Factors Affecting the Price of Property in King's County
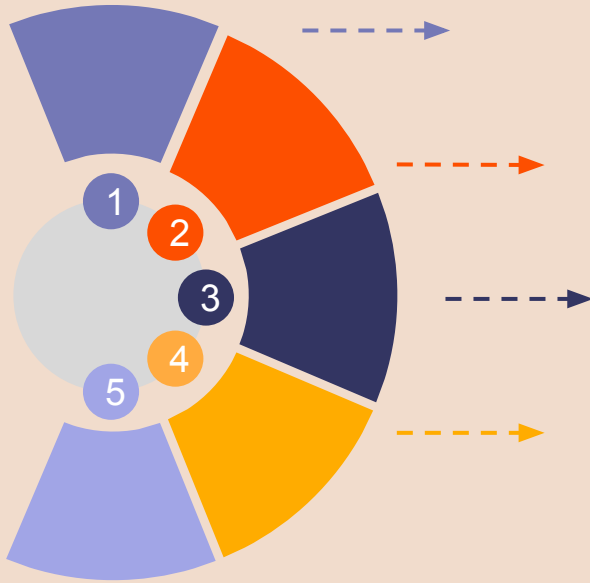
Presentation By Group 23;

1. Sophy Owuor
2. Ndanu Mwatu
3. Denis Ochieng
4. Laban Kariuki

Table of contents

01 Introduction

02 Business Understanding

03 Findings

04 Conclusion

05 Recommendations

01

Introduction

Overview

01 **Project Goal**

Identifying gaps in market knowledge on real estate pricing and giving insights on the key determinants of these prices
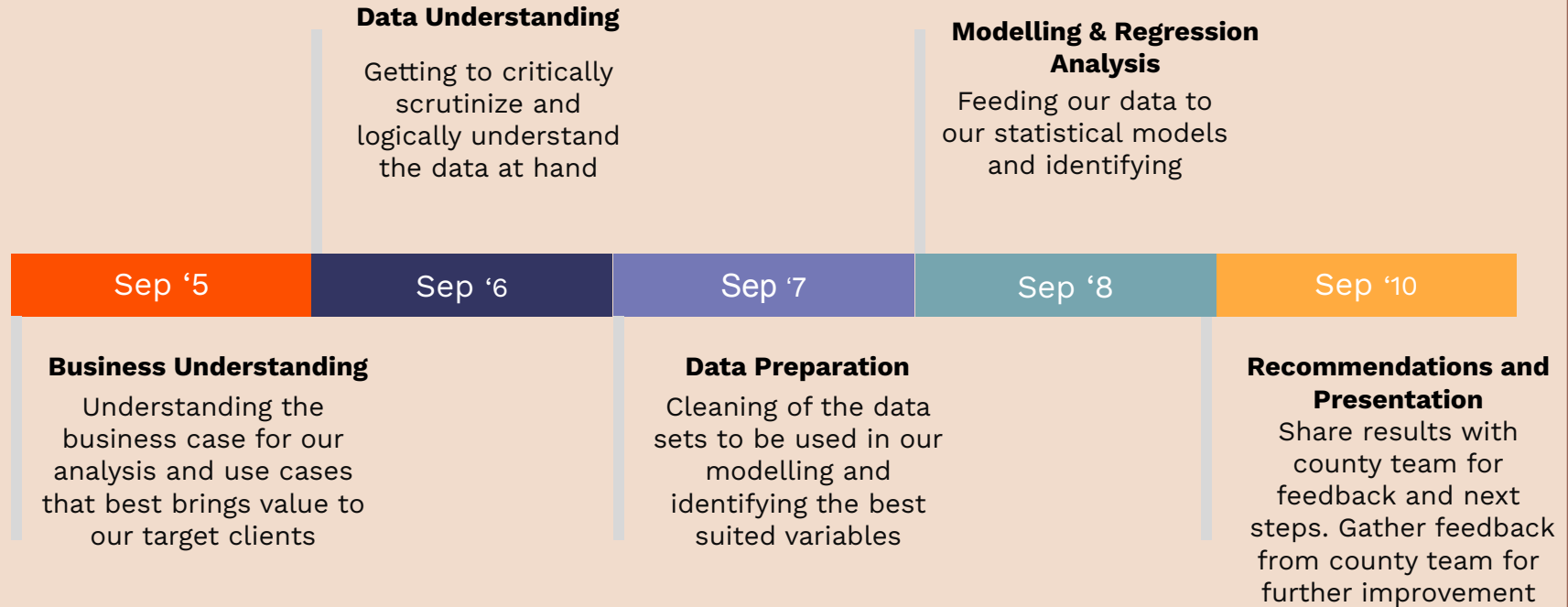
02 **Audience**

Investors eyeing the real estate market, home buyers and sellers alike

03 **Dataset**

Our analysis is based on house data from king County, Washington, USA.

Customers have been buying houses that depreciate in value over time. Seeing a **business opportunity**, we realized that **lack of information** on the real value of these properties made it easy for them to be **shortchanged**. As investment bankers and real estate market analyzer, we **provide insights** based on statistics models advicing potential homeowners on the **best value** for their money.

Time Scope

Data Understanding

Getting to critically scrutinize and logically understand the data at hand

Modelling & Regression Analysis

Feeding our data to our statistical models and identifying

Sep '5	Sep '6	Sep '7	Sep '8	Sep '10

Business Understanding

Understanding the business case for our analysis and use cases that best brings value to our target clients

Data Preparation

Cleaning of the data sets to be used in our modelling and identifying the best suited variables

Recommendations and Presentation

Share results with county team for feedback and next steps. Gather feedback from county team for further improvement

Background

King County is a popular real estate investment market. Population of approximately 2.2 million people and 13th most populated county.

Housing prices in King County are some of the fastest growing with median price increases of 53% and mean rent increase of 43% between 2012 to 2017.

Its economy depends majorly on businesses such as: Amazon, Boeing Commercial Airplanes, Microsoft, Starbucks and the University of Washington.

O2

Business Understanding

Business Understanding

Understanding the problem

Customers have been buying houses that depreciate with value over time, they realized that lack of information on the real value of the property made it easy for them to be shortchanged.

Deliverables

- Understand how price is affected by location
- Understand how various numeric variables (bedrooms, bathrooms, sqft_living, sqft_lot, floors, sqft_above, yr_built, sqft_living15 and sqft_lot15) affect property price.
- Understand how condition of houses affect price.
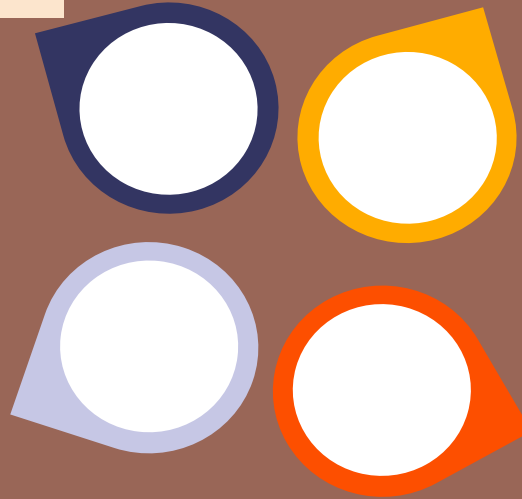
## Tailored solution to the problem

As investment bankers and real estate market analyzer, we are tasked with providing a model that will be used to advice homeowners the kind of home they might purchase based on their needs.

- Price
- Size
- Flipping - buying and selling at appreciating value
- Buying as a capital investment - speculators

Target Objective

- Create models that accurately estimate the best predictors of price value of residential properties.

Data Process

Data Cleaning

Data Analysis

Presentation
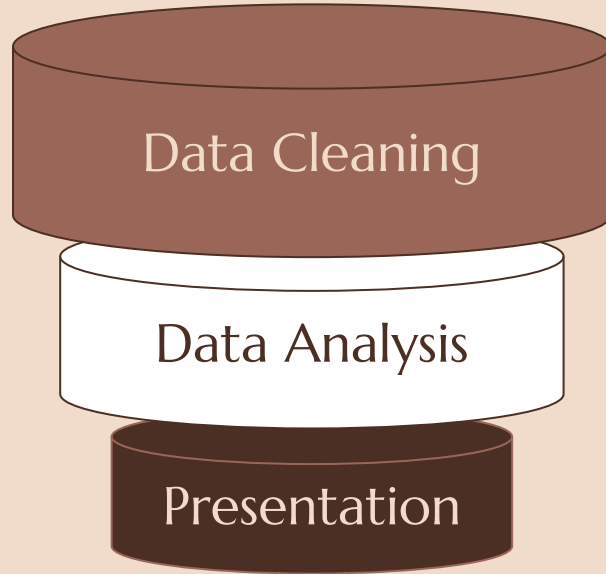
Data Cleaning

This stage is crucial in ensuring that data is accurate and complete. It involves identifying and correcting errors, outliers, inconsistencies, and missing values in the data.
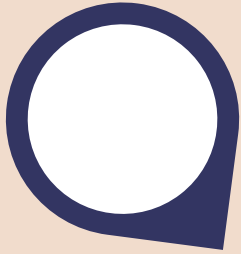
Data Analysis

In this stage we explored the data and ran predictive models on our variables.
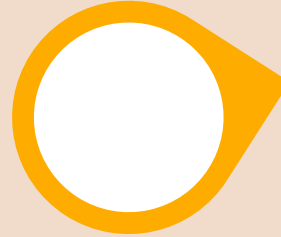
Presentation

Here we presented the gist of the findings. The data here is small when compared to the data at the cleaning stage but it contains the needed information.

Data Preparation

Loading and transforming

Reviewed the data using python DataFrame library methods .info() and .describe() and identified missing values in the data fields.
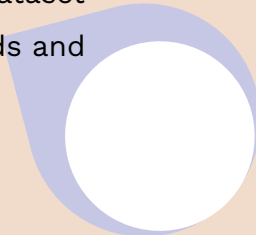
Critical Categorization

Identified null values in three of the data fields i.e waterfront, view and year renovated.

Dropped the three fields from the dataset leaving 18 fields.

Different Perspectives

Further review of unique identifiers revealed instances of duplicates which were dropped from the dataset reducing the data to 21420 records and 17 fields.

Clear Vision

Used the method .value_counts() to further review the data and identify field containing 454 records with question marks (?) thereby dropping the field.
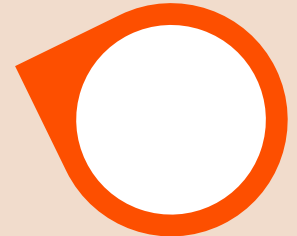
Data Cleaning in Summary

Duplicated values

We dropped the duplicated values based on the 'id' field.

Null values

The study dropped all null values that were present in entire row or could affect the main variables.

Outliers

The study identified outliers in the price field but did not remove.

03

Findings

Overall Findings

Descriptive

- The average price of a property in King's County is USD 540,296.
- Findings also show that 75% of the houses cost above USD 322,000 and 25% of the properties cost above USD 645,000.
- A wide variation in terms of house prices can be denoted from the standard deviation of USD 367,368; with the minimum cost of a house being USD 78,000 and the highest priced house USD 7,700,000.

Geospatial Analysis

A look into the location distribution of the data points indicate that the most expensive properties are found along the Coastlines.

Regression Analysis

Correlation analysis showed that price of property had a positive correlation with all the independent variables. However, the strongest correlations were observed between price and square foot living (0.7). The weakest correlation were observed in sqrft_lot (0.089) and the year property was built (0.05)

Models Prediction Findings

The findings showed that third model explained about 57% of the variance in Prices of houses. From the findings we observe that out of the 9 (numeric) independent variables only sqft_lot and sqft_above did not have a significant relationship with the prices of properties in King's county. We used the condition of the house which was a categorical variable in this third model.

Multiple Linear

1

2

3

Simple Linear

The first model was a simple linear regression between the price of a property and the square footage of the living space in the property. This model which explained 49% of the variance in price showed a statistically significant relationship between price and square foot living of a property

Multiple Linear Regression

The second model explained 56.9% of the variance in Prices of houses which is 7.6% more than the variance explained in the first model (indicating a better fit).

Out of the 9 independent variables used in the model, only Sqft_lot was not statistically significant with the price of a property. The other independent variables showed a significant relationship with the price of property.

04

Conclusion

Regression Conclusion

On predictive ability of the model, we can conclude that Bedrooms, Bathrooms, Sqft_living, Floors, Year built, Sqft_living15, and Sqft_lot15 are the best predictors for price of a property, whereas sqft_lot and sqft_above do not have significant relationships with price of property. It is however important to note that even though the year built and sqft_lot15 shows statistical significant relationship with the price of a property, these two variables have very weak relationship with price of property.

Geospatial

From the geospatial analysis we observe properties found around Coastline areas to be expensive as compared to those in the mainland. This finding shows the location of a given property to be a key factor in determining the price of a property.

Descriptive

Though the high variation makes it a challenge to estimate the real average price value of properties in King's county, we can conclude based on the upper quartile value of USD 645,000 and the average value of USD 540,296, that properties in King's county are sold above the average USA property price, indicating that these are properties with high market value.

05

Recommendations

Recommendation

- As investment bankers and real estate market analyzer offering insights to customers we recommend the following based on our findings;
- That customers should highly consider investing in real estate in King's county as property here fetches above average sale prices compared to the general American market.
- That customers consider the number of bedrooms, number of bathrooms, square footage of living space in the home as well as square footage of interior housing living space for the nearest 15 neighbors and number of floors (levels); as the key factors when selecting a property to invest in. In particular, square footage of living space is noted to be the most significant factor as shows the highest strongest correlation with price of property.
- Customers aiming for highest earnings should consider investing in property along the coastline areas, as the geospatial analysis showed properties along Coastline areas in Kings County to have a higher purchasing value as compared to those on the mainland.

Reference

U.S. Census Bureau and U.S. Department of Housing and Urban Development,

Average Sales Price of Houses Sold for the United States [ASPUS], retrieved from

FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/ASPUS,

September 9, 2023.

Pictures picked from various online sources.

Thanks!

Do you have any questions?
denis.ochieng@student.moringaschool.com
+254 720 664 599

 Denis Omondi