



# GROUP 14

# PHASE 2

HOUSING SALES PROJECT



# CONTENT

**01**

PROJECT  
OVERVIEW

**02**

BUSINESS  
UNDERSTANDING

**03**

DATA  
UNDERSTANDING

**04**

DATA  
MODELLING

**05**

RESULTS

**06**

RECOMMENDATION

**07**

CONCLUSION



# PROJECT OVERVIEW

## **Real world problem:**

Fanaka Group, a Real Estate Investment Group is the identified stakeholders for this project. This group deals with buying, selling, renovating and financing properties. The group is seeking for ways it can provide better housing facilities for their clients. They are seeking to improve their efficiency and effectiveness in delivering their facilities to their clients.

# BUSINESS UNDERSTANDING

01

## Project Statement

The real estate group needs us to provide insights that speak to how home renovations might increase the estimated values of their homes

02

## Project Solution

To cultivate the use of data in attempting to predict whether or nota property renovation in affects the house sale price



# OBJECTIVES



## Property Features

How do the various house features affect the pricing of a property?



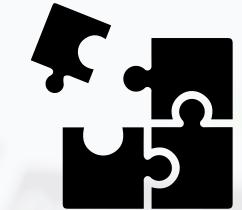
## Property Valuation

How does renovation affect the valuation of a property?



## Customer Satisfaction

What house features are most likely to attract customers?



## Recommendations

*Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.*

# DATA UNDERSTANDING

We are using the King County House Sales dataset, King County House Sales refer to the real estate market and transactions involving residential properties in King County, Washington, including the greater Seattle area.

The Primary source of the Dataset King County , is placed under the `data` folder file path for the purpose of our analysis. The data folder includes the following files:



**column\_names.md** - provides a description of the column names



**kc\_house\_data.csv** - provides list of attributes for the houses and their prices



The methods used in handling the data set given include:

### **Data Preparation**

This process is used to load the data set provided for analysis. Here we are able to find out information such as:

- Shape of the dataset
- Characteristics of dataset
- Columns associated
- Data Types present

### **Data Cleaning**

For this process, we determine action to take in terms of:

- Handling missing data
- Handling duplicated values
- Handling wrong data types

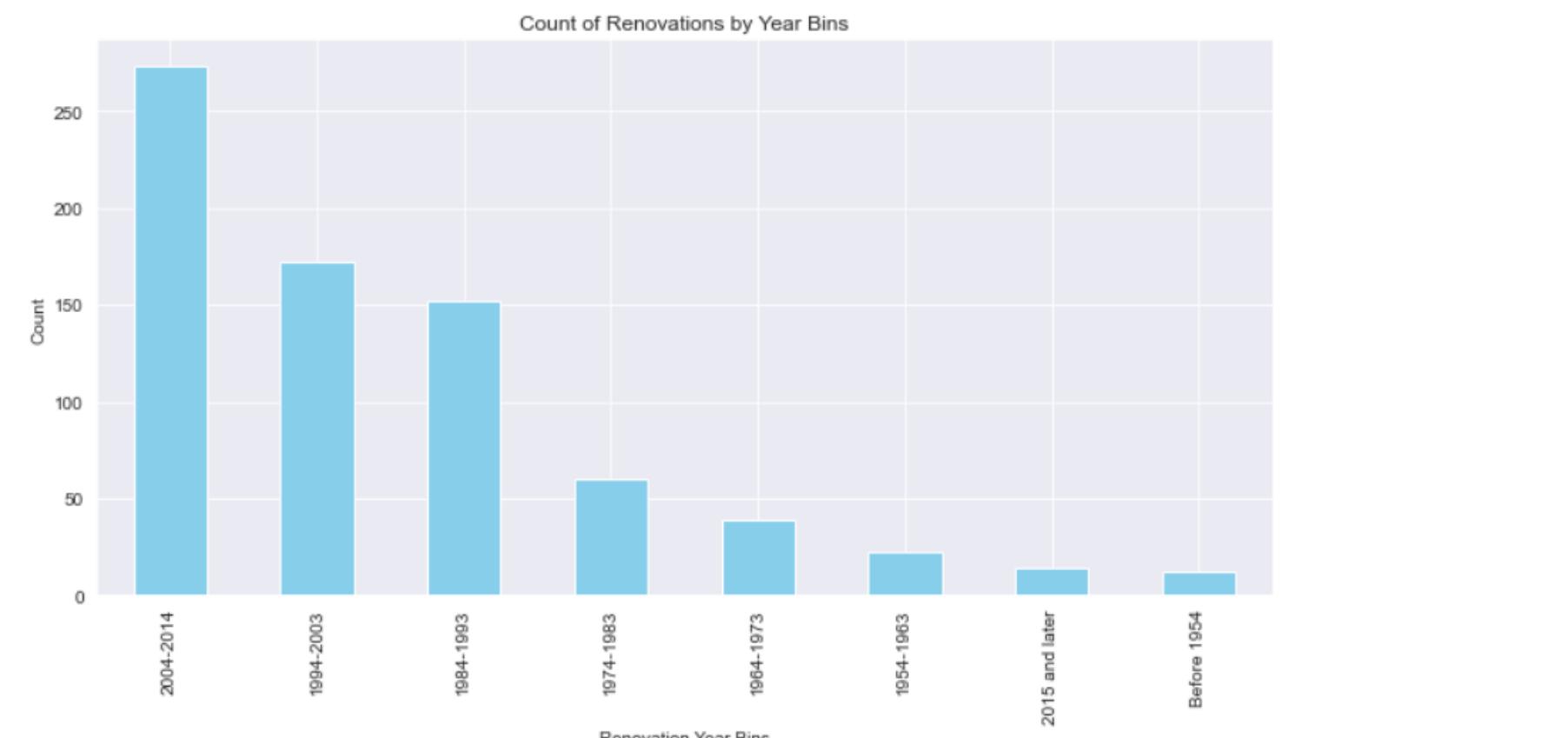
### **Data Visualization**

This process enables us to visually represent relationships between the different columns in the dataset

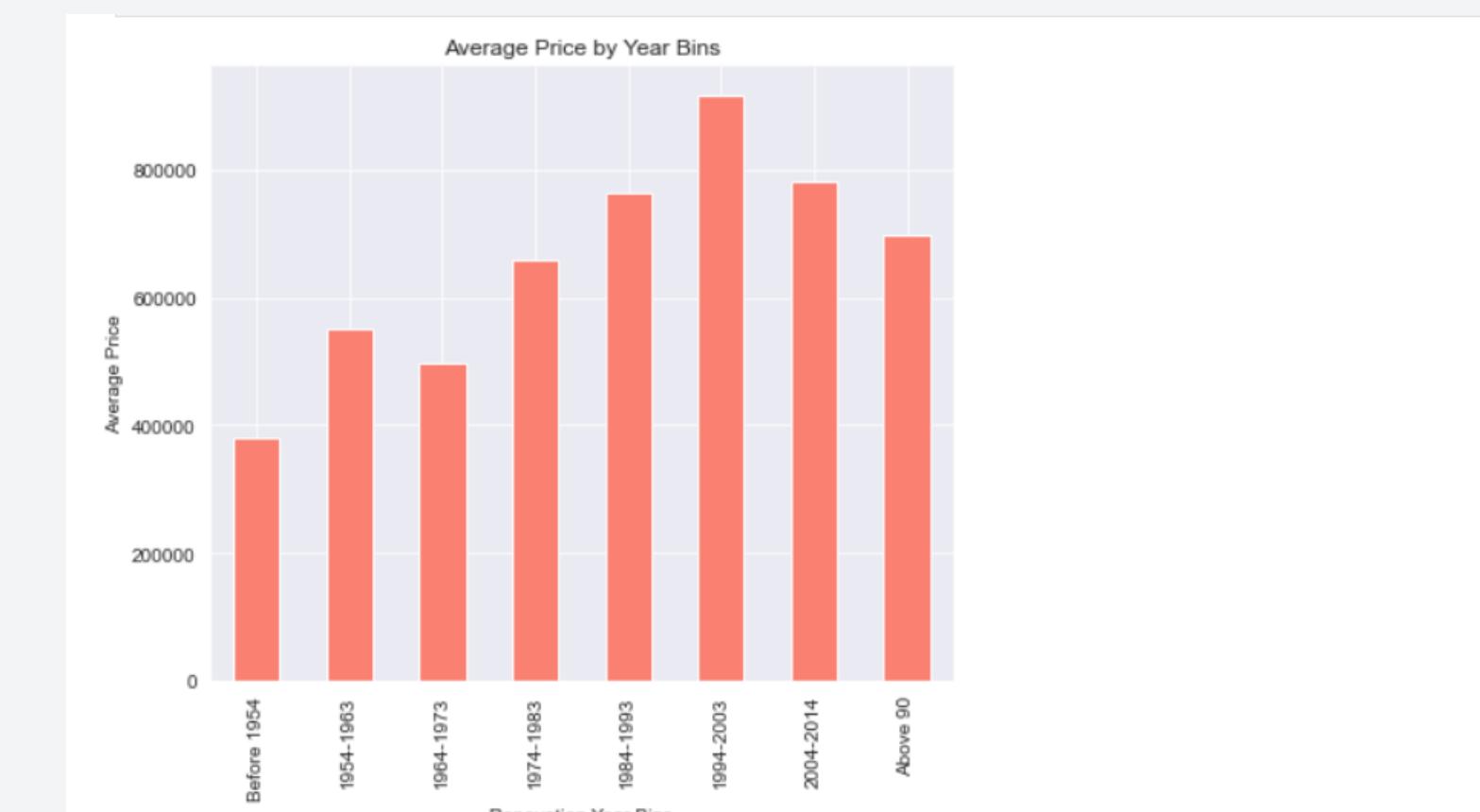




From the above scatter plot, we observe that sqft\_living has a continuous positive correlation with price.



From the analysis above we can see that there was a high number of house renovations between 2004 - 2014 as compared to other years.



During the 1994 - 2003 period there were more expensive renovations that were done as compared to other years.



# MODELLING

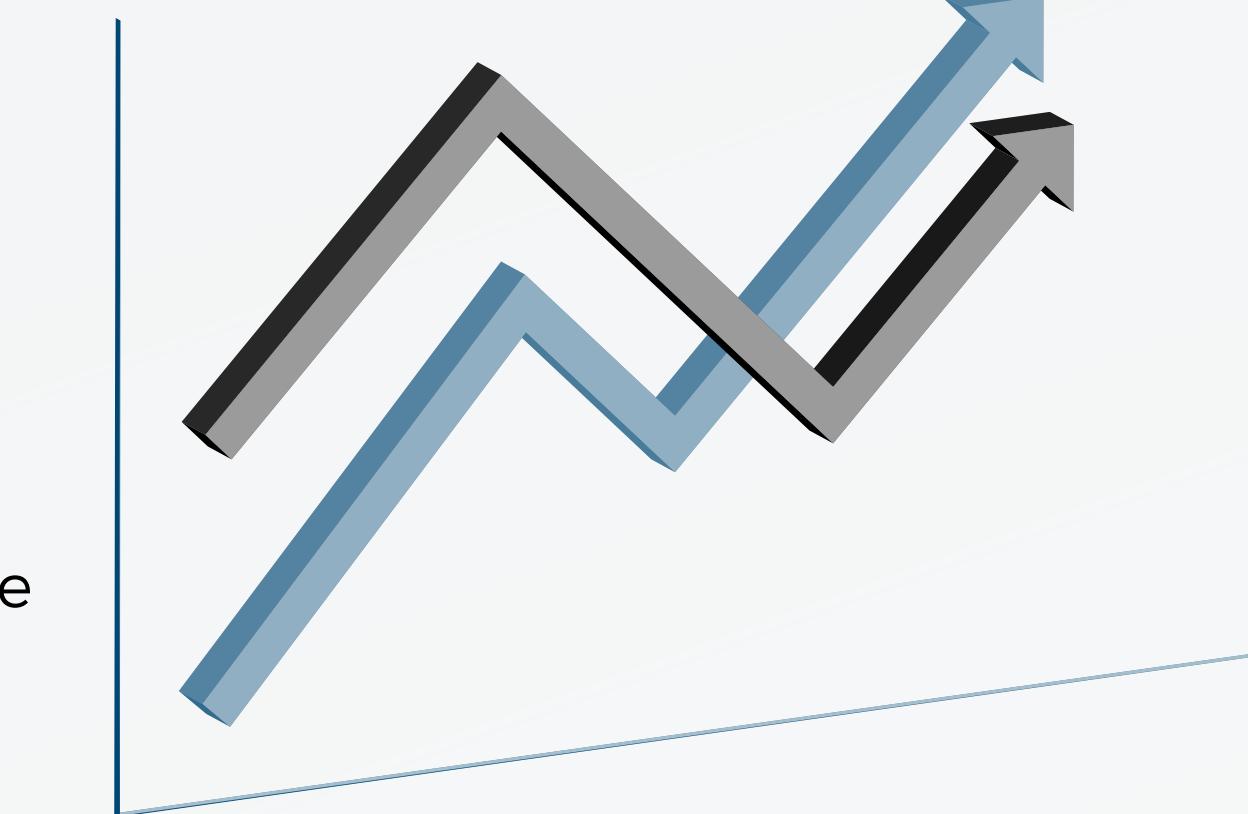
For the dataset provided, we are able to conduct regression modelling in order to draw conclusions and predictions from the data.

- Univariate Analysis
- Multivariate Analysis

**80%**

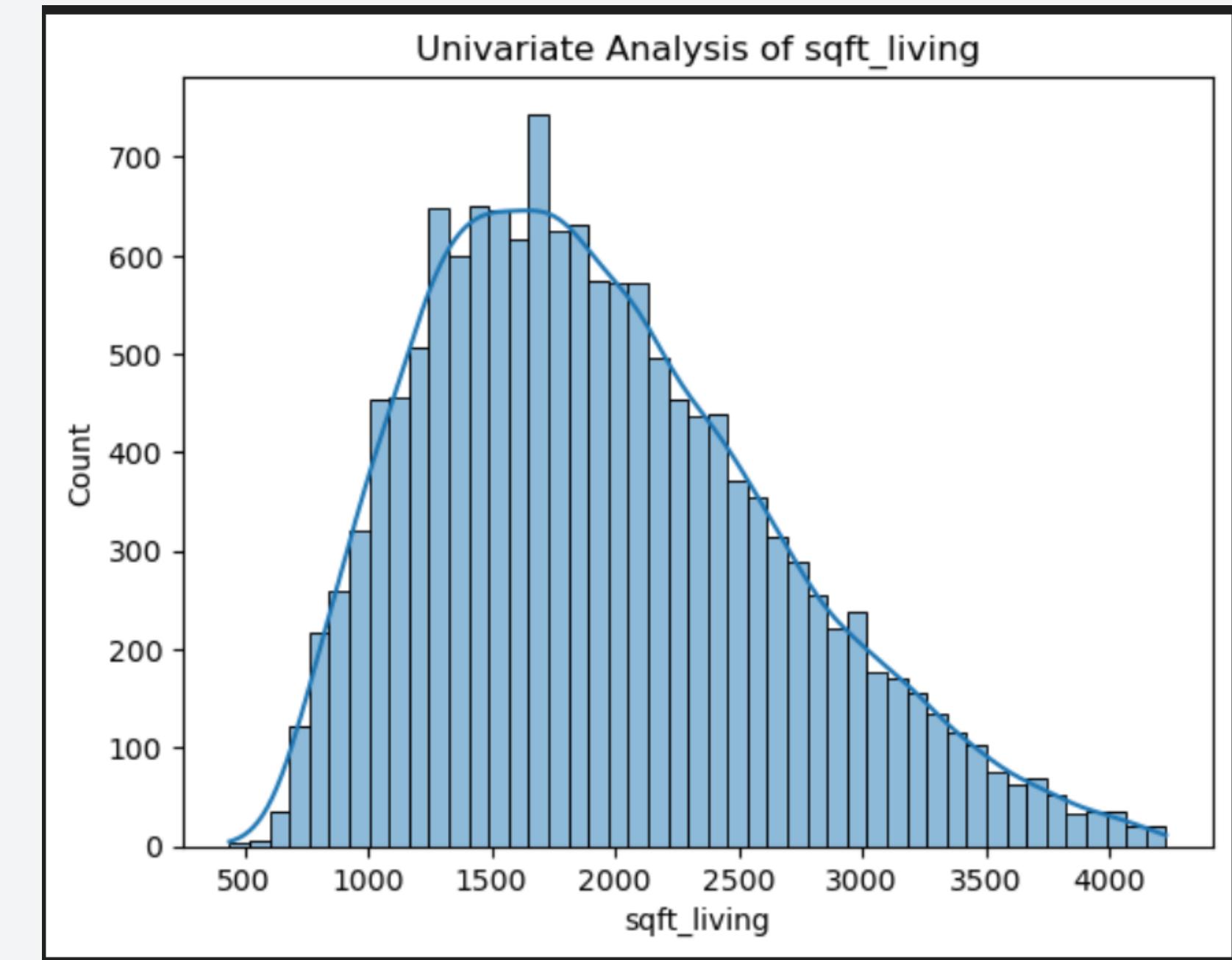
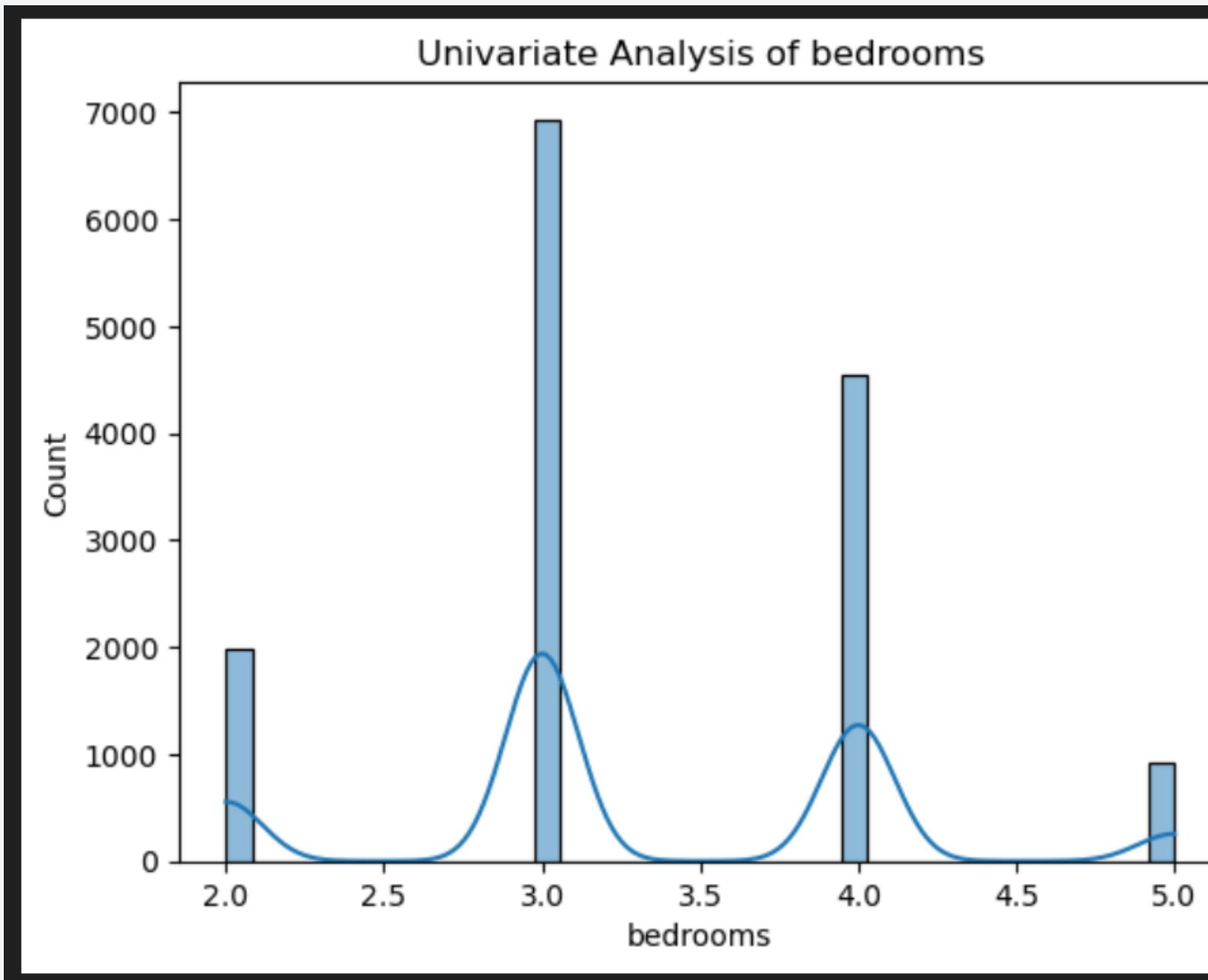


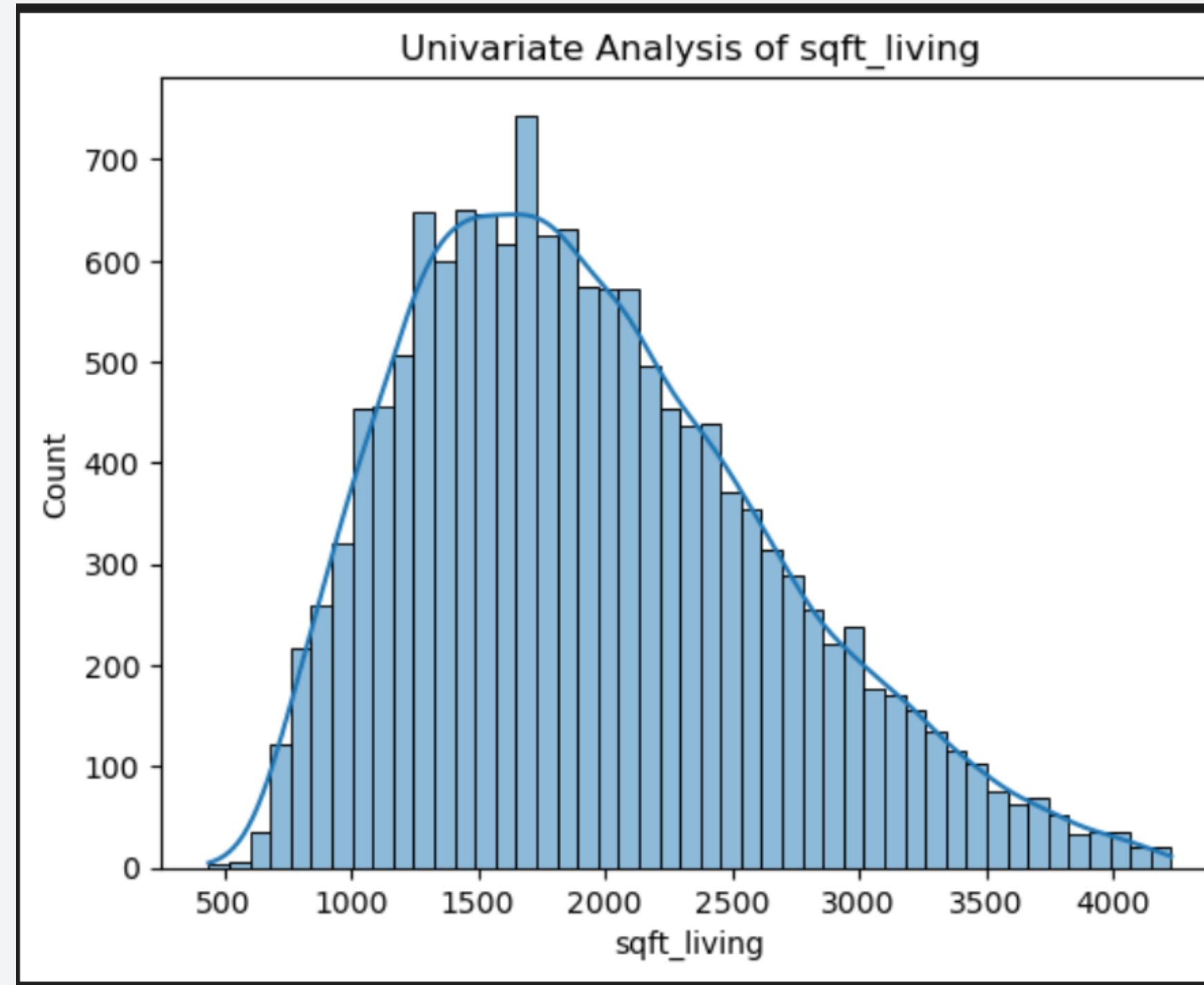
**20%**



# UNIVARIATE ANALYSIS

We conducted univariate analysis to understand the distribution and characteristics of individual variables.





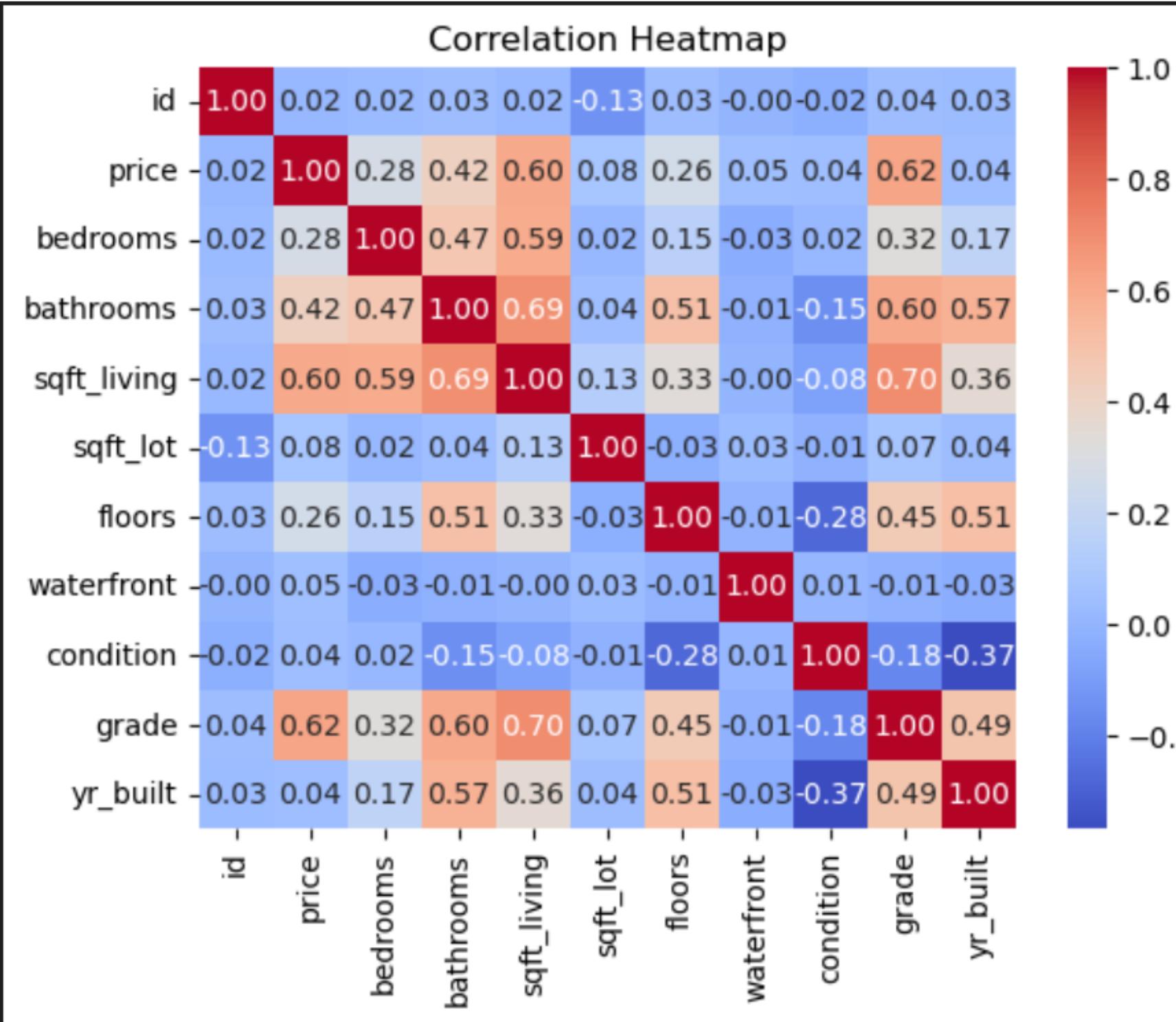
## Results

For the univariate analysis of the dataset, the categorical data put into consideration include the 3 listed:

- Bedrooms
- Bathroom
- sqft\_living.

We are able to conclude that from the count from this, the major features that attract customers to buy house are like the above listed

# MULTIVARIATE ANALYSIS



We conducted multivariate analysis to understand the relationship between the house features and the price.

## Results

The size of the living area, indicated by sqft\_living, has the strongest positive impact on property prices.

The number of bathrooms also significantly influences prices. Bedrooms, while still positively correlated, have a comparatively weaker impact. Other attributes like floors, sqft\_living15, and yr\_built show moderate correlations with various features.

# RECOMMENDATIONS

01

## Planning

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.

02

## Competitive

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.

03

## Analysis

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.

04

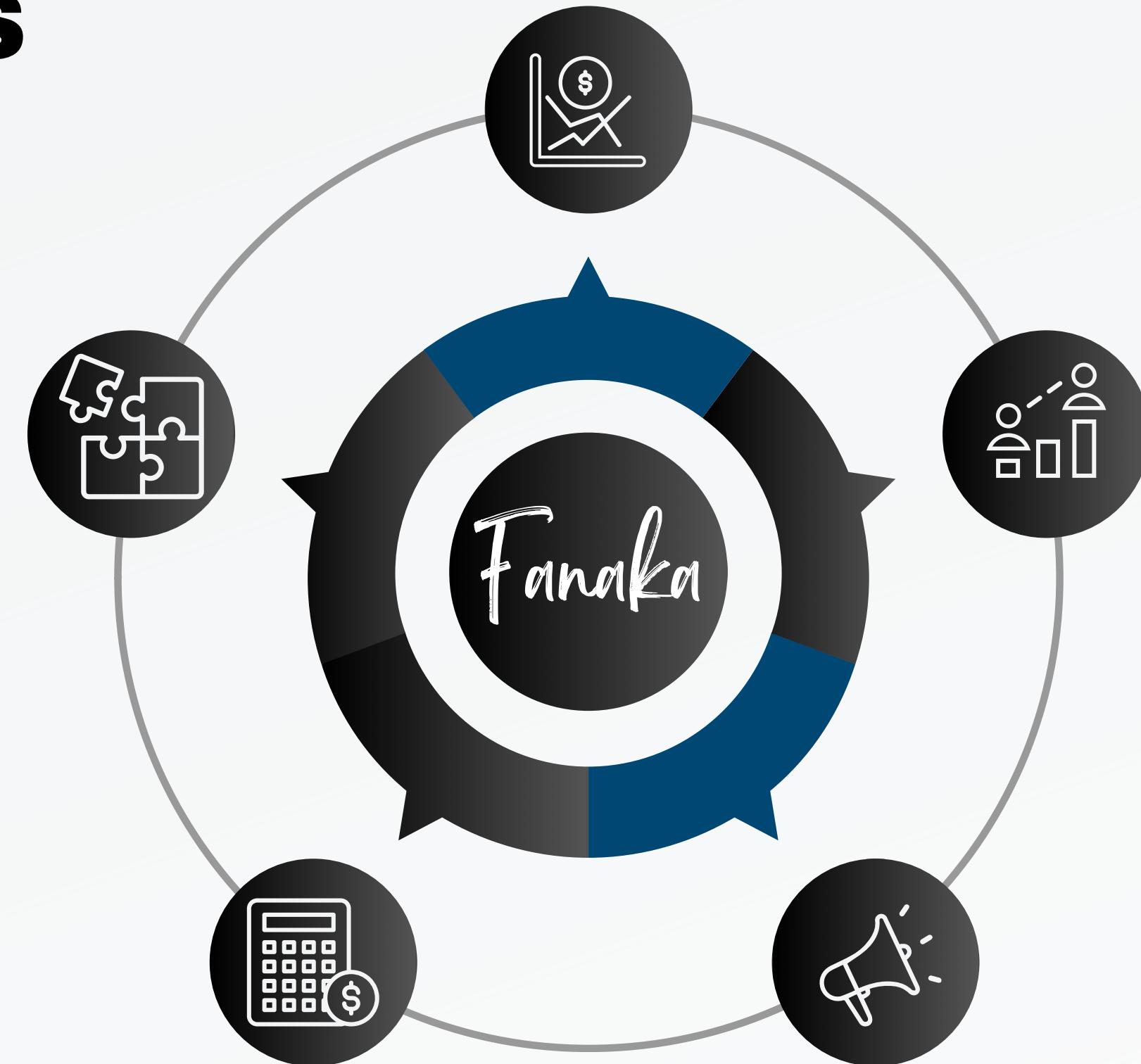
## Execution

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.

05

## Growth

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.



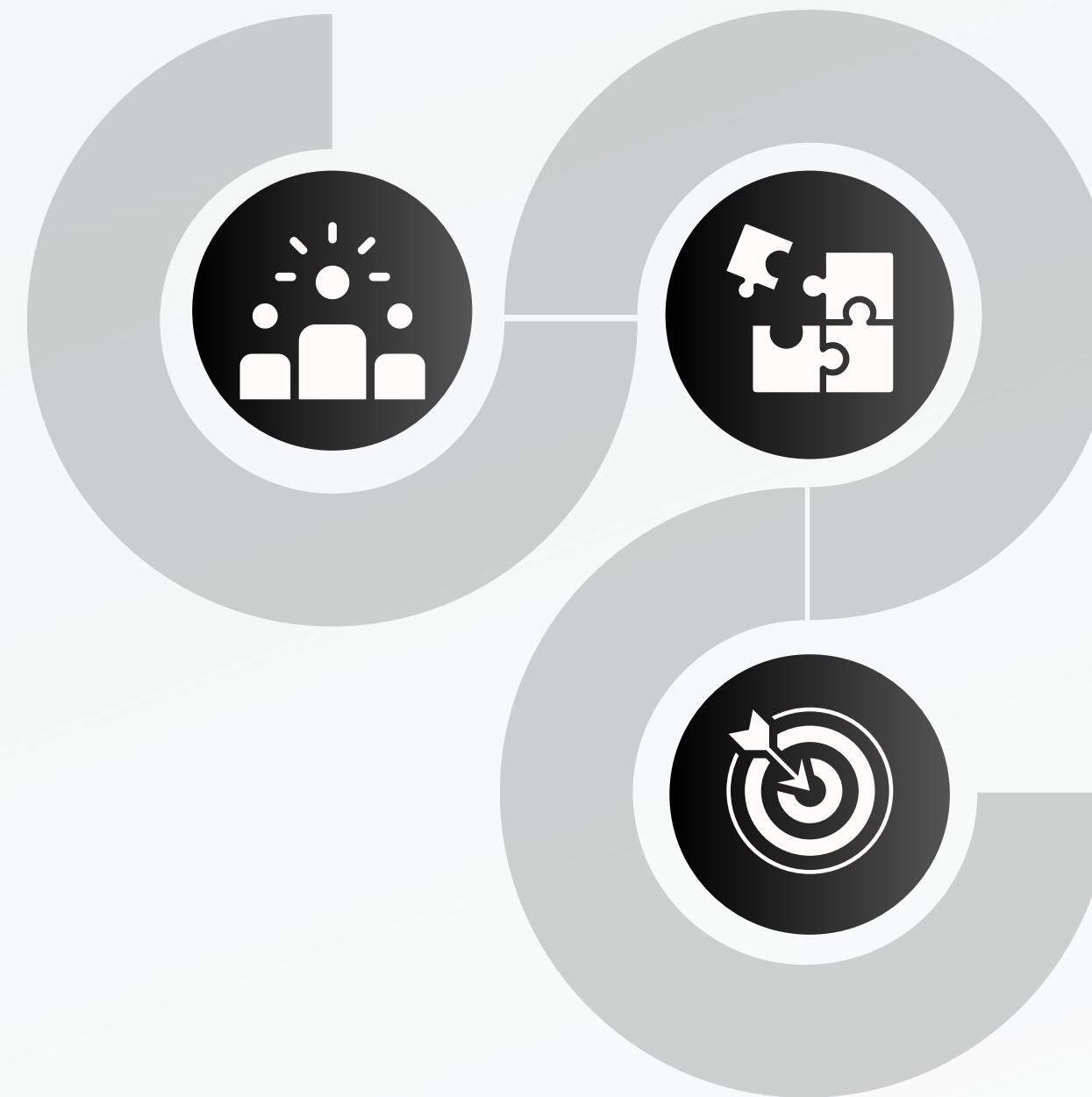
# SUMMARY

- 01**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.
- 02**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.
- 03**

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis vulputate nulla at ante rhoncus, vel efficitur felis condimentum. Proin odio odio.





# **OUR TEAM**

- 1. BUSHRA MOHAMMED**
- 2. PENNINAH GITUKU**
- 3. LESLEY KAMAMO**
- 4. CYNTHIA ADISA**

**GROUP 14**