# Phase 2 Project


**Group 13:**

Sharon Kaliku

Ezra Kipchirchir

Paul Kamau

Heri Kimotho

Kipkosgei Kiptui

# King County, Washington

# The Business Case

- A real estate company’s need to help its clients understand how prices of houses vary
- The clients: 1. Homeowners
  2. Potential house buyers

# Objectives

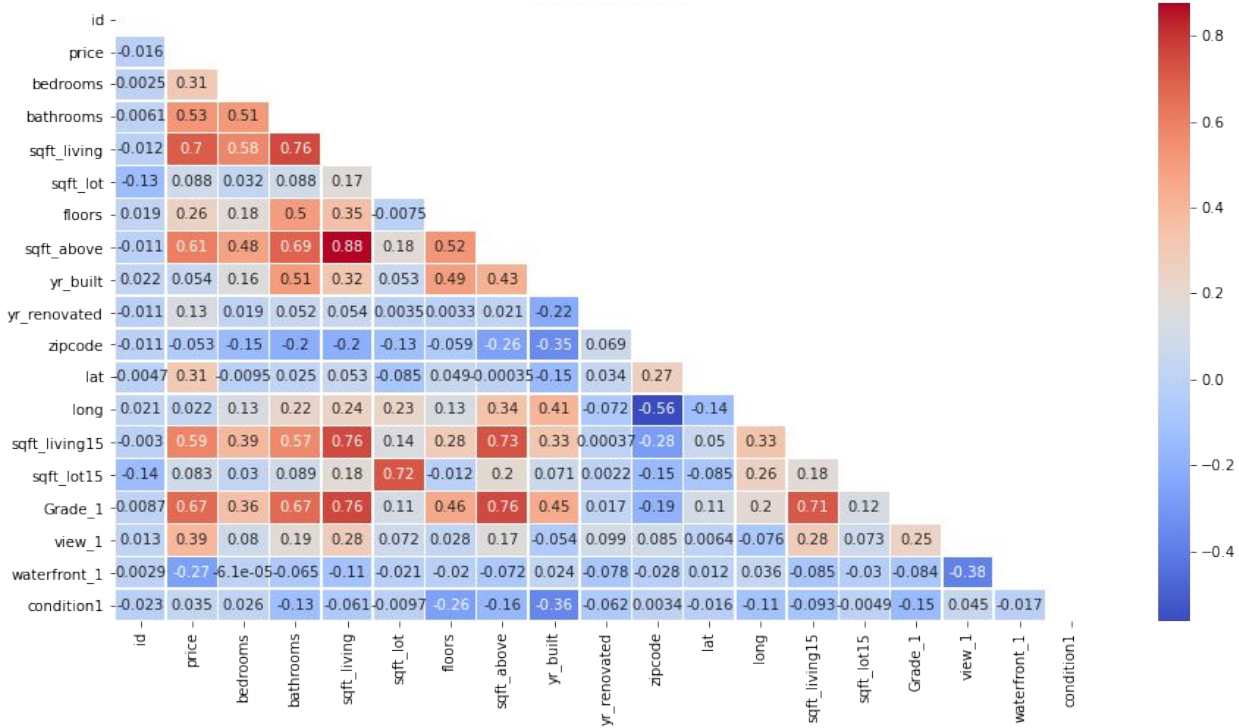
- Analyze relationship between location and price of houses
- Analyze seasonal trends of price
- Predict prices of houses depending on features

# Data Understanding

- Sources of data for analysis:
    1. KC house data - Various features against price
    2. Delivery locations - Zip-codes and their corresponding cities

# Data Visualization

Correlation Matrix

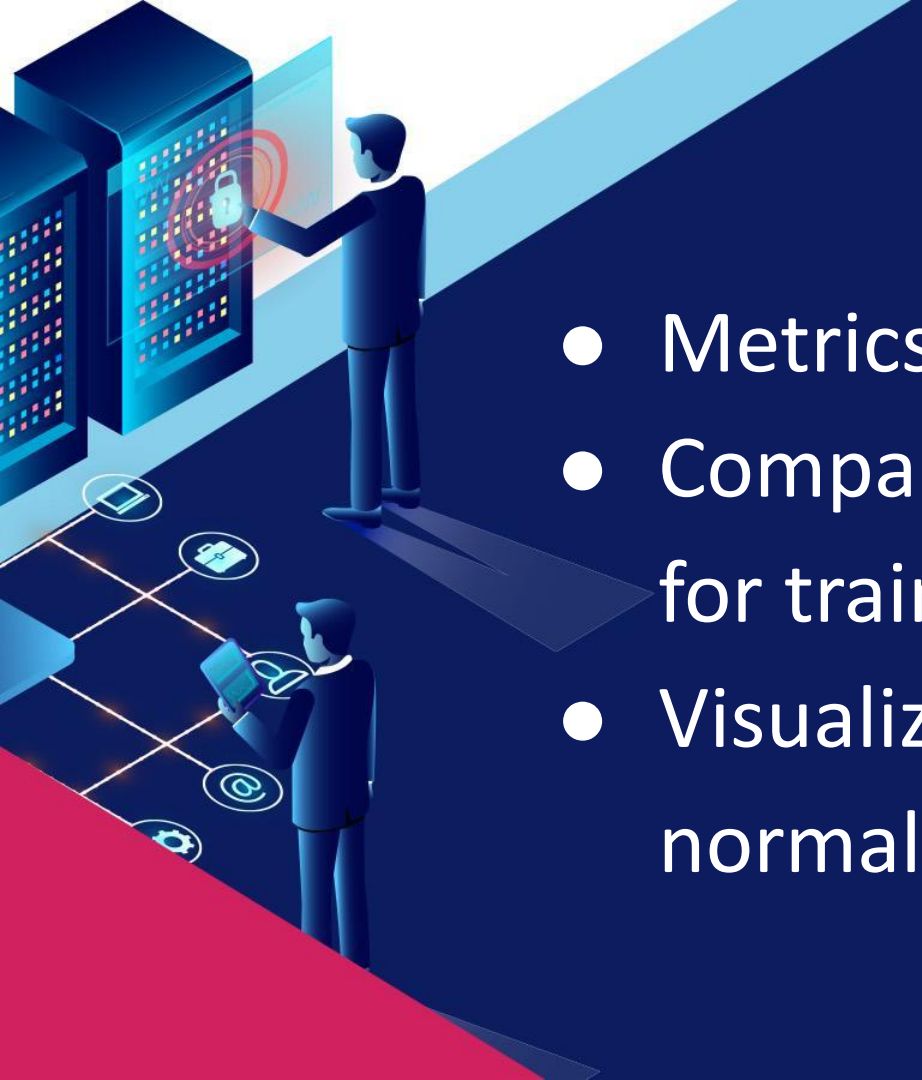


# Data Analysis

- Visualized top cities with the most houses.
- Showed the top ten cities with the highest prices.
- Analyzed seasonal trends in prices
  - Created seasonal variables
  - Visualized trends and differences
- Performed ANOVA test for seasonality

# Linear Regression -Baseline Model

- Detailing the selection of the target and features for the baseline model.
- Explaining the process of splitting the data into training and test sets.
- Providing the model summary and interpreting the findings:
  - F-statistic and R-squared.
  - Interpretation of coefficients.
- Displaying the evaluation metrics:
  - Mean Absolute Error (MAE).
  - Mean Squared Error (MSE).
  - R-squared values for training and test sets.



# Model Evaluation

- Metrics for model evaluation
- Compared R-squared values for training and test models
- Visualized residuals for normality

# Log Transformation

- Discussed the need for target transformation
- Performed a log transformation on the target
- Visualized the transformed target
- Created a new model with the log-transformed target
- Model summary and evaluation

## 2nd Model (Multiple Linear Regression)

- Feature selection
- Standardization of data
- Model creation and summary
- Model evaluation and metrics

# Polynomial Transformation

- Introduction to polynomial transformation
- Transformation of features
- Model creation and summary for polynomial regression
- Analysis of coefficients and p-values

CONCLUSION RECOMMENDATIONS

1. The agency should be on the lookout for features such as square footage of the living area, square footage above,when advising and valuing house for homeowners because they have strong correlations to price.

2. The agency should be on the lookout for houses in the areas: Seattle, Tacoma, Olympia, City, Seatac, Lakewood, Bellevue, JBLM, Lacey, Tukwila because they have the highest number of houses

3. When advising homeowners, the agency should be aware that the areas: Dupont, Lacey, Startup, Port Gamble and Eatonville.

CONCLUSION RECOMMENDATIONS

4. The agency should be aware that the Spring season generally demands higher prices for the houses.

5. The agency should be aware that the Summer season generally demands lower prices for the houses.

NEXT STEPS

1. The agency should look for more data in regards to other house features.

2. The agency should conduct surveys to look find specific factors that cause this seasonal variation so as to understand the market better.

3. The agency should conduct research to find location specific data, such as social amenities, neighborhoods and political stability to understand why certain areas command higher prices as compared to other