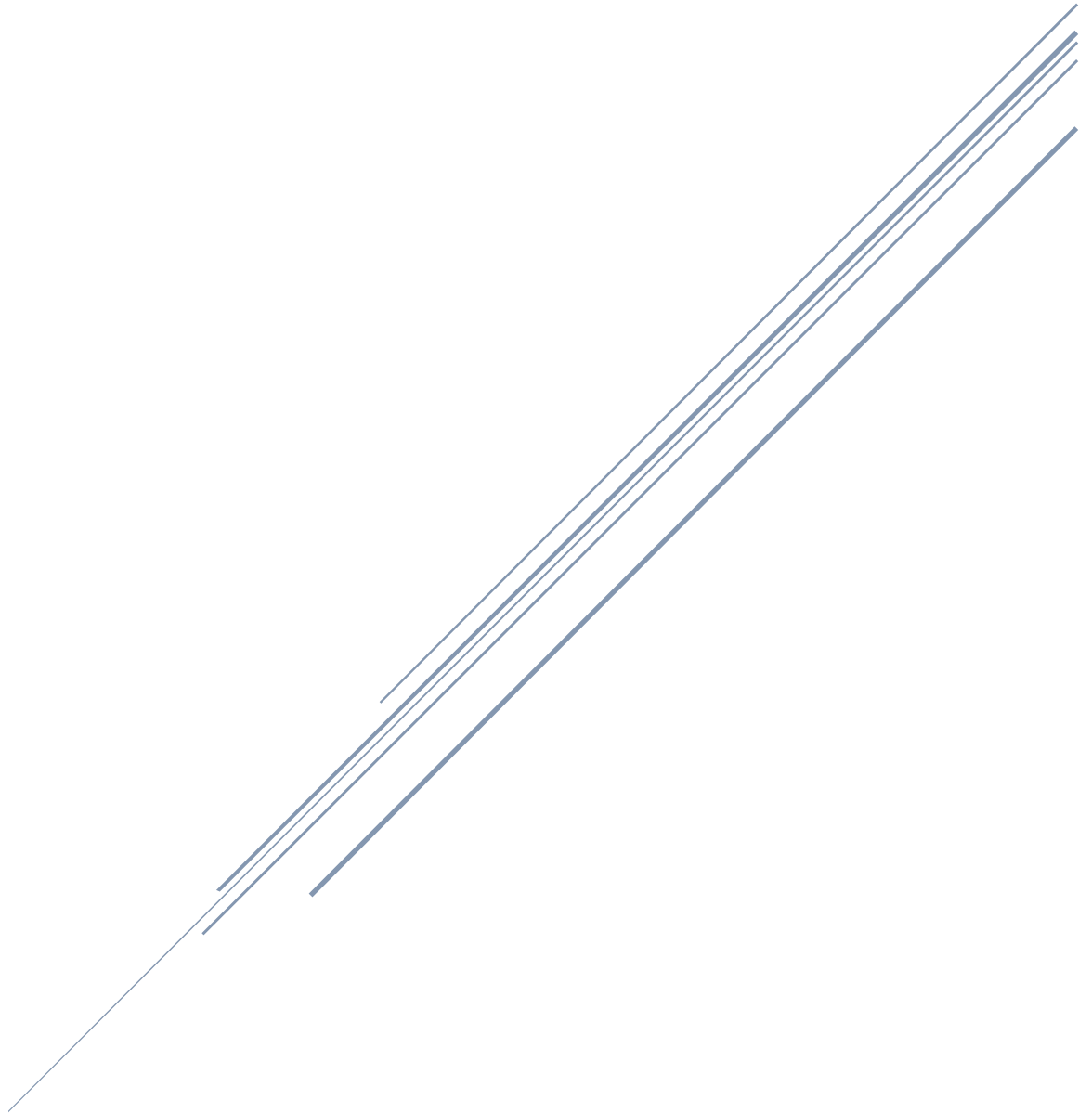


CAPSTONE PROJECT REPORT ON EXPLORATORY ANALYSIS OF HOUSE PRICES IN KING COUNTY, WASHINGTON



Group 7 – Data Analysis

Table of Contents

Executive Summary 2

Business Outcome Highlights..... 2

Background & Motivation 2

Problem Statement 2

Project Objectives 3

Tools & Technologies Used 3

Data Overview 3

ROCCC Framework..... 3

Data Analysis Process 4

 Data Cleaning & Preparation 4

 Grouping for Interpretability..... 4

 Exploratory Data Analysis (EDA) 4

 Power BI Visualization Workflow 5

 Core Measures Created (DAX)..... 5

 Dashboard Design & Layout 5

 Design Specifications..... 6

 Final Output 6

 Dashboards..... 6

Key Insights from the Dashboard 8

Executive Summary

This report presents a comprehensive analysis of residential house prices in King County, Washington. The project involved a complete data analytics workflow, beginning with data acquisition and cleaning, through exploratory analysis, culminating in a professional interactive dashboard built using Microsoft Power BI.

The analysis leveraged tools such as Excel, SQL, and Power BI to uncover trends, patterns, and actionable insights that support strategic real estate decision-making for investors, agents, and other stakeholders.

Business Outcome Highlights

This project:

- Identified high-value zones such as Seattle, Bellevue, and Kirkland
- Highlighted the pricing premium of waterfront and renovated properties
- Revealed seasonal trends in housing prices to support sales timing
- Established the importance of house size, condition, and architectural grade in pricing

Background & Motivation

King County is a significant real estate hub in Washington State. Despite the availability of sales data, valuable insights are often masked by the volume and complexity of the information. This project was initiated to bridge that gap by transforming raw historical data into accessible insights that can inform pricing, marketing, and investment strategies.

Problem Statement

Stakeholders in the King County real estate market frequently rely on subjective judgments for property valuation. The absence of a data-driven framework leads to inconsistent pricing, undervalued properties, and missed investment opportunities. This report addresses that challenge by conducting an exploratory analysis of housing sales data to reveal underlying price influencers and build tools for strategic decision-making.

Project Objectives

- a) Identify the most influential factors affecting house prices
- b) Analyze price trends across locations, house features, and time
- c) Build a dynamic, interactive dashboard to support real estate decision-making
- d) Deliver actionable insights for pricing strategy, investment, and property evaluation

Tools & Technologies Used

- a) **Microsoft Excel:** Data cleaning, preprocessing, and preliminary exploration
- b) **SQL:** Data querying, aggregations, and statistical computations
- c) **Microsoft Power BI:** Visualization and dashboard development

Data Overview

- a) **Source:** King County House Sales Dataset (via Kaggle - https://1drv.ms/x/c/1ff33ac6bbc8d182/EeDV-5FO_o5PhZT8Ea-mbfoBCqK--wSF3GRzPoVgnEPZlw?e=RGhhtA)
- b) **Record Count:** 21,613 property sales
- c) **Date Range:** 2014–2015

ROCCC Framework

To ensure data reliability and relevance, the dataset was evaluated using the ROCCC framework:

- a) **Reliable:** Sourced from official government records
- b) **Original:** Based on King County's proprietary data
- c) **Comprehensive:** Contains over 21 variables including pricing, location, size, and features
- d) **Current:** Reflects market activity during a recent and stable two-year period
- e) **Cited:** Properly attributed and published through Kaggle

Data Analysis Process

Data Cleaning & Preparation

Data preprocessing was performed using Microsoft Excel. Key steps included:

- Removal of null and duplicate records
- Standardization of data formats, particularly date and numerical columns
- Creation of new variables such as price per square foot, renovation flags, and seasonal categories
- Elimination of extreme outliers to enhance data quality and reduce skew

Grouping for Interpretability

To facilitate understanding and make the dataset more usable, several columns were grouped or re-labeled:

Original Column	New Label	Transformation Applied
sqft_living	House Size	Binned into small, medium, and large
waterfront	Are There Waterfronts?	Converted to binary (Yes/No)
view	How Many Views?	Grouped into tiers
condition	Habitability	Re-labeled to reflect livability
grade	Architecture	Interpreted as structural quality
yr_renovated	Renovated/Not Renovated	Binary flag based on renovation year

This transformation enhanced the dashboard's clarity and relevance for both technical and non-technical users.

Exploratory Data Analysis (EDA)

Excel Pivot Tables (with Power Pivot) were used to:

- Examine the relationship between house prices and various features
- Segment data by location, time, and renovation status

- c) Identify seasonal pricing trends through extracted month variables

SQL was used to generate summary statistics such as average, median, and mode for numerical variables, supporting more robust pattern discovery.

Power BI Visualization Workflow

Setup and Data Modeling

- a) Imported the cleaned CSV dataset into Power BI
- b) Verified data types and formatting for all fields
- c) Defined calculated columns such as Renovation Status and House Size
- d) Created DAX measures to support interactive visuals

Core Measures Created (DAX)

- a) Total Listings
- b) Average Price
- c) Waterfront Property Percentage
- d) Average Living Area (sqft)
- e) Aggregated Price Metrics by City, House Size, and Renovation Status

Dashboard Design & Layout

Visual Type	Purpose
KPI Cards	Summary of total listings, average price, and key metrics
Filled Map	Price distribution by zip code
Clustered Bar Chart	Price comparisons by city, house size, and renovation
Stacked Column	Bedroom distribution by size or renovation
Slicers	Filters for city, house features, renovation, and condition

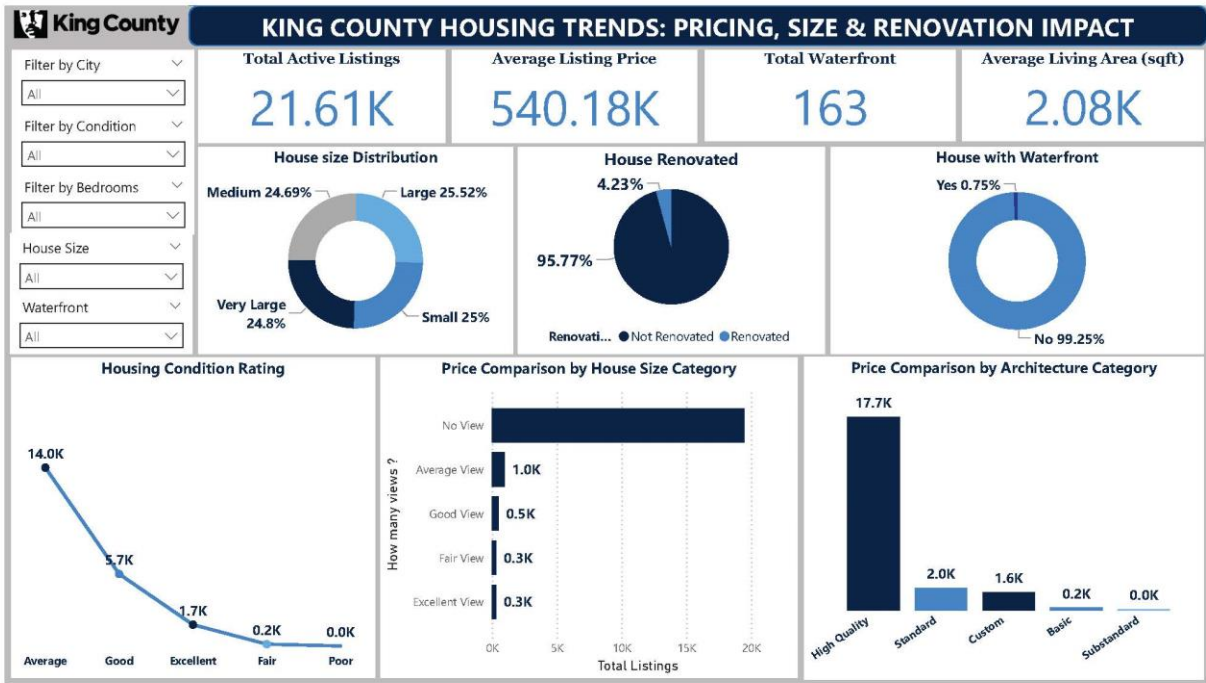
Design Specifications

- a) **Theme Colors:** Navy (#002147), Sky Blue (#5DADEC), Grey (#A9A9A9), White (#FFFFFF)
- b) **Typography:** Segoe UI and Calibri
- c) **Layout Principles:** Grid-based structure with intuitive filtering and tooltips

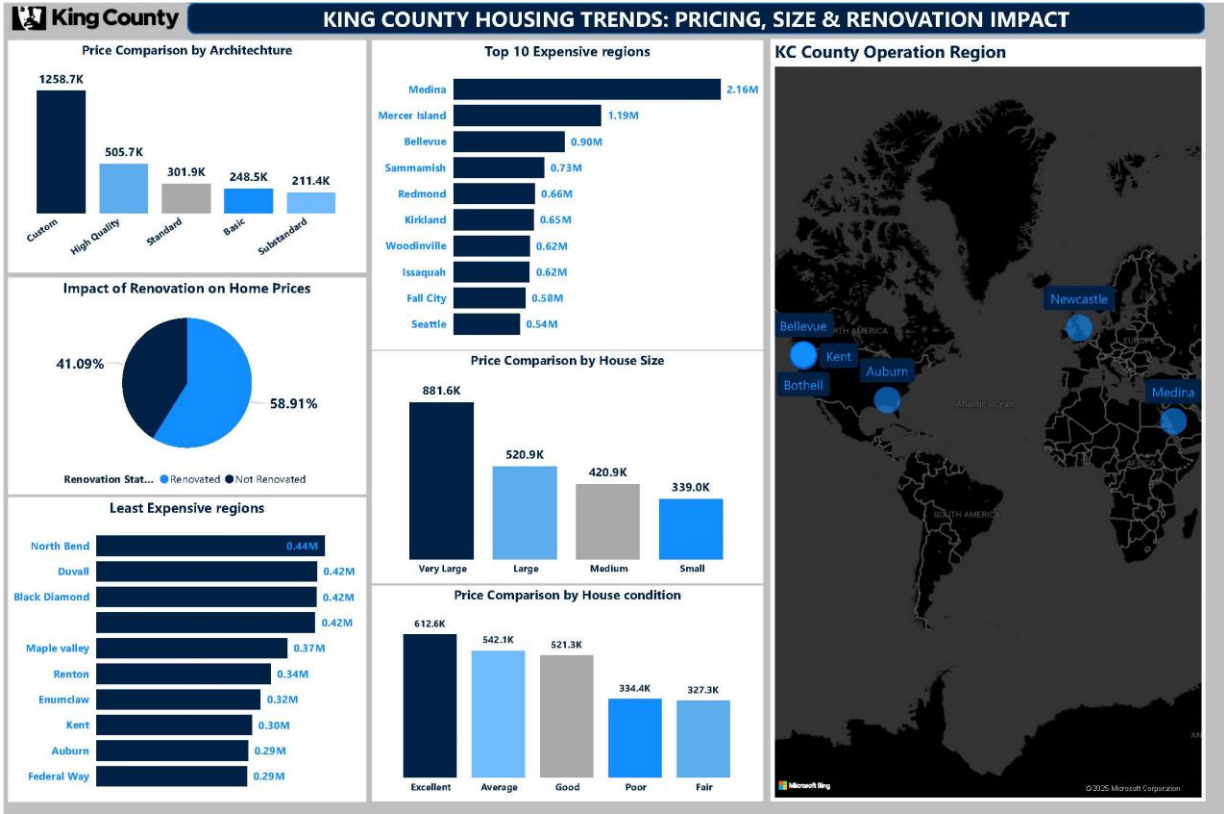
Final Output

- a) Dashboard optimized for interactivity and published in both PDF and Power BI Service formats
- b) Provides a single view for comparative analysis and strategic insight

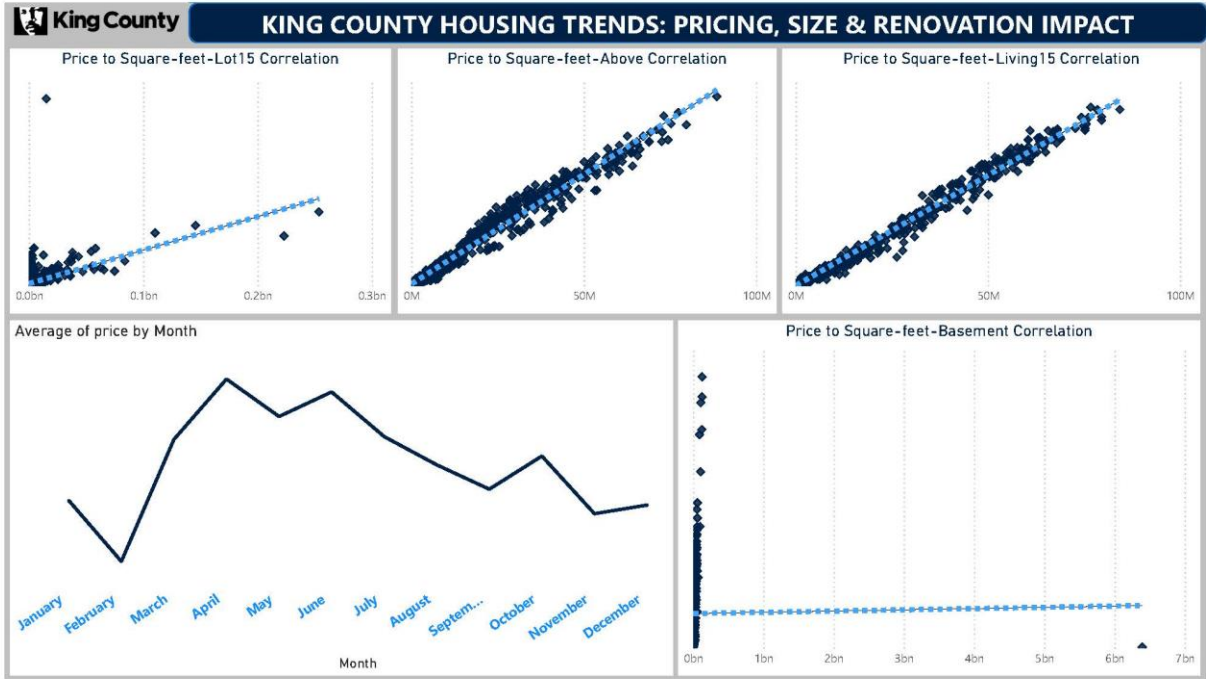
Dashboards



This dashboard reveals that pricing varies drastically by zip code, with waterfront and renovated properties showing premium pricing. Medium and large houses dominate urban centres like Seattle and Bellevue, while smaller homes cluster in more affordable zones like Kent and Auburn. This insight helps us position listings, tailor marketing, and optimize investments based on area-specific demand.

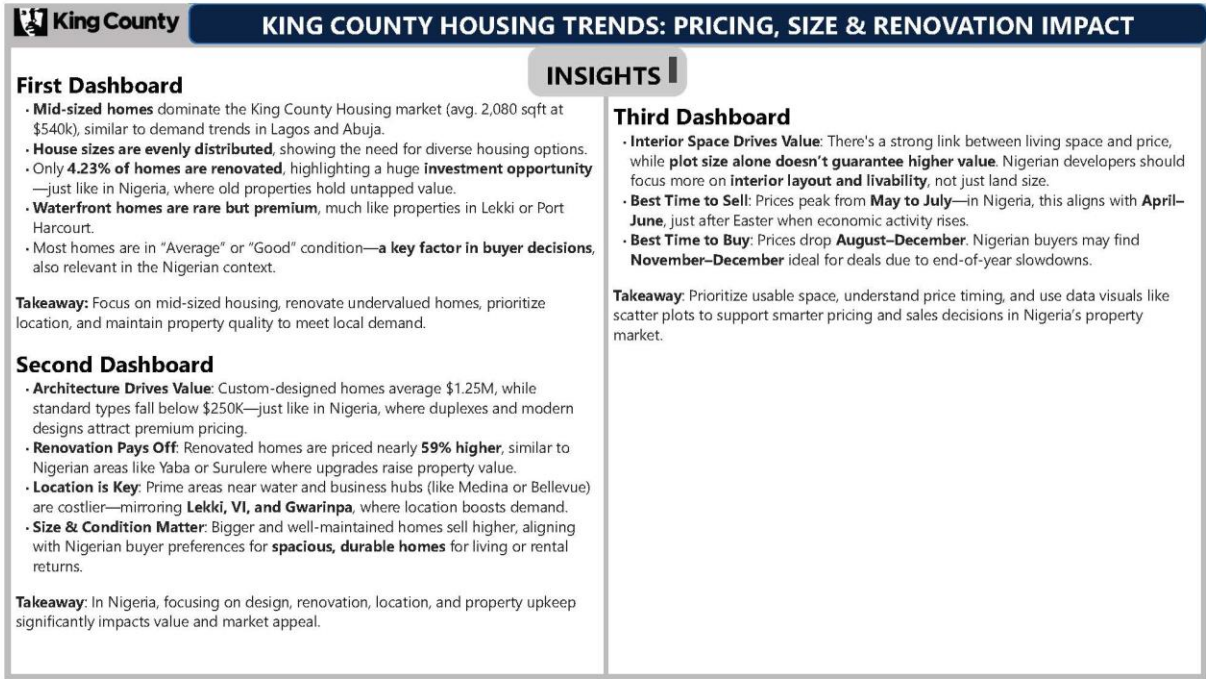


This second dashboard shows the impact of Design, Renovation, location(regions), Size, Condition and the map shows the coverage of the location covered during data collection and also shows the pricing impact when expanded further.



This image shows the correlation analysis between some of the house features with graphical representations of how strongly or weakly a house feature may impact house pricing.

Key Insights from the Dashboard



Conclusion

This project demonstrates how structured data analytics can turn raw information into valuable insights for real estate professionals. By integrating data cleaning, statistical exploration, and visualization, the report provides a reliable framework for strategic pricing, investment targeting, and performance evaluation. The Power BI dashboard further enhances accessibility and engagement through its interactive features and executive-style design.