses for stable profit.
- Buy non-renovated houses in good areas and renovate them.
- Focus on medium-sized houses (2000–4000 sqft).
- Study zipcodes carefully before investing.

This will help increase profit and reduce risk.

# Sector & Business Context

## Sector Overview

The real estate market depends on:

- House size
- Location
- Condition
- Renovation
- Neighborhood

Some features increase price a lot, others only a little.

## Current Challenges

- Investors sometimes overpay.
- They may ignore renovation opportunities.
- They may not study neighborhoods carefully.
- They may buy very large houses with low return.

## Why This Problem Was Chosen

If we understand what truly increases house price, investors can make smarter decisions and earn more profit.

# Problem Statement & Objectives

## Problem Statement

Find which house features affect price the most and create a strategy to help investors buy the right properties.

## Project Scope

We included:

- Size (living area)
- Bedrooms & bathrooms
- Condition
- Renovation
- Waterfront
- Zipcode
- Age
- Sale date

## Success Criteria

The project is successful if:

- We identify the most important price drivers.
- We create a clear investment strategy.
- We build a dashboard for decision making.

# Data Description

## Dataset Information

- 21,607 house sales
- Around 20 columns
- Data from 2016–2017

## Important Columns

- |                        |                    |            |
|------------------------|--------------------|------------|
| • Sale Price           | Living Area (sqft) | Lot Area   |
| • Bedrooms<br>(Yes/No) | Bathrooms          | Waterfront |
| • Condition            | Renovation Year    | Zipcode    |
| • Sale Date            | Age of house       |            |

## Data Limitations

- No data about renovation cost
- No data about interest rates
- Only 2 years of data
- No rental income data

## Data Cleaning & Preparation

- All columns were renamed to clean,machine-learning-friendly format
- One column was removed
- The data format have been changed.
- Changed above\_basement\_sqft,basement\_sqft format to number

## Missing Values Handling

- Missing values were handled (probably filled or removed)
- Dataset is now much more ML-ready

## New Columns Created

- Renovation Status (Renovated / Not Renovated)

- Price per Sqft = Sale Price ÷ Living Area
- Living Area Groups
- Age Groups
- Year & Month from Sale Date

## **Assumptions**

# KPI & Metric Framework

These are the main measures we used:

## **1. Average Sale Price**

Shows general market price.

## **2. Median Sale Price**

Shows middle price without luxury effect.

## **3. Price per Sqft**

Helps compare houses of different sizes.

## **4. Waterfront Premium**

Difference between waterfront and non-waterfront price.

## **5. Renovation Premium**

Difference between renovated and non-renovated price.

## **6. Condition Impact**

Shows how quality affects price.

# Exploratory Data Analysis

## 1. Overall Market

- Average Price: \$511,627
- Median Price: \$450,000

Since average is higher than median, it means there are some very expensive houses.

---

## 2. Waterfront Effect

- Non-waterfront: \$508,228
- Waterfront: \$958,761

Waterfront houses are about 88% more expensive.

This is the strongest price driver.

---

## 3. Size Effect

0–1999 sqft → \$385K  
2000–3999 sqft → \$627K  
4000–5999 sqft → \$1M

Price increases with size.  
But after 4000 sqft, extra size gives smaller benefit.

---

## **4. Renovation Effect**

- Not renovated: \$505K
- Renovated: \$656K

Renovated houses sell for about \$151K more.

Renovation clearly increases value.

---

## **5. Condition Effect**

Excellent houses sell for much higher prices than bad-condition houses.

Good maintenance increases selling price.

---

## **6. Zipcode Effect**

Some zipcodes have average prices above \$1 million.

Some are below \$300K.

Location is very important.

---

## 7. Age Effect

Very old and very new houses are expensive.

Middle-aged houses are cheaper.

This shows age alone does not decide price — location matters too.

# Advanced Analysis

We divided properties into 3 types:

## 1. Premium Houses

- Waterfront
- Prime zipcode
- High quality

High cost but stable returns.

---

## 2. Renovation Strategy (Best Profit)

- Not renovated
- Good location
- Medium size (2000–4000 sqft)

Buy cheap → Renovate → Sell higher.

### 3. Undervalued Areas

- Low price per sqft
- Stable sales volume
- Buy and hold for future growth.

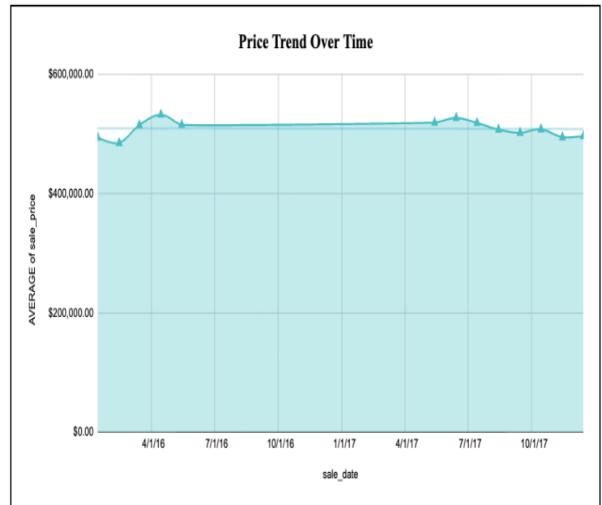
# Dashboard Design

## Housing Market Intelligence Dashboard

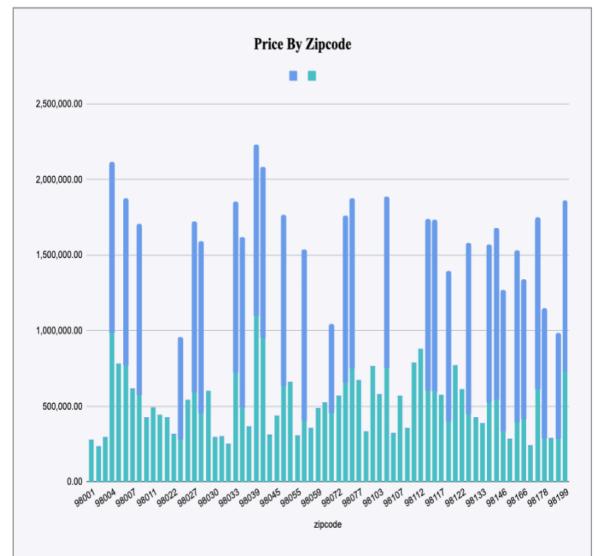
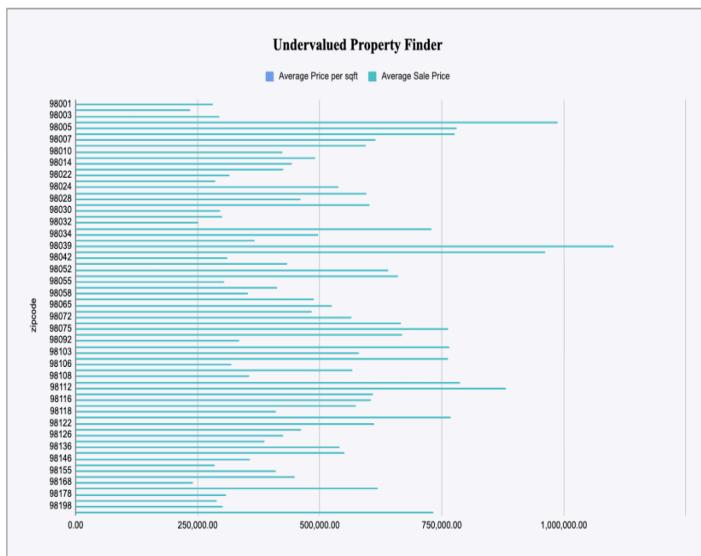
Analyzing Property Prices, Market Trends, and Buyer Behavior to Identify Key Investment Insights.



# Market Trends



# Opportunity & Geography



# Insights Summary

## 1. Waterfront has the biggest impact on price:-

Waterfront homes sell for about 88% more than non-waterfront homes, showing location near water creates a very strong price premium.

## 2. Bigger houses sell for more money :-

As living area increases, average sale price also increases, proving size is strongly linked to higher valuation.

## 3. Renovation increases price significantly:-

Renovated houses sell for around \$151K more than non-renovated ones, showing upgrades clearly add market value.

#### 4. Location (zipcode) is very important:-

Some zipcodes have average prices 4–5 times higher than others, proving neighborhood greatly affects property price.

#### 5. Very large houses give less extra return:-

After 4000 sqft, price increases slower compared to earlier size ranges, showing diminishing returns on extra space.

#### 6. Good condition improves selling price:-

Houses in excellent condition sell much higher than poor-condition houses, showing buyers pay more for quality.

#### 7. Luxury houses increase average price:-

The average price is much higher than the median, meaning a few expensive luxury homes are pulling the average upward.

# Recommendations

Insight	Recommendation	Impact
Waterfront premium	Buy selective waterfront houses	Stable returns
Renovation impact	Buy non-renovated homes in good areas	Higher profit
Size trend	Focus on 2000–4000 sqft homes	Better ROI
Zipcode variation	Study neighborhood before buying	Smarter investment

## Impact Estimation

### 1) Profit Increase (Revenue Impact)

If 20 houses are bought at \$450K and improved by 20%:

New value  $\approx \$540K$

Profit per house  $\approx \$90K$

Total gain  $\approx \$1.8M$  (before renovation cost)

#### Impact:

Using renovation strategy increases resale value and improves overall portfolio returns.

## 2) Better Capital Allocation (Efficiency Impact)

Houses between 2000–3999 sqft sell for about \$627K, while houses above 6000 sqft sell for about \$1.12M.

But price increase from 4000+ sqft is not proportional to size increase (diminishing returns).

Example:

- 2000–3999 sqft → ~\$627K
- 4000–5999 sqft → ~\$1.01M
- 6000+ sqft → ~\$1.12M

The jump from 4000 to 6000+ sqft gives only ~\$110K extra despite huge size increase.

### Impact:

By focusing on 2000–4000 sqft homes instead of 6000+ sqft luxury homes, the firm can save approximately \$400K–\$500K per property in acquisition cost while still capturing strong resale value.

This improves ROI per property and allows purchasing more units with the same capital.

### 3) Risk Reduction (With Numbers)

Zipcode analysis shows:

- Premium zipcode (e.g., 98039) → Avg price above **\$1M**
- Lower-tier zipcode (e.g., 98002) → Avg price around **\$234K**

That's almost **5x difference**.

Also:

- Waterfront homes: ~\$958K
- Non-waterfront homes: ~\$508K

If the firm randomly buys without zipcode analysis, it risks buying in weak areas with low appreciation.

#### **Impact:**

Avoiding low-performing zipcodes can prevent potential underperformance of **\$200K-\$700K per property** compared to premium areas.

Using data-driven selection reduces probability of mispricing and protects millions in capital allocation.

# Limitations

## 1 No Renovation Cost Data

The dataset shows that renovated homes sell for about **\$151K more**, but we do not know how much renovation actually cost.

Without cost data, we cannot calculate exact profit or true ROI.

---

## 2 No Economic Data (Interest Rates, Inflation, Jobs)

House prices are influenced by interest rates and economic conditions.

Since this data is missing, we cannot fully explain why prices moved during certain months.

---

## 3 Limited Time Period (2016–2017 Only)

The dataset covers only about 2 years.

This is not enough to study long-term market cycles or recession effects.

---

## 4 No Rental Income Analysis

The dataset focuses only on selling prices.

Investors who want rental income cannot evaluate rental yield or cash flow performance.

# Future Scope

## 1 Add Regression Analysis

Using statistical models can measure exactly how much each factor (size, location, renovation) affects price.

This will make predictions more accurate and scientific.

---

## 2 Include Interest Rate & Economic Data

Adding mortgage rates and economic indicators will help explain market trends and improve forecasting accuracy.

---

## 3 Study Rental Income

Adding rental data will allow investors to compare buy-to-sell vs buy-to-rent strategies.

---

## 4 Predict Future Prices

Using machine learning or forecasting models can help estimate future property values and support long-term investment planning.

# Conclusion

The analysis clearly shows that **waterfront location, living area size, zipcode, renovation status, and house condition** are the strongest drivers of property prices.

Waterfront homes sell for almost **88% higher**, renovated homes sell for about **\$151K more**, and mid-sized homes (2000–4000 sqft) provide the best balance between price and investment efficiency.

The housing market during 2016–2017 remained stable, making it suitable for strategic portfolio expansion.

Using data-driven insights instead of guesswork allows real estate firms to:

- Increase profit margins
- Reduce investment risk
- Allocate capital more efficiently

A balanced strategy combining premium waterfront properties and value-add renovation opportunities will maximize long-term returns.

# Appendix

## 17.1 Data Dictionary

Column Name	Description	Why It Is Important
sale_price	Final selling price of the house	Main target variable
living_area_sqft	Total interior area of the house	Measures size impact
bedrooms	Number of bedrooms	Affects family demand
bathrooms	Number of bathrooms	Influences comfort & price
waterfront	Whether house faces water (Yes/No)	Measures location premium
house_condition	Quality rating (Bad–Excellent)	Impacts buyer perception
renovation_year	Year of renovation	Used to create renovation status
zipcode	Area code of property	Measures neighborhood values
sale_date	Date of transaction	Used for time trend analysis
price_per_sqft	sale_price / living_area	Standardized price comparison

# Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Naveen Kumar	Done				Done		Strategy Lead
Satyam Kumar			Done				Analysis Lead
Satya Ya...	Done			Done			Project Lead
Kapil Karan Mathur		Done		Done			Dashboard Lead
Hemamdharnath		Done					Data Lead
Saksham Sharma						Done	PPT and Quality Lead

Declaration: We confirm that the above contribution details are accurate and verifiable through version history and submitted artifacts.

Team Signature Block: Kapil Karan Mathur

Naveen Kumar

Satyam Kumar

Satya Yadav

Saksham Sharma

Hemamdharnath

