

RIYADH REAL ESTATE!

Done by:

Leyan saud al samhan - 431203178 ,

Farah Mohammed Alsubaheen - 431204185,

Futun fahad alqudayri - 431202679 ,

Shadan Khalid Aldhawi - 431202748,

Mozn ali alzunidi - 431202856

supervised :

Dr.Amal aljohny

Machine Learning Meets Riyadh Real Estate: An Empirical Perspective

عقاري.ai

2024-11-13

Machine Learning Meets Riyadh Real Estate: An Empirical Perspective!

Our project aims to analyze the real estate market in Saudi Arabia, specifically in the city of Riyadh, focusing on understanding the price differences between residential and commercial properties as well as various residential units across different areas of the city. Through this project, we collect comprehensive data on rental prices in Riyadh, categorized by region, and investigate the reasons behind price fluctuations and trends in Saudi Arabia's real estate market. This analysis provides investors and renters with a clear overview of price trends and market variations based on location and property type, ultimately supporting informed investment decisions.

Data collection

The data was collected from the official Aqar website using web scraping through the Aqar API. We utilized Python's requests library to create a request header that mimics the official Aqar website, enabling us to access the data.

Title	Description
Title	Akar2024
Purpose	For an empirical study of Riyadh real estate
Owner	عقاري.ai
Target	price
Restriction	Only for members of عقاري.ai
Date of collection	2024/10/24
Scraped from	/https://sa.aqar.fm
Number of attributes	37
Number of records	111297
Category	Real estate
Location	Riyadh, Saudi Arabia

Alt Text

This dataset provides detailed information about property listings, including unique identifiers (id), location (city, district, location.lat, location.lng), and attributes like price, area, age, and category (property type). It also includes amenities such as WC, furnished, pool, and lift, along with user details (user.name, user.review, user.am verified). Binary columns indicate the availability of features like water, electricity, and drainage, while room counts (beds, kitchen, living) provide additional context. This structured metadata supports comprehensive analysis of property features and user preferences. - these methods may be unethical as they access data without explicit permission, potentially violating the website's terms of service.

```
library(dplyr)
library(ggplot2)
library(scales)
library(readr)
library(gbm)
library(randomForest)
library(caret)
library(gridExtra)
library(showtext)
library(tidyr)

library(knitr)
```

Hypotheses

Focusing on The Area, number of rooms and their kind, and the location is sufficient to make a predictive model that predict the price range of the property.

```

#import pandas as pd
#from sklearn.model_selection import train_test_split, RandomizedSearchCV
#from sklearn.ensemble import GradientBoostingRegressor
#from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
#import numpy as np
#import gradio as gr
#import joblib

# Load the dataset
#df = pd.read_csv('#combine.csv')

# Define features and target variable
#X = df[['livings', 'kitchen', 'rooms', 'wc', 'area', 'category', 'encoded_direction']]
#y = df['price']

# Split the dataset into training and testing sets
#X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Hyperparameter tuning using RandomizedSearchCV
#param_grid = {
#    'n_estimators': [100, 200, 300],
#    'learning_rate': [0.01, 0.05, 0.1, 0.2],
#    'max_depth': [3, 5, 7]
#}

# Initialize the model
#model = GradientBoostingRegressor(random_state=42)

# Perform randomized search with cross-validation
#random_search = RandomizedSearchCV(model, param_grid, n_iter=10, scoring='neg_mean_absolute_error', cv=5, random_state=42)
#random_search.fit(X_train, y_train)

# Best model after hyperparameter tuning
#best_model = random_search.best_estimator_

# Train the best model
#best_model.fit(X_train, y_train)

# Quantile Regression to predict price range (lower and upper bounds)
#lower_model = GradientBoostingRegressor(loss="quantile", alpha=0.1, random_state=42)
#upper_model = GradientBoostingRegressor(loss="quantile", alpha=0.9, random_state=42)

# Train the quantile models
#lower_model.fit(X_train, y_train)
#upper_model.fit(X_train, y_train)

# Save the best model and quantile models
#joblib.dump(best_model, "Best_Gradient_Model.joblib")
#joblib.dump(lower_model, "Lower_Quantile_Model.joblib")
#joblib.dump(upper_model, "Upper_Quantile_Model.joblib")

# Make predictions for the test set
#y_pred = best_model.predict(X_test)
#lower_pred = lower_model.predict(X_test)
#upper_pred = upper_model.predict(X_test)

# Calculate the accuracy within the predicted range
#within_range = (y_test >= lower_pred) & (y_test <= upper_pred)
#range_accuracy = within_range.mean() * 100 # Convert to percentage

#print(f"Accuracy within predicted range: {range_accuracy:.2f}%")

```

++++++

: (f"Accuracy within predicted range: {range_accuracy:.2f}%")

Accuracy within predicted range: 79.21%

- The results support the hypothesis that focusing on property area, room count and type, and location can effectively predict property price range. using a Gradient Boosting Regressor The model achieved an accuracy of 79.2%, highlighting its effectiveness in capturing key determinants of pricing trends. ————— As properties age, the impact of age on price is moderated by other factors like area and location; older properties in high-demand locations retain value better than those in less desirable areas

```

akar2 <- read_csv("akar2024.csv")

remove_outliers <- function(data, columns) {
  for (column in columns) {

    Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
    Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
    IQR_value <- Q3 - Q1

    lower_bound <- Q1 - 1.5 * IQR_value
    upper_bound <- Q3 + 1.5 * IQR_value

    data <- data %>%
      filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
  }

  return(data)
}

city_data <- akar2 %>%
  filter(city == "الرياض")

city_data_no_outliers <- remove_outliers(city_data, c("age", "price"))
head(city_data_no_outliers)

```

```

## # A tibble: 6 × 37
##   ...1     id rent_period title  price city   beds district category furnished
##   <dbl>   <dbl>       <dbl> <chr> <dbl> <chr> <dbl> <chr>      <dbl>
## 1     1 5840917        3 1.20 ... ٩٦٢٢ ... حي الرمـ... 4 حـي الـرمـ... NA
## 2     2 5954514        3 6.5 ... ٤٠  حـي العـا... 3 حـي العـا... الـريـ...
## 3     3 5885280        3 1.35 ... ٦٥٠  حـي العـقـ... 4 حـي العـقـ... الـريـ...
## 4     4 5867744        3 7.5 ... ٤٠  حـي العـا... 4 حـي العـا... الـريـ...
## 5     5 5919112        3 4.40 ... ٨٤١  حـي طـوـيقـ... 4 حـي طـوـيقـ... الـريـ...
## 6     8 5779951        2 3.3 ... ٤٣  حـي المـوـ... 1 حـي المـوـ... الـريـ...
## # i 27 more variables: ketchen <dbl>, livings <dbl>, playground <dbl>,
## #   pool <dbl>, rooms <dbl>, water_availability <lgl>,
## #   electrical_availability <lgl>, drainage_availability <lgl>,
## #   private_roof <lgl>, apartment_in_villa <lgl>, daily_rentable <lgl>,
## #   age <dbl>, area <dbl>, wc <dbl>, content <chr>, direction <chr>,
## #   lift <dbl>, verified <dbl>, two_entrances <lgl>, special_entrance <lgl>,
## #   user.name <chr>, user.review <dbl>, user.iam_verified <lgl>, ...

```

```

top_5_categories <- city_data_no_outliers %>%
  count(category, sort = TRUE) %>%
  slice_head(n = 5)

city_data2 <- city_data %>%
  filter(category %in% c(22, 3, 6)) %>%
  select(category, price, age, district, direction, area)

city_data2 <- city_data2 %>%
  mutate(category = as.character(category)) %>%
  mutate(category = case_when(
    category == "3" ~ "فلا للبيع",
    category == "6" ~ "شقـه للـبيـع",
    category == "22" ~ "دور للـبيـع",
    TRUE ~ category
  ))

city_data2_no_outliers <- remove_outliers(city_data2, c("price"))
head(city_data2_no_outliers)

```

```

## # A tibble: 6 × 6
##   category     price   age district      direction     area
##   <chr>       <dbl> <dbl> <chr>       <chr>       <dbl>
## 1 220        ٠ حـي الـرمـا... شـرق الـرمـا... 1200000  دور للـبيـع
## 2 170        ٠ حـي طـوـيقـ... غـرب الـرمـا... 599000   شـقـه للـبيـع
## 3 140        ٠ حـي السـوـديـ... غـرب الـرمـا... 650000   شـقـه للـبيـع
## 4 136        ٠ حـي بـدرـ... جـنـوب الـرمـا... 625000   دور للـبيـع
## 5 196        ٠ حـي طـهـرـهـ... نـماـرـ غـرب الـرمـا... 530000   شـقـه للـبيـع
## 6 205        ٠ حـي الشـفـاـ... جـنـوب الـرمـا... 720000   فيـلا للـبيـع

```

the first part of the hypothesis (As properties age, the impact of age on price is moderated by other factors like area and location for sale)

```

# Function to generate plots for age impact analysis
property_age_analysis <- function(data, category) {
  # Filter data for the specific category and select relevant columns
  data <- data %>%
    mutate(age_group = case_when(
      age <= 2 ~ "0-2 years",
      age <= 4 ~ "2-4 years",
      age <= 6 ~ "4-6 years",
      age <= 8 ~ "6-8 years",
      age <= 10 ~ "8-10 years",
      age <= 12 ~ "10-12 years",
      age <= 14 ~ "12-14 years",
      age <= 16 ~ "14-16 years",
      age <= 18 ~ "16-18 years",
      age <= 20 ~ "18-20 years",
      age <= 22 ~ "20-22 years",
      age <= 24 ~ "22-24 years",
      age <= 26 ~ "24-26 years",
      age <= 28 ~ "26-28 years",
      age <= 30 ~ "28-30 years",
      age > 30 ~ "30+ years"
    ))
  category_data <- data %>%
    filter(category == !!rlang::enquo(category)) %>%
    select(category, price, age, district, area, direction, age_group)

  category_data <- na.omit(category_data)

  demand_classification <- category_data %>%
    group_by(district) %>%
    summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop") %>%
    mutate(demand = ifelse(avg_price >= median(avg_price, na.rm = TRUE), "مُنْهَجٌ لِلِّيَابَانِ", "مُنْهَجٌ لِلِّيَابَانِ"))

  category_data <- category_data %>%
    left_join(demand_classification %>% select(district, demand), by = "district")

  summary_stats <- category_data %>%
    group_by(demand, area, age) %>%
    summarize(
      mean_price = mean(price, na.rm = TRUE),
      median_price = median(price, na.rm = TRUE),
      count = n(),
      .groups = "drop"
    )
  print("Summary Statistics:")
  print(summary_stats)

  correlation_data <- category_data %>%
    mutate(
      district_numeric = as.numeric(factor(district)),
      direction_numeric = as.numeric(factor(direction))
    ) %>%
    select(price, age, area, district_numeric, direction_numeric)

  correlation_matrix <- cor(correlation_data, use = "complete.obs")
  print("Correlation Matrix:")
  print(correlation_matrix)

  linear_model <- lm(price ~ age + area + district, data = category_data)
  print("Linear Regression Model Summary:")
  print(summary(linear_model))

  # Creating plots

  # Plot 2: Bar Plot of Average Price by Age Group
  p2 <- ggplot(category_data %>% group_by(age_group) %>%
                summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop") %>%
                mutate(age_group = factor(age_group, levels = c("0-2 years", "2-4 years", "4-6 years", "6-8 years", "8-10 years", "10-12 years", "12-14 years", "14-16 years", "16-18 years", "18-20 years", "20-22 years", "22-24 years", "24-26 years", "26-28 years", "28-30 years", "30+ years"))),
                aes(x = age_group, y = avg_price)) +
    geom_bar(stat = "identity", fill = "lightcoral", width = 0.7) +
    labs(title = paste("السعر مقابل العمر للفئة:", category),

```

```

x = "عمر العقار (بالسنوات",
y = "السعر"
scale_y_continuous(labels = scales::comma) +
theme_minimal() +
theme(text = element_text(family = "IBM")) +
coord_flip()

print(p2)

# Facet Grid Plot of Price vs Age by Direction
p3 <- ggplot(category_data, aes(x = age, y = price)) +
  geom_point(alpha = 0.7) +
  labs(title = paste("السعر حسب العمر حسب الاتجاه للفئة:", category),
       x = "عمر العقار",
       y = "السعر") +
  scale_y_continuous(labels = scales::comma) +
  facet_wrap(~ direction) +
  theme_minimal() +
  theme(text = element_text(family = "IBM"))

print(p3)

p5 <- ggplot(category_data %>% group_by(district, demand) %>%
               summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop") %>%
               top_n(50, avg_price),
               aes(x = reorder(district, avg_price), y = avg_price, fill = demand)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = paste("متوسط السعر حسب الحي للفئة:", category),
       x = "الحي",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  scale_fill_manual(values = c("lightcoral", "grey")) +
  coord_flip() +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), text = element_text(family = "IBM"))

print(p5)

# Bar Plot of Average Price by Direction
p6 <- ggplot(category_data %>% group_by(direction) %>%
               summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop"),
               aes(x = direction, y = avg_price, fill = direction)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = paste("متوسط السعر حسب الاتجاه للفئة:", category),
       x = "الاتجاه",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(text = element_text(family = "IBM"))

print(p6)

}

property_age_analysis(data = city_data2_no_outliers, category = "دور للبيع")

```

```

## [1] "Summary Statistics:"
## # A tibble: 510 x 6
##   demand    area    age mean_price median_price count
##   <chr>     <dbl>  <dbl>      <dbl>        <dbl> <int>
## 1 1      1500000  1500000    1      56 طلب عالي
## 2 1      1300000  1300000    0      58 طلب عالي
## 3 1      950000   950000     0      64 طلب عالي
## 4 1      1100000  1100000    0      75 طلب عالي
## 5 1      799000   799000     0      77 طلب عالي
## 6 2      1199500  1199500    0      80 طلب عالي
## 7 1      1150000  1150000    1      85 طلب عالي
## 8 1      1550000  1550000    0      95 طلب عالي
## 9 1      1600000  1600000    3      95 طلب عالي
## 10 4     1050000  1325000   5     105 طلب عالي
## # i 500 more rows
## [1] "Correlation Matrix:"
##                   price        age       area district_numeric
## price      1.0000000  0.11281499  0.258295444 -0.199210833
## age       0.1128150  1.00000000  0.066267434 -0.115397515

```

```

## area          0.2582954  0.06626743 1.000000000  0.002375823
## district_numeric -0.1992108 -0.11539751 0.002375823 1.000000000
## direction_numeric  0.2381083 -0.03188859 0.052003364 0.119058497
##               direction_numeric
## price           0.23810833
## age            -0.03188859
## area            0.05200336
## district_numeric  0.11905850
## direction_numeric 1.000000000
## [1] "Linear Regression Model Summary:"
##
## Call:
## lm(formula = price ~ age + area + district, data = category_data)
##
## Residuals:
##    Min      1Q   Median      3Q      Max
## -1284342 -77553 -14348   64868 1834315
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                703162.97   38447.94 18.289 < 2e-16 ***
## age                      -1356.48    2219.89 -0.611 0.541217
## area                     170.56     26.86  6.350 2.55e-10 ***
## district2.01 5.213 83220.38 433808.32   6.07*** حي الشيشان
## district3.12 4.673 103782.82 484992.24   7.40*** حي العارض
## district0.346783 0.941 141453.17 133110.56   0.44*** حي البرية
## district0.001240 3.233 196707.59 635972.88   0.23*** حي التماعون
## district1.77 5.652 65879.03 372317.66   0.98*** حي الخطاويه
## district0.563765 0.577- 42666.31 24632.69-   0.09*** حي الحرم
## district1.03 6.128 118655.86 727172.54   0.11*** حي الخالدية
## district0.605181 0.517- 71754.67 37099.25-   0.05*** حي الظاهر
## district8.59 5.776 69336.44 400480.61   0.07*** حي الدار البيضاء
## district0.001427 3.193- 44740.80 142837.05-   0.02*** حي الدربيه
## district0.942363 0.072 141473.92 10229.64   0.00*** حي العلوي
## district2 > 20.058 65501.76 1313852.69   0.00*** حي العلوي
## district5.48 6.229 196481.15 1223837.25   0.00*** حي العلوي
## district2 > 9.372 39637.41 371480.96   0.00*** حي العلوي
## district2.73 5.968 104166.07 621693.11   0.00*** حي العلوي
## district4.52 4.086 103535.83 423050.34   0.00*** حي العلوي
## district0.080257 1.750 71776.31 125601.24   0.00*** حي العلوي
## district2.62 5.975 45157.34 269813.85   0.00*** حي العلوي
## district0.668846 0.428- 196441.76 84034.37-   0.00*** حي العلوي
## district2 > 18.623 198266.00 3692366.15   0.00*** حي العلوي
## district0.453249 0.750- 52533.35 39406.61-   0.00*** حي العلوي
## district0.855284 0.182- 63918.23 11658.63-   0.00*** حي العلوي
## district1.42 4.350 50562.38 219935.91   0.00*** حي العلوي
## district0.256708 1.134- 53463.06 60651.79-   0.00*** حي العلوي
## district2.02 7.694 117649.99 905252.57   0.00*** حي العلوي
## district3.46 7.624 117613.01 896640.99   0.00*** حي العلوي
## district0.029119 2.183 94779.23 206913.55   0.00*** حي العلوي
## district2 > 16.814 42774.48 719199.99   0.00*** حي العلوي
## district0.360622 0.914 103539.35 94670.67   0.00*** حي العلوي
## district0.039717 2.058- 117581.59 241952.51-   0.00*** حي العلوي
## district0.341474 0.951- 41326.32 39319.29-   0.00*** حي العلوي
## district2.79 8.237 82136.37 676569.38   0.00*** حي العلوي
## district0.177019 1.350- 117831.06 159115.05-   0.00*** حي العلوي
## district0.008590 2.630 196442.97 516647.87   0.00*** حي العلوي
## district6.33 8.136 46030.89 374500.08   0.00*** حي العلوي
## district1.16 6.472 141526.88 915949.63   0.00*** حي العلوي
## district2 > 17.699 87805.10 1554059.88   0.00*** حي العلوي
## district2 > 8.433 141458.94 1192891.55   0.00*** حي العلوي
## district0.318878 0.997- 117554.38 117197.14-   0.00*** حي العلوي
## district2 > 9.864 74721.84 737022.07   0.00*** حي العلوي
## district3.10 5.553 58226.90 323317.08   0.00*** حي العلوي
## district2 > 9.531 117569.39 1120570.76   0.00*** حي العلوي
## district0.498863 0.676 196453.82 132876.95   0.00*** حي العلوي
## district2.21 7.979 87570.31 698734.64   0.00*** حي العلوي
## district0.053396 1.933 201897.71 390187.71   0.00*** حي العلوي
## district2 > 28.587 64190.08 1834996.45   0.00*** حي العلوي
## district2 > 13.057 42947.35 560771.27   0.00*** حي العلوي
## district0.007776 2.664 219315.06 584202.50   0.00*** حي العلوي
## district2 > 20.694 44357.54 917914.03   0.00*** حي العلوي
## district1.96 4.277 197178.29 843332.82   0.00*** حي العلوي
## district0.716435 0.363 141445.83 51382.73   0.00*** حي النسيم الشرقي
## district0.246258 1.160 87682.65 101690.20   0.00*** حي النسيم الغربي
## district0.009427 2.598 117815.63 306100.65   0.00*** حي النظيم
## district6.31 6.905 196431.05 1356414.26   0.00*** حي العلوي
## district2.30 7.370 63051.26 464686.74   0.00*** حي العلوي
## district8.31 7.193 141738.43 1019500.89   0.00*** حي العلوي

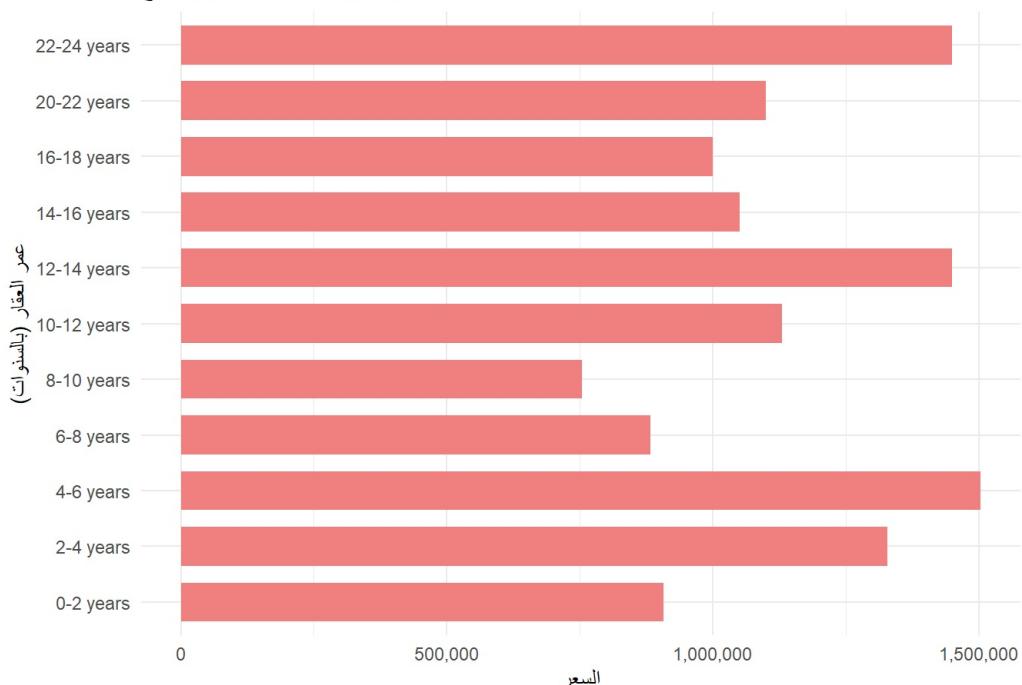
```

```

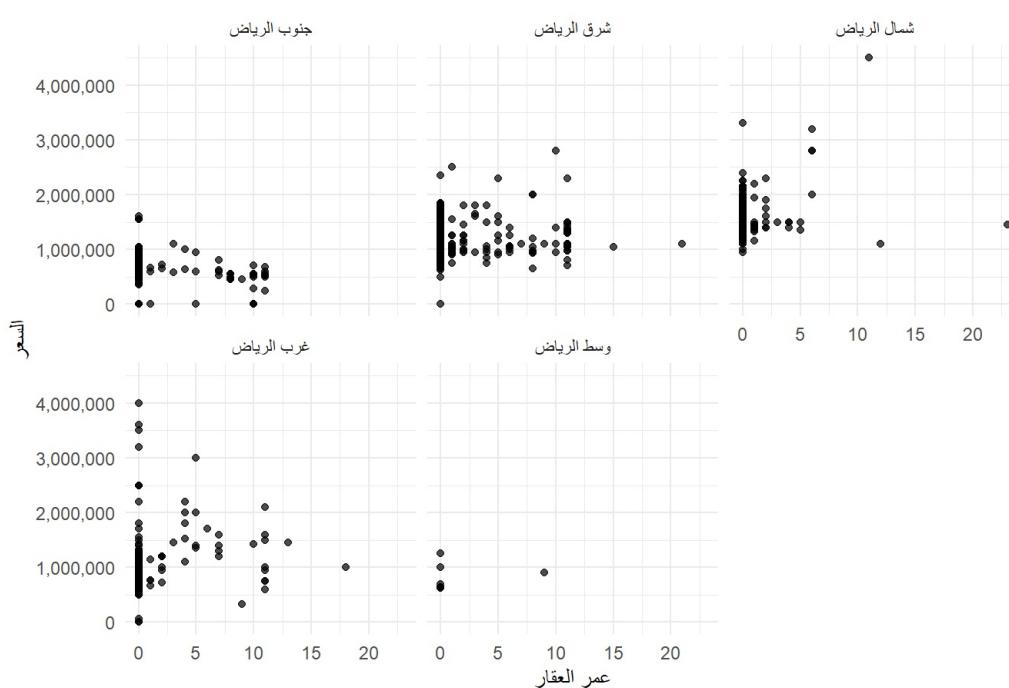
## district2 > 13.906 78040.04 1085227.68 حي #البيهقي
## district8.06 8.106 53510.10 433739.76 حي #البيهقي
## district0.524029 0.637 196476.37 125201.73 حي ام الحمام الشرقي
## district7.79 3.957- 38837.83 153690.27- حي #البيهقي
## district2.82 4.195 145115.03 608801.21 حي #البيهقي
## district0.001876 3.113 62536.60 194646.90 حي ديراب
## district0.438551 0.775- 61299.30 47492.47- حي سلطانة
## district0.000228 3.691 87382.20 322537.99 حي شبرا
## district1.78 5.650 47862.98 270421.39 حي صالحية
## district0.036037 2.098 38938.30 81677.93 حي طوبق
## district0.127779 1.523- 87304.15 133000.62- حي طيبة
## district0.252484 1.145 69431.69 79471.25 حي طهرة البدعية
## district3.21 5.124 142377.02 729580.29 حي #البيهقي
## district0.040114 2.054 42870.00 88038.86 حي طهرة نمار
## district5.99 7.551 143614.91 1084409.25 حي #البيهقي
## district0.859736 0.177- 196440.62 34716.61- حي عريص
## district0.033736 2.124- 40996.08 87090.52- حي عكاظ
## district6.83 8.126 94255.40 765953.20 حي #البيهقي
## district2 > 15.392 53123.21 817695.64 حي #البيهقي
## district0.005588 2.773 117539.39 325980.45 حي مطار الملك خالد الدولي
## district0.812657 0.237 61291.32 14527.55 حي نمار
## ---
## Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 192800 on 2539 degrees of freedom
## Multiple R-squared: 0.7759, Adjusted R-squared: 0.769
## F-statistic: 111.3 on 79 and 2539 DF, p-value: < 2.2e-16

```

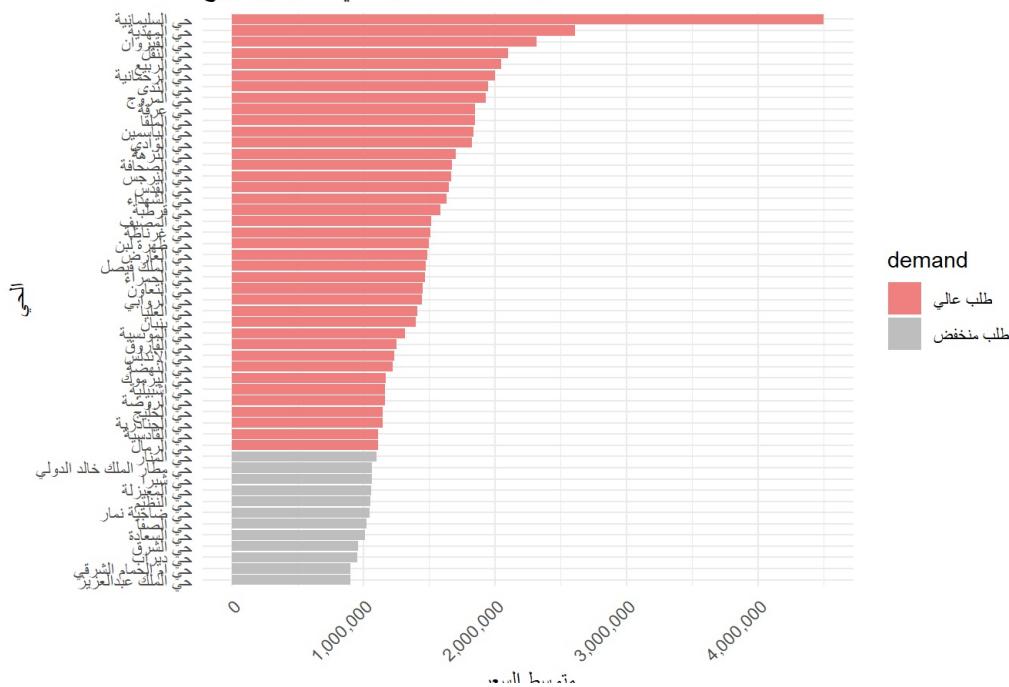
السعر مقابل العمر للفئة: دور للبيع



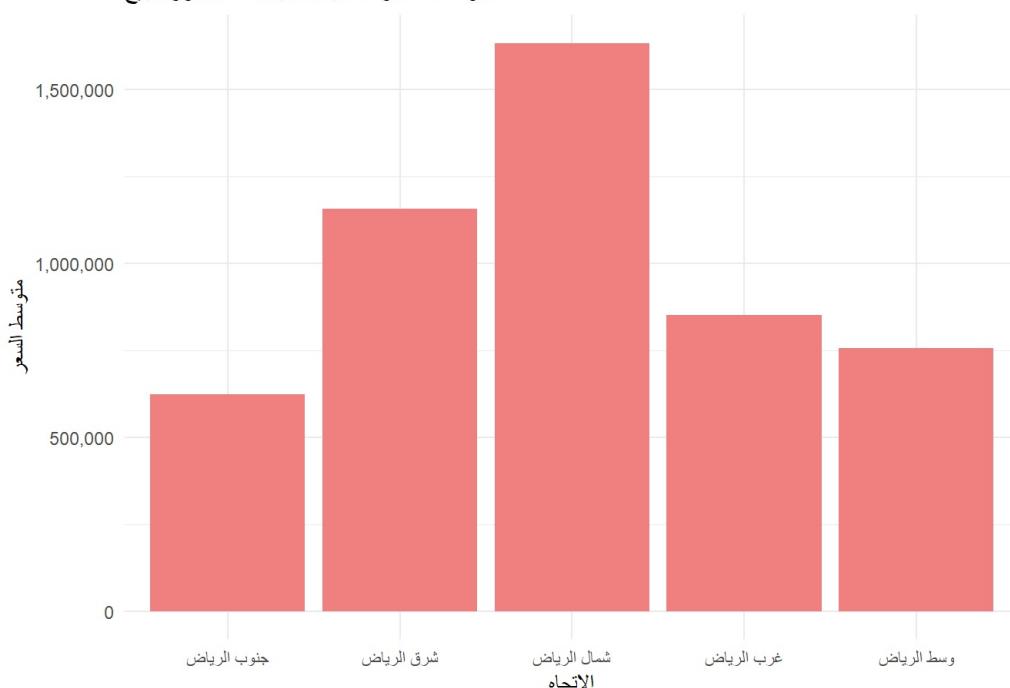
السعر حسب العمر حسب الاتجاه للفئة: دور للبيع



متوسط السعر حسب الحي للفئة: دور للبيع



متوسط السعر حسب الاتجاه للبيع: دور للبيع



```
property_age_analysis(data = city_data2_no_outliers, category = "فيلا للبيع")
```

```
## [1] "Summary Statistics:"
## # A tibble: 1,791 × 6
##   demand    area    age mean_price median_price count
##   <dbl>    <dbl>  <dbl>      <dbl>        <dbl> <int>
## 1 2       3475000  3475000  11       37      طلب عالي
## 2 1       780000   780000   0        91      طلب عالي
## 3 5       3000000  3320000  0        119     طلب عالي
## 4 1       5000000  5000000  2        137     طلب عالي
## 5 1       2300000  2300000  0        139     طلب عالي
## 6 2       3375000  3375000  3        139     طلب عالي
## 7 3       3650000  3350000  11       148     طلب عالي
## 8 3       1250000  1250000  0        150     طلب عالي
## 9 1       2000000  2000000  36       150     طلب عالي
## 10 1      4500000  4500000  0        154     طلب عالي
## # i 1,781 more rows
## [1] "Correlation Matrix:"
##            price      age      area district_numeric
## price 1.00000000 0.160236869 0.010638064 -0.064578191
## age   0.16023687 1.000000000 -0.005795679 -0.131868406
## area  0.01063806 -0.005795679 1.000000000 -0.009956989
## district_numeric -0.06457819 -0.131868406 -0.009956989 1.000000000
## direction_numeric 0.11401680 -0.134802961 0.001162506 0.221047780
##            direction_numeric
## price 0.114016798
## age  -0.134802961
## area  0.001162506
## district_numeric 0.221047780
## direction_numeric 1.000000000
## [1] "Linear Regression Model Summary:"
## 
## Call:
## lm(formula = price ~ age + area + district, data = category_data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -3661881 -440987 -90997  330053  3604326 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 8.782e+05  1.151e+05  7.627 2.65e-14 ***
## age         1.794e+04  1.789e+03  10.027 < 2e-16 ***
## area        9.927e-02  1.729e-01   0.574 0.565764    
## district2.126  ٤٠٦١٣١٠e+05 16.237 < 2e-16 ***
## district2.341  ٩٤٠٦٢١٢٩٠e+05 10.993 < 2e-16 ***
## district4.217  ٥٤٥٥٧١٧٠٠٣e+05  0.602 0.547046    
## district2.061  ٥٤٠٦١٩٣٥e+05 10.652 < 2e-16 ***
## district1.224  ٤٠٦٣٦٤٥e+05  3.359 0.000785 ***
```

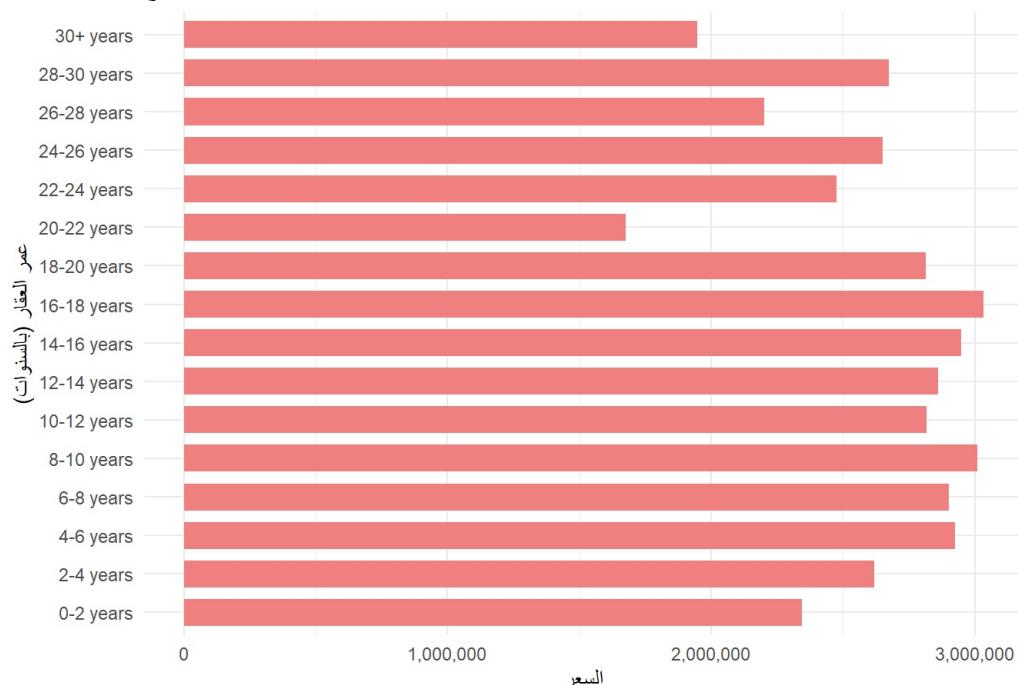
district2.173 +0417.903e+05 0.031 0.975250
 ## district2.201 05062.387e+05 9.218 < 2e-16 ***
 ## district9.615 +04703e+05 0.137 0.890807
 ## district1.012 +0472.588e+05 3.912 9.21e-05 ***
 ## district8.893 +0505367e+05 6.507 8.09e-11 ***
 ## district7.373- +0413.644e+05 -0.202 0.839645
 ## district4.702 +05051.482e+05 3.172 0.001518 **
 ## district2.690 +0511.597e+05 16.847 < 2e-16 ***
 ## district3.191 +0531300e+05 0.967 0.333632
 ## district1.720 +0611.313e+05 13.101 < 2e-16 ***
 ## district1.370 +0511.351d+05 1.014 0.310578
 ## district1.944 +0521859e+05 0.680 0.496609
 ## district2.545 +0611.379e+05 14.220 < 2e-16 ***
 ## district3.098 +0612.940e+05 15.190 < 2e-16 ***
 ## district3.565 +062304e+05 15.472 < 2e-16 ***
 ## district2.894 +0611.824e+05 15.868 < 2e-16 ***
 ## district1.147 +0611.191e+05 9.631 < 2e-16 ***
 ## district4.068 +0617.993e+05 5.809 6.51e-09 ***
 ## district1.693 +062.048e+05 8.266 < 2e-16 ***
 ## district2.403 +0611.645e+05 14.606 < 2e-16 ***
 ## district1.976 +0612.708e+05 7.299 3.15e-13 ***
 ## district9.244 +054.155e+05 2.225 0.026111 *
 ## district4.600 +0511.730e+05 2.659 0.007843 **
 ## district1.252 +0511.400e+05 8.941 < 2e-16 ***
 ## district1.772 +051.2133e+05 8.308 < 2e-16 ***
 ## district2.217 +051.7003e+05 0.317 0.751522
 ## district2.569 +051.474e+05 10.384 < 2e-16 ***
 ## district2.604 +051.451e+05 1.794 0.072878 .
 +05051.085e+05 3.139 0.001703 **
 +051.161e+05 1.844 0.065263 .
 +051.386e+05 10.432 < 2e-16 ***
 +051.371e+05 0.632 0.527458
 +052.580e+05 8.947 < 2e-16 ***
 +051.005e+05 -1.179 0.238580
 +051.362e+05 21.684 < 2e-16 ***
 +051.235e+05 5.694 1.28e-08 ***
 +051.7005e+05 3.461 0.000541 ***
 +051.187e+05 19.652 < 2e-16 ***
 +051.047e+05 2.428 0.015211 *
 +051.563e+05 2.001 0.045470 *
 +051.782e+05 2.287 0.022193 *
 +051.564e+05 1.690 0.091049 .
 +051.454e+05 19.436 < 2e-16 ***
 +051.376e+05 15.262 < 2e-16 ***
 +051.5033e+05 1.173 0.240949
 +051.7003e+05 3.616 0.000301 ***
 +051.4355e+05 8.644 < 2e-16 ***
 +051.1151e+05 -0.301 0.763449
 +051.005e+05 1.891 0.058711 .
 +051.665e+05 5.234 1.69e-07 ***
 +051.2180e+05 12.823 < 2e-16 ***
 +051.7005e+05 -0.179 0.857661
 +051.623e+05 9.433 < 2e-16 ***
 +051.336e+05 9.919 < 2e-16 ***
 +051.708e+05 17.506 < 2e-16 ***
 +051.1352e+05 17.901 < 2e-16 ***
 +051.1003e+05 2.601 0.009299 **
 ## district7.826- +0417.003e+05 11210.911019
 ## district1.717 +0612.474e+05 6.939 4.24e-12 ***
 ## district2.859 +0613.052e+05 9.369 < 2e-16 ***
 ## district9.040 +0512.709e+05 3.337 0.000852 ***
 ## district3.196 +0612.984e+05 15.336 < 2e-16 ***
 ## district2.308 +0611.833e+05 12.586 < 2e-16 ***
 ## district2.880 +051.158e+05 6.925 4.65e-12 ***
 ## district1.817 +0611.453e+05 12.469 < 2e-16 ***
 ## district2.453 +0611.163e+05 5.894 3.90e-09 ***
 +061.879e+05 7.284 3.52e-13 ***
 +061.1471e+05 23.663 < 2e-16 ***
 +0613.050e+05 8.127 4.99e-16 ***
 +062.382e+05 11.964 < 2e-16 ***
 +0611.872e+05 16.658 < 2e-16 ***
 +061.018e+05 11.268 < 2e-16 ***
 +0612.309e+05 8.954 < 2e-16 ***
 +0613.037e+05 4.936 8.13e-07 ***
 +0417.003e+05 -0.140 0.888412
 +061.164e+05 12.142 < 2e-16 ***
 +0612.204e+05 14.084 < 2e-16 ***
 +0613.051e+05 11.794 < 2e-16 ***
 +061.381e+05 12.908 < 2e-16 ***

```

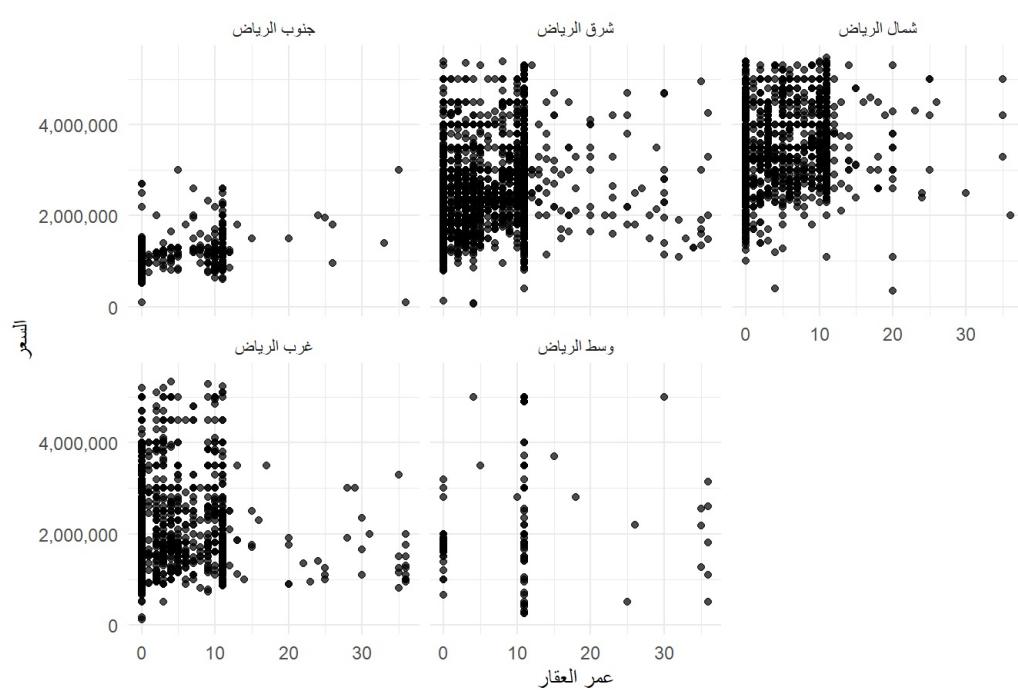
## district2.634               ٢٧٨٠٥ < 2e-16 ***
## district2.545               ٥٨٣٠٥ < 2e-16 ***
## district9.245              ٧٦٣٠٥  5.245 1.60e-07 ***
## district5.773              ٤٠٥٠٥  4.107 4.04e-05 ***
## district8.496               ٣٤٧٠٥  4.864 1.17e-06 ***
## district2.141              ١٣٣١٠٥ 16.084 < 2e-16 ***
## district2.474              ٠٢١٠٥  4.928 8.47e-07 ***
## district1.616               ٣٤٠٥ < 2e-16 ***
## district3.519               ٤١٥١ < 2e-16 ***
## district2.775               ٢٨٦١ < 2e-16 ***
## district2.325               ١٦٦٩ < 2e-16 ***
## district2.862               ٣٩٥٢ < 2e-16 ***
## district2.758               ٣٧٥٣ < 2e-16 ***
## district8.256               ٣٥٥٧ < 2e-16 ***
## district5.619               ٣٥٥٣ < 2e-16 ***
## district2.913               ١٢٥٦ < 2e-16 ***
## district1.742               ١٣٦٦ < 2e-16 ***
## district3.388-               ٣٦٤٣ < 2e-16 ***
## district1.341               ١٥٩٤ < 2e-16 ***
## district1.381               ٣١١٤ < 2e-16 ***
## district4.923               ٩٥٤١ < 2e-16 ***
## district1.595               ١١٩٤ < 2e-16 ***
## district5.999-               ٥٣٤٥ < 2e-16 ***
## district1.860               ١١٣٤ < 2e-16 ***
## district1.267               ٣٧٠٦ < 2e-16 ***
## district3.848               ٣٣٦٦ < 2e-16 ***
## district2.124               ٣٥٥٦ < 2e-16 ***
## district1.327               ٣٥٥٣ < 2e-16 ***
## district5.590               ٣٦٥٣ < 2e-16 ***
## district7.346               ٣٧٥٣ < 2e-16 ***
## district5.173               ٣٣٥٣ < 2e-16 ***
## district3.858               ٣٣٥٣ < 2e-16 ***
## district5.939               ٣٦٥٣ < 2e-16 ***
## district1.177               ٣٣٥٣ < 2e-16 ***
## district7.244               ٣٦٥٣ < 2e-16 ***
## district3.309               ٣٣٥٣ < 2e-16 ***
## district6.669               ٣٦٥٣ < 2e-16 ***
## district2.444               ٣٦٥٣ < 2e-16 ***
## district1.344               ٣٣٥٣ < 2e-16 ***
## district7.422-               ٣٦٥٣ < 2e-16 ***
## district2.106               ٣٦٥٣ < 2e-16 ***
## district2.674               ٣٦٥٣ < 2e-16 ***
## district2.416               ٣٦٥٣ < 2e-16 ***
## district1.526               ٣٦٥٣ < 2e-16 ***
## district6.506-               ٣٦٥٣ < 2e-16 ***
## district6.711               ٣٦٥٣ < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 690700 on 8731 degrees of freedom
## Multiple R-squared:  0.6255, Adjusted R-squared:  0.6199
## F-statistic: 110.5 on 132 and 8731 DF, p-value: < 2.2e-16

```

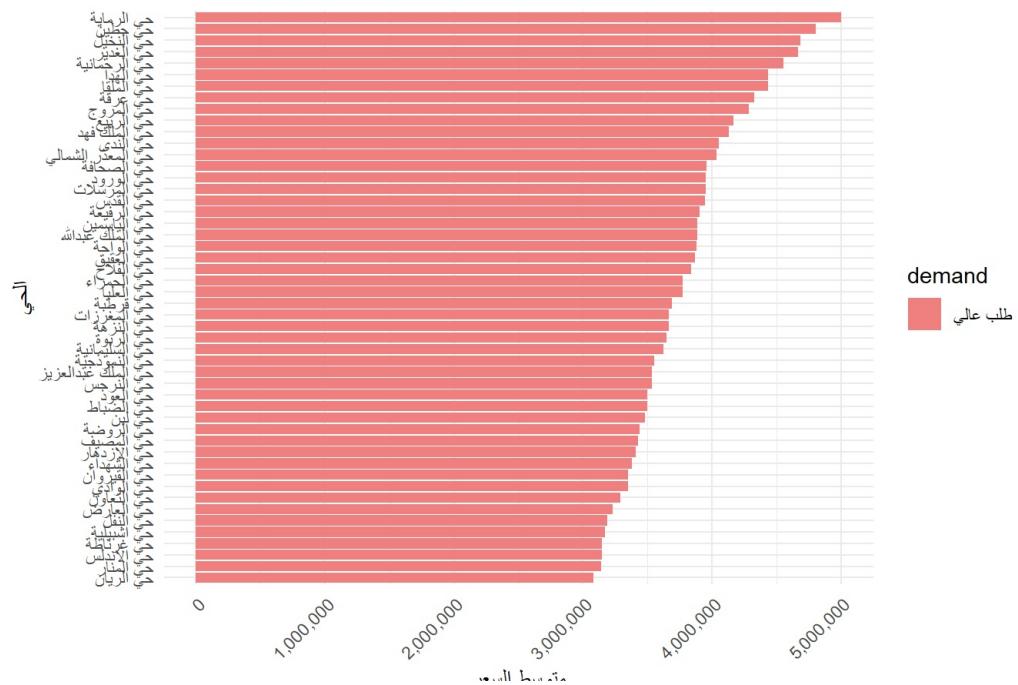
السعر مقابل العمر للفئة: فيلا للبيع



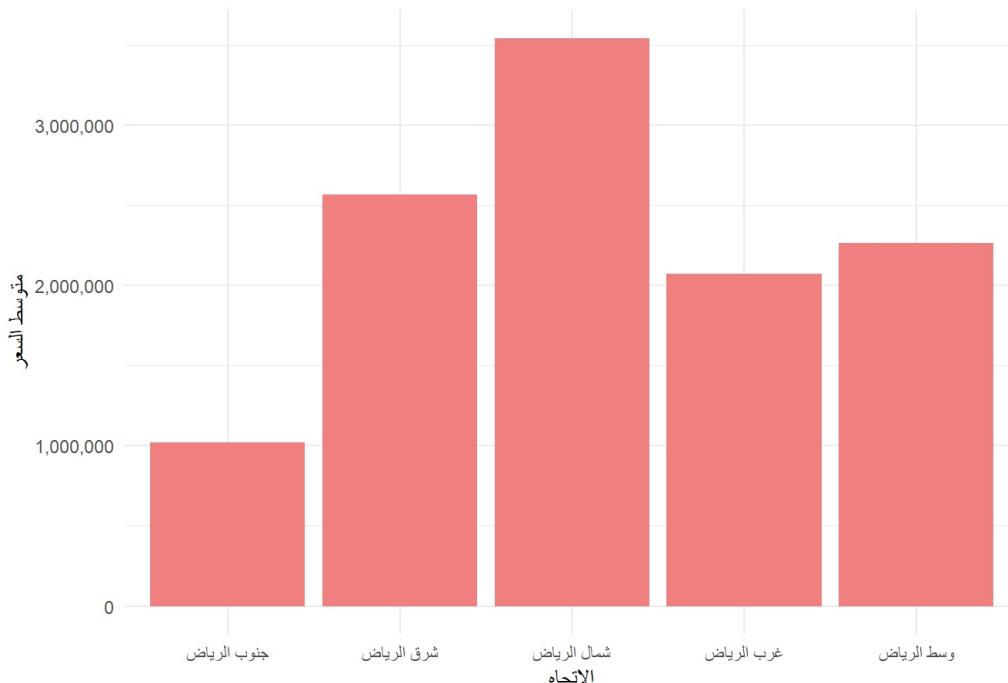
السعر حسب العمر حسب الاتجاه للفئة: فيلا للبيع



متوسط السعر حسب الحي للفئة: فيلا للبيع



متوسط السعر حسب الاتجاه للفئة: فيلا للبيع



```
property_age_analysis(data = city_data2_no_outliers, category = "شقه للبيع")
```

```
## [1] "Summary Statistics:"
## # A tibble: 1,106 × 6
##   demand    area    age mean_price median_price count
##   <dbl>    <dbl> <dbl>      <dbl>       <dbl> <int>
## 1  2      2575000  2575000  3        38     طلب عالي
## 2  1      2160000  2160000  0        43     طلب عالي
## 3  1      1685000  1685000  2        44     طلب عالي
## 4  3      1600000  .2316667  4        44     طلب عالي
## 5  1      1500000  1500000  2        56     طلب عالي
## 6  1      950000   950000   0        58     طلب عالي
## 7  1      900000   900000   1        64     طلب عالي
## 8  2      870000   870000   2        64     طلب عالي
## 9  1      870000   870000   3        64     طلب عالي
## 10 1     1180000  1180000  0        76     طلب عالي
## # i 1,096 more rows
## [1] "Correlation Matrix:"
##            price         age       area district_numeric
## price 1.0000000000  0.027937036  0.0002914597 -0.0760862093
## age   0.0279370357  1.0000000000 -0.0073862886 -0.0746999611
## area  0.0002914597 -0.0073862899  1.0000000000  0.0003264577
## district_numeric -0.0760862093 -0.0746999611  0.0003264577  1.0000000000
## direction_numeric  0.0031573242 -0.1173965420  0.0175172760  0.2526821262
##                  direction_numeric
## price          0.003157324
## age           -0.117396542
## area           0.017517276
## district_numeric  0.252682126
## direction_numeric  1.000000000
## [1] "Linear Regression Model Summary:"
##
## Call:
## lm(formula = price ~ age + area + district, data = category_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1532015 -104913  -18124   70165  4569234 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 5.300e+05  1.697e+05   3.122  0.001809 ** 
## age         -1.338e+04  1.405e+03  -9.523 < 2e-16 ***  
## area          9.365e-02  2.760e-01   0.339  0.734367    
## district4.019  6.051747e+05  2.301  0.021432 *   
## district7.152  9.405186e+05  3.844  0.000123 ***  
## district2.945  8.405196e+05  0.993  0.320779    
## district5.840  8.0512192e+05  2.665  0.007745 **  
## district8.176  8.4051730e+05  4.620  3.98e-06 *** 
```

```

## district1.556   ٤٥٣٠٨٥e+٥  ٠.٧٧٥  ٠.٤٣٨٥٣٨
## district5.318   ٦٩٤١٧٢٩e+٥  ٠.٣٠٨  ٠.٧٥٨٣٦٢
## district5.741   ٦٥٥١٧٣٦e+٥  ٣.٣٠٧  ٠.٠٠٩٥٤ ***  

## district4.094   ٦٥٥٢٠٠٨e+٥  ٢.٠٣٨  ٠.٠٤١٥٨٧ *
## district2.370-  ٦٤٠٤١٧١٨d+٥  -٠.١٣٨  ٠.٨٩٠٣٠٢  

## district2.213   ٦٥٥٢١٩٢e+٥  ١.٠٠٩  ٠.٣١٢٨١٣
## district1.361   ٦٥٥٢٠٧٩e+٥  ٦.٥٤٦  ٦.٧٨e-١١ ***
## district6.698   ٦٥٥٢٩٤٠e+٥  ٢.٢٧٨  ٠.٠٢٢٧٧٧ *
## district9.553   ٦٥٥٢٨٦٠e+٥  ٥.١٣٧  ٢.٩٤e-٠٧ ***
## district8.136   ٦٥٥٢٩٤٢e+٥  ٢.٧٦٦  ٠.٠٠٥٧٠٥ **  

## district2.961   ٦٥٥٢١٣٠٩e+٥  ١.٧٣٢  ٠.٠٨٣٣٠٩ .
## district4.223   ٦٥٣١٨٦٠e+٥  ٢.٢٧٠  ٠.٠٢٣٢٦٢ *
## district5.204   ٦٥٥٢١٨٠٧e+٥  ٢.٨٨٠  ٠.٠٠٤٠٠٦ **
## district6.461   ٦٥٥٢١٩٦٠e+٥  ٣.٢٩٦  ٠.٠٠٠٩٩٢ ***
## district3.678   ٦٥٤٢١٣٥٧e+٥  ٠.٢٠٩  ٠.٨٣٤٢١٩  

## district1.644   ٦٥٥٢١٧٤١e+٥  ٠.٩٤٤  ٠.٣٤٥١٨٤
## district2.899   ٦٥٥٢٢٠٠٩e+٥  ١.٤٤٣  ٠.١٤٩٢٢١
## district3.690   ٦٥٥٢٢٩٤٠e+٥  ١.٢٥٥  ٠.٢٠٩٥١٦  

## district6.385   ٦٥٥٢٣٤٠٠e+٥  ٢.٦٦٠  ٠.٠٠٧٨٥١ **  

## district9.458   ٦٥٥٢١٧٦٤e+٥  ٠.٥٣٦  ٠.٥٩١٩٥٤  

## district2.195   ٦٥٥٢٤٠٠٥٠٥٢٠٥  ٠.٩١٤  ٠.٣٦٠٥٧٠  

## district3.640   ٦٥٥٢١٨٠٧e+٥  ٢.٠١٥  ٠.٠٤٤٠٣٤ *
## district6.667   ٦٥٤١٧٨١٥e+٥  ٠.٣٦٧  ٠.٧١٣٣١٣  

## district6.100   ٦٥٥٢١٧٩٠e+٥  ٣.٤٠٨  ٠.٠٠٠٦٦٢ ***  

## district1.192   ٦٥٦٢١٧٥٠e+٥  ٦.٨١٣  ١.١٣e-١١ ***  

## district1.700   ٦٥٥٢٢٩٤٠e+٥  ٠.٥٧٨  ٠.٥٦٣١٥٤  

## district5.155   ٦٥٥٢٢٠٠٩e+٥  ٢.٥٦٧  ٠.٠١٠٣١٣ *
## district6.413   ٦٥٥٢١٣١٠e+٥  ٣.٧٥٠  ٠.٠٠٠١٨٠ ***  

## district1.550   ٦٥٥٢١٧٩٤e+٥  ٠.٧٤٦  ٠.٤٥٥٩٤٥  

## district3.132-  ٦٥٤١١٧٨٥e+٥  -٠.١٧٥  ٠.٨٦٠٧١١  

## district8.205   ٦٥٥٢١٨٦٢e+٥  ٤.٤٠٨  ١.٠٨e-٠٥ ***
## district2.206   ٦٥٥٢١٣٧٤e+٥  ١٢.٤٣٦ < ٢e-١٦ ***  

## district1.219   ٦٥٥٢١٩٢٠e+٥  ٥.٥٦٢  ٢.٨٨e-٠٨ ***  

## district1.922   ٦٥٥٢٢٤٠٥e+٥  ٠.٧٩٩  ٠.٤٢٤٤٤٦  

## district2.402   ٦٥٥٢١٨٨٠e+٥  ١.٢٧٧  ٠.٢٠١٥٤٥  

## district2.514   ٦٥٥٢١٧٣١e+٥  ١.٤٥٣  ٠.١٤٦٤٣٠  

## district6.246   ٦٥٥٢١٨٤٥e+٥  ٣.٣٨٥  ٠.٠٠٠٧٢٠ ***  

## district8.331   ٦٥٥٢١٧٤٣e+٥  ٤.٧٨٠  ١.٨٢e-٠٦ ***  

## district1.374   ٦٥٦٢١٨٩٨e+٥  ٧.٢٣٨  ٥.٦١e-١٣ ***  

## district6.668   ٦٥٥٢١٨٥٠e+٥  ٣.٦٠٤  ٠.٠٠٠٣١٧ ***  

## district1.144   ٦٥٦٢١٤٠٧e+٥  ٤.٧٥٢  ٢.١٠e-٠٦ ***  

## district2.935-  ٦٥٤١٢٠٧٩٩e+٥  -٠.١٤١  ٠.٨٨٧٧٤٣  

## district4.583   ٦٥٥٢١٩١٠e+٥  ٢.٠٩١  ٠.٠٣٦٥٧٩ *  

## district6.827   ٦٥٥٢١٩٨٠e+٥  ٣.٢٨٣  ٠.٠٠١٠٣٨ **  

## district1.495   ٦٥٦٢٢٤٠٥٥٥٢٠٥  ٦.٢٢٨  ٥.٣٠e-١٠ ***  

## district2.763   ٦٥٥٢١٧٥٩e+٥  ١.٥٧١  ٠.١١٦٣٣٥  

## district8.584   ٦٥٥٢١٤٠٩e+٥  ٣.٥٧٦  ٠.٠٠٠٣٥٤ ***  

## district2.672   ٦٥٥٢١٩٤٤e+٥  ٠.٩٠٧  ٠.٣٦٤٢٣٤  

## district9.162   ٦٥٥٢١٧٠٨e+٥  ٥.٣٦٦  ٨.٦١e-٠٨ ***  

## district4.815   ٦٥٥٢١٨٨١d+٥  ٢.٥٦٠  ٠.٠١٠٥٥٥ *  

## district3.307   ٦٥٥٢١٨٤٦e+٥  ١.٧٩٢  ٠.٠٧٣٢٦٧ .  

## district3.449   ٦٥٥٢١٩٦٠e+٥  ١.٧٥٩  ٠.٠٧٨٦٤٨  

## district3.117   ٦٥٦٢١٩٤٤e+٥  ١٠.٥٨٨ < ٢e-١٦ ***  

## district1.038   ٦٥٥٢١٩٤٣e+٥  ٠.٣٥٣  ٠.٧٢٤٤٤٧  

## district2.798   ٦٥٥٢١٧٠١e+٥  ١.٦٤٥  ٠.١٠٠٠٦٠  

## district5.303   ٦٥٥٢١٧٠٨e+٥  ٣.١٠٥  ٠.٠٠١٩١٩ **  

## district1.198   ٦٥٦٢١٨٤٦e+٥  ٦.٤٩٠  ٩.٧٩e-١١ ***  

## district7.241   ٦٥٥٢١٧٤٢e+٥  ٤.١٥٧  ٣.٣٠e-٠٥ ***  

## district7.329   ٦٥٥٢١٧١٠e+٥  ٤.٢٨٦  ١.٨٧e-٠٥ ***  

## district9.883   ٦٥٥٢١٩٦١e+٥  ٥.٠٤١  ٤.٨٧e-٠٧ ***  

## district1.932   ٦٥٥٢١٧٨٧d+٥  ١.٠٨١  ٠.٢٧٩٦٩٨  

## district7.000   ٦٥٤١٢٠٩٠٨d+٥  ٠.٢٣٨  ٠.٨١١٨٠٦  

## district1.815   ٦٥٥٢١٨٠٧e+٥  ١.٠٠٤  ٠.٣١٥٢٨٨  

## district8.936   ٦٥٥٢١٧٤٠e+٥  ٣.٧٢١  ٠.٠٠٠٢٠٢ ***  

## district2.252   ٦٥٥٢١٧٣٦e+٥  ١.٢٦٨  ٠.٢٠٤٨٧٩  

## district6.626   ٦٥٥٢١٩٢٦e+٥  ٣.٤٤١  ٠.٠٠٠٥٨٧ ***  

## district7.562   ٦٥٥٢١٧٠٩e+٥  ٤.٤٤٢  ١.٠٠e-٠٥ ***  

## district4.907   ٦٥٥٢١٧١٦e+٥  ٢.٨٦٠  ٠.٠٠٤٢٦٥ **  

## district3.285-  ٦٥٤١٢٠٩٤٤e+٥  -٠.١١٢  ٠.٩١١١٥٣  

## district4.460-  ٦٥٥٢١٧٠٩e+٥  -٠.٠٢٦  ٠.٩٧٩١٨٠  

## district1.550   ٦٥٥٢١٩٤٠e+٥  ٠.٥٢٧  ٠.٥٩٨٠٤٣  

## district1.042   ٦٥٥٢١٧٣٥e+٥  ٦.٠٠٦  ٢.١٠e-٠٩ ***  

## district3.187-  ٦٥٤١٢٠٨٠٠e+٥  -٠.١٧٧  ٠.٨٥٩٥٠٤  

## district1.356   ٦٥٥٢١٨٧٧e+٥  ٠.٧٢٢  ٠.٤٧٠٠٩٨  

## district8.134   ٦٥٥٢١٨٩٤٠e+٥  ٢.٧٦٧  ٠.٠٠٥٦٩٤ **  

## district6.930   ٦٥٤١٢٢٤٠٥٥٢٠٥  ٠.٠٠٠  ٠.٩٩٩٩٧٧  

## district6.705   ٦٥٤١٢١٧٠٤e+٥  ٠.٣٩٣  ٠.٦٩٤٠٥١  

## district3.107-  ٦٥٤١٢١٩١٠e+٥  -٠.١٤٢  ٠.٨٨٧٢٤٠  

## district1.510   ٦٥٥٢١٧٣٩٤e+٥  ٠.٧٢٦  ٠.٤٦٧٨٨٠

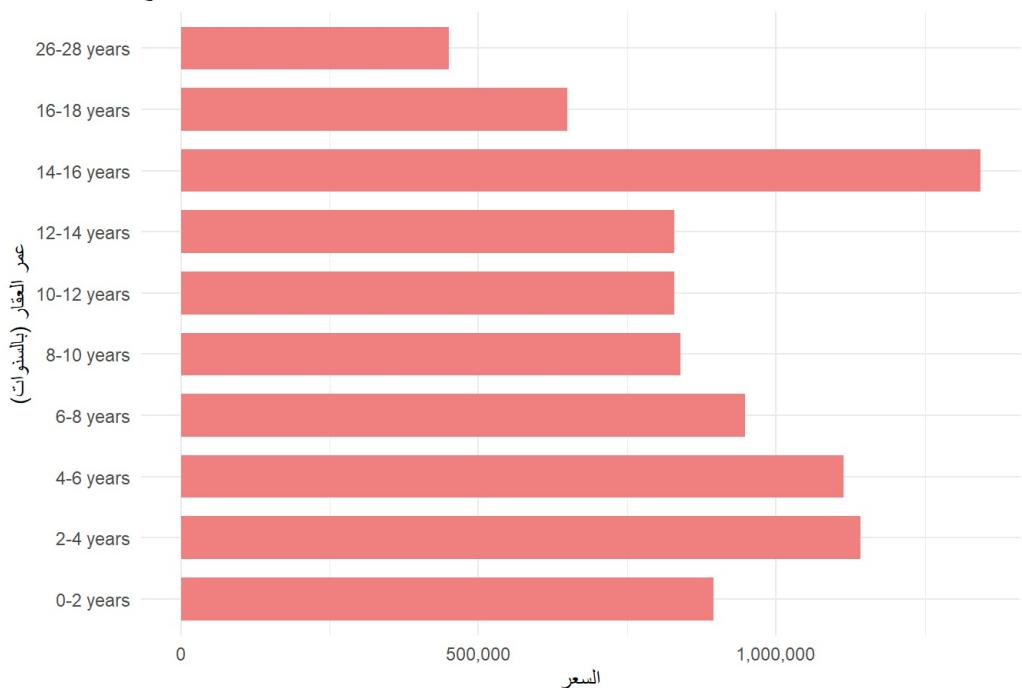
```

```

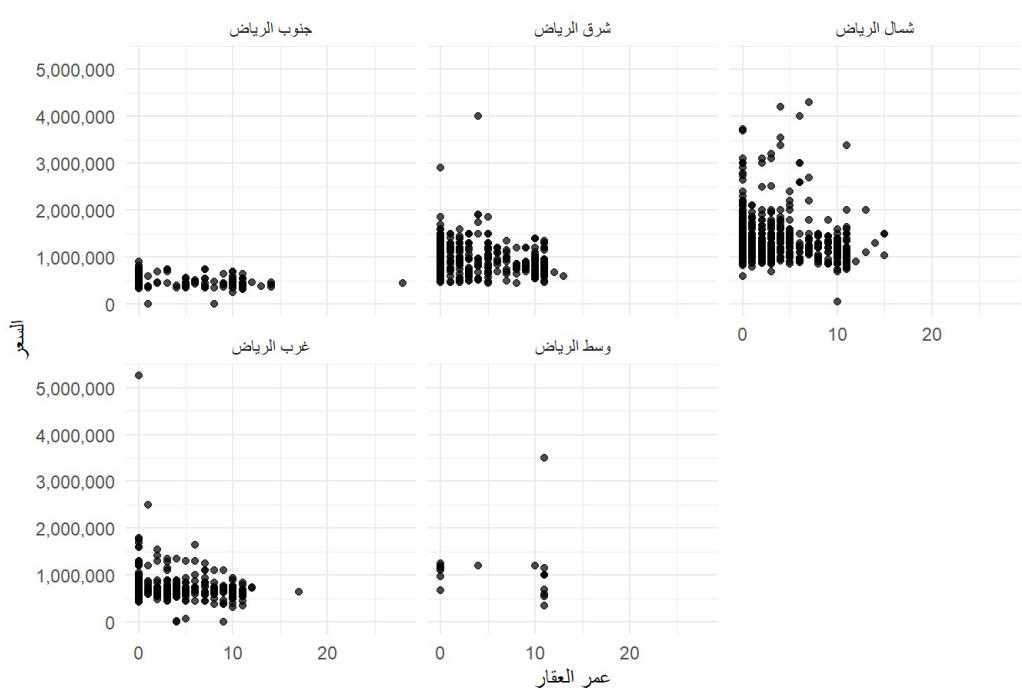
## district2.108      8±05هـ ١٦٧٥e+05    1.236  0.216430
## district1.508      9±05هـ ١٦٩٥e+05    0.882  0.377662
## district7.384      8±05هـ ١٨٠٨e+05    4.085  4.51e-05 ***
## district5.283      8±05هـ ١٧٢٦e+05   -0.031  0.975576
## district5.512      8±05هـ ١٨٤٧e+05    2.985  0.002856 **
## district6.399      8±05هـ ١٧٧٥e+05    3.750  0.000180 ***
## district4.540      8±05هـ ١٨٢٦e+05    2.486  0.012960 *
## district7.001      8±04هـ ١٢٩٤٩e+05   0.1828هـ 0.811804
## district1.273      9±05هـ ١٧٩٤e+05    0.709  0.478193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 240000 on 3428 degrees of freedom
## Multiple R-squared:  0.6707, Adjusted R-squared:  0.6616
## F-statistic: 73.49 on 95 and 3428 DF,  p-value: < 2.2e-16

```

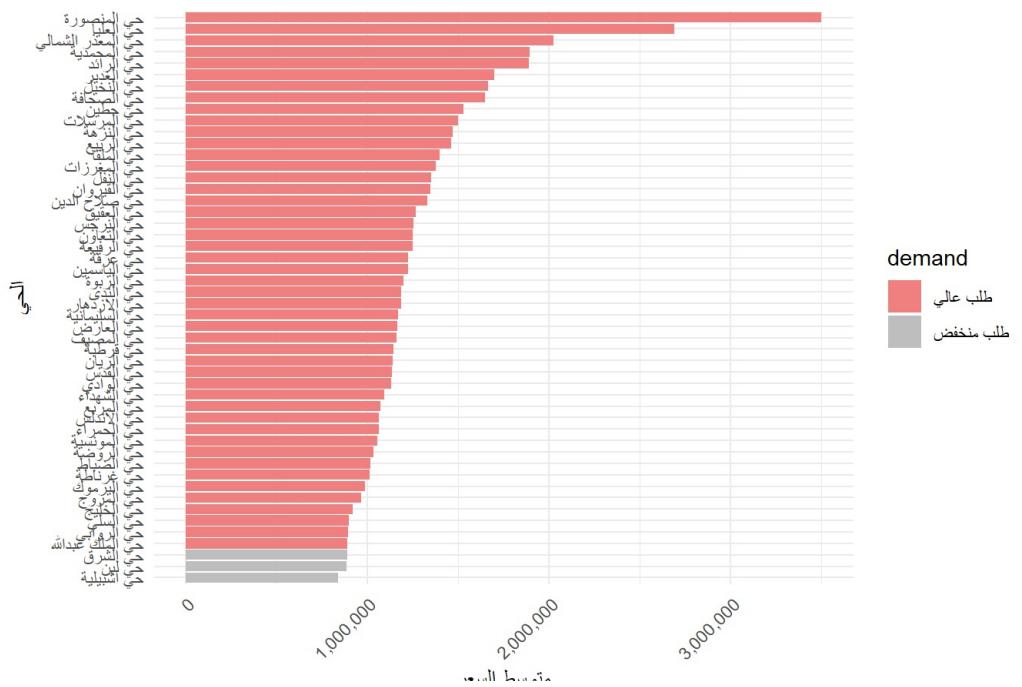
السعر مقابل العمر للفئة: شقه للبيع



السعر حسب العمر حسب الاتجاه للفئة: شقه للبيع



متوسط السعر حسب الحي للفئة: شقه للبيع



متوسط السعر حسب الاتجاه للفئة: شقه للبيع



second part (older properties in high-demand locations retain value better than those in less desirable areas for sale)

```
# Function to analyze and visualize property value retention based on demand
property_demand_analysis <- function(data, category) {
  # Filter and prepare data
  data <- data %>%
    filter(category == !!rlang::enquo(category)) %>%
    select(category, price, age, district, area, direction) %>%
    mutate(
      age_group = case_when(
        age <= 2 ~ "0-2 years",
        age <= 4 ~ "2-4 years",
        age <= 6 ~ "4-6 years",
        age <= 8 ~ "6-8 years",
        age <= 10 ~ "8-10 years",
        age > 10 ~ "10+ years"
      )
    )

  # Remove rows with any missing values
  data <- na.omit(data)

  demand_classification <- data %>%
    group_by(district) %>%
    summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop") %>%
    mutate(demand = ifelse(avg_price >= median(avg_price, na.rm = TRUE), "طلب اعلى القيمة", "طلب اقل القيمة"))

  # Merge demand classification back into the data
  data <- data %>%
    left_join(demand_classification %>% select(district, demand), by = "district")

  # Step 1: Summary statistics
  summary_stats <- data %>%
    group_by(demand, age_group) %>%
    summarize(
      mean_price = mean(price, na.rm = TRUE),
      median_price = median(price, na.rm = TRUE),
      count = n(),
      .groups = "drop"
    )
  print("Summary Statistics:")
  print(summary_stats)

  # Step 2: Correlation Analysis
  correlation_data <- data %>%
    mutate(district_numeric = as.numeric(factor(district))) %>%
    select(price, age, area, district_numeric)
  correlation_matrix <- cor(correlation_data, use = "complete.obs")
  print("Correlation Matrix:")
}
```

```

print(correlation_matrix)

# Step 3: Linear Regression Model
linear_model <- lm(price ~ age + area + district, data = data)
print("Linear Regression Model Summary:")
print(summary(linear_model))

# Visualization

# Plot 2: Bar Plot of Average Price by Age Group and Demand
p2 <- ggplot(data %>% group_by(age_group, demand) %>%
              summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop"),
              aes(x = age_group, y = avg_price, fill = demand)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "متوسط السعر حسب الفئة العمرية والطلب",
       x = "الفئة العمرية",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(text = element_text(family = "IBM"))

print(p2)

# Plot 4: Line Plot of Average Price by Age Group, Faceted by Demand
p4 <- ggplot(data %>% group_by(age_group, demand) %>%
              summarize(avg_price = mean(price, na.rm = TRUE), .groups = "drop"),
              aes(x = age_group, y = avg_price, group = demand, color = demand)) +
  geom_line() +
  geom_point() +
  labs(title = "متوسط السعر حسب الفئة العمرية مفصولة بالطلب",
       x = "الفئة العمرية",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(text = element_text(family = "IBM"))

print(p4)
}

property_demand_analysis(data = city_data2_no_outliers, category = "(فيلا للبيع"

```

```

## [1] "Summary Statistics:"
## # A tibble: 12 × 5
##   demand    age_group  mean_price median_price count
##   <chr>      <chr>        <dbl>        <dbl> <int>
## 1 2-0 طلبيات 3257208.     3200000  1964
## 2 10 طلبيات 3329171.     3200000  1083
## 3 4-2 طلبيات 3286867.     3100000  285
## 4 6-4 طلبيات 3475476.     3370000  233
## 5 8-6 طلبيات 3330166.     3050000  175
## 6 10-8 طلبيات 3461801.     3450000  361
## 7 2-0 years طلبيات 1817597.     1800000  3416
## 8 10 years طلبيات 1698472.     1575000  538
## 9 4-2 years طلبيات 2096415.     1900000  364
## 10 6-4 years طلبيات 2148409.     1900000  165
## 11 8-6 years طلبيات 2231411.     2200000  112
## 12 10-8 years طلبيات 2038732.     1875000  168
## [1] "Correlation Matrix:"
##                price         age        area district_numeric
## price 1.000000000 0.160236869 0.010638064 -0.064578191
## age   0.16023687 1.000000000 -0.005795679 -0.131868406
## area  0.01063806 -0.005795679 1.000000000 -0.009956989
## district_numeric -0.06457819 -0.131868406 -0.009956989 1.000000000
## [1] "Linear Regression Model Summary:"
## 
## Call:
## lm(formula = price ~ age + area + district, data = data)
## 
## Residuals:
##    Min     1Q   Median     3Q    Max 
## -3661881 -440987 -90997  330053  3604326 
## 
## Coefficients:
## (Intercept)            Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            8.782e+05 1.151e+05 7.627 2.65e-14 ***
## age                  1.794e+04 1.789e+03 10.027 < 2e-16 ***
## area                 9.927e-02 1.729e-01  0.574 0.565764  
## district2.126          ٤٦٠٦٣١.٣١٠e+05 16.237 < 2e-16 ***
## district2.341          ٩٤٦٠٦٢.١٣٩e+05 10.993 < 2e-16 ***

```

## district4.217	٦٤٥٧٠٩٣٥e+٥	٠.٦٠٢	٠.٥٤٧٠٤٦
## district2.061	٦١٩٣٥e+٥	١٠.٦٥٢	< 2e-16 ***
## district1.224	٦٣٦٤٥e+٥	٣.٣٥٩	٠.٠٠٠٧٨٥ ***
## district2.173	٦٤٩١٧٠٩٣٥e+٥	٠.٠٣١	٠.٩٧٥٢٥٠
## district2.201	٦٣٠٦٢٣٨٧e+٥	٩.٢١٨	< 2e-16 ***
## district9.615	٦٤٠٤٧١٠٩٣٥e+٥	٠.١٣٧	٠.٨٩٠٨٠٧
## district1.012	٦٣٤٢٥٨٨e+٥	٣.٩١٢	٩.٢١e-05 ***
## district8.893	٦٣٥٣٦٧٥e+٥	٦.٥٠٧	٨.٠٩e-11 ***
## district7.373-	٦٤٠٤١٣٦٤٤e+٥	-0.202	0.839645
## district4.702	٦٣٥٥١٤٤٨٢e+٥	٣.١٧٢	0.001518 **
## district2.690	٦٤٠٦١٥٩٧e+٥	١٦.٨٤٧	< 2e-16 ***
## district3.191	٦٤٠٥١٣٣٠٩e+٥	٠.٩٦٧	٠.٣٣٣٦٣٢
## district1.720	٦٤٠٦١١٣١٣١٣e+٥	١٣.١٠١	< 2e-16 ***
## district1.370	٦٤٠٥١٣٣١٣٥١e+٥	١.٠١٤	٠.٣١٠٥٧٨
## district1.944	٦٤٠٥٢٢٨٥٩e+٥	٠.٦٨٠	٠.٤٩٦٦٠٩
## district2.545	٦٤٠٥١١٣٨٩e+٥	١٤.٢٢٠	< 2e-16 ***
## district3.098	٦٤٠٦٢٠٩٤٠e+٥	١٥.١٩٠	< 2e-16 ***
## district3.565	٦٤٠٦٢٣٣٠٤٠e+٥	١٥.٤٧٢	< 2e-16 ***
## district2.894	٦٤٠٦١٨٢٤e+٥	١٥.٨٦٨	< 2e-16 ***
## district1.147	٦٤٠٦١١٩١e+٥	٩.٦٣١	< 2e-16 ***
## district4.068	٦٤٠٦١٧٠٩٣٥e+٥	٥.٨٠٩	٦.٥١e-09 ***
## district1.693	٦٤٠٦٢٠٤٨٠e+٥	٨.٢٦٦	< 2e-16 ***
## district2.403	٦٤٠٦١١٦٤٥e+٥	١٤.٦٠٦	< 2e-16 ***
## district1.976	٦٤٠٦١٢٣٠٨e+٥	٧.٢٩٩	٣.١٥e-13 ***
## district9.244	٦٤٠٥٤١٥٥٥e+٥	٢.٢٢٥	٠.٠٢٦١١١ *
## district4.600	٦٤٠٥١١٧٣٠٩e+٥	٢.٦٥٩	٠.٠٠٧٨٤٣ **
## district1.252	٦٤٠٥١١٤٩٠٠e+٥	٨.٩٤١	< 2e-16 ***
## district1.772	٦٤٠٥١٢١٣٣٣e+٥	٨.٣٠٨	< 2e-16 ***
## district2.217	٦٤٠٥١٣٠٣٣e+٥	٠.٣١٧	٠.٧٥١٥٢٢
## district2.569	٦٤٠٥١٣٤٧٤٩e+٥	١٠.٣٨٤	< 2e-16 ***
## district2.604	٦٤٠٥١٣٤٥١e+٥	١.٧٩٤	٠.٠٧٢٨٧٨ .
## district6.543	٦٤٠٥١٣٤٥٥٥e+٥	٣.١٣٩	٠.٠٠١٧٥٣ **
## district7.671	٦٤٠٥١٤٦١e+٥	١.٨٤٤	٠.٠٦٥٢٦٣ .
## district1.446	٦٤٠٥١٤٣٨٦e+٥	١٠.٤٣٢	< 2e-16 ***
## district8.667	٦٤٠٥١٤٣٧١e+٥	٠.٦٣٢	٠.٥٢٧٤٥٨
## district2.308	٦٤٠٥١٤٣٨٠e+٥	٨.٩٤٧	< 2e-16 ***
## district8.256-	٦٤٠٥١٤٣٥٥e+٥	-1.١٧٩	٠.٢٣٨٥٨٠
## district2.953	٦٤٠٥١٤٣٦٢e+٥	٢١.٦٨٤	< 2e-16 ***
## district1.273	٦٤٠٥١٤٢٣٥e+٥	٥.٦٩٤	١.٢٨e-08 ***
## district2.424	٦٤٠٥١٤٢٠٥e+٥	٣.٤٦١	٠.٠٠٠٥٤١ ***
## district2.334	٦٤٠٥١٤٢١٨٧e+٥	١٩.٦٥٢	< 2e-16 ***
## district7.397	٦٤٠٥١٤٢٣٤٣e+٥	٢.٤٢٨	٠.٠١٥٢١١ *
## district3.127	٦٤٠٥١٤٢٦٣٥٥e+٥	٣.٢٠١	٠.٠٤٥٤٧٠ *
## district4.076	٦٤٠٥١٤٢٧٢٥٥e+٥	٢.٢٨٧	٠.٠٢٢١٩٣ *
## district2.643	٦٤٠٥١٤٥٦٤e+٥	١.٦٩٠	٠.٠٩١٠٤٩ .
## district2.826	٦٤٠٥١٤٤٤e+٥	١٩.٤٣٦	< 2e-16 ***
## district2.697	٦٤٠٥١٤٣٦٧e+٥	١٥.٢٦٢	< 2e-16 ***
## district5.902	٦٤٠٥١٤٣٥٣e+٥	١.١٧٣	٠.٢٤٠٩٤٩
## district2.532	٦٤٠٥١٤٢٠٣e+٥	٣.٦١٦	٠.٠٠٠٣٠١ ***
## district3.591	٦٤٠٥١٤٢١٤٥e+٥	٨.٦٤٤	< 2e-16 ***
## district1.249-	٦٤٠٥١٤١١٥٣e+٥	-0.٣٠١	٠.٧٦٣٤٤٩
## district1.324	٦٤٠٥١٤١٠٥٥e+٥	١.٨٩١	٠.٥٨٧١١ .
## district8.714	٦٤٠٥١٤١٦٦٥e+٥	٥.٢٣٤	١.٦٩e-07 ***
## district2.796	٦٤٠٥١٤١٨١٠e+٥	١٢.٨٢٣	< 2e-16 ***
## district1.256-	٦٤٠٥١٤١٧٩٥٥e+٥	-0.١٧٩	٠.٨٥٧٦٦١
## district1.531	٦٤٠٥١٤١٦٢٣e+٥	٩.٤٣٣	< 2e-16 ***
## district1.325	٦٤٠٥١٤١٣٣٦e+٥	٩.٩١٩	< 2e-16 ***
## district2.991	٦٤٠٥١٤١٧٠٨٠e+٥	١٧.٥٠٦	< 2e-16 ***
## district2.420	٦٤٠٥١٤١٣٥٣e+٥	١٧.٩٠١	< 2e-16 ***
## district1.822	٦٤٠٥١٤١٠٣٣e+٥	٢.٦٠١	٠.٠٠٩٢٩٩ **
## district7.826-	٦٤٠٥١٤١٠٩٣٤٥١	٥.٩١١٠٩١١٠١٩	
## district1.717	٦٤٠٥١٤١٢٤٧٤٠e+٥	٦.٩٣٩	٤.٢٤e-12 ***
## district2.859	٦٤٠٥١٤١٣٥٢e+٥	٩.٣٦٩	< 2e-16 ***
## district9.040	٦٤٠٥١٤١٢٧٣٩e+٥	٣.٣٣٧	٠.٠٠٠٨٥٢ ***
## district3.196	٦٤٠٥١٤١٢٠٨٤٠e+٥	١٥.٣٣٦	< 2e-16 ***
## district2.308	٦٤٠٥١٤١١٨٣٣e+٥	١٢.٥٨٦	< 2e-16 ***
## district2.880	٦٤٠٥١٤١٥٨٤٠e+٥	٦.٩٢٥	٤.٦٥e-12 ***
## district1.817	٦٤٠٥١٤١٤٥٣e+٥	١٢.٤٦٩	< 2e-16 ***
## district2.453	٦٤٠٥١٤١٤٦٣e+٥	٥.٨٩٤	٣.٩٠e-09 ***
## district1.369	٦٤٠٥١٤١٨٧٩٠e+٥	٧.٢٨٤	٣.٥٢e-13 ***
## district3.482	٦٤٠٥١٤١١٤٧١e+٥	٢٣.٦٦٣	< 2e-16 ***
## district2.479	٦٤٠٥١٤١٣٥٥٤٥٥	٨.١٢٧	٤.٩٩e-16 ***
## district2.852	٦٤٠٥١٤١٢٣٨٣٤٥٥	١١.٩٦٤	< 2e-16 ***
## district3.119	٦٤٠٥١٤١٢٨٧٢٥e+٥	١٦.٦٥٨	< 2e-16 ***
## district1.823	٦٤٠٥١٤١٢٩١٨٥٥	١١.٢٦٨	< 2e-16 ***
## district2.067	٦٤٠٥١٤١٢٣٠٩٠e+٥	٨.٩٥٤	< 2e-16 ***
## district1.632	٦٤٠٥١٤١٣٠٧٣e+٥	٤.٩٣٦	٨.١٣e-07 ***
## district9.826-	٦٤٠٥١٤١٢٠٠٣٩e+٥	-0.١٤٠	٠.٨٨٨٤١٢
## district1.413	٦٤٠٥١٤١٦٤٤e+٥	١٢.١٤٢	< 2e-16 ***

```

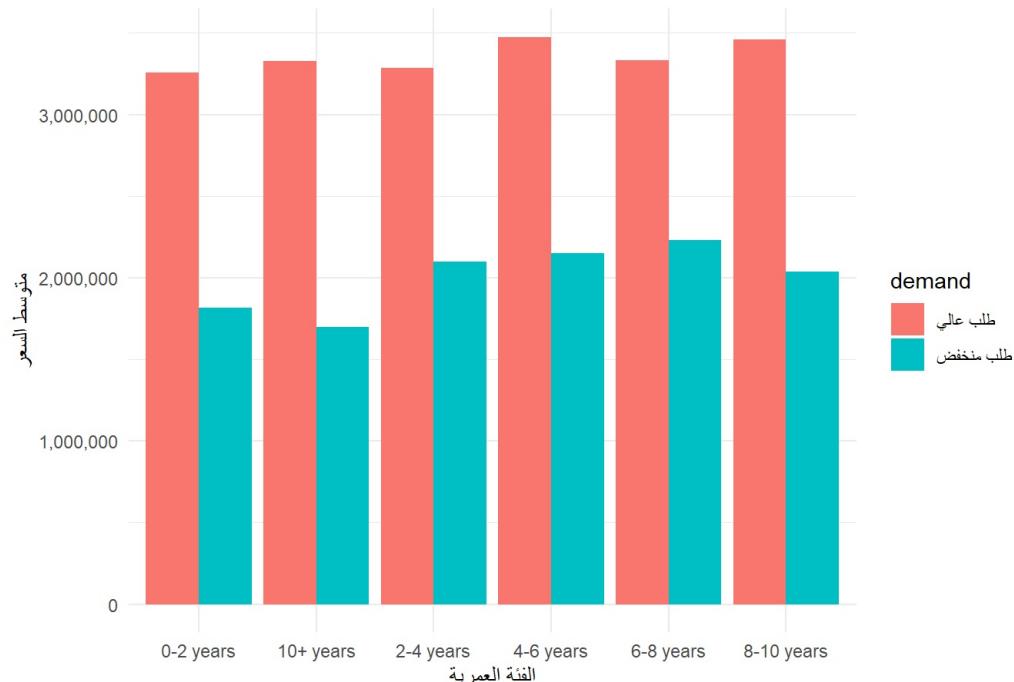
## district1.696               ±0.61204e+05 14.084 < 2e-16 ***
## district3.599               ±0.613051e+05 11.794 < 2e-16 ***
## district3.074               ±0.61381e+05 12.908 < 2e-16 ***
## district2.634               ±0.61178e+05 22.349 < 2e-16 ***
## district2.545               ±0.612583e+05 9.851 < 2e-16 ***
## district9.245               ±0.611763e+05 5.245 1.60e-07 ***
## district5.773               ±0.611405e+05 4.107 4.04e-05 ***
## district8.496               ±0.611347e+05 4.864 1.17e-06 ***
## district2.141               ±0.611331e+05 16.084 < 2e-16 ***
## district2.474               ±0.61021e+05 4.928 8.47e-07 ***
## district1.616               ±0.611240e+05 13.028 < 2e-16 ***
## district3.519               ±0.611151e+05 8.478 < 2e-16 ***
## district2.775               ±0.611286e+05 9.700 < 2e-16 ***
## district2.325               ±0.6111669e+05 13.931 < 2e-16 ***
## district2.862               ±0.6113052e+05 9.380 < 2e-16 ***
## district2.758               ±0.611030e+05 0.392 0.694837
## district8.256               ±0.6117.005e+05 -1.179 0.238577
## district5.619               ±0.6113.645e+05 1.542 0.123216
## district2.913               ±0.6111206e+05 24.144 < 2e-16 ***
## district1.742               ±0.6111.266e+05 13.760 < 2e-16 ***
## district3.388-              ±0.6113.643e+05 -0.930 0.352379
## district1.341               ±0.6114.159d+05 3.224 0.001271 **
## district1.381               ±0.6112.311d+05 5.974 2.40e-09 ***
## district4.923               ±0.6114.194e+05 0.412 0.680187
## district1.595               ±0.6111.741e+05 9.162 < 2e-16 ***
## district5.999-              ±0.6115.034e+05 -1.192 0.233410
## district1.860               ±0.6111.1341 < 2e-16 ***
## district1.267               ±0.6112.706e+05 4.682 2.88e-06 ***
## district3.848               ±0.6112.236e+05 17.213 < 2e-16 ***
## district2.124               ±0.6111.587e+05 1.338 0.180863
## district1.327               ±0.6112.243e+05 0.592 0.554139
## district5.590               ±0.6111.854e+05 3.015 0.002579 **
## district7.346               ±0.6111.471e+05 4.996 5.97e-07 ***
## district5.173               ±0.6111.221e+05 4.235 2.30e-05 ***
## district3.858               ±0.6111.926e+05 0.200 0.841257
## district5.939               ±0.6111.650e+05 3.600 0.000320 ***
## district1.177               ±0.6111.232e+05 9.552 < 2e-16 ***
## district7.244               ±0.6111.838e+05 5.342 9.40e-08 ***
## district3.309               ±0.6111.997e+05 16.566 < 2e-16 ***
## district6.669               ±0.6111.003e+05 0.095 0.924129
## district2.444               ±0.6111.374e+05 1.779 0.075199 .
## district1.344               ±0.6111.301e+05 0.407 0.683994
## district7.422-              ±0.6111.471e+05 -1.780 0.075187 .
## district2.106               ±0.6111.902e+05 11.074 < 2e-16 ***
## district2.674               ±0.6111.318e+05 20.291 < 2e-16 ***
## district2.416               ±0.6111.2579e+05 9.367 < 2e-16 ***
## district1.526               ±0.6111.844e+05 8.1827 < 2e-16 ***
## district6.506-              ±0.6111.023e+05 -1.296 0.195099
## district6.711               ±0.6111.125e+05 3.159 0.001589 **
## ---

## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

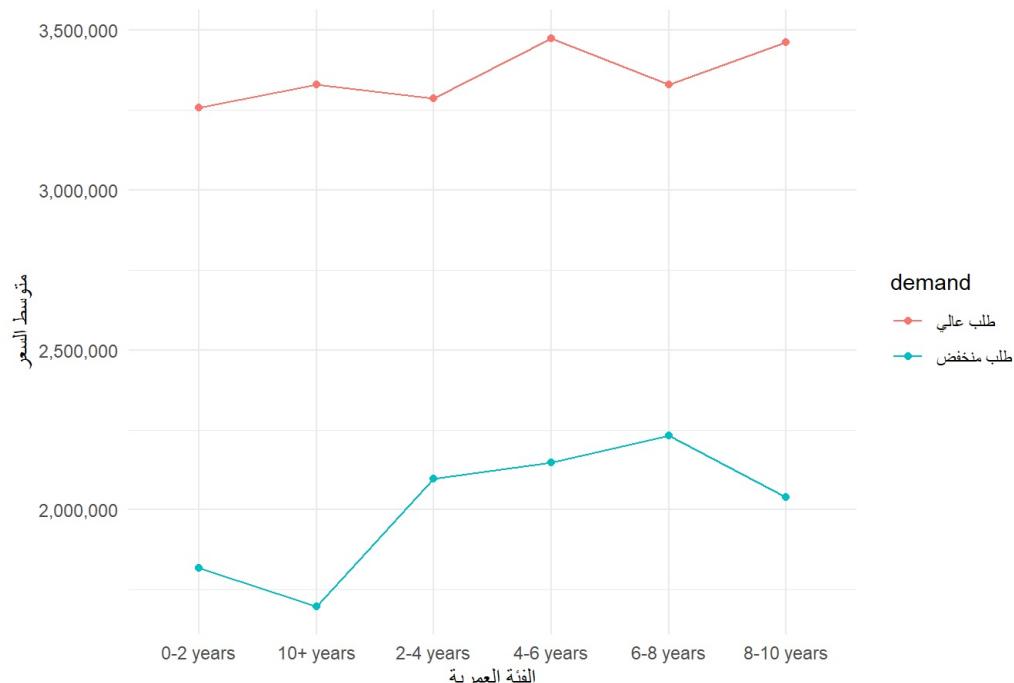
## Residual standard error: 690700 on 8731 degrees of freedom
## Multiple R-squared: 0.6255, Adjusted R-squared: 0.6199
## F-statistic: 110.5 on 132 and 8731 DF, p-value: < 2.2e-16

```

متوسط السعر حسب الفئة العمرية والطلب



متوسط السعر حسب الفئة العمرية مفصولة بالطلب



```
property_demand_analysis(data = city_data2_no_outliers, category = "دور للبيع")
```

```
## [1] "Summary Statistics:"
## # A tibble: 12 × 5
##   demand    age_group  mean_price median_price count
##   <chr>      <chr>        <dbl>       <dbl>   <int>
## 1 طلب عالي  1-2 years  1335880.    1250000     707
## 2 طلب عالي  2-4 years  1494348.    1350000     23
## 3 طلب عالي  4-6 years  1438947.    1500000     19
## 4 طلب عالي  6-8 years  1639818.    1375000     22
## 5 طلب عالي  8-10 years 1259091.    1100000     11
## 6 طلب عالي  10+ years  1470000.    1100000      5
## 7 طلب عالي  2-4 years  734900.     720000    1767
## 8 طلب عالي  4-6 years  797500.     725000     24
## 9 طلب عالي  6-8 years  976667.     1050000     6
## 10 طلب عالي  8-10 years 1000000.    1150000     6
## 11 طلب عالي  10+ years 624375.     540000    16
## 12 طلب عالي  2-4 years 478508.     500000    13
## [1] "Correlation Matrix:"
##          price      age      area district_numeric
## price  1.0000000  0.11281499  0.258295444 -0.199210833
## age    0.1128150  1.00000000  0.066267434 -0.115397515
## area   0.2582954  0.06626743  1.000000000  0.002375823
## district_numeric -0.1992108 -0.11539751  0.002375823  1.000000000
```

```

## [1] "Linear Regression Model Summary:"
##
## Call:
## lm(formula = price ~ age + area + district, data = data)
##
## Residuals:
##    Min      1Q  Median      3Q     Max
## -1284342   -77553   -14348    64868   1834315
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)               703162.97   38447.94 18.289 < 2e-16 ***
## age                      -1356.48    2219.89 -0.611 0.541217
## area                     170.56     26.86  6.350 2.55e-10 ***
## district2.01 5.213    83220.38   433808.32   حي الشيشان
## district3.12 4.673    103782.82   484992.24   حي العلا
## district0.346783 0.941   141453.17   133110.56   حي البرية
## district0.001240 3.233    196707.59   635972.88   حي التعاون
## district1.77 5.652    65879.03    372317.66   حي الخطا
## district0.563765 0.577-  42666.31    24632.69-   حي الحرم
## district1.03 6.128    118655.86   727172.54   حي الحادمة
## district0.605181 0.517-  71754.67   37099.25-   حي الولفي
## district8.59 5.776    69336.44    400480.61   حي الدار البيضاء
## district0.001427 3.193-  44740.80   142837.05-   حي الدريهمية
## district0.942363 0.072    141473.92   10229.64   حي الظهرة
## district2 > 20.058   65501.76   1313852.69   حي السلام
## district5.48 6.229    196481.15   1223837.25   حي الشاهزاد
## district2 > 9.372    39637.41    371480.96   حي العزيزية
## district2.73 5.968    104166.07   621693.11   حي العزيز
## district4.52 4.086    103535.83   423050.34   حي العزيز
## district0.080257 1.750   71776.31   125601.24   حي الزهرة
## district2.62 5.975   45157.34   269813.85   حي السليماني
## district0.668846 0.428-  196441.76   84034.37-   حي السوادي
## district2 > 18.623   198266.00   3692366.15   حي السوادي الغربي
## district0.453249 0.750-  52533.35   39406.61-   حي العزيز
## district0.855284 0.182-  63918.23   11658.63-   حي العزيز
## district1.42 4.350   50562.38   219935.91   حي الشفا
## district0.256708 1.134-  53463.06   60651.79-   حي الشفا
## district2.02 7.694   117649.99   905252.57   حي المكان
## district3.46 7.624   117613.01   896640.99   حي المكان
## district0.029119 2.183   94779.23   206913.55   حي العريجاء الغربية
## district2 > 16.814   42774.48   719199.99   حي العريجاء الوسطى
## district0.360622 0.914   103539.35   94670.67   حي العريجاء
## district0.039717 2.058-  117581.59   241952.51-   حي العزيزية
## district0.341474 0.951-  41326.32   39319.29-   حي العزيز
## district2.79 8.237   82136.37   676569.38   حي المصانع
## district0.177019 1.350-  117831.06   159115.05-   حي الغنامية
## district0.008590 2.630   196442.97   516647.87   حي الفازوق
## district6.33 8.136   46030.89   374500.08   حي العزيز
## district1.16 6.472   141526.88   915949.63   حي العزيز
## district2 > 17.699   87805.10   1554059.88   حي العزيز
## district2 > 8.433   141458.94   1192891.55   حي العزيز
## district0.318878 0.997-  117554.38   117197.14-   حي الملك عبد العزيز
## district2 > 9.864   74721.84   737022.07   حي المنوار
## district3.10 5.553   58226.90   323317.08   حي المكان
## district2 > 9.531   117569.39   1120570.76   حي المكان
## district0.498863 0.676   196453.82   132876.95   حي العزيز
## district2.21 7.979   87570.31   698734.64   حي العزيز
## district0.053396 1.933   201897.71   390187.71   حي العزيز
## district2 > 28.587   64190.08   1834996.45   حي العزيز
## district2 > 13.057   42947.35   560771.27   حي العزيز
## district0.007776 2.664   219315.06   584202.50   حي العزيز
## district2 > 20.694   44357.54   917914.03   حي العزيز
## district1.96 4.277   197178.29   843332.82   حي العزيز
## district0.716435 0.363   141445.83   51382.73   حي النسيم الشرقي
## district0.246258 1.160   87682.65   101690.20   حي النسيم الغربي
## district0.009427 2.598   117815.63   306100.65   حي الناظم
## district6.31 6.905   196431.05   1356414.26   حي العزيز
## district2.30 7.370   63051.26   464686.74   حي العزيز
## district8.31 7.193   141738.43   1019500.89   حي العزيز
## district2 > 13.906   78040.04   1085227.68   حي العزيز
## district8.06 8.106   53510.10   433739.76   حي العزيز
## district0.524029 0.637   196476.37   125201.73   حي ام الحمام الشرقي
## district7.79 3.957-  38837.83   153690.27-   حي العزيز
## district2.82 4.195   145115.03   608801.21   حي العزيز
## district0.001876 3.113   62536.60   194646.90   حي العزيز
## district0.438551 0.775-  61299.30   47492.47-   حي سلطنة
## district0.000228 3.691   87382.20   322537.99   حي العزيز
## district1.78 5.650   47862.98   270421.39   حي العزيز

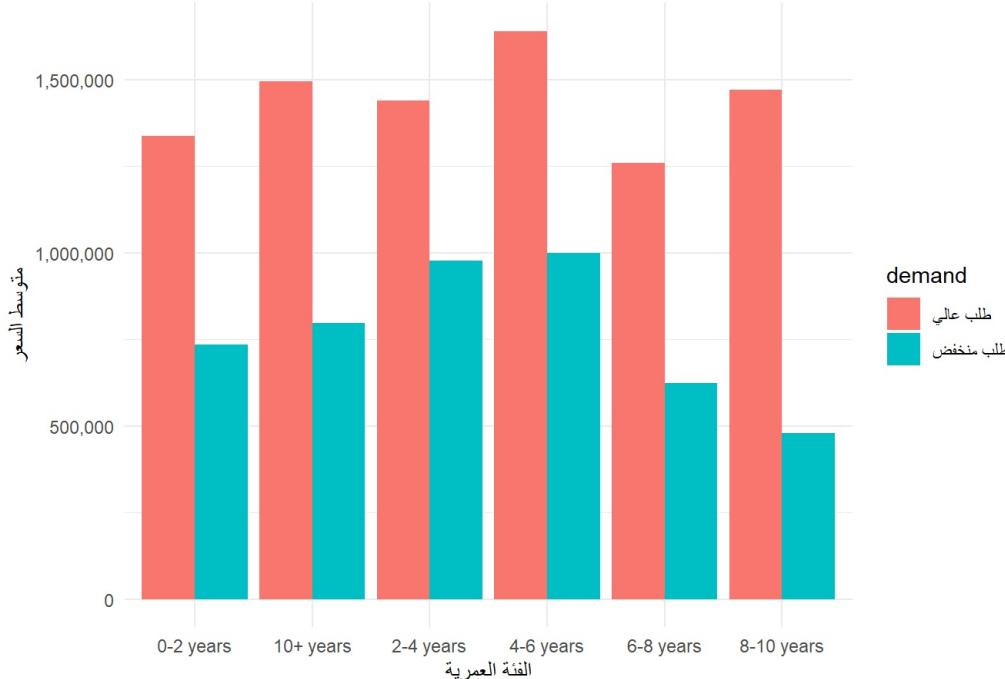
```

```

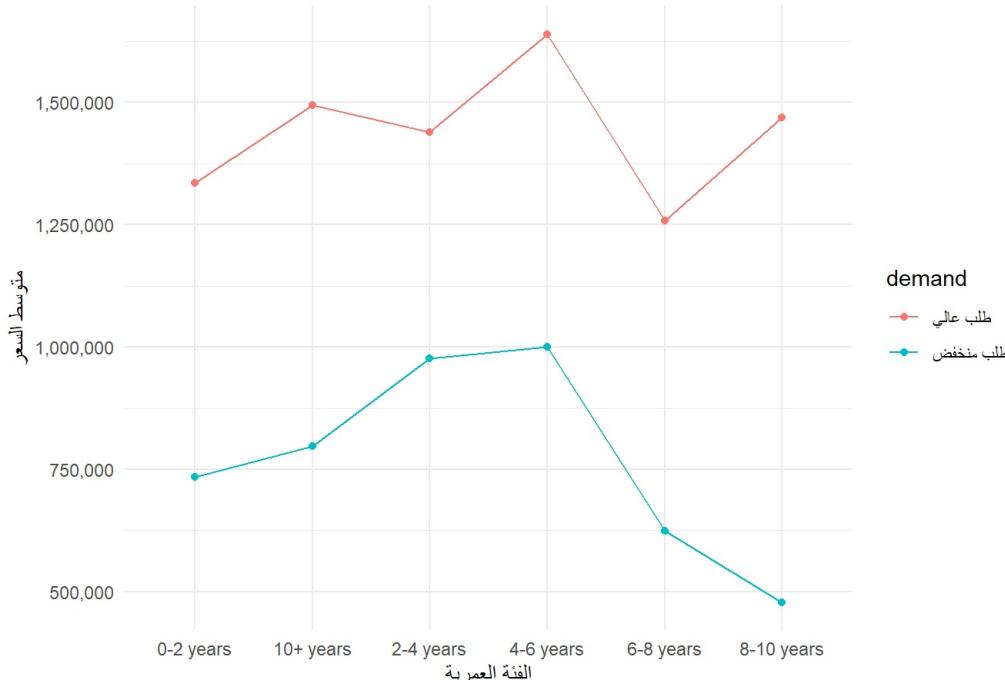
## district0.036037 2.098   38938.30   81677.93   حي طويق
## district0.127779 1.523-  87304.15   133000.62-  حي طيبة
## district0.252484 1.145   69431.69   79471.25   حي طهرة البديةعه
## district3.21 5.124    142377.02   729580.29   حي ظهرة7لبيع
## district0.040114 2.054   42870.00   88038.86   حي ظهرة نمار
## district5.99 7.551    143614.91   1084409.25   **414**فقط
## district0.859736 0.177-  196440.62   34716.61-  حي عريض
## district0.033736 2.124-  40996.08   87090.52-  حي عكاظ
## district6.83 8.126    94255.40   765953.20   حي*غرافط
## district2 > 15.392   53123.21   817695.64   حي*قطابة
## district0.005588 2.773   117539.39   325980.45   حي مطار الملك خالد الدُّوَّلي
## district0.812657 0.237   61291.32    14527.55   حي نمار
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 192800 on 2539 degrees of freedom
## Multiple R-squared:  0.7759, Adjusted R-squared:  0.769
## F-statistic: 111.3 on 79 and 2539 DF,  p-value: < 2.2e-16

```

متوسط السعر حسب الفئة العمرية والطلب



متوسط السعر حسب الفئة العمرية مفصولة بالطلب



```
property_demand_analysis(data = city_data2_no_outliers, category = "شقة للبيع")
```

```

## [1] "Summary Statistics:"
## # A tibble: 12 × 5

```

```

##    demand    age_group  mean_price median_price count
##    <chr>      <chr>        <dbl>       <dbl> <int>
## 1 2-0 طلبات<br>عالي 1231293.    1190000  985
## 2 10 طلبات<br>عالي 1081922.    1000000  107
## 3 4-2 طلبات<br>عالي 1395806.    1300000  155
## 4 6-4 طلبات<br>عالي 1346383.    1250000  115
## 5 8-6 طلبات<br>عالي 1214316.    1150000  57
## 6 10-8 طلبات<br>عالي 1092330.    1100000  94
## 7 2-0 طلبات<br>متوسط 691068.     690000  1622
## 8 10 طلبات<br>متوسط 541363.     550000  91
## 9 4-2 طلبات<br>متوسط 709978.     650000  91
## 10 6-4 طلبات<br>متوسط 696230.     650000  64
## 11 8-6 طلبات<br>متوسط 643143.     650000  50
## 12 10-8 طلبات<br>متوسط 583480.     600000  93
## [1] "Correlation Matrix:"
##          price         age       area district_numeric
## price 1.0000000000 0.027937036 0.0002914597 -0.0760862093
## age   0.0279370357 1.000000000 -0.0073862886 -0.0746999611
## area  0.0002914597 -0.007386289 1.0000000000 0.0003264577
## district_numeric -0.0760862093 -0.074699961 0.0003264577 1.0000000000
## [1] "Linear Regression Model Summary:"
##
## Call:
## lm(formula = price ~ age + area + district, data = data)
##
## Residuals:
##    Min      1Q  Median      3Q      Max 
## -1532015 -104913  -18124   70165  4569234
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 5.300e+05 1.697e+05 3.122 0.001809 **  
## age         -1.338e+04 1.405e+03 -9.523 < 2e-16 *** 
## area         9.365e-02 2.760e-01  0.339 0.734367    
## district4.019  e+05 شعبه+05 2.301 0.021432 *   
## district7.152  e+05 رياض+05 3.844 0.000123 *** 
## district2.945  e+05 عياده+05 0.993 0.320779    
## district5.840  e+05 عياده+05 2.665 0.007745 **  
## district8.176  e+05 تلبيه+05 4.620 3.98e-06 *** 
## district1.556  e+05 حي+05 0.775 0.438538    
## district5.318  e+04 ابراهيم+05 0.308 0.758362    
## district5.741  e+05 حي+05 3.307 0.000954 *** 
## district4.094  e+05 ابراهيم+05 2.038 0.041587 *  
## district2.370- e+04 حي+05 -0.138 0.890302    
## district2.213  e+05 حي+05 1.009 0.312813    
## district1.361  e+05 عياده+05 6.546 6.78e-11 *** 
## district6.698  e+05 عياده+05 2.278 0.022777 *  
## district9.553  e+05 عياده+05 5.137 2.94e-07 *** 
## district8.136  e+05 عياده+05 2.766 0.005705 **  
## district2.961  e+05 عياده+05 1.732 0.083309 .  
## district4.223  e+05 عياده+05 2.270 0.023262 *  
## district5.204  e+05 عياده+05 2.880 0.004006 **  
## district6.461  e+05 عياده+05 3.296 0.000992 *** 
## district3.678  e+04 عياده+05 0.209 0.834219    
## district1.644  e+05 عياده+05 0.944 0.345184    
## district2.899  e+05 عياده+05 1.443 0.149221    
## district3.690  e+05 عياده+05 1.255 0.209516    
## district6.385  e+05 عياده+05 2.660 0.007851 **  
## district9.458  e+04 عياده+05 0.536 0.591954    
## district2.195  e+05 حي+05 0.914 0.360570    
## district3.640  e+05 حي+05 2.015 0.044034 *  
## district6.667  e+04 حي+05 0.367 0.713313    
## district6.100  e+05 حي+05 3.408 0.000662 *** 
## district1.192  e+05 حي+05 6.813 1.13e-11 *** 
## district1.700  e+05 حي+05 0.578 0.563154    
## district5.155  e+05 حي+05 2.567 0.010313 *  
## district6.413  e+05 حي+05 3.750 0.000180 *** 
## district1.550  e+05 حي+05 0.746 0.455945    
## district3.132- e+04 حي+05 -0.175 0.860711    
## district8.205  e+05 حي+05 4.408 1.08e-05 *** 
## district2.206  e+05 حي+05 12.436 < 2e-16 *** 
## district1.219  e+05 حي+05 5.562 2.88e-08 *** 
## district1.922  e+05 حي+05 0.799 0.424446    
## district2.402  e+05 حي+05 1.277 0.201545    
## district2.514  e+05 حي+05 1.453 0.146430    
## district6.246  e+05 حي+05 3.385 0.000720 *** 
## district8.331  e+05 حي+05 4.780 1.82e-06 *** 
## district1.374  e+06 حي+05 7.238 5.61e-13 *** 
## district6.668  e+05 حي+05 3.604 0.000317 ***

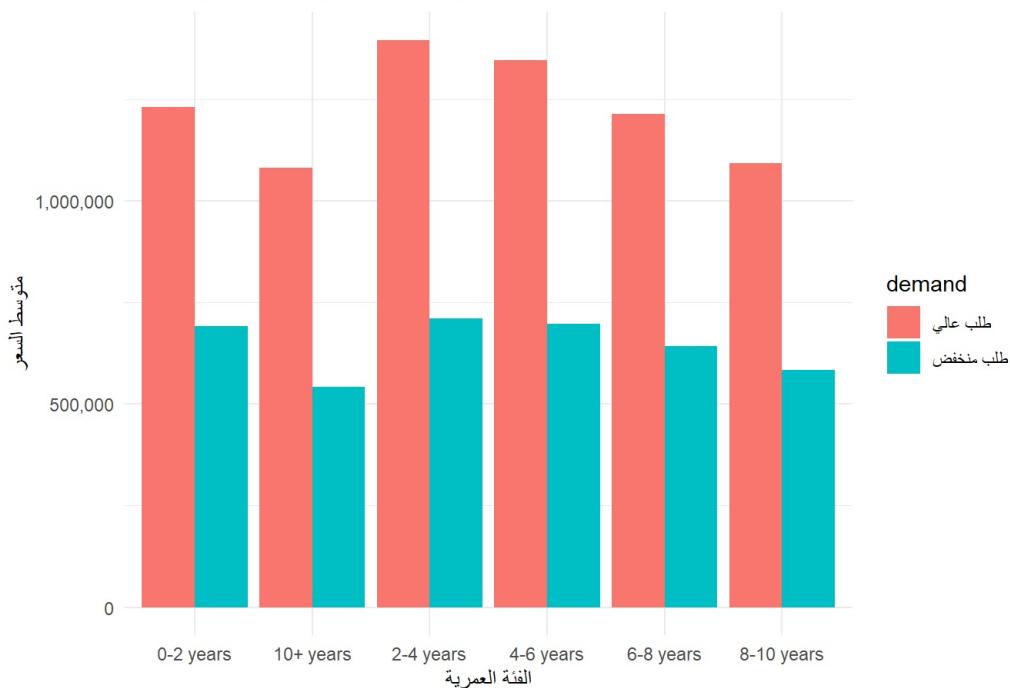
```

```

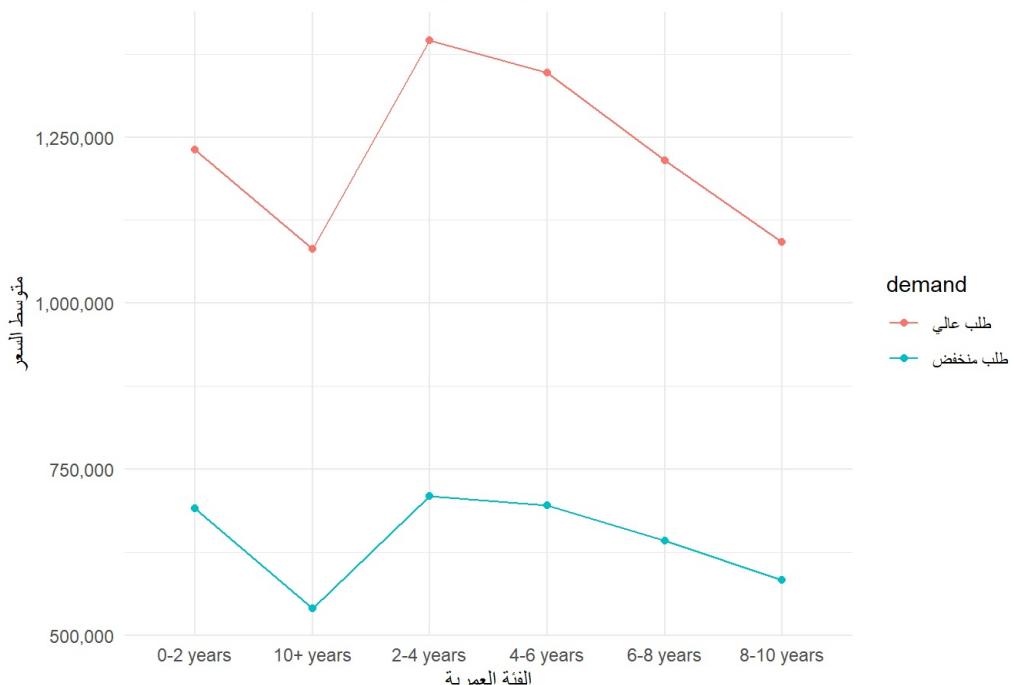
## district1.144      ٦٥٠٦٢.٤٩٧e+05  4.752 2.10e-06 ***
## district2.935      ٦٣٤١٢.٥٧٩e+05 -0.141 0.887743
## district4.583      ٦٣٥١٢.١٩١e+05  2.091 0.036579 *
## district6.827      ٦٣٥١٢.٥٨٠e+05  3.283 0.001038 **
## district1.495      ٦٣٦٢.٤٠٠e+05  6.228 5.30e-10 ***
## district2.763      ٦٣٧١١٧٥٩e+05  1.571 0.116335
## district8.584      ٦٣٩١٤٠٠e+05  3.576 0.000354 ***
## district2.672      ٦٣٩١٢.٩٤٤e+05  0.907 0.364234
## district9.162      ٦٣٩١٣٧٠٨e+05  5.366 8.61e-08 ***
## district4.815      ٦٤٥١٨٨١e+05  2.560 0.010505 *
## district3.307      ٦٤٥١٨٤٦e+05  1.792 0.073267 .
## district3.449      ٦٤٦١٩٦٠e+05  1.759 0.078648 .
## district3.117      ٦٤٦١٩٤٤e+05  10.588 < 2e-16 ***
## district1.038      ٦٤٥٢٩٤٣e+05  0.353 0.724447
## district2.798      ٦٤٥١٧٠١e+05  1.645 0.100060
## district5.303      ٦٤٥١٧٠٨e+05  3.105 0.001919 **
## district1.198      ٦٤٦١٨٤٦e+05  6.490 9.79e-11 ***
## district7.241      ٦٤٧١٧٤٢e+05  4.157 3.30e-05 ***
## district7.329      ٦٤٥١٧١٠e+05  4.286 1.87e-05 ***
## district9.883      ٦٤٥١٩٦١e+05  5.041 4.87e-07 ***
## district1.932      ٦٤٥١٧٨٧٤e+05  1.081 0.279698
## district7.000      ٦٤٤١٩٤٠٤e+05  0.238 0.811806
## district1.815      ٦٤٥١٨٠٧e+05  1.004 0.315288
## district8.936      ٦٤٥١٤٠٢e+05  3.721 0.000202 ***
## district2.252      ٦٤٥١٣٧٦e+05  1.268 0.204879
## district6.626      ٦٤٥١٩٢٦e+05  3.441 0.000587 ***
## district7.562      ٦٤٥١١٧٩٩e+05  4.424 1.00e-05 ***
## district4.907      ٦٤٥١١.٣١٦e+05  2.860 0.004265 **
## district3.285-     ٦٤٤٢.٩٤٤e+05 -0.112 0.911153
## district4.460-     ٦٤٣١.٧٠٩e+05 -0.026 0.979180
## district1.550      ٦٤٥٢.٩٤٠e+05  0.527 0.598043
## district1.042      ٦٤٦١.٧٣٥e+05  6.006 2.10e-09 ***
## district3.187-     ٦٤٩١٨٠٠e+05 -0.177 0.859504
## district1.356      ٦٤٥١.٨٧٧e+05  0.722 0.470098
## district8.134      ٦٤٥١٦٨٤٩e+05  2.767 0.005694 **
## district6.930      ٦٤٦٢٤٠٠e+05  0.000 0.999977
## district6.705      ٦٤٦١.٧٠٤e+05  0.393 0.694051
## district3.107-     ٦٤٦١.١٩١e+05 -0.142 0.887240
## district1.510      ٦٤٥٢.٠٧٩e+05  0.726 0.467880
## district2.108      ٦٤٥١٦٧٥e+05  1.236 0.216430
## district1.508      ٦٤٥١٤٠٩e+05  0.882 0.377662
## district7.384      ٦٤٥١.٨٠٨e+05  4.085 4.51e-05 ***
## district5.283-     ٦٤٦١.٧٢٦e+05 -0.031 0.975576
## district5.512      ٦٤٥١٨٤٧e+05  2.985 0.002856 **
## district6.399      ٦٤٥١٧٠٦e+05  3.750 0.000180 ***
## district4.540      ٦٤٥١.٨٢٦e+05  2.486 0.012960 *
## district7.001      ٦٤٤١٢.٩٤٠e+05  ٦١٢.٨١٨٠٤
## district1.273      ٦٤٥١.٧٩٤e+05  0.709 0.478193
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 240000 on 3428 degrees of freedom
## Multiple R-squared:  0.6707, Adjusted R-squared:  0.6616
## F-statistic: 73.49 on 95 and 3428 DF,  p-value: < 2.2e-16

```

متوسط السعر حسب الفئة العمرية والطلب



متوسط السعر حسب الفئة العمرية مفصولة بالطلب



```
# Filter data for Riyadh city
dataset <- read_csv("akar2024.csv", locale = locale(encoding = "UTF-8"))
data_riyadh <- dataset %>%
  filter(city == "الرياض")

# Function to filter by category and rent period
filter_data <- function(data, category_value, rent_period_value) {
  data %>%
    filter(category == category_value, rent_period == rent_period_value) # %>%
    #filter(is.na(price) | is.na(age) | is.na(district))
}

# Filter data for category 5 and rent period 3 means renting villa for a year
filtered_data5 <- filter_data(data_riyadh, 5, 3)
# Filter data for category 1 and rent period 3 means renting apartment for a year
filtered_data1 <- filter_data(data_riyadh, 1, 3)
```

```

# Remove rows with missing values in 'age' or 'price'
filtered_data5_clean <- filtered_data5 %>%
  filter(!is.na(age), !is.na(price))

filtered_data1_clean <- filtered_data1 %>%
  filter(!is.na(age), !is.na(price))

# Check for extreme outliers in 'age' and 'price'
# We'll use the IQR method to detect outliers for both 'age' and 'price'

# IQR function to detect outliers
IQR_outlier_removal <- function(data, column) {
  Q1 <- quantile(data[[column]], 0.25)
  Q3 <- quantile(data[[column]], 0.75)
  IQR_value <- Q3 - Q1
  lower_bound <- Q1 - 1.5 * IQR_value
  upper_bound <- Q3 + 1.5 * IQR_value
  data %>%
    filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
}

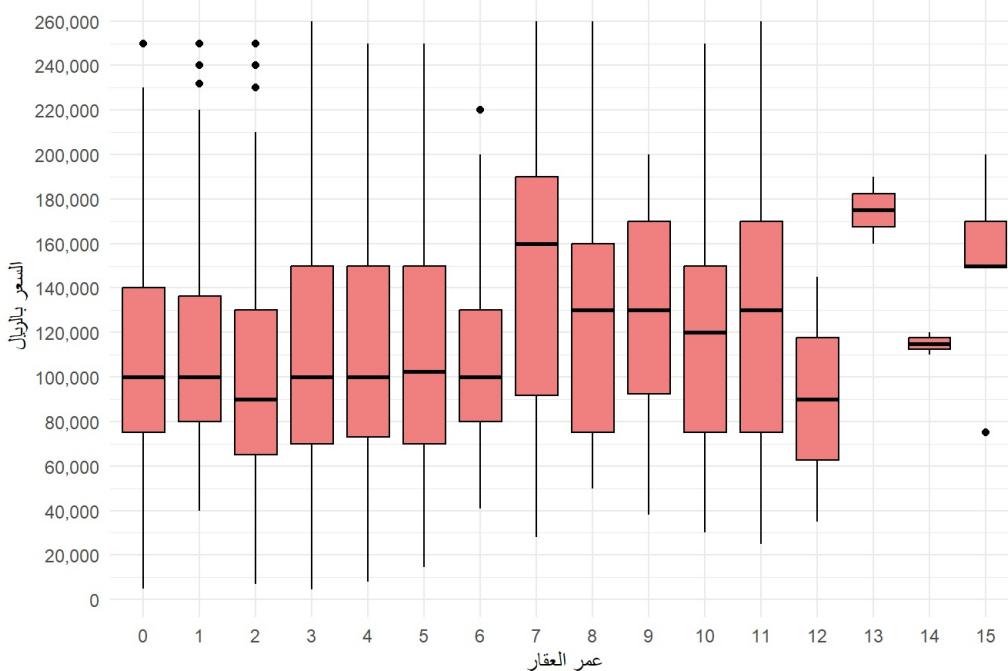
# Remove outliers from 'age' and 'price' using IQR
filtered_data5_clean <- filtered_data5_clean %>%
  IQR_outlier_removal("age") %>%
  IQR_outlier_removal("price")

filtered_data1_clean <- filtered_data1_clean %>%
  IQR_outlier_removal("age") %>%
  IQR_outlier_removal("price")

# Create a boxplot for price by age
ggplot(filtered_data5_clean, aes(x = factor(age), y = price)) +
  geom_boxplot(fill = "lightcoral", color = "black") +
  labs(title = "توزيع الاسعار حسب عمر العقار (فيلا للإيجار)",
       x = "عمر العقار",
       y = "السعر بالریال" +
  theme_minimal() +
  scale_y_continuous(labels = label_comma(), breaks = seq(0, max(filtered_data5_clean$price, na.rm = TRUE), by = 20000))

```

توزيع الاسعار حسب عمر العقار (فيلا للإيجار)

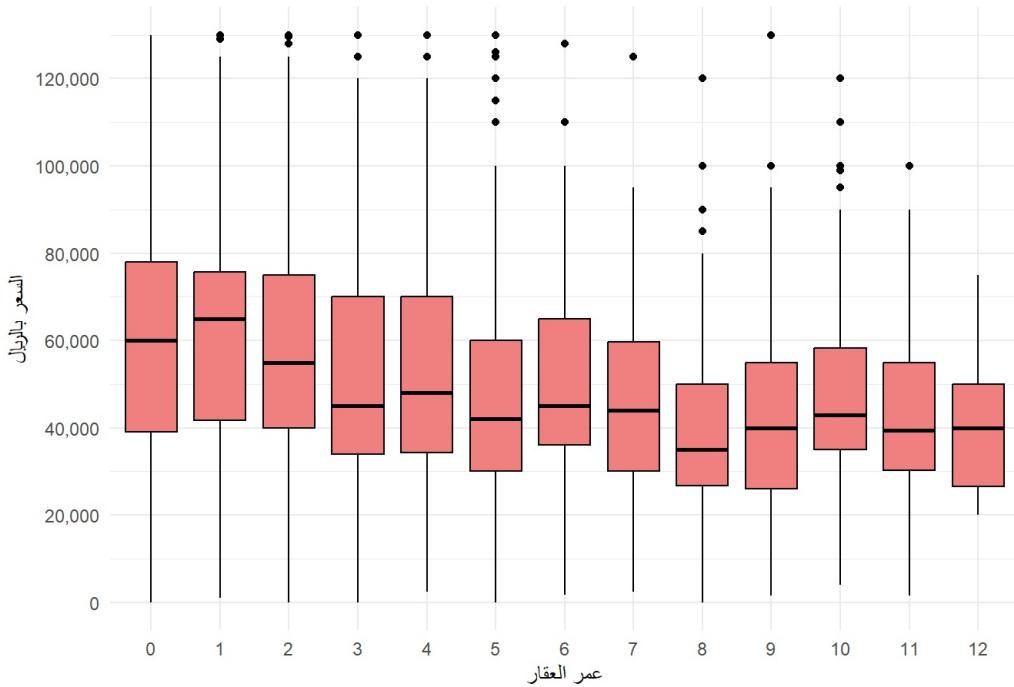


```

ggplot(filtered_data1_clean, aes(x = factor(age), y = price)) +
  geom_boxplot(fill = "lightcoral", color = "black") +
  labs(title = "توزيع الاسعار حسب عمر العقار (شقة للإيجار)",
       x = "عمر العقار",
       y = "السعر بالریال" +
  theme_minimal()+
  scale_y_continuous(labels = label_comma(),
                     breaks = seq(0, max(filtered_data1_clean$price, na.rm = TRUE), by = 20000) # Adjust the step size as needed
)

```

توزيع الاسعار حسب عمر العقار (شقة للإيجار)



```
# Select the top 5 districts with the most properties
top_districts5 <- filtered_data5_clean %>%
  count(district, sort = TRUE) %>%
  top_n(5) %>%
  pull(district)

top_districts1 <- filtered_data1_clean %>%
  count(district, sort = TRUE) %>%
  top_n(5) %>%
  pull(district)

# Filter the data for these top districts
filtered_top_districts5 <- filtered_data5_clean %>%
  filter(district %in% top_districts5)

filtered_top_districts1 <- filtered_data1_clean %>%
  filter(district %in% top_districts1)

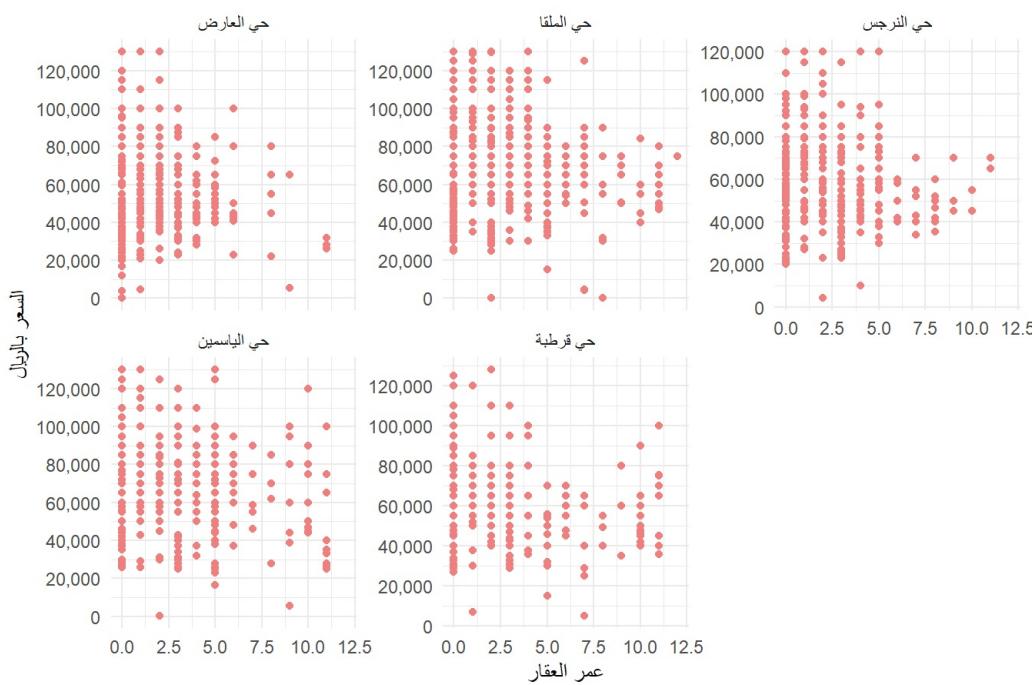
# Create the faceted scatter plot with only the top districts
ggplot(filtered_top_districts5, aes(x = age, y = price)) +
  geom_point(color = "lightcoral") +
  facet_wrap(~ district, scales = "free_y") +
  labs(title = "السعر مقابل المعلمات في المدن",
       x = "عمر العقار",
       y = "(السعر بالريل" +
  theme_minimal() +
  scale_y_continuous(
    labels = label_comma(),
    breaks = seq(0, max(filtered_top_districts5$price, na.rm = TRUE), by = 40000))
```

السعر مقابل العمر حسب المنطقة (أفضل 5 مناطق في الفئة فلة للايجار)



```
ggplot(filtered_top_districts1, aes(x = age, y = price)) +
  geom_point(color = "lightcoral") +
  facet_wrap(~ district, scales = "free_y") +
  labs(title = "السعر مقابل العمر حسب المنطقة (أفضل 5 مناطق في الفئة فلة للايجار)",
       x = "عمر العقار",
       y = "السعر بالريال") +
  theme_minimal() +
  scale_y_continuous(
    labels = label_comma(),
    breaks = seq(0, max(filtered_top_districts1$price, na.rm = TRUE), by = 20000))
```

السعر مقابل العمر حسب المنطقة (أفضل 5 مناطق في الفئة شقة للايجار)



```
# Correlation between age and price
correlation_age_price5 <- cor(filtered_data5_clean$age, filtered_data5_clean$price, use = "complete.obs")
print(paste("Correlation between age and price in category 5: ", correlation_age_price5))
```

```
## [1] "Correlation between age and price in category 5:  0.132087024937789"
```

```
print(paste("Correlation between age and price seems weak in category 5 and does not effect renting much, age alone has minimal effect on price."))
```

```
## [1] "Correlation between age and price seems week in category 5 and does not effect renting much, age alone has minimal effect on price."
```

```
correlation_age_price1 <- cor(filtered_data1_clean$age, filtered_data1_clean$price, use = "complete.obs")
print(paste("Correlation between age and price in category 1: ", correlation_age_price1))
```

```
## [1] "Correlation between age and price in category 1: -0.240745180832964"
```

```
print(paste("Correlation between age and price seems week in category 1 and does not effect renting much, age alone has minimal effect on price."))
```

```
## [1] "Correlation between age and price seems week in category 1 and does not effect renting much, age alone has minimal effect on price."
```

```
# Linear regression: Price ~ Age + Area + District
# For category 5
model_cat5 <- lm(price ~ age + area + district, data = filtered_data5_clean)
summary_cat5 <- summary(model_cat5)

cat("Category 5:\n")
```

```
## Category 5:
```

```
cat("District (location) and area strongly affect price, while age has a smaller and less significant influence on its own.\n")
```

```
## District (location) and area strongly affect price, while age has a smaller and less significant influence on its own.
```

```
cat("Model Summary:\n")
```

```
## Model Summary:
```

```
cat(paste0("Adjusted R-squared: ", round(summary_cat5$adj.r.squared, 4), "\n"))
```

```
## Adjusted R-squared: 0.5021
```

```
cat("Coefficients:\n")
```

```
## Coefficients:
```

```
coeff_cat5 <- summary_cat5$coefficients
print(coeff_cat5)
```

	Estimate	Std. Error	t value
## (Intercept)	1.353993e+05	11822.468080	11.4527095
## age	-9.497031e+02	369.665295	-2.5690890
## area	6.216283e+00	3.056775	2.0336084
## district2.700406	9.497031e+02	3.380176	1.2471059
## district3.024081-	6.216283e+00	0.779761	-1.5113772
## district8.546423-	6.216283e+00	0.070094	-2.2291681
## district3.534860	6.216283e+00	0.582527	1.7595486
## district4.945640-	6.216283e+00	0.160068	-2.0486341
## district7.484750-	6.216283e+00	0.643214	-5.3243274
## district8.681919-	6.216283e+00	0.173276	-5.0104502
## district4.798500	6.216283e+00	0.267087	0.2219651
## district1.126115	6.216283e+00	0.422723	2.9343163
## district4.282413-	6.216283e+00	0.508213	-1.9765130
## district7.413754-	6.216283e+00	0.346626	-3.9277444
## district2.559647	6.216283e+00	0.340780	1.1842905
## district1.720660-	6.216283e+00	0.793260	-1.0230580
## district2.861902	6.216283e+00	0.324251	1.8295888
## district1.766225	6.216283e+00	0.002712	0.8624145
## district3.652851	6.216283e+00	0.677358	1.9365496
## district5.721270-	6.216283e+00	0.268611	-4.7258738
## district1.704123-	6.216283e+00	0.198668	-0.6012883
## district2.033017	6.216283e+00	0.771480	1.2106498

```

## district4.050782          e+03 28345.614580  0.1429068
## district4.927622          e+04 24056.968396  2.0483136
## district3.664255-          e+04 38378.456682 -0.9547687
## district3.806264-          e+04 21736.026789 -1.7527441
## district3.474996-          e+04 124045.127793 -1.4451975
## district2.366942          e+04 21637.848880  1.0938897
## district8.034010-          e+04 21625.260026 -3.7151043
## district6.925336-          e+04 17326.800493 -3.9968926
## district1.446333          e+04 20765.968730  0.3547893
## district5.335576-          e+04 116407.818542 -3.2518494
## district7.259559-          e+04 124059.151768 -3.0173793
## district1.203752          e+04 24056.950824  0.5003760
## district2.109994          e+04 13394.310608  1.5752915
## district8.774189          e+04 28331.654655  3.0969562
## district1.272053-          e+04 12011.903793 -1.0589937
## district6.131757-          e+04 20006.226587  3.0649245
## district3.287340-          e+04 28348.695104 -1.1596090
## district8.535204-          e+04 21673.254840 -3.9381276
## district3.622137          e+03 20053.486144  0.1806238
## district5.996318          e+03 14727.194729  0.4071596
## district1.007607-          e+05 138350.503237 -2.6273625
## district8.315958          e+04 24046.524432  3.4582786
## district1.109534-          e+05 38373.599796 -2.8910970
## district9.020385          e+03 24101.923764  0.3742600
## district1.695336-          e+04 38373.599796 -0.4417515
## district5.116587          e+04 121612.905307  2.3673757
## district7.035725          e+04 28302.844279 -2.4858722
## district4.324554-          e+04 113366.682187 -3.2353231
## district1.541230-          e+04 124088.491039 -0.6398203
## district8.290704          e+03 13421.774897  0.6177055
## district5.981550          e+03 24055.190452  0.2486594
## district2.276917          e+04 28340.807899  0.8034059
## district2.274789          e+04 16333.416966  1.3927209
## district6.933848-          e+04 38316.810483 -1.8096099
## district6.593363          e+04 24191.564078  2.7356578
## district2.153135          e+04 24130.314961  0.8922947
## district2.873746-          e+04 28348.062565 -1.0137365
## district1.153798          e+05 28325.252767  3.0105423
## district3.231210-          e+04 124042.700107 -1.3439465
## district3.473544          e+04 13025.607229  2.6667044
## district7.729218          e+04 38439.462574  2.0107509
## district2.821079          e+04 17458.817098  1.6158479
## district3.510666          e+04 110001.375433  1.1701762
## district8.303257          e+03 18938.339918  0.4384354
## district1.448525-          e+04 21640.249581 -0.6693663
## district5.963536-          e+04 24160.034199 -2.4683473
## district3.894524-          e+04 13233.001747 -2.9430386
## district2.371289-          e+04 12980.601609 -1.8267947
## district4.253745          e+04 38329.794865  1.1097751
## district2.190413          e+04 124043.183635  0.9110330
## district5.376687-          e+03 12047.818223 -0.4462789
## district2.346787          e+04 20009.146902  1.1728569
## district6.648654-          e+04 38306.232012 -1.7356585
## district3.792400-          e+03 20058.218602 -0.1890696
## district6.648339-          e+04 17332.717875 -3.8357162
## district1.150960          e+04 18874.383706  0.6098001
## district7.123285          e+04 38328.528679  1.8584813
## district3.659689-          e+04 17413.462270 -2.1016436
## district9.141256          e+04 124045.110642  3.8017111
## district7.040366          e+04 24037.223327  2.9240771
## district2.789941          e+04 15123.360614  1.8447891
## district8.385210          e+04 18930.120460  4.4295599
## district3.875851          e+04 12575.343364  3.0821036
## district3.047450-          e+04 14942.502213 -2.0394506
## district5.885001-          e+03 38348.284459  0.1534579
## district2.137172-          e+04 24141.321207  0.8852756
## district8.738635-          e+04 14196.522180 -6.1554757
## district6.025218-          e+04 24107.407633 -2.4993220
## district1.379432-          e+04 12319.1016985  3.1203039
## district2.378932          e+04 24094.311891  0.9873418
## district5.806011          e+04 18030.835892  3.2200456
## district8.171525-          e+04 16855.734408 -4.8479200
## district6.382636-          e+04 28345.430020 -2.2517336
## district8.593367-          e+04 16112.459243 -5.3333679
## district7.645881-          e+04 13065.374401 -5.8520185
## district7.562710-          e+04 18947.472887 -3.9914081
## district6.501799-          e+04 18867.242078 -3.4460784
## district5.521843-          e+04 12074.397010 -4.2234017
## district5.986971-          e+04 16248.414255 -3.6621109

```

district4.159463- ٤٥٤ ٢٨٢٩٠.٠٨٣٨٨٩ -1.٤٧٠٢٩٠٠
district8.096178- ٤٥٤ ٣٥٨٦.٨٨٨٧٣٨ -5.٩٥٨١٦٩
district2.740725- ٤٥٤ ٢٤٠٤٣.٠٦٩٦٨١ -1.١٣٩٩٢٣٠
district1.336935 ٤٥٤ ١٤٩١٤.٤٢٥٢٩٠ ٠.٨٩٦٤٠٤٣
district2.089327- ٤٥٤ ٢٤٠٨٠.٩٨٠٤١٩ -0.٨٦٧٦٢٥٤
district6.945304- ٤٥٤ ١٣٩١٧.٦٤٥٩٨٩ ٤٦٩٩٠٢٨٦٥
district8.531299- ٤٥٤ ٣٨٣٢٨.٨٥٥١٣٧ -2.٢٢٥٨١٦٣
Pr(> t |)
(Intercept) 5.726202e-29
age 1.031016e-02
area 4.219896e-02
district2.125890 حي الازدياد
district1.309416 حي الاعلام
district2.597737 حي الالبيه
district7.872549 حي الندوه
district4.070341 حي الجفره
district1.197469 حي الحناوه
district6.198381 حي افلاطون
district8.243768 حي الحلوه
district3.402995 حي الخروج
district4.831312 حي القصيم
district9.036504 حي الدار الباهة
district2.365200 حي الزئه
district3.064756 حي الظواه
district6.754615 حي الطويق
district3.886225 حي الرحمانية
district5.302238 حي الرقة
district2.546960 حي الدهنه
district5.477555 حي الروابي
district2.262553 حي الزفافه
district8.863865 حي الزهاويه
district4.073484 حي الزهراء
district3.398767 حي الزهره
district7.988761 حي السداده
district1.486495 حي الملايم
district2.742111 حي السليمانيه
district2.119551 حي السعده
district6.787788 حي السوادي القويه
district7.228064 حي الشفافيه
district1.176732 حي افلاطون
district2.600542 حي افلاطون
district6.168972 حي الشهاده
district1.154386 حي العبدالله
district1.998099 حي العقاده
district2.898043 حي العريجاء الغربيه
district2.223234 حي العريجاء الوسطى
district2.464264 حي العزقيه
district8.658869 حي العزقيه
district8.566917 حي النميره
district6.839595 حي النميره
district8.708763 حي افلاطون
district5.616553 حي العباس
district3.904440 حي الغنمه
district7.082735 حي الفاطمه
district6.587443 حي الفاوده
district1.806418 حي افلاطون
district1.305145 حي المهدى
district1.246416 حي القافية
district5.224050 حي العدد
district5.368804 حي القبراء
district8.036645 حي المحمله
district4.218907 حي النور
district1.639484 حي المازلاط
district7.059301 حي الحلوه
district6.312411 حي القروع
district3.724043 حي المدهنه
district3.109019 حي المعلقه
district2.659380 حي المغره
district1.792061 حي الملعون
district7.757361 حي المقهه
district4.456315 حي الملك عبد العزيز
district1.063760 حي الملك عبد الله
district2.421500 حي الملك فهد
district6.611453 حي الملك سلمان
district5.033837 حي المنصور
district1.370499 حي المهيجه
district3.309209 حي الموهبة
district6.796563 حي الموهبة
district2.673064 حي الطيف

```

## district3.624512   حي العدد
## district6.554720  حي التلوج
## district2.410735  حي التلوجه
## district8.286718  حي النسيم الـ٩٢
## district8.500685  حي النسيم الـ٩٣
## district1.313541  حي الـ٩٤
## district5.421035  حي الـ٩٥
## district6.333228  حي النموذـ١
## district3.578139  حي الـ٩٦
## district1.505275  حي الـ٩٧
## district3.516178  حي الـ٩٨
## district6.530131  حي الـ٩٩
## district1.025633  حي الـ٩٩
## district2.099750  حي الـ٩٩
## district4.161203  حي الـ٩٩
## district8.780616  حي ام الحمام الـ٩٩
## district3.761759  حي ام الحمام الـ٩٩
## district1.002449  حي الـ٩٩
## district1.256875  حي بـ١٠٢
## district2.627964  حي جامـة الأمـرة بـ١٠٢
## district3.236635  حي بـ١٠٣
## district1.314209  حي بـ١٠٣
## district1.401938  حي دـ١٠٦
## district2.450969  حي شـ١٠٧
## district1.140575  حي صـ١٠٧
## district6.175666  حي بـ١٠٨
## district6.944715  حي بـ١٠٨
## district5.873856  حـ١٠٩
## district2.578321  حـ١٠٩
## district2.604457  حـ١٠٩
## district1.417313  حـ١١٠
## district3.287481  حـ١١٠
## district2.545337  حـ١١٠
## district3.702070  حـ١١٠
## district3.857636  حـ١١٠
## district6.867822  حـ١١٠
## district2.620175  حـ١١٠

```

```
cat("\n")
```

```
# For category 1
model_cat1 <- lm(price ~ age + area + district, data = filtered_data_clean)
summary_cat1 <- summary(model_cat1)
```

```
cat("Category 1:\n")
```

```
## Category 1:
```

```
cat("Age and area are strong predictors, while certain districts have significant positive impacts on price, but many are weak predictors individually.\n")
```

```
## Age and area are strong predictors, while certain districts have significant positive impacts on price, but many are weak predictors individually.
```

```
cat("Model Summary:\n")
```

```
## Model Summary:
```

```
cat(paste0("Adjusted R-squared: ", round(summary_cat1$adj.r.squared, 4), "\n"))
```

```
## Adjusted R-squared: 0.3427
```

```
cat("Coefficients:\n")
```

```
## Coefficients:
```

```
coeff_cat1 <- summary_cat1$coefficients
print(coeff_cat1)
```

	Estimate	Std. Error	t value
## (Intercept)	35790.683939	2.112567e+04	1.694179835
## age	-1420.557191	9.007182e+01	-15.771383011
## area	1.902873	4.362725e-01	4.361662065
## district2.120585	6410.451453	٦٤١٠٤٥١٤٥٣	٣٠٢٢٩٦٢٩٩
## district2.134491	23482.310093	٩٤٠١٣٦١٢٩	١٠٠١٣٦١٢٩
## district2.987953	4121.814686	٨٤٠٤٧٥	١٣٧٩٤٧٧٥
## district2.154656	12698.749109	٨٤٠٤٧٥	٥٥٨٩٣٦٣٠٨٥
## district2.987613	34003.805747	٨٤٠٤٧٥	١٣٨١٥٩٥٨٤
## district2.150241	29803.282762	٨٤٠٤٧٥	٣٨٦٠٤٣٨٤١
## district2.439743	16348.599024-	٨٤٠٤٧٥	٦٣٠٩٥٩٢١٣
## district2.362213	1201.231054-	٨٤٠٤٧٥	٥٥٥٠٨٥١٩٣٤
## district2.192443	13943.804435-	٨٤٠٤٧٥	١٦٣٥٩٩٤٠٣٧
## district2.170588	160.142625-	٨٤٠٤٧٥	٠٠٧٣٧٧٨٤٤
## district2.147884	24420.121250	٨٤٠٤٧٥	٣٣٦٩٣٨٦٩١
## district2.988963	8083.788016	٨٤٠٤٧٥	٠٢٣٠٤٥٤٦١٥
## district2.124509	2933.624691	٨٤٠٤٧٥	١٣٨٠٨٤٨٥٤
## district2.147272	2240.805700	٨٤٠٤٧٥	١٠٤٣٥٥٩٣٢
## district2.588910	14629.456974	٨٤٠٤٧٥	١٥٦٥٠٨١٦٥١
## district2.989237	1178.953876	٨٤٠٤٧٥	٠٣٩٤٣٩٩٦٦
## district2.258577	49555.704356	٨٤٠٤٧٥	٢١٩٤١١١٨٧١
## district2.362178	29707.228494	٨٤٠٤٧٥	٢٥٧٦٢٠٢٦٢
## district2.138706	30870.087337	٨٤٠٤٧٥	٤٤٣٤٣٠٤٥٣
## district2.281898	68464.966399	٨٤٠٤٧٥	٣١٠٩٣٥٢٠٨٣
## district2.314373	42390.889125	٨٤٠٤٧٥	٨٣١٦٣٦٢٧٢
## district2.120045	6074.425573	٨٤٠٤٧٥	٢٨٦٥٢٣٤٦٩
## district2.207344	16230.942110	٨٤٠٤٧٥	٣٣٥٣١٥٥٨١
## district2.138608	8351.696866	٨٤٠٤٧٥	٣٩٠٥٢٠٢٣١
## district2.156210	26600.051550	٨٤٠٤٧٥	٢٣٣٦٤٨٣٢
## district2.989240	11556.317702	٨٤٠٤٧٥	٣٨٦٥٩٧١٦٢
## district2.441486	5252.539041	٨٤٠٤٧٥	٢١٥١٣٦٩٥٥
## district2.149263	2213.007067	٨٤٠٤٧٥	٠١٠٢٩٦٥٨٣٤
## district2.207299	12625.159264	٨٤٠٤٧٥	٥٧١٩٧٣٢٠٨
## district2.587949	16627.089829	٨٤٠٤٧٥	٦٤٢٤٨١٤٦٣
## district2.158164	29734.921003	٨٤٠٤٧٥	١٣٧٣٧٨٧٧٧
## district2.282824	10570.948269	٨٤٠٤٧٥	٤٦٣٠٦٤٦٣٢
## district2.315749	1804.175534	٨٤٠٤٧٥	٦٧٩٠٨٩٥٥
## district2.989256	24626.129096	٨٤٠٤٧٥	٨٢٣٨٢١٤٥٣
## district2.215710	2965.460055-	٨٤٠٤٧٥	١٣٣٨٣٧٩١٩
## district2.258774	3870.601239-	٨٤٠٤٧٥	١٧١٣٥٨٤٦٦
## district2.989250	442.374342-	٨٤٠٤٧٥	٠١٤٧٩٨٨٣٨
## district2.141873	30237.493181	٨٤٠٤٧٥	٤١١٧٣١١٦٩
## district2.123244	27898.261086	٨٤٠٤٧٥	٣١٣٩٤٤٨٧١
## district2.987613	4034.251719	٨٤٠٤٧٥	١٣٥٣٢٥٩٤
## district2.315219	17222.526373	٨٤٠٤٧٥	٧٤٣٨٨٣٢٩٩
## district2.114327	23106.180024	٨٤٠٤٧٥	٠٩٢٨٣٨٥٧٧
## district2.988964	3106.622495	٨٤٠٤٧٥	٠٦٣٩٣٦٤٢٦
## district2.241820	472.881614	٨٤٠٤٧٥	٢٢٦٨٥٢٣٣
## district2.315929	5253.738655	٨٤٠٤٧٥	٠٩٦٠٦١٦٤
## district2.215895	2428.757397-	٨٤٠٤٧٥	٧٤٧٠٨٨١٠٦
## district2.116984	15815.736245	٨٤٠٤٧٥	٠٤٧٢٨٢١٢٨
## district2.129069	22297.360295	٨٤٠٤٧٥	٥٥٨٦٥٤١٢٦
## district2.989058	16698.496268	٨٤٠٤٧٥	٢١٨٢٩١٥٦٦
## district2.989248	6525.276812	٨٤٠٤٧٥	٢٨٥٦٥٦٦٠٣٠
## district2.182010	62319.656469	٨٤٠٤٧٥	٥٥٩٢٨٥٩٢٥٩
## district2.989257	17645.157829	٨٤٠٤٧٥	٧٠٩٥٩٩٦٩٤
## district2.151660	15268.170685	٨٤٠٤٧٥	٣٤٢٢٦٨٤٢٧
## district2.988973	10230.309258	٨٤٠٤٧٥	٣٤٦٠٢٠٠٥٠
## district2.160889	7477.110185	٨٤٠٤٧٥	١٥٥٤٩٠٠٢١
## district2.124913	3304.027316	٨٤٠٤٧٥	٧٩٢٤٣٨٢٢٤
## district2.130137	16880.020202	٨٤٠٤٧٥	٥٣١٢١٤٠٢٧
## district2.118230	32010.992206	٨٤٠٤٧٥	٥٥٩٩٢١٢٧٨
## district2.173907	66519.838300	٨٤٠٤٧٥	١٨٧٦٠٧٦٧٠٧٠١
## district2.314862	43428.584493	٨٤٠٤٧٥	١٥٣٥١٩٣٠٨
## district2.177856	25121.990661	٨٤٠٤٧٥	٠١٦٣٤٦٠٩٨
## district2.227645	364.133058	٨٤٠٤٧٥	٨١٢٧٩٤٩٥٣
## district2.177643	17699.769931	٨٤٠٤٧٥	٦٥٥٥٠٤٠٦٢٧
## district2.138292	14006.683343	٨٤٠٤٧٥	٦٥٣٣٧٩٢٣٣
## district2.314296	61407.053809	٨٤٠٤٧٥	٢٦٣٣٥٠٣٧٧
## district2.167521	5708.175564-	٨٤٠٤٧٥	٤٠٥٢٥١١٥٦
## district2.154606	8731.564166	٨٤٠٤٧٥	٢٩٣٢٣٤٠٤٠٤
## district2.199806	6450.580357	٨٤٠٤٧٥	٩١٨٩٤٧٠١٠
## district2.114484	40575.829940	٨٤٠٤٧٥	٢٤٤٠٣٠٧٤٩٤٣
## district2.587353	62176.026876	٨٤٠٤٧٥	١٦١٥٢٦٢٦٨
## district2.174071	38296.835734	٨٤٠٤٧٥	٦٣١٧٤٦٣٢
## district2.216350	13767.400889	٨٤٠٤٧٥	٥٩٣٢١٩٧١٢
## district2.150290	12841.956137	٨٤٠٤٧٥	٨٨٦٩٧٥٦٤٦
## district2.216213	19657.272705	٨٤٠٤٧٥	

```

## district2.988965 24129.456974   ٦٤٩٤ لمن٠٨٩٧٢٨٤٥٦٧
## district2.124853 5045.238583   ٦٤٩٤ ٢٣٧٤٣٩٤٦٠
## district2.587789 67669.426609   ٦٤٩٤ ٦١٤٩٥١٣١٨
## district2.117118 15287.054640   ٦٤٩٤ ٧٢٢٠٦٩٠٣٠
## district2.150306 53593.654645   ٦٤٩٤ ح٤٩٢٣٧٣٦١٨
## district2.128033 46054.930878   ٦٤٩٤ ٢.١٦٤٢٠٢٣٥٩
## district2.114032 29016.096420   ٦٤٩٤ ح٣٧٢٥٤٧٥٤٧
## district2.151959 22346.197111   ٦٤٩٤ ٣٠٣٨٤١١٨٩٥
## district2.150728 11100.533850   ٦٤٩٤ ٥٥١٦١٢٩٦٧١
## district2.142264 2841.973484-  ٦٤٩٤ ٥١٣٢٦٦٣١٣٤
## district2.362155 2372.530587   ٦٤٩٤ ٥١٠٤ ١٠٠٤٣٩٢٦٢
## district2.130486 12320.891600   ٦٤٩٤ ٥٥٧٨٣١٣٧٤٧
## district2.587598 24397.691915-  ٦٤٩٤ ٥٩٤٢٨٧٠٣٨٦
## district2.125143 2509.466503   ٦٤٩٤ ٦١١٨٠٨٤٥٩٠
## district2.241121 39381.474738   ٦٤٩٤ ح٧٥٧٢٢٢١٠٧
## district2.125534 15201.463063   ٦٤٩٤ ٦٧١٥١٨٣١٤٤
## district2.206989 36825.489455   ٦٤٩٤ ح٦٦٨٥٨٥٤٢٩
## district2.199061 10020.586906   ٦٤٩٤ ٥٤٥٥٦٧٥٧٩٧
## district2.588633 1633.121642-  ٦٤٩٤ ٥٠٦٣٠٨٨١٨٧
## district2.115438 36089.648601   ٦٤٩٤ ٥٦٥٠١٣٤٥٢
## district2.117858 8422.059631   ٦٤٩٤ ٥٣٩٧٦٦٨٧٩٠
## district2.363490 6722.414043-  ٦٤٩٤ ٥٢٨٤٤٢٧٣٩٩
## district2.241495 11071.290648   ٦٤٩٤ ٥١٤٩٣٩٢٤٢٨٧
## district2.314197 1482.021857-  ٦٤٩٤ ٥١٥٦٤٠٤٠٤٤٣
## district2.987953 19121.814686   ٦٤٩٤ ٥١٥٦٤٠١١٤٩
## district2.154848 2614.365393-  ٦٤٩٤ ٥٠١٢١٣٢٤٨٠٣
## district2.361951 5163.202527   ٦٤٩٤ ٥٠٢١٨٥٩٩٠٨٨
## district2.989244 49454.870501   ٦٤٩٤ ٥٠١٦٥٤٤٢٧٢٦٥
## district2.987670 66574.712122   ٦٤٩٤ ٥٢٨٣١٥٦٢٤ ح٢٠١٢١٢١٢٢
## district2.282172 10914.380378   ٦٤٩٤ ٥٠٤٧٨٢٤٥٣٢٠
## district2.119419 46704.478321   ٦٤٩٤ ٥٢٠٣٦٤٥٣١٦
## district2.361954 1984.763766-  ٦٤٩٤ ٥٠٠٨٤٠٣٥٨٢
## district2.988965 4129.456974   ٦٤٩٤ ٥٠١٣٨١٥٦٧٣١
## district2.440428 16678.458343   ٦٤٩٤ ٥٠٦٨٣٤٢٣٥٩٩
## district2.363017 35891.573469   ٦٤٩٤ ٥١٥١٨٨٧٨٨٤
## district2.314255 6842.261569   ٦٤٩٤ ٥٢٩٥٥٥٧٢٦٥
## district2.125325 378.908640   ٦٤٩٤ ٥٠٠١٧٨٢٨٢٧٠
## district2.282133 14554.597726   ٦٤٩٤ ٥٠١٠٦٦٢٧٦٣٠٣٢
## district2.120924 7141.000629   ٦٤٩٤ ٥٠٣٣٦٦٩٢٨٧١
## district2.152854 1922.782972-  ٦٤٩٤ ٥٠٣٦٥٩٣١٣١٩٩
## district2.989245 17535.615133-  ٦٤٩٤ ٥٠٥٨٦٦٢٣٦١٥
## district2.145637 50033.990827   ٦٤٩٤ ٥٠٣٣١٨٩٤٣١٤
## district2.987615 23359.187377-  ٦٤٩٤ ٥٠٧٨١٨٦٧٤٠٨
## district2.167477 3075.574816-  ٦٤٩٤ ٥٠١٤١٨٩٦٥١٦
## district2.216162 10640.305679-  ٦٤٩٤ ٥٠٤٨٠١٢٢٩٩٥
## district2.135145 31006.277978   ٦٤٩٤ ٥٠٤٤٥٢١٨٦٥٦١
## district2.116696 32199.913691   ٦٤٩٤ ٥٠٥٢١٢٣٤٧٠٤
## district2.178244 11505.872707   ٦٤٩٤ ٥٠٥٢٨٢١٧٨٢٥
## district2.588898 4831.657306   ٦٤٩٤ ٥٠١٨٦٦٢٩٨٦٣
## district2.587450 29603.463360-  ٦٤٩٤ ٥٠١١٤١١٧٣٧١
##                                     Pr(>|t|)

## (Intercept) 9.028304e-02
## age          6.148813e-55
## area         1.312400e-05

## district7.624367   ح٢٠١٢١٢١٢٢
## district2.713170   ح٢٠١٢١٢١٢٢
## district8.902863   ح٢٠١٢١٢١٢٢
## district5.556399   ح٢٠١٢١٢١٢٢
## district2.550993   ح٢٠١٢١٢١٢٢
## district1.657851   ح٢٠١٢١٢١٢٢
## district5.028229   ح٢٠١٢١٢١٢٢
## district9.594452   ح٢٠١٢١٢١٢٢
## district5.248046   ح٢٠١٢١٢١٢٢
## district9.941136   ح٢٠١٢١٢١٢٢
## district2.556093   ح٢٠١٢١٢١٢٢
## district7.868198   ح٢٠١٢١٢١٢٢
## district8.901780   ح٢٠١٢١٢١٢٢
## district9.168904   ح٢٠١٢١٢١٢١٢٢
## district5.720394   ح٢٠١٢١٢١٢١٢٢
## district9.685409   ح٢٠١٢١٢١٢١٢٢
## district2.826560   ح٢٠١٢١٢١٢١٢٢
## district2.085781   ح٢٠١٢١٢١٢١٢٢
## district1.489599   ح٢٠١٢١٢١٢١٢٢
## district2.707766   ح٢٠١٢١٢١٢١٢٢
## district6.705520   ح٢٠١٢١٢١٢١٢٢
## district7.744871   ح٢٠١٢١٢١٢١٢٢
## district4.621761   ح٢٠١٢١٢١٢١٢٢
## district6.961658   ح٢٠١٢١٢١٢١٢٢
## district2.173823   ح٢٠١٢١٢١٢١٢٢

```

district6.990682 حي الزهراء
district8.296679 حي الظهرة
district9.179935 حي السيدة
district5.673616 حي العطام
district5.205852 حي العجلة
district1.683202 حي السليمانية
district6.433348 حي السعيدية
district9.379031 حي السويدي العوالي
district4.100738 حي الشفوية
district8.935352 حي العبرة
district8.639477 حي العطف
district9.881932 حي الشهاد
district1.580810 حي الشهداء
district1.889150 حي الصناعة
district8.925906 حي العلامة
district4.569762 حي العذراء
district2.745086 حي العذر
district9.172233 حي العريجاء الغربية
district9.831717 حي العريجاء المطلقة
district8.205463 حي العزيزية
district9.127254 حي العزيز
district4.550397 حي العزيز
district2.950117 حي العزم
district5.764187 حي العزم
district8.272094 حي العزم
district4.304001 حي العزم
district5.550214 حي الفاو
district4.779800 حي العلاج
district7.321609 حي العطاء
district7.293398 حي العفاف
district8.764403 حي القابضة
district4.281365 حي القديم
district1.307866 حي القبلة
district2.223735 حي المحاجة
district6.069328 حي المطبع
district2.487433 حي المولدة
district9.869588 حي الملعون
district4.163679 حي الملعون
district5.124667 حي المطير
district7.990039 حي المعذر الشفاف
district7.922896 حي المعياطة
district6.853074 حي المغارة
district7.693534 حي العلوى
district5.503845 حي العلوى
district1.628804 حي الملك عبد العظيم
district7.820032 حي الملك عبد الله
district5.345083 حي الملك فهد
district5.503832 حي الملك فهد
district3.751276 حي العلوى
district4.195346 حي المنارة
district8.123240 حي المطير
district8.946353 حي المؤتمنة
district4.702802 حي الموالية
district1.271603 حي العزيز
district3.048817 حي العبد
district1.699444 حي العلوى
district2.991202 حي العلوى
district6.057833 حي النسيم الشفاف
district8.944650 حي النسيم العظيم
district9.199990 حي العليم
district5.630740 حي العلوى
district3.457852 حي النموذجية
district9.060046 حي العلوى
district7.893098 حي العلوى
district4.745237 حي العلوى
district9.525177 حي العلوى
district6.486396 حي الوزير
district9.496984 حي العظام
district8.805735 حي البارحة
district6.908885 حي البارحة
district7.760927 حي البارحة
district6.213777 حي ام الحمام الشفاف
district9.489402 حي ام الحمام العظيم
district5.222207 حي ام نافع
district9.034378 حي العلوى
district8.269698 حي العلوى
district9.809306 حي العلوى
district2.589641 حي جامعة الأميرة سارة
district6.324930 حي العلوى

```

## district2.758717 حي طيبة
## district9.330349 حي دارالبيضاء
## district8.901212 حي سلطنة
## district4.943656 حي نهرة
## district1.288434 حي صلاح الدين
## district7.675021 حي ماحية
## district9.857764 حي طويق
## district5.236523 حي طهرا البداع
## district7.363602 حي طهرا بن جعفر
## district9.288360 حي طهرا بن جعفر
## district5.574786 حي عاليه
## district1.973919 حي قرطبة
## district4.343234 حي بن جعفر
## district8.871665 حي بن حماد
## district6.311574 حي عاليه
## district1.465022 حي عمر بن الخطاب
## district1.282537 حي فاطمة
## district5.973677 حي بن بشر
## district8.519572 حي منور
## district2.526206 حي اقبال

```

```

# Calculate median price for each district
district_medians <- aggregate(price ~ district, data = filtered_data5_clean, median)

district_medians <- aggregate(price ~ district, data = filtered_data1_clean, median)

# Define high-demand as districts above the overall median price
overall_median <- median(filtered_data5_clean$price)
district_medians$demand_level <- ifelse(district_medians$price > overall_median, "عاليه متخصص", "عاليه متوسط")

overall_median <- median(filtered_data1_clean$price)
district_medians$demand_level <- ifelse(district_medians$price > overall_median, "عاليه متخصص", "عاليه متوسط")

# Merge the demand_level back into your main dataset
filtered_data5_clean <- merge(filtered_data5_clean, district_medians[, c("district", "demand_level")], by = "district")

filtered_data1_clean <- merge(filtered_data1_clean, district_medians[, c("district", "demand_level")], by = "district")

# Remove duplicate 'demand_level' columns by selecting only unique column names
filtered_data5_clean <- filtered_data5_clean[, !duplicated(colnames(filtered_data5_clean))]
filtered_data1_clean <- filtered_data1_clean[, !duplicated(colnames(filtered_data1_clean))]

#colnames(filtered_data1_clean)

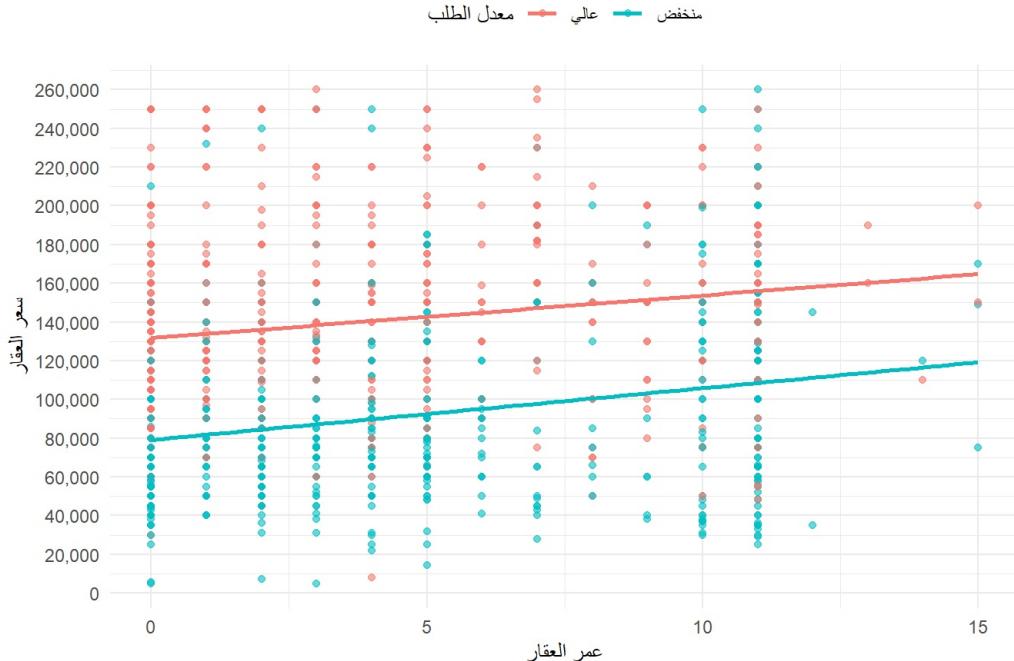
```

```

ggplot(filtered_data5_clean, aes(x = age, y = price, color = demand_level)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) + # Adds trend lines
  labs(title = "تمثيل العمر مقابل السعر لفيلا للإيجار",
       x = "عمر العقار",
       y = "سعر العقار",
       color = "معدل الطلب") +
  theme_minimal() +
  theme(legend.position = "top") +
  scale_y_continuous(
    labels = label_comma(),
    breaks = seq(0, max(filtered_data5_clean$price, na.rm = TRUE), by = 20000) # Adjust the step size as needed
  )

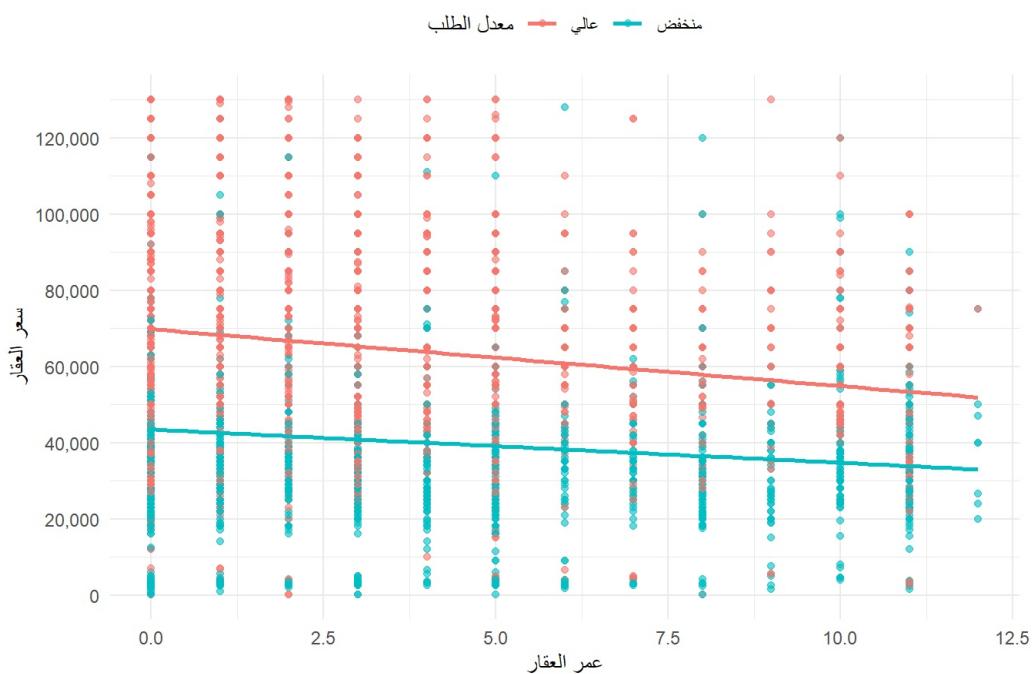
```

تمثيل العمر مقابل السعر لفيلا للإيجار



```
ggplot(filtered_data1_clean, aes(x = age, y = price, color = demand_level)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) + # Adds trend lines
  labs(title = "تمثيل العمر مقابل السعر لشقة للايجار",
       x = "عمر العقار",
       y = "سعر العقار",
       color = "معدل الطلب") +
  theme_minimal() +
  theme(legend.position = "top")+
  scale_y_continuous(
    labels = label_comma(),
    breaks = seq(0, max(filtered_data1_clean$price, na.rm = TRUE), by = 20000) # Adjust the step size as needed
  )
```

تمثيل العمر مقابل السعر لشقة للايجار



- For Renting Properties in Riyadh: After analyzing the data and trying various plots and calculations, the hypothesis is partially confirmed. While age itself is a weak predictor of price, area and specific district locations significantly impact price, supporting the idea that location moderates the effect of age. The weak correlations indicate that age does not independently have a strong influence on price, and if the analysis were applied to other cities, it might yield different results. This finding highlights the distinction between high and low demand locations in Riyadh.^{**} - **For Buying Properties in Riyadh: The hypothesis is confirmed. Older properties in high-demand areas retain more value, emphasizing that the desirability of the location plays a critical role in maintaining property value over time, even as the property ages.**

the number and type of rooms in a property (e.g., bedrooms, living rooms, etc.) significantly affect its attributes or value, but that this effect decreases beyond a certain point.

```
df9 <- read_csv("akar2024.csv", locale = locale(encoding = "UTF-8"))
df9 <- df9 %>%
  filter(rent_period == 3.0, category == 1, city == "الرِّبَاط") %>%
  filter(!is.na(price) & !is.na(area)) %>%
  mutate(price_per_meter=price/area)

print(head(df9$price_per_meter))

## [1] 433.3333 767.0455 543.4783 729.9270 333.3333 614.0351

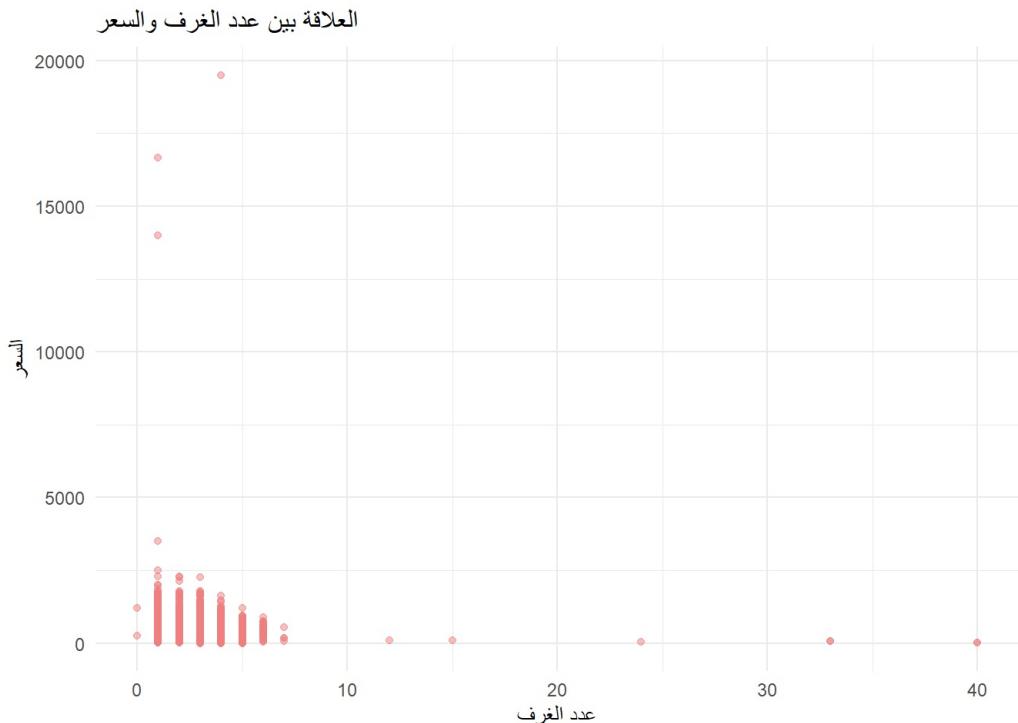
room_summary <- df9 %>%
  group_by(beds, livings, rooms, wc) %>%

  summarize(mean_price = mean(price_per_meter, na.rm = TRUE),
            price_per_meter = list(price_per_meter)) %>%
  ungroup() %>%
  na.omit()

room_summary <- room_summary %>%
  unnest(price_per_meter)

room_summary$total_rooms <- room_summary$rooms + room_summary$livings + room_summary$beds + room_summary$wc

create_plot <- function(data, x_var, y_var, x_label, y_label, plot_title) {
  ggplot(data, aes_string(x = x_var, y = y_var)) +
  geom_point(color = "lightcoral", alpha = 0.5) +
  labs(
    title = plot_title,
    x = x_label,
    y = y_label
  ) +
  theme_minimal()
}
create_plot(room_summary, "rooms", "price_per_meter", "العلاقة بين عدد الغرف والسعر"
"
```



```
room_price_group <- aggregate(price_per_meter ~ rooms, data = room_summary, FUN = mean)
cat("متوسط السعر بناءً على عدد الغرف\n")
```

متوسط السعر بناءً على عدد الغرف ##

```
print(room_price_group)
```

```
##   rooms price_per_meter
## 1      0    721.62162
## 2      1    445.16173
## 3      2    355.94705
## 4      3    369.15453
## 5      4    379.97836
## 6      5    320.84153
## 7      6    425.20981
## 8      7    238.88498
## 9     12     93.65559
## 10    15    106.66667
## 11    24    31.42857
## 12    33    66.66667
## 13    40    24.61538
```

```
total_room_price_group <- aggregate(price_per_meter ~ total_rooms, data = room_summary, FUN = mean)
cat("متوسط السعر بناءً على مجموع الغرف \n")
```

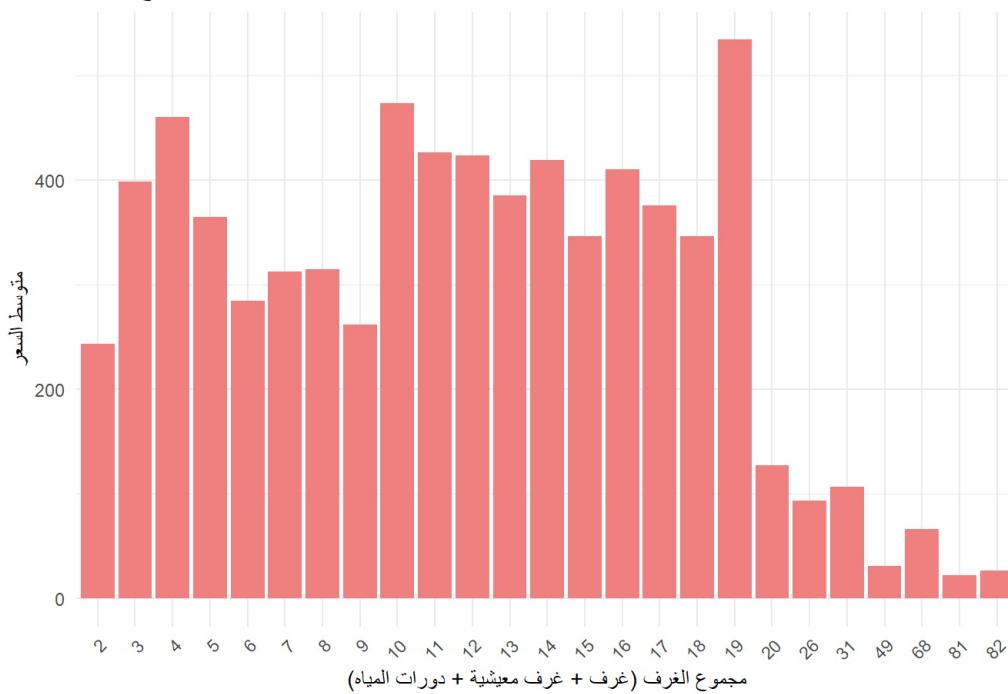
متوسط السعر بناءً على مجموع الغرف #:

```
print(total_room_price_group)
```

```
##   total_rooms price_per_meter
## 1            2    243.24324
## 2            3    399.01965
## 3            4    460.71601
## 4            5    365.01158
## 5            6    285.10053
## 6            7    313.02946
## 7            8    314.80278
## 8            9    262.23201
## 9           10    473.63723
## 10          11    426.28117
## 11          12    423.63100
## 12          13    385.64595
## 13          14    419.56549
## 14          15    346.34346
## 15          16    410.70092
## 16          17    375.53857
## 17          18    346.52555
## 18          19    534.59119
## 19          20    127.14103
## 20          26     93.65559
## 21          31    106.66667
## 22          49    31.42857
## 23          68    66.66667
## 24          81    22.30769
## 25          82    26.92308
```

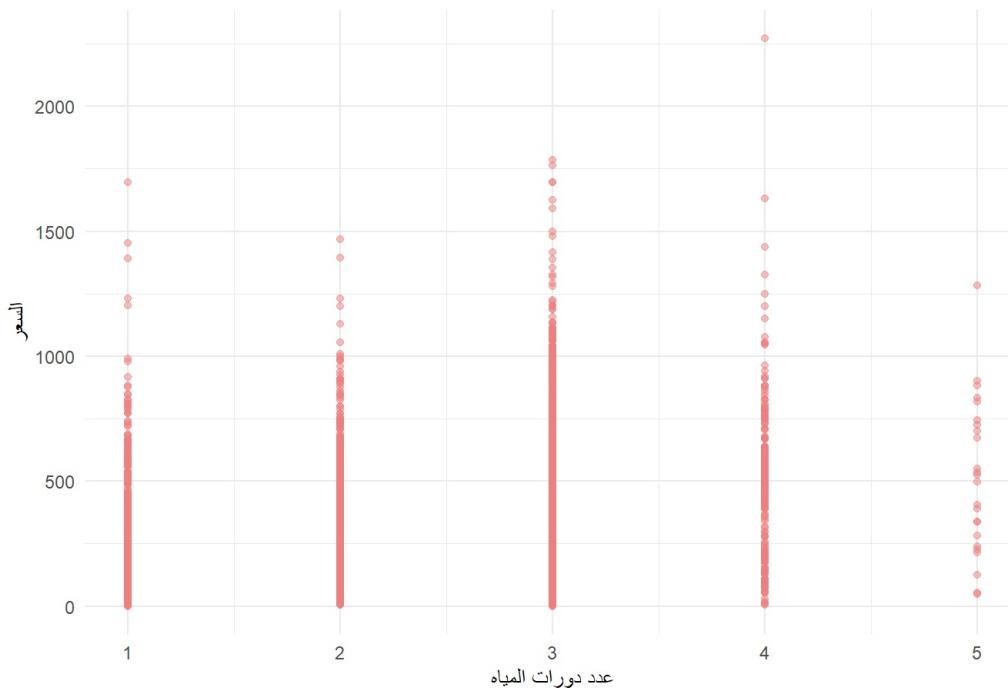
```
ggplot(total_room_price_group, aes(x = factor(total_rooms), y = price_per_meter)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    title = "متوسط السعر بناءً على مجموع الغرف",
    x = "(مجموع الغرف (غرف + غرف معيشية + دورات المياه",
    y = "متوسط السعر"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

متوسط السعر بناء على مجموع الغرف



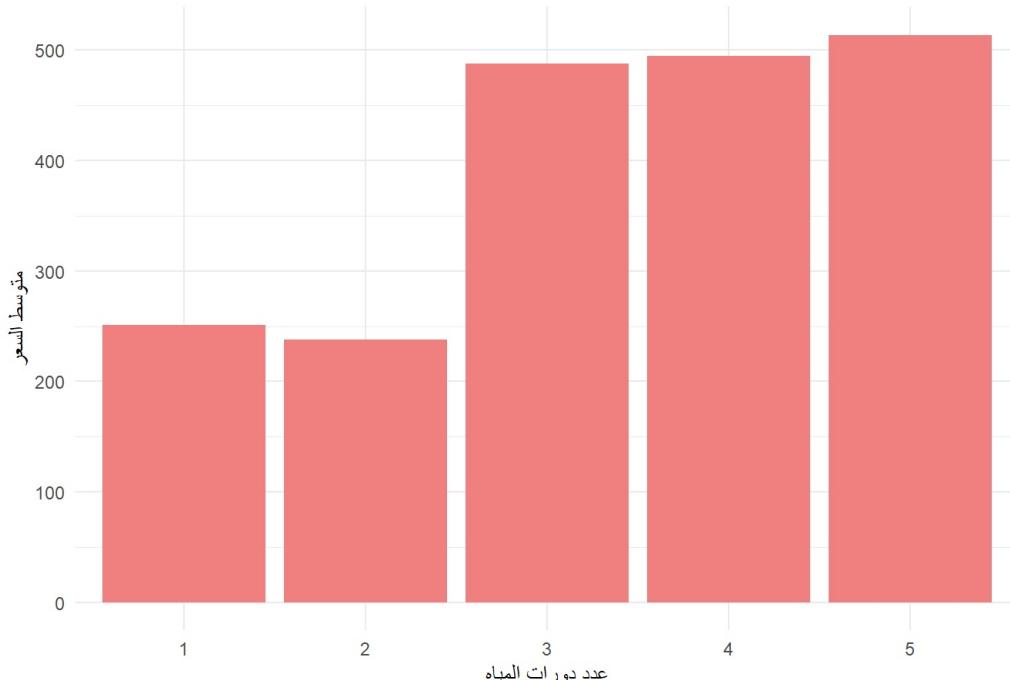
```
filtered_data <- subset(room_summary, rooms == 3)
create_plot(filtered_data, "wc", "price_per_meter", "المباهة", "السعر", "العلاقة بين عدد دورات المياه والسعر لعدد غرف 3 غرف
```

العلاقة بين عدد دورات المياه والسعر لعدد 3 غرف



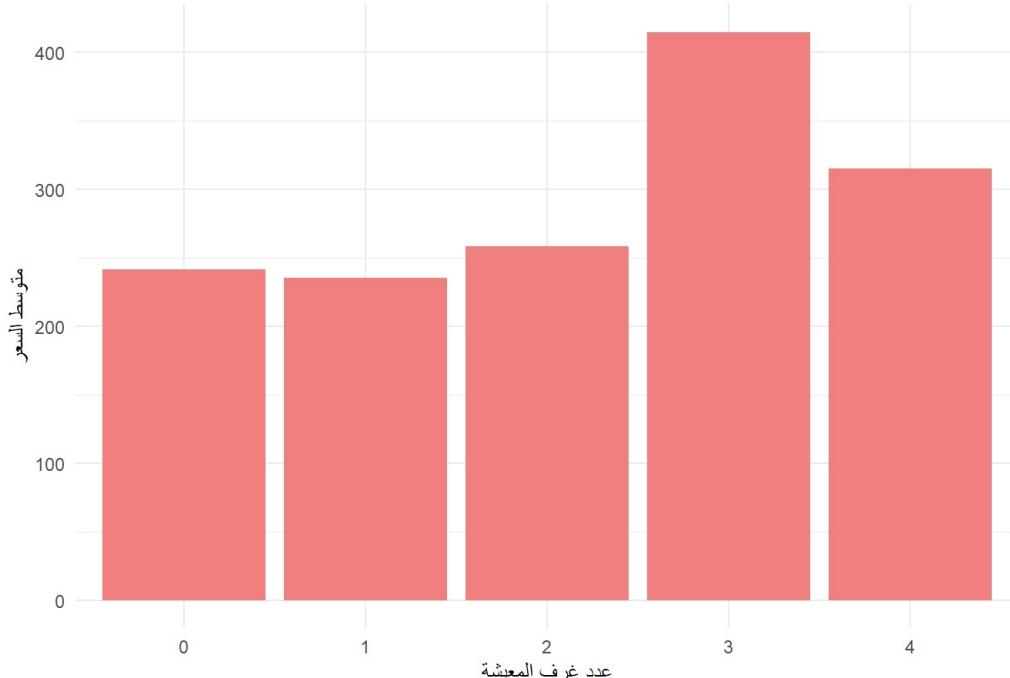
```
wc_price_group <- aggregate(price_per_meter ~ wc, data = filtered_data, FUN = mean)
ggplot(wc_price_group, aes(x = factor(wc), y = price_per_meter)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    title = "متوسط السعر بناء على عدد دورات المياه لعدة غرف",
    x = "عدد دورات المياه",
    y = "متوسط السعر"
  ) +
  theme_minimal()
```

متوسط السعر بناءً على عدد دورات المياه لعدد 3 غرف



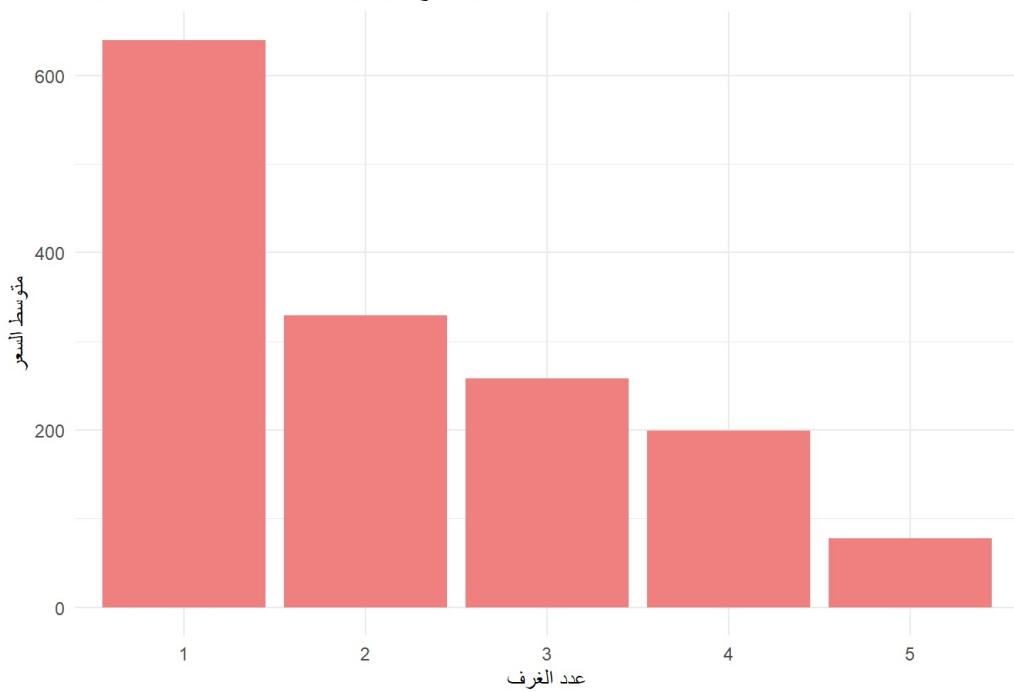
```
filtered_data <- subset(room_summary, rooms == 3 & wc == 2)
livings_price_group <- aggregate(price_per_meter ~ livings, data = filtered_data, FUN = mean)
ggplot(livings_price_group, aes(x = factor(livings), y = price_per_meter)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    title = "متوسط السعر بناءً على عدد غرف المعيشة لعدد ثلاثة غرف واثنتان من دورات مياه",
    x = "عدد غرف المعيشة",
    y = "متوسط السعر"
  ) +
  theme_minimal()
```

متوسط السعر بناءً على عدد غرف المعيشة لعدد ثلاثة غرف واثنتان من دورات مياه



```
filtered_data <- subset(room_summary, livings == 2 & wc == 2)
rooms_price_group <- aggregate(price_per_meter ~ rooms, data = filtered_data, FUN = mean)
ggplot(rooms_price_group, aes(x = factor(rooms), y = price_per_meter)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    title = "متوسط السعر بناءً على عدد الغرف مع غرفتين معيشة واثنتان من دورات مياه",
    x = "عدد الغرف",
    y = "متوسط السعر"
  ) +
  theme_minimal()
```

متوسط السعر بناء على عدد الغرف مع غرفتين معيشة واثنتان من دورات مياه



- After analyzing the plots, we concluded that certain amenities have varying impacts on property prices. The data reveals that having zero to two living rooms does not significantly affect the price, whereas properties with three or more living rooms tend to command higher prices. Additionally, it is noticeable that the number of bathrooms plays a significant role in shaping tenant preferences. Renters tend to favor properties where the number of bathrooms matches or exceeds the number of bedrooms, reflecting a preference for convenience and privacy. This suggests that ensuring an adequate number of bathrooms can enhance the appeal and value of a property. Interestingly, properties with fewer rooms tend to be more expensive, indicating a potential premium on compact, well-designed spaces.

EDA

part one

Which city - region - neighborhood is the most and least expensive?

This question focuses on identifying the areas with the highest and lowest property prices. By analyzing rental prices across various cities, regions, and neighborhoods, we aim to pinpoint specific locations that are the most and least expensive. This information will be valuable for investors, renters, and market analysts, offering insights into price distribution and helping in making informed real estate decisions.

```

df1 <- read.csv("akar2024.csv")
cityprice <- df1 %>% select(city, price, rent_period, rooms, category)
cityprice <- cityprice %>% filter(rooms == 5, rent_period == 3, category == 1)
cityprice <- cityprice %>%
  filter(city %in% ("الرِّبَاط", "جدة", "الدمام", "الخبر", "المنورة"))
cityprice <- cityprice %>%
  select(-rent_period, -rooms, -category)

riyadh <- cityprice %>% filter(city == "الرِّيَاض")
Q1 <- quantile(riyadh$price, 0.25)
Q3 <- quantile(riyadh$price, 0.75)
IQR <- Q3 - Q1
riyadh <- riyadh %>%
  filter(price >= (Q1 - 1.5 * IQR) & price <= (Q3 + 1.5 * IQR))
riyadh <- riyadh %>% filter(price <= 700000)

jeddah <- cityprice %>% filter(city == "جدة")
jeddah <- jeddah %>% filter(price <= 800000)

dammam <- cityprice %>% filter(city == "الدَّمَم")
dammam <- dammam %>% mutate(z_score = (price - mean(price)) / sd(price))
dammam <- dammam %>% filter(abs(z_score) <= 3) %>% select(-z_score)

khobar <- cityprice %>% filter(city == "الْخَبْر")
Q1 <- quantile(khobar$price, 0.25)
Q3 <- quantile(khobar$price, 0.75)
IQR <- Q3 - Q1
khobar <- khobar %>%
  filter(price >= (Q1 - 1.5 * IQR) & price <= (Q3 + 1.5 * IQR))

makkah <- cityprice %>% filter(city == "مَكْهُوكَةُ الْمَكْرَمَة")
Q1 <- quantile(makkah$price, 0.25)
Q3 <- quantile(makkah$price, 0.75)
IQR <- Q3 - Q1
makkah <- makkah %>%
  filter(price >= (Q1 - 1.5 * IQR) & price <= (Q3 + 1.5 * IQR))

madinah <- cityprice %>% filter(city == "الْمَدِينَةُ الْمُنُورَة")
Q1 <- quantile(madinah$price, 0.25)
Q3 <- quantile(madinah$price, 0.75)
IQR <- Q3 - Q1
madinah <- madinah %>%
  filter(price >= (Q1 - 1.5 * IQR) & price <= (Q3 + 1.5 * IQR))

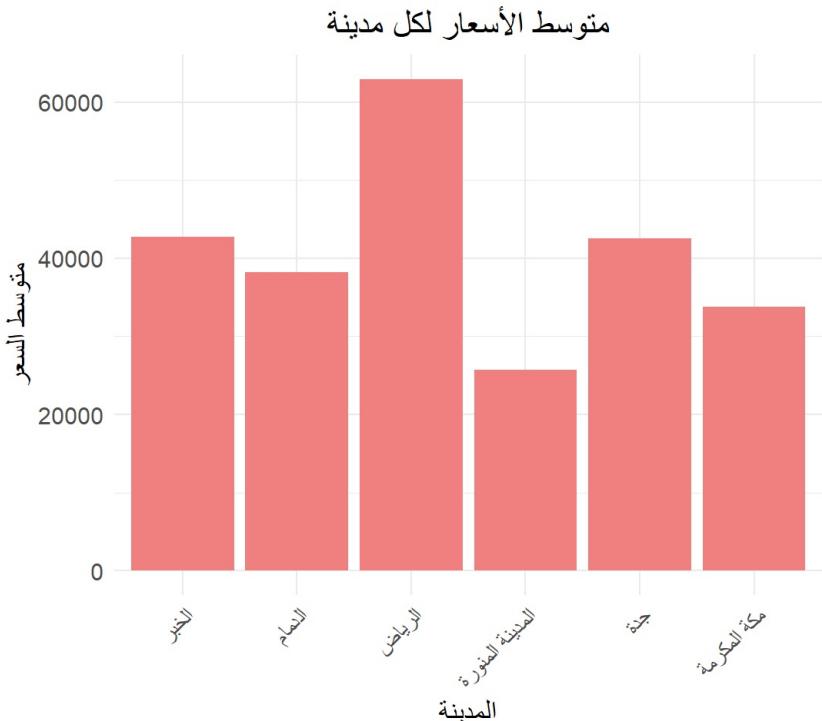
mean_price_riyadh <- mean(riyadh$price)
mean_price_jeddah <- mean(jeddah$price)
mean_price_dammam <- mean(dammam$price)
mean_price_khobar <- mean(khobar$price)
mean_price_makkah <- mean(makkah$price)
mean_price_madinah <- mean(madinah$price)

#combine
citypricel <- rbind(
  data.frame(city = "الرِّيَاض", price = riyadh$price),
  data.frame(city = "جدة", price = jeddah$price),
  data.frame(city = "الدَّمَم", price = dammam$price),
  data.frame(city = "الْخَبْر", price = khobar$price),
  data.frame(city = "مَكْهُوكَةُ الْمَكْرَمَة", price = makkah$price),
  data.frame(city = "الْمَدِينَةُ الْمُنُورَة", price = madinah$price)
)

average_price <- citypricel %>%
  group_by(city) %>%
  summarize(price = mean(price)) %>%
  mutate(price = as.integer(price))

ggplot(average_price, aes(x = city, y = price, fill = city)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = rep("lightcoral", length(unique(average_price$city)))) +
  labs(title = "متوسط السعر في كل مدينة", x = "المدينة", y = "متوسط السعر") +
  theme_minimal() +
  theme(
    text = element_text(family = "Arial", size = 14),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5)
  )

```



Average Price by Region in Riyadh

This metric provides the average rental price across different regions in Riyadh, allowing for a clearer understanding of price variations within the city. By calculating the average prices for each region—such as North, East, West, South, and Central Riyadh—we can identify which areas are more affordable or more expensive, helping stakeholders make informed choices based on regional price trends.

```

df2=read.csv("akar2024.csv")

riyadh1 <- df2 %>%
  filter(city == "الرياض") %>%
  select(city, price, district, direction, rent_period, rooms, category)

riyadh1 <- riyadh1[riyadh1$rent_period == 3.0, ]
riyadh1 <- riyadh1[riyadh1$category == 1, ]
riyadh1 <- riyadh1[riyadh1$rooms == 3.0, ]

northriyadh <- riyadh1 %>% filter(direction == "شمال الرياض")
eastriyadh <- riyadh1 %>% filter(direction == "شرق الرياض")
Westriyadh <- riyadh1 %>% filter(direction == "غرب الرياض")
Southriyadh <- riyadh1 %>% filter(direction == "جنوب الرياض")
Centralriyadh <- riyadh1 %>% filter(direction == "وسط الرياض")

Centralriyadh <- Centralriyadh[, !(colnames(Centralriyadh) %in% c('rent_period', 'city', 'rooms'))]
eastriyadh <- eastriyadh[, !(colnames(eastriyadh) %in% c('rent_period', 'city', 'rooms'))]
Westriyadh <- Westriyadh[, !(colnames(Westriyadh) %in% c('rent_period', 'city', 'rooms'))]
Southriyadh <- Southriyadh[, !(colnames(Southriyadh) %in% c('rent_period', 'city', 'rooms'))]
northriyadh <- northriyadh[, !(colnames(northriyadh) %in% c('rent_period', 'city', 'rooms'))]

library(dplyr)
library(ggplot2)

Westriyadh <- riyadh1 %>%
  filter(direction == "غرب الرياض" & price <= 50000 & price >= 20000)

Centralriyadh <- riyadh1 %>%
  filter(direction == "وسط الرياض" & price <= 70000)

eastriyadh <- riyadh1 %>%
  filter(direction == "شرق الرياض" & price <= 100000)

Southriyadh <- riyadh1 %>%
  filter(direction == "جنوب الرياض" & price <= 50000 & price >= 10000)

northriyadh <- riyadh1 %>%
  filter(direction == "شمال الرياض" & price <= 100000)

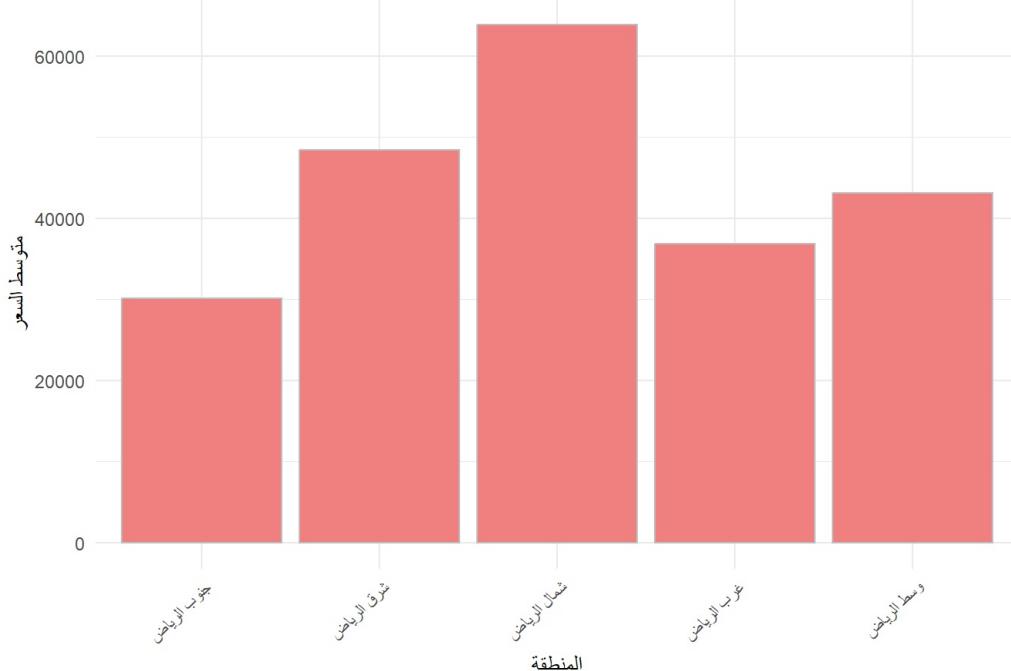
means <- mean(Centralriyadh$price, na.rm = TRUE)
متوسط_الوسط <- mean(eastriyadh$price, na.rm = TRUE)
متوسط_الشرق <- mean(Westriyadh$price, na.rm = TRUE)
متوسط_الغرب <- mean(Southriyadh$price, na.rm = TRUE)
متوسط_الجنوب <- mean(northriyadh$price, na.rm = TRUE)

dfmeans <- data.frame(
  وسط_الرياض = means,
  "شمال الرياض" = northriyadh$price,
  "شرق الرياض" = eastriyadh$price,
  "غرب الرياض" = Westriyadh$price,
  "جنوب الرياض" = Southriyadh$price
) = المنطقة

#plot
ggplot(dfmeans, aes(x = (متوسط_السعر + متوسط_السعر)/2, y = متوسط_السعر)) +
  geom_bar(stat = "identity", color = "gray", fill = "lightcoral") +
  labs(title = "متوسط سعر إيجار الشقة حسب المنطقة", subtitle = "المدينة = x, y = متوسط_السعر") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5, size = 14)
  )

```

متوسط سعر إيجار الشقة حسب المنطقة



Average Price in Each Neighborhood in Riyadh

This metric calculates the average rental price in each neighborhood across Riyadh. By analyzing neighborhood-level data, we can gain detailed insights into price variations within the city. This allows us to identify specific neighborhoods that are either more affordable or more expensive, supporting renters, investors, and analysts in making well-informed decisions based on local price trends.

```

Westriyadhmeans <- Westriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

Centralriyadhmeans <- Centralriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

Eastriyadhmeans <- eastriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

Southriyadhmeans <- Southriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

Northriyadhmeans <- northriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

# Prepare the data for barplot (districts and mean prices)
mean_prices_west <- Westriyadhmeans$mean_price
districts_west <- Westriyadhmeans$district

mean_prices_central <- Centralriyadhmeans$mean_price
districts_central <- Centralriyadhmeans$district

mean_prices_east <- Eastriyadhmeans$mean_price
districts_east <- Eastriyadhmeans$district

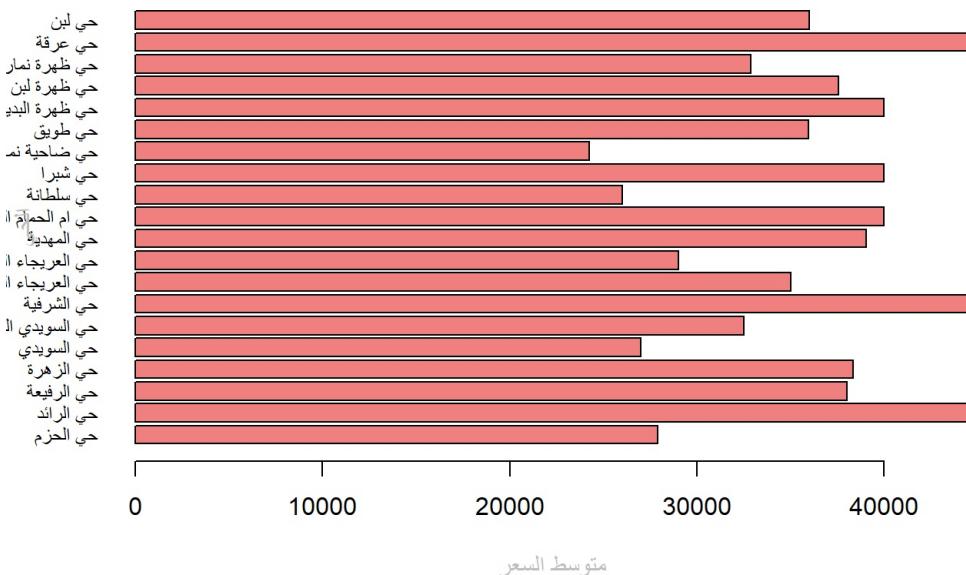
mean_prices_south <- Southriyadhmeans$mean_price
districts_south <- Southriyadhmeans$district

mean_prices_north <- Northriyadhmeans$mean_price
districts_north <- Northriyadhmeans$district
options(scipen = 999)

# Create the barplots
barplot(mean_prices_west,
        names.arg = districts_west,
        col = "lightcoral",
        main = "متوسط السعر حسب الحي في غرب الرياض",
        xlab = "متوسط السعر",
        ylab = "الحي",
        las = 1,
        cex.names = 0.8,
        col.main = "gray",
        col.lab = "gray",
        horiz = TRUE)

```

متوسط السعر حسب الحي في غرب الرياض

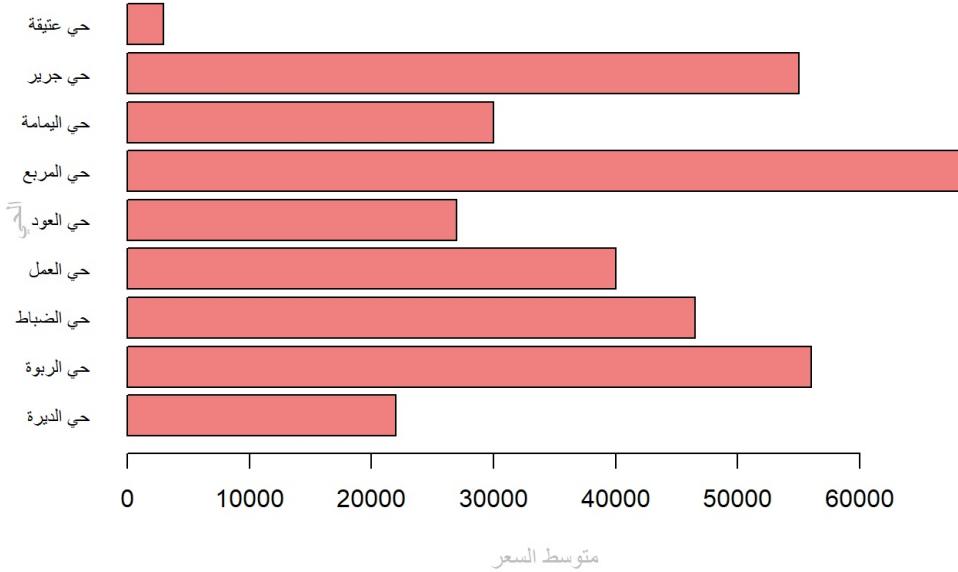


```

barplot(mean_prices_central,
        names.arg = districts_central,
        col = "lightcoral",
        main = "متوسط السعر حسب الحي في وسط الرياض",
        xlab = "متوسط السعر",
        ylab = "الحي",
        las = 1,
        cex.names = 0.8,
        col.main = "gray",
        col.lab = "gray",
        horiz = TRUE)

```

متوسط السعر حسب الحي في وسط الرياض

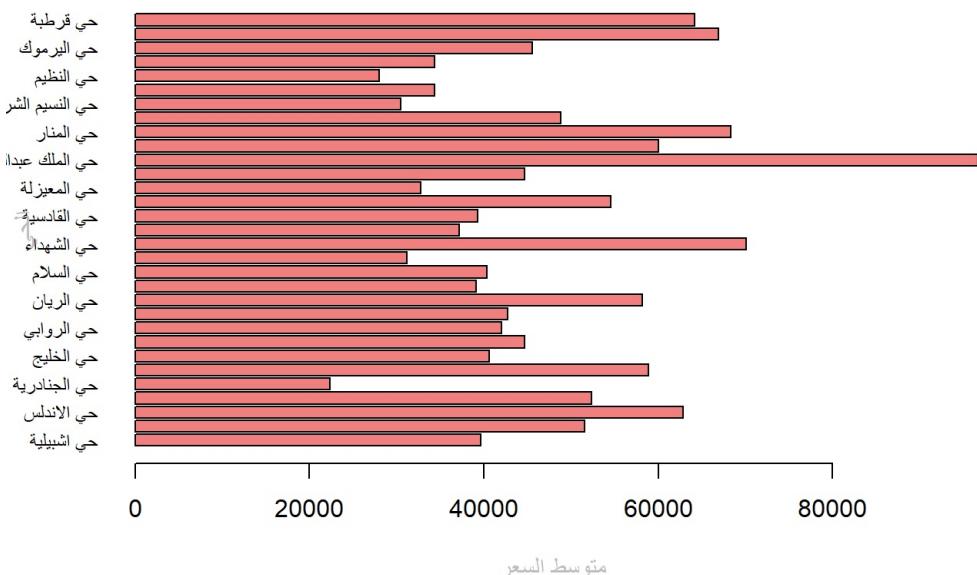


```

barplot(mean_prices_east,
        names.arg = districts_east,
        col = "lightcoral",
        main = "متوسط السعر حسب الحي في شرق الرياض",
        xlab = "متوسط السعر",
        ylab = "الحي",
        las = 1,
        cex.names = 0.8,
        col.main = "gray",
        col.lab = "gray",
        horiz = TRUE)

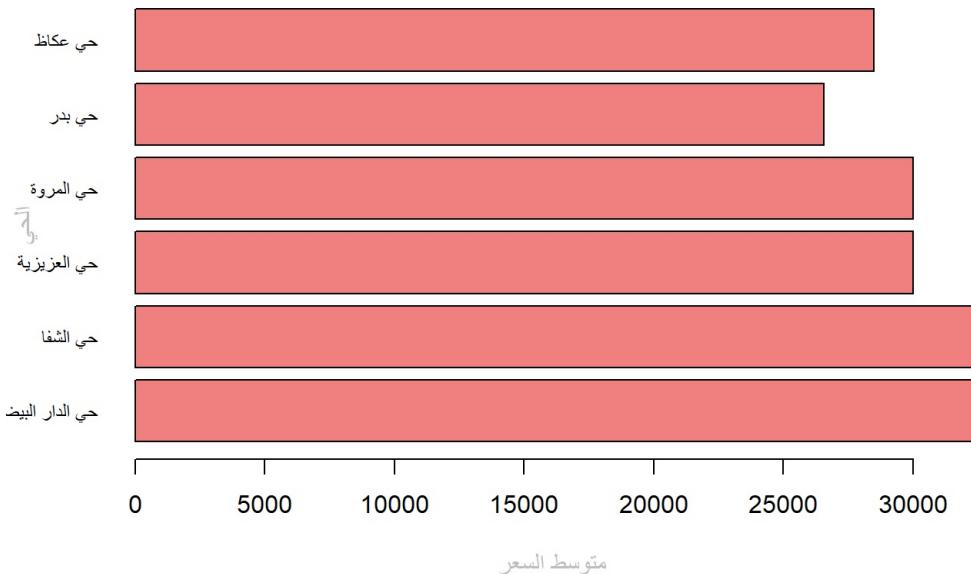
```

متوسط السعر حسب الحي في شرق الرياض



```
barplot(mean_prices_south,
       names.arg = districts_south,
       col = "lightcoral",
       main = "متوسط السعر حسب الحي في جنوب الرياض",
       xlab = "متوسط السعر",
       ylab = "الحي",
       las = 1,
       cex.names = 0.8,
       col.main = "gray",
       col.lab = "gray",
       horiz = TRUE)
```

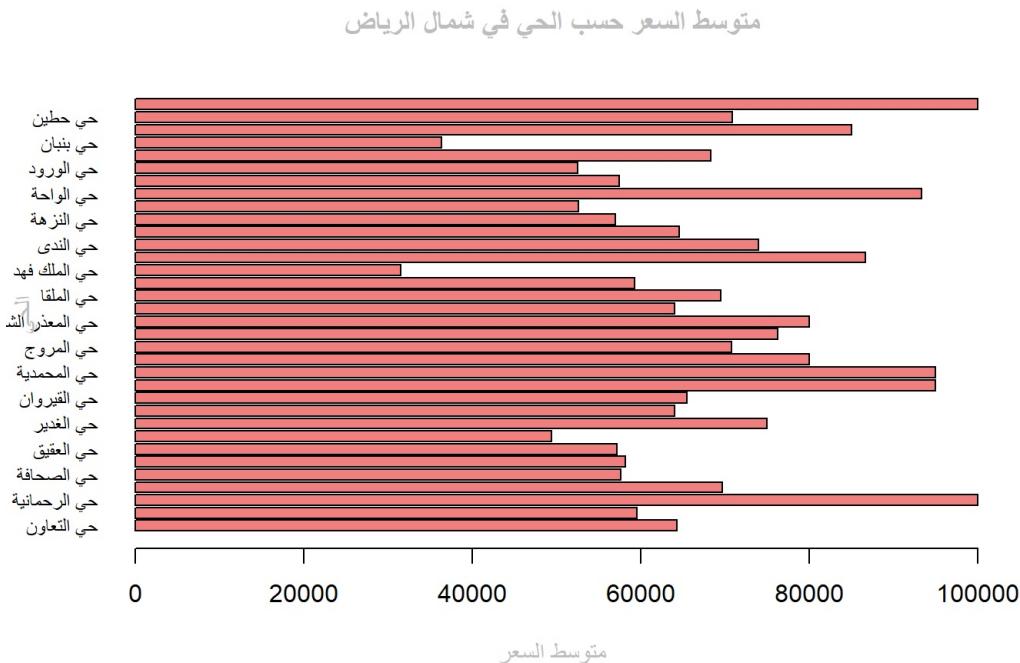
متوسط السعر حسب الحي في جنوب الرياض



```

barplot(mean_prices_north,
       names.arg = districts_north,
       col = "lightcoral",
       main = "متوسط السعر حسب الحي في شمال الرياض",
       xlab = "متوسط السعر",
       ylab = "الحي",
       las = 1,
       cex.names = 0.8,
       col.main = "gray",
       col.lab = "gray",
       horiz = TRUE)

```



The Most and Least Expensive Neighborhoods in Saudi Arabia

This analysis focuses on identifying the neighborhoods with the highest and lowest property prices across Saudi Arabia. By examining real estate data from various neighborhoods, we aim to highlight specific areas that are considered the most expensive and the most affordable. This information provides valuable insights for investors, renters, and market analysts, helping them make informed decisions based on neighborhood-level pricing trends throughout the country.

```

#all cities before filtering
#####library(reader)
data <- read.csv("akar2024.csv")
topcities=data %>%
  group_by(city)%>%
  summarise(count=n())%>%
  arrange(desc(count))
# top6 cities and districts with at least 10 records, the previous work is just to understand the data
topcities=filter(data,city%in% c("الرياض", "جدة", "الدمام", "الخبر", "المجمعة", "الظهران", "المنطقة الشرقية"))

price_summaryall <- topcities %>%
  group_by(city, district) %>%
  summarise(avg_price = mean(price, na.rm = TRUE), count = n()) %>%
  filter(count >= 10) %>% # Filter out districts with less than 10 records
  arrange(desc(avg_price))

# Get the top 10 and bottom 10 districts after filtering
top10all <- head(price_summaryall, 10)
top10all <- top10all %>%
  mutate(city_district = paste(city, district, sep = " - ")) # Combine city and district names

bottom10all <- tail(price_summaryall, 10)
bottom10all <- bottom10all %>%
  mutate(city_district = paste(city, district, sep = " - ")) # Combine city and district names

```

```

#cities after filtering
#take category by account
exp_dis <- function(data, catg, m_l = "m") {
  # Function to remove outliers based on IQR
  remove_outliers <- function(data, column) {
    Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
    Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
    IQR_value <- Q3 - Q1
    lower_bound <- Q1 - 1.5 * IQR_value
    upper_bound <- Q3 + 1.5 * IQR_value
    filtered_data <- data %>%
      filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
    return(filtered_data)
  }

  # Filter data for selected cities and category
  topcities <- filter(data, city %in% c("جدة", "الدمام", " الخبر", "الخبر", "المنطقة المقدمة للفحوصات", "المنطقة المقدمة للفحوصات")) %>%
    filter(category == catg)

  # Remove outliers from 'price' column
  topcities_no_outliers <- remove_outliers(topcities, "price")

  # Summarize data by city and district
  price_summaryall <- topcities_no_outliers %>%
    group_by(city, district) %>%
    summarise(avg_price = mean(price, na.rm = TRUE), count = n(), .groups = "drop") %>%
    filter(count >= 10) %>% # Filter out districts with less than 10 records
    arrange(desc(avg_price))

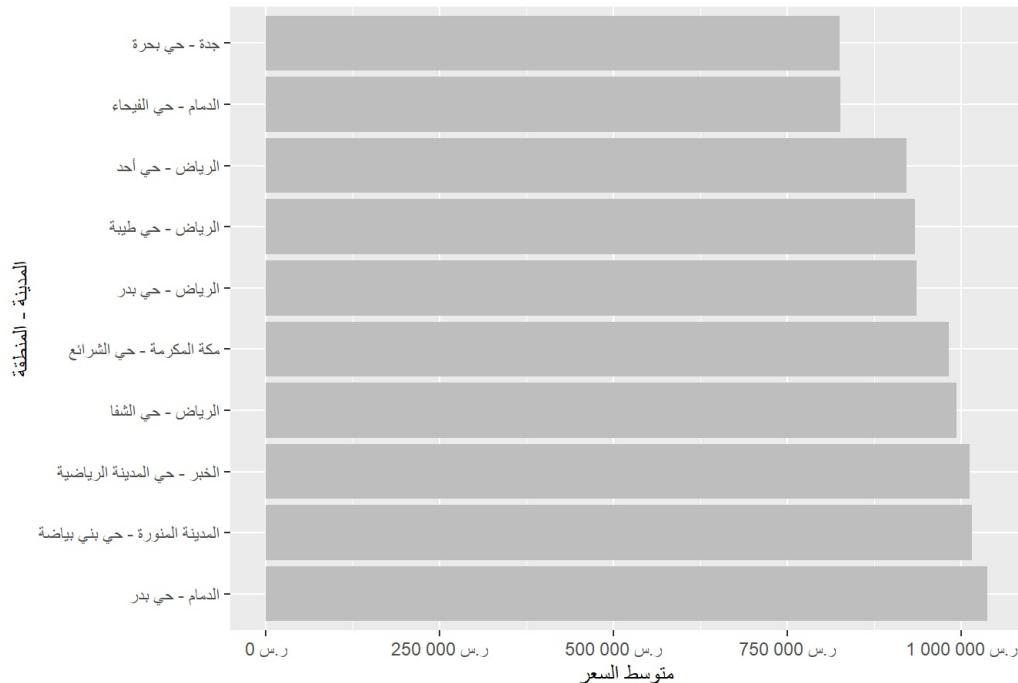
  # Get the top 10 and bottom 10 districts
  top10all <- head(price_summaryall, 10) %>%
    mutate(city_district = paste(city, district, sep = " - "))
  bottom10all <- tail(price_summaryall, 10) %>%
    mutate(city_district = paste(city, district, sep = " - "))

  # Plotting
  if (m_l == "m") {
    # Top 10 most expensive districts
    ggplot(top10all, aes(x = reorder(city_district, avg_price), y = avg_price)) +
      geom_bar(stat = "identity", fill = "lightcoral") +
      labs(title = "المدينة - المنطقه معايير 100%", y = "متوسط السعر") +
      coord_flip() +
      scale_y_continuous(labels = scales::label_number(scale = 1, suffix = " .ر.", decimal.mark = "."))
  } else if (m_l == "l") {
    # Bottom 10 least expensive districts
    ggplot(bottom10all, aes(x = reorder(city_district, -avg_price), y = avg_price)) +
      geom_bar(stat = "identity", fill = "gray") +
      labs(title = "المدينة - المنطقه معايير 100%", y = "متوسط السعر") +
      coord_flip() +
      scale_y_continuous(labels = scales::label_number(scale = 1, suffix = " .ر.", decimal.mark = "."))
  }
}

# Example usage:
exp_dis(data, 3, "l") # For least expensive

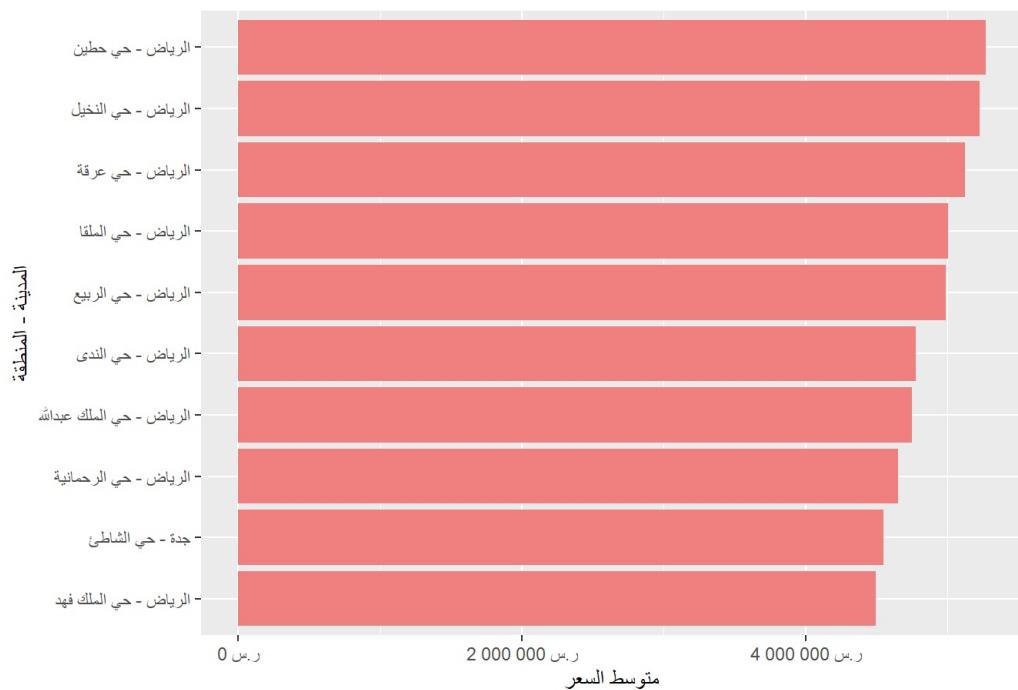
```

أقل 10 مناطق سكنية



```
exp_dis(data, 3, "m") # For most expensive
```

أعلى 10 مناطق سكنية



The Most and Least Expensive Neighborhoods in Riyadh

This analysis identifies the neighborhoods with the highest and lowest rental prices within the city of Riyadh. By examining rental data across various neighborhoods, we aim to pinpoint specific areas that stand out as the most expensive and the most affordable. These insights are valuable for investors, renters, and market analysts, offering a clear view of price distribution at the neighborhood level and assisting in making well-informed real estate decisions in Riyadh.

```

#riyadh
data <- read_csv("akar2024.csv", locale = locale(encoding = "UTF-8"))
topcities_riyadh <- filter(data, city == "الرياض")

# Function to remove outliers based on IQR
remove_outliers <- function(data, column) {
  Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
  Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
  IQR_value <- IQR(data[[column]], na.rm = TRUE)
  lower_bound <- Q1 - 1.5 * IQR_value
  upper_bound <- Q3 + 1.5 * IQR_value
  filtered_data <- data %>% filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
  return(filtered_data)
}

# Remove outliers from 'price' column for Riyadh
topcities_riyadh_no_outliers <- remove_outliers(topcities_riyadh, "price")

# Summarize data by city and district after removing outliers
price_summary_riyadh <- topcities_riyadh_no_outliers %>%
  group_by(city, district) %>%
  summarise(avg_price = mean(price, na.rm = TRUE), count = n()) %>%
  filter(count >= 10) %>% # Filter out districts with less than 10 records
  arrange(desc(avg_price))

# Get the top 10 and bottom 10 districts in Riyadh after filtering
top10_riyadh <- head(price_summary_riyadh, 10)
top10_riyadh <- top10_riyadh %>%
  mutate(city_district = paste(city, district, sep = " - ")) # Combine city and district names

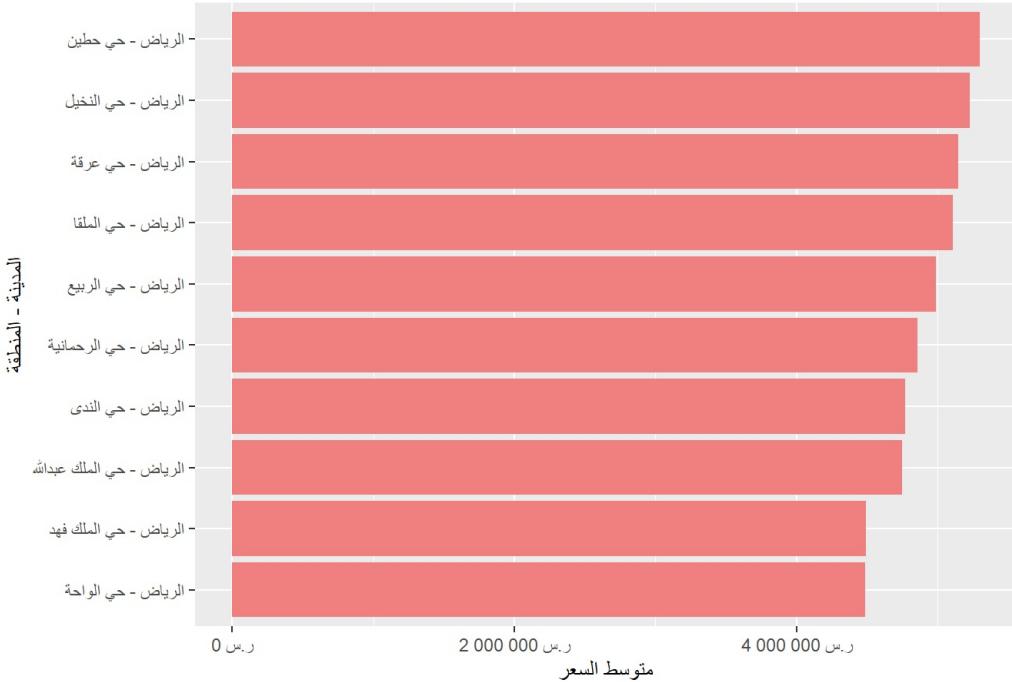
bottom10_riyadh <- tail(price_summary_riyadh, 10)
bottom10_riyadh <- bottom10_riyadh %>%
  mutate(city_district = paste(city, district, sep = " - ")) # Combine city and district names

topcities_riyadh <- filter(data, city == "الرياض")

exp_dis(topcities_riyadh, 3, "m")

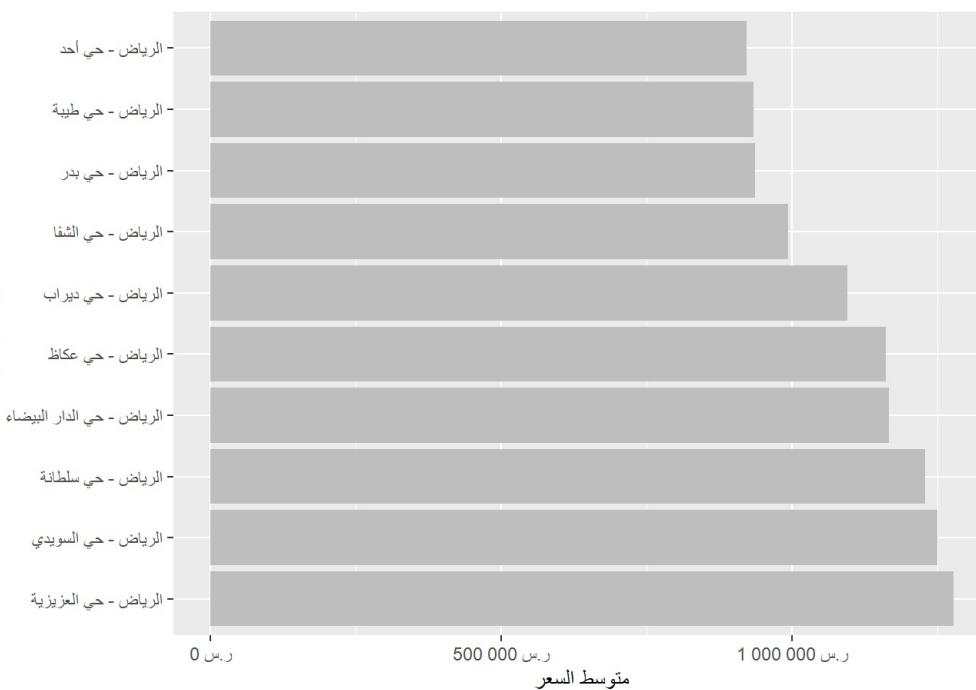
```

أعلى 10 مناطق سكنية



```
exp_dis(topcities_riyadh, 3, "l")
```

أقل 10 مناطق سكنية



Which cities have the highest and lowest prices for apartments for sale?

We will examine the range of apartment prices across various cities, focusing on identifying the cities with the highest and lowest price averages. By ranking the cities from the most expensive to the least, we can assess patterns and trends in the real estate market and pinpoint regions with notable price extremes. This analysis will be helpful for potential buyers, investors, and real estate market analysts looking to understand which areas have premium pricing and which offer more affordable options.

```
#function to remove outliers
remove_outliers <- function(data, columns) {
  for (column in columns) {
    Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
    Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
    IQR_value <- Q3 - Q1
    lower_bound <- Q1 - 1.5 * IQR_value
    upper_bound <- Q3 + 1.5 * IQR_value
    data <- data %>%
      filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
  }
  return(data)
}
akar2 <- read_csv("akar2024.csv")

akar2_no_outliers <- remove_outliers(akar2,c("price","age") )
summary(akar2_no_outliers$price)
```

```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##        1    87000  590000  791383 1100000 3510000
```

```
summary(akar2$price)
```

```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##        1    160000  670000  2052049 1500000 4900000000
```

```
city_count <- akar2_no_outliers %>%
  group_by(city) %>%
  summarize(count = n())
```

```
akar_floor <- akar2_no_outliers %>%
  filter(category == "6")
summary(akar_floor$price)
```

```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##        275    480000  600000  652007  750000 3500000
```

```
akar_floor_outliers <- remove_outliers(akar_floor,c( "price","age"))
summary(akar_floor_outliers$price)
```

```
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## 90000 470000 580000 595994 700000 1155000
```

```
city_count2 <- akar_floor_outliers %>%
  group_by(city) %>%
  summarize(count = n())

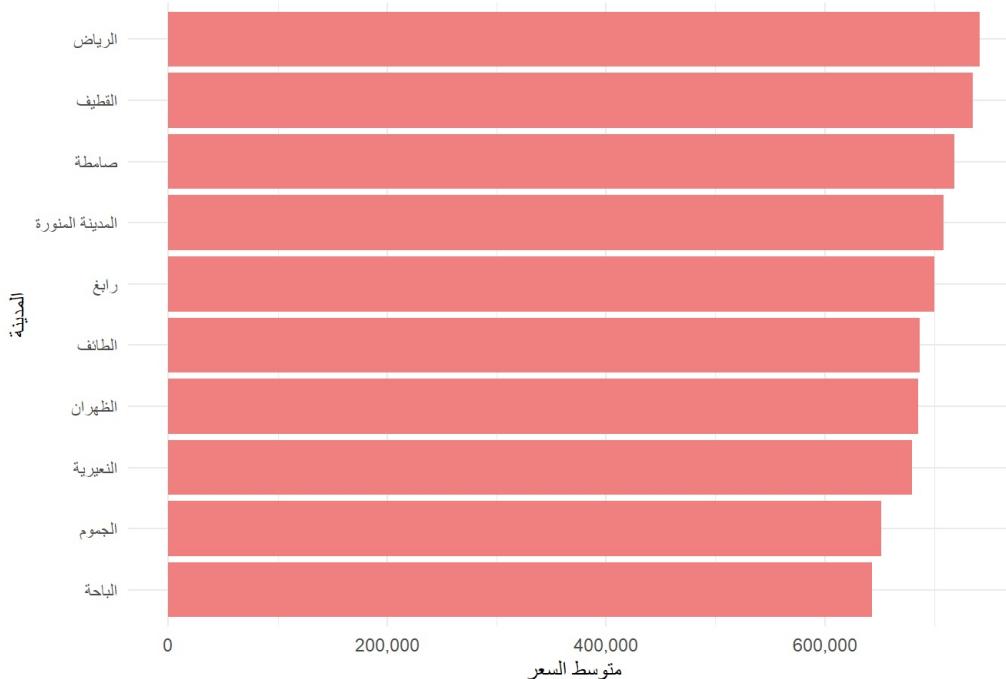
akar_floor_outliers%>%
  group_by(city) %>%
  summarize(count = n(), average_price = mean(price, na.rm = TRUE)) %>%
  arrange(desc(average_price))
```

```
## # A tibble: 41 x 3
##   city           count average_price
##   <chr>        <int>      <dbl>
## 1 741649        1909      ١٩٠٩
## 2 735000          4      ٤٠٠٠
## 3 718571          7      ٧٠٠٠
## 4 708452        293      ٢٩٣
## 5 700000          1      ١٠٠٠
## 6 686502        287      ٢٨٧
## 7 685353        51      ٥١
## 8 680000          1      ١٠٠٠
## 9 651667          6      ٦٠٠٠
## 10 643333         3      ٣٠٠٠
## # ... i 31 more rows
```

```
top_10_cities <- akar_floor_outliers %>%
  filter(category == "6") %>%
  group_by(city) %>%
  summarize(
    average_price = mean(price, na.rm = TRUE)
  ) %>%
  arrange(desc(average_price)) %>%
  slice_head(n = 10)

ggplot(top_10_cities, aes(x = reorder(city, average_price), y = average_price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "مدن في متوسط أسعار الشقق لـ ٢٠١٤",
       x = "المدينة",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  coord_flip() +
  theme_minimal() +
  theme(
    text = element_text(family = "IBM"),
    legend.position = "none"
  )
```

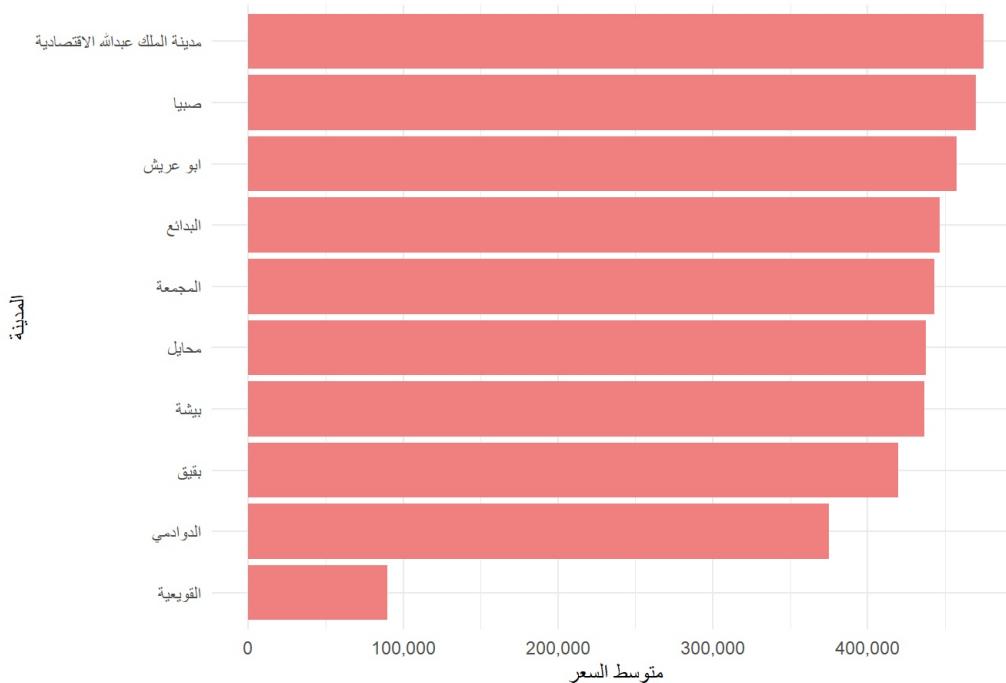
أكثر 10 مدن في متوسط أسعار الشقق للبيع



```
bottom_10_cities <- akar_floor_outliers %>%
  filter(category == "6") %>%
  group_by(city) %>%
  summarize(
    average_price = mean(price, na.rm = TRUE)
  ) %>%
  arrange(average_price) %>%
  slice_head(n = 10)

ggplot(bottom_10_cities, aes(x = reorder(city, average_price), y = average_price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "مدن في متوسط أسعار أقل لـ 10 مدن",
       x = "المدينة",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  coord_flip() +
  theme_minimal() +
  theme(
    text = element_text(family = "IBM"),
    legend.position = "none"
  )
```

أقل 10 مدن في متوسط أسعار الشقق



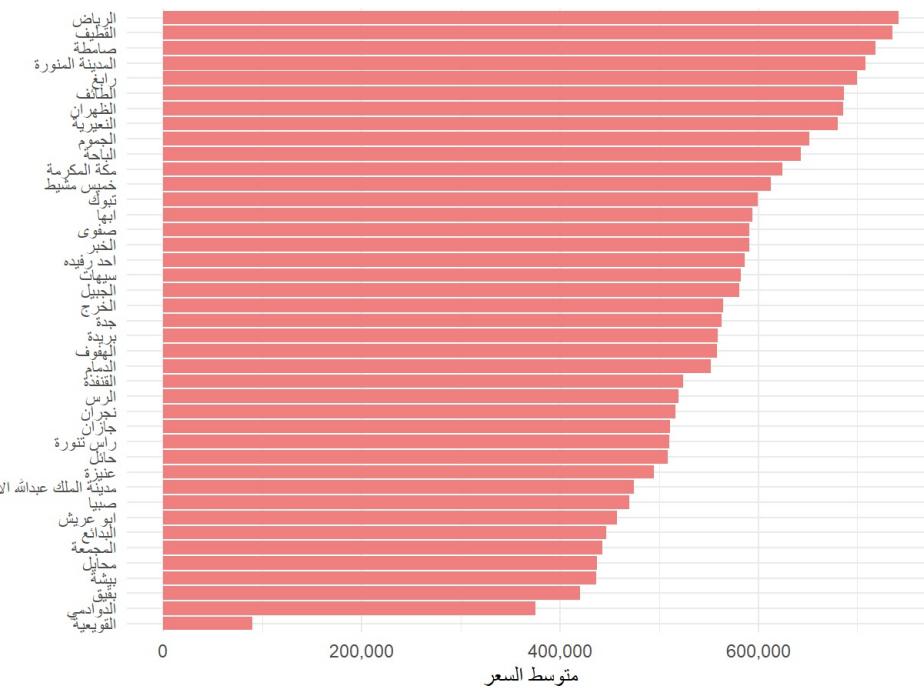
```

avg_city <- akar_floor_outliers %>%
  filter(category == "6") %>%
  group_by(city) %>%
  summarize(
    average_price = mean(price, na.rm = TRUE)
  ) %>%
  arrange(average_price)

ggplot(avg_city, aes(x = reorder(city, average_price), y = average_price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "مقارنة متوسط سعر الشقة للبيع حسب المدينة",
       x = "المدينة",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  coord_flip() +
  theme_minimal() +
  theme(text = element_text(family = "IBM"))

```

مقارنة متوسط سعر الشقة للبيع حسب المدينة



What is the average sale price of apartments in cities across the Kingdom? This analysis aims to determine the average prices of apartments for sale in various cities within the Kingdom. By calculating these averages, we can gain insights into the real estate market, understand regional pricing trends, and identify areas with comparatively higher or lower apartment prices.

```

akar_r_P<- akar2_no_outliers %>%
  filter(city == "الرياض" & category == "6") %>%
  select(city, category, district, direction,price)

akar_r_p_c<- remove_outliers(akar_r_P, "price" )
summary(akar_r_p_c$price)

```

```

##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   360  650000  820000  892075 1125000 1900000

```

```

district_price <- akar_r_p_c %>%
  group_by(district) %>%
  summarize(average_price = mean(price, na.rm = TRUE))

str(district_price)

```

```

## #tibble [87 x 2] (S3:tbl_df/tbl/data.frame)
## $ district      : chr [1:87] "حي اشبيلية" "حي الازادهار" "حي الاندلس" ...
## $ average_price: num [1:87] 530000 887458 1212500 947500 1280933 ...

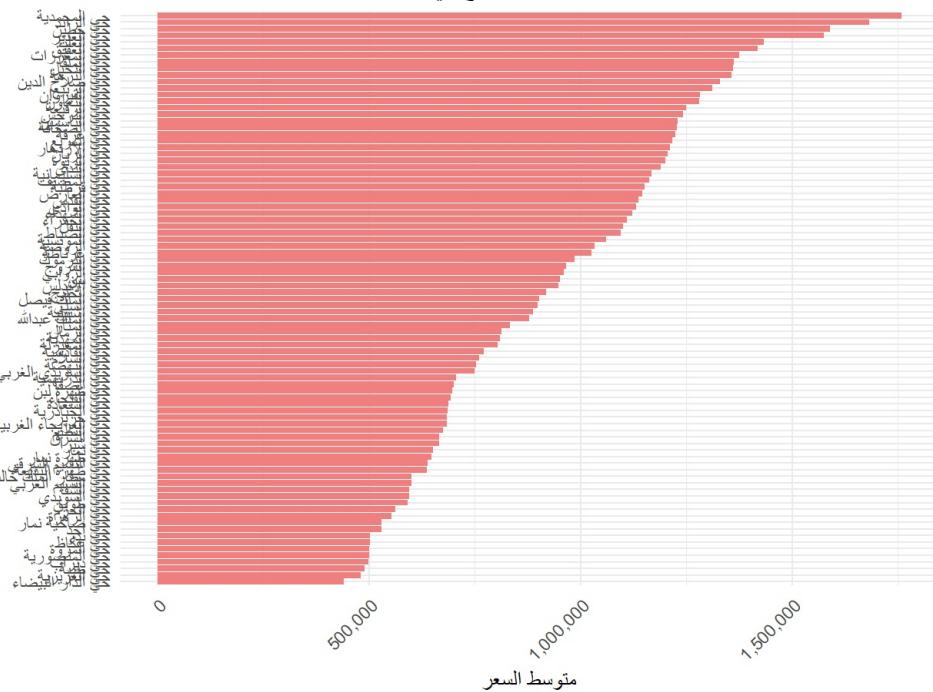
```

```

ggplot(district_price, aes(x = reorder(district, average_price), y = average_price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "شقق للبيع في احياء الرياض مقارنة متوسط السعر",
       x = "الحي",
       y = "متوسط السعر") +
  scale_y_continuous(labels = scales::comma) +
  coord_flip() +
  theme_minimal() +
  theme(text = element_text(family = "IBM"),
        axis.text.x = element_text(angle = 45, hjust = 1))

```

شقق للبيع في احياء الرياض مقارنة متوسط السعر

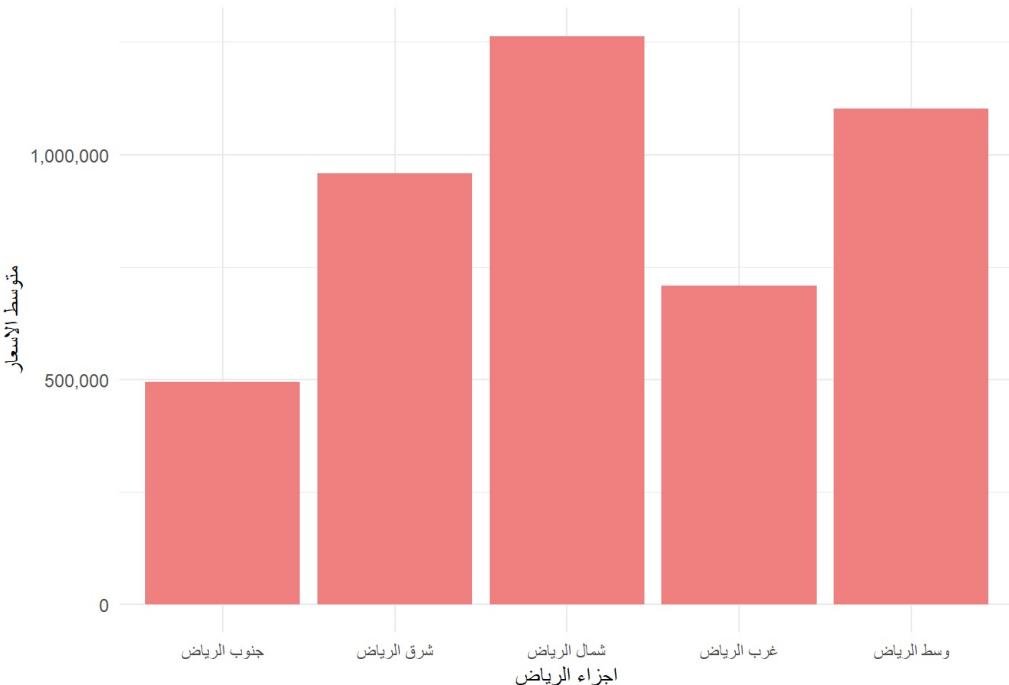


```

direction_price <- akar_r_p_c %>%
  group_by(direction) %>%
  summarize(average_price = mean(price, na.rm = TRUE))
ggplot(direction_price, aes(x = direction, y = average_price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "مقارنة بين مناطق الرياض ومتوسط الاسعار",
       x = "اجزاء الرياض",
       y = "متوسط الاسعار") +
  scale_y_continuous(labels = scales::comma) +
  theme_minimal() +
  theme(text = element_text(family = "IBM"))

```

مقارنة بين مناطق الرياض ومتوسط الاسعار



```

akar_r_d<- akar2_no_outliers %>%
  filter(city == "مكة المكرمة" & category == "6") %>%
  select(city, category, district, price)
#View(akar_r_d)

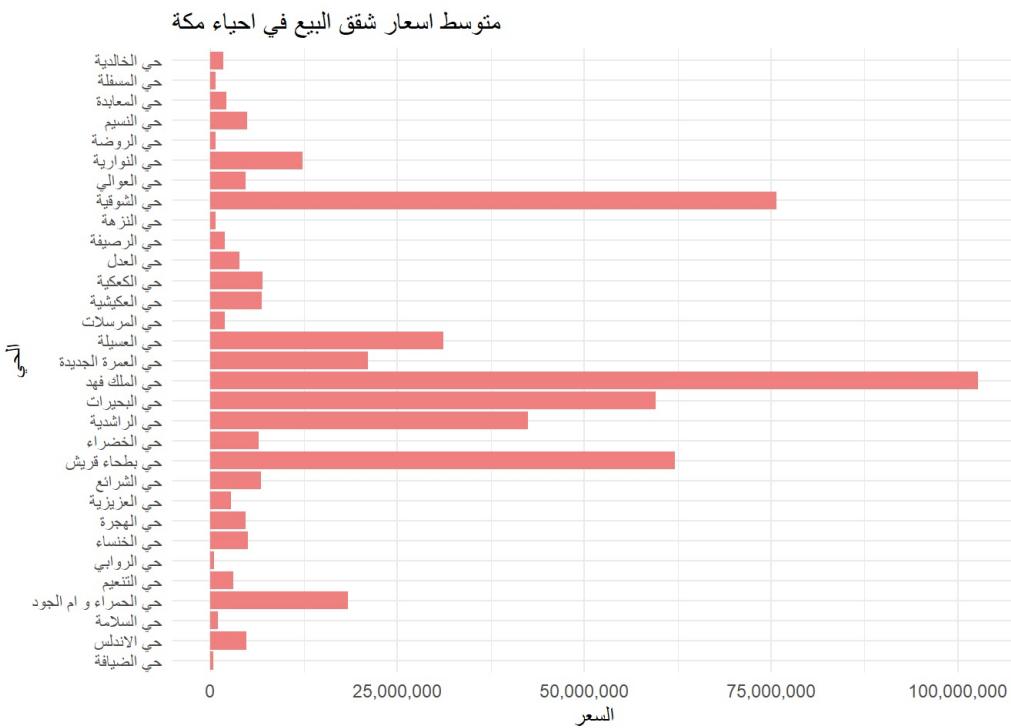
```

```
akar_r_d_c<- remove_outliers(akar_r_d,"price" )
```

```

ggplot(akar_r_d_c, aes(x = reorder(district, price), y = price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "متوسط اسعار شقق البيع في احياء مكة",
       x = "الحي",
       y = "السعر") +
  scale_y_continuous(labels = scales::comma) +
  coord_flip() +
  theme_minimal() +
  theme(text = element_text(family = "IBM"),
        legend.position = "none")

```



```

akar_r_dd<- akar2_no_outliers %>%
  filter(city == "الدمام" & category == "6") %>%
  select(city, category, district, price)
#View(akar_r_dd)

```

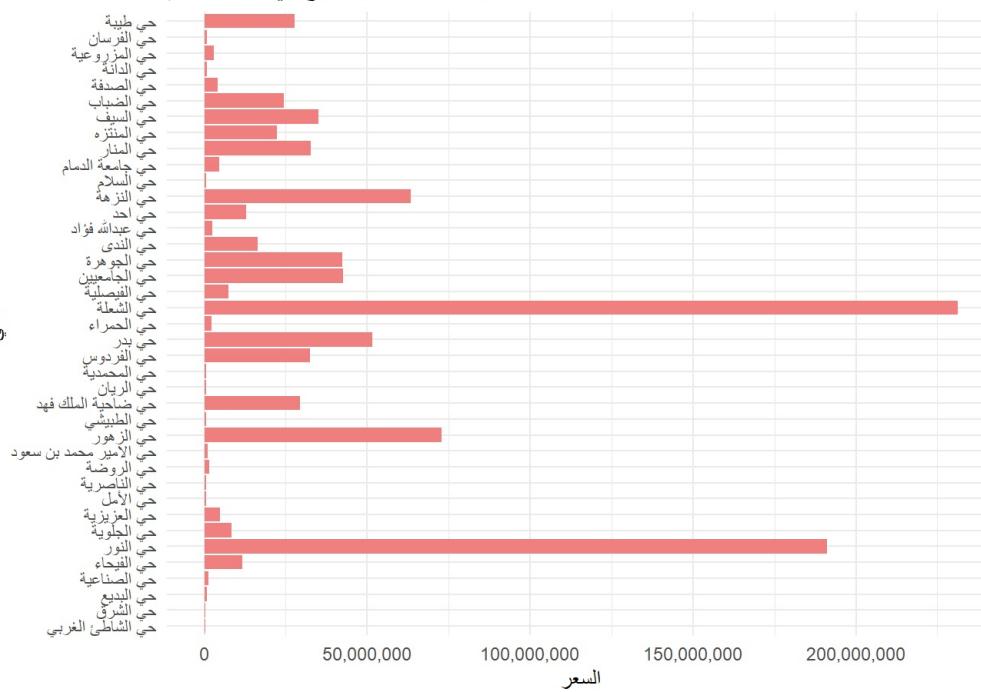
```
akar_r_ddc<- remove_outliers(akar_r_dd, "price" )
```

```

ggplot(akar_r_ddc, aes(x = reorder(district, price), y = price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "متوسط اسعار شقق البيع في احياء الدمام",
       x = "الحي",
       y = "السعر") +
  scale_y_continuous(labels = scales::comma) +
  coord_flip() +
  theme_minimal() +
  theme(text = element_text(family = "IBM"),
        legend.position = "none")

```

متوسط اسعار شقق البيع في احياء الدمام



part two

Is Commercial or Residential Property More Expensive? Generally, commercial properties tend to be more expensive than residential ones, largely due to their potential for generating income. However, prices vary significantly based on the location, property size, and demand in each area. In Riyadh, the analysis can consider average rental prices in various districts to offer insights into the differences.

```
df4=read_csv("akar2024.csv")
df3 <- df4[, c('category', 'price', 'rent_period', 'city', 'district', 'direction', 'rooms', 'area')]
values_to_remove <- c(3, 6, 11, 18, 21, 22, 13, 17, 20, 15, 5, 2, 19, 16, 7, 10, 12, 4, 9, 23)
df3 <- df3[!df3$category %in% values_to_remove, ]

df3 <- df3[df3$city == 'الرياض' ]
df3 <- df3[df3$rent_period == 3.0, ]
df3 <- df3[df3$category %in% c(1, 8, 14), ]
print(df3)
```

```
## # A tibble: 9,188 × 8
##   category price rent_period city   district      direction rooms area
##   <dbl>   <dbl>     <dbl> <chr> <chr>        <chr>    <dbl> <dbl>
## 1       1   65000     3 150   شمال الرياض   حي العارض      3   150
## 2       1  135000     3 176   شمال الرياض   حي العقيق      4   176
## 3       1   75000     3 138   شمال الرياض   حي العارض      4   138
## 4       1  200000     3 150   شمال الرياض   حي المعذر الشمالي 3   150
## 5       1   75000     3 176   الرياض   حي الرمال      3   176
## 6       1   70000     3 114   شمال الرياض   حي النرجس      3   114
## 7       1  125000     3 150   شمال الرياض   حي الياسمين 3   150
## 8       1  130000     3 176   شمال الرياض   حي العارض      3   176
## 9       1   85000     3 150   شمال الرياض   حي الصحافة      3   150
## 10      1  110000     3 150   الرياض   حي الملك عبدالله 3   150
## # ... i 9,178 more rows
```

```
#remove
df3 <- df3[, !(names(df3) %in% c('rent_period', 'city', 'rooms'))]
print(df3)
```

```

## # A tibble: 9,188 × 5
##   category  price district      direction    area
##   <dbl>     <dbl> <chr>        <chr>       <dbl>
## 1 1       65000 150 شمال الرياض   حي العارض
## 2 1       135000 176 شمال الرياض   حي العقيق
## 3 1       75000 138 شمال الرياض   حي العارض
## 4 1       200000 274 حي المعدن الشمالي شمال الرياض حي العارض
## 5 1       75000 225 شرق الرياض   حي الرمال
## 6 1       70000 114 شمال الرياض   حي النرجس
## 7 1       125000 130 شمال الرياض   حي الياسمين
## 8 1       130000 210 شمال الرياض   حي العارض
## 9 1       85000 202 شمال الرياض   حي الصحافة
## 10 1      110000 149 حي الملك عبدالله شمال الرياض حي العارض
## # i 9,178 more rows

```

```

df3$type <- ifelse(df3$category == 1, 'سكنى', ifelse(df3$category %in% c(8, 14), 'جبل(جفروف', 'جبل(جفروف'
mean_by_type <- aggregate(price ~ type, data = df3, FUN = mean)
print(mean_by_type)

```

```

##   type     price
## 1 تجاري 355250.83
## 2 سكنى  59047.04

```

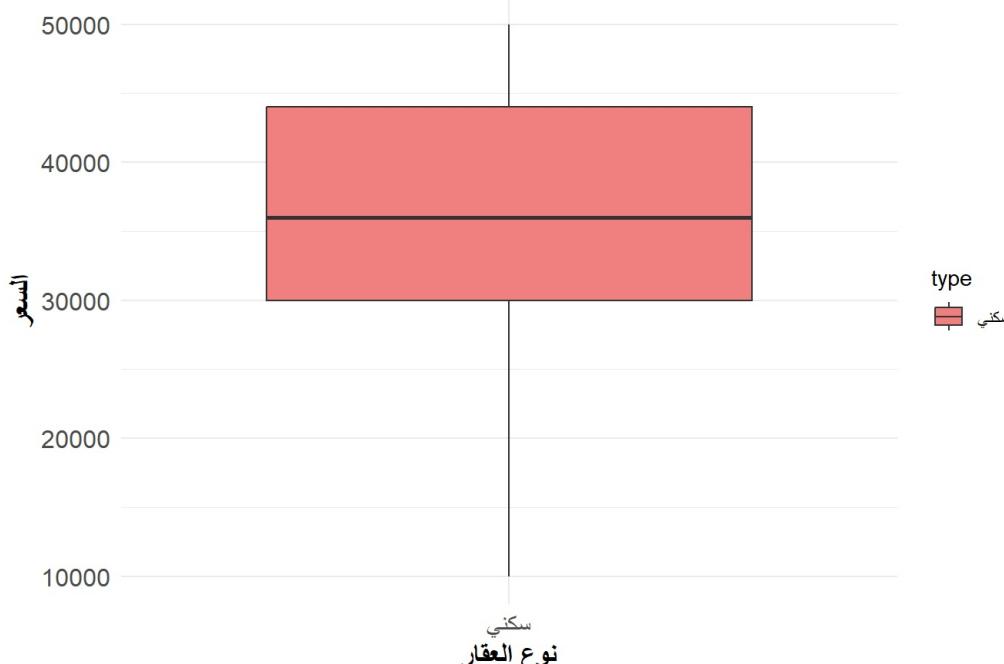
```

df3$type <- ifelse(df3$category == 1, 'سكنى', ifelse(df3$category %in% c(8, 14), 'جبل(جفروف', 'جبل(جفروف'
skeny_data <- df3[df3$type == 'سكنى', ]
Q1 <- quantile(skeny_data$price, 0.25)
Q3 <- quantile(skeny_data$price, 0.75)
IQR <- Q3 - Q1
lower_bound <- Q1 - 1.5 * IQR
upper_bound <- Q3 + 1.5 * IQR
skeny_data_filtered <- skeny_data[skeny_data$price >= lower_bound & skeny_data$price <= upper_bound, ]
skeny_data_filtered <- skeny_data_filtered[skeny_data_filtered$price >= 10000 & skeny_data_filtered$price <= 5000
0, ]
library(ggplot2)

ggplot(skeny_data_filtered, aes(x = type, y = price, fill = type)) +
  geom_boxplot() +
  scale_fill_manual(values = c("سكنى" = "lightcoral")) +
  labs(
    title = "توزيع الأسعار للعقارات السكنية",
    x = "نوع العقار",
    y = "السعر"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12)
  )

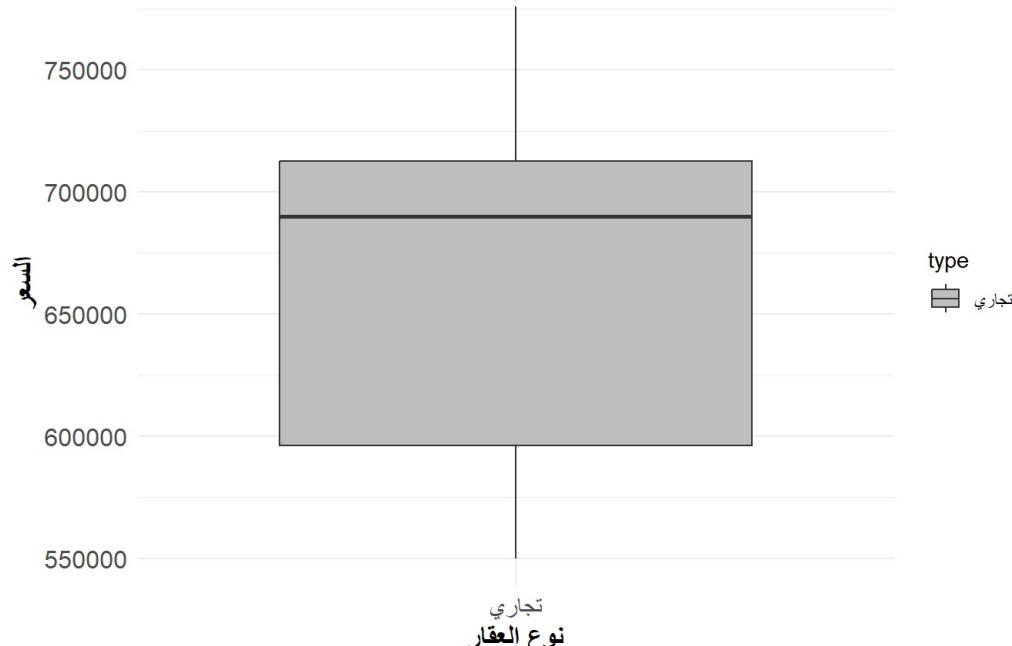
```

توزيع الأسعار للعقارات السكنية



```
df3$type <- ifelse(df3$category == 1, 'سكنى', ifelse(df3$category %in% c(8, 14), 'تجاري', 'غير مفروض'))  
tagari_data <- df3[df3$type == 'تجاري', ]  
Q1 <- quantile(tagari_data$price, 0.25)  
Q3 <- quantile(tagari_data$price, 0.75)  
IQR <- Q3 - Q1  
lower_bound <- Q1 - 1.5 * IQR  
upper_bound <- Q3 + 1.5 * IQR  
tagari_data_filtered <- tagari_data[tagari_data$price >= lower_bound & tagari_data$price <= upper_bound, ]  
manual_outliers <- tagari_data_filtered[tagari_data_filtered$price < 700000, ]  
tagari_data_filtered <- tagari_data_filtered[tagari_data_filtered$price >= 550000, ]  
  
library(ggplot2)  
  
ggplot(tagari_data_filtered, aes(x = type, y = price, fill = type)) +  
  geom_boxplot() +  
  scale_fill_manual(values = c("تجاري" = "gray")) +  
  labs(  
    title = "توزيع الأسعار للعقارات التجارية بدون قيم شاذة",  
    x = "نوع العقار",  
    y = "السعر")  
  ) +  
  theme_minimal() +  
  theme(  
    plot.title = element_text(size = 16, face = "bold"),  
    axis.title = element_text(size = 14, face = "bold"),  
    axis.text.x = element_text(size = 12),  
    axis.text.y = element_text(size = 12)  
  )
```

توزيع الأسعار للعقارات التجارية بدون قيم شاذة



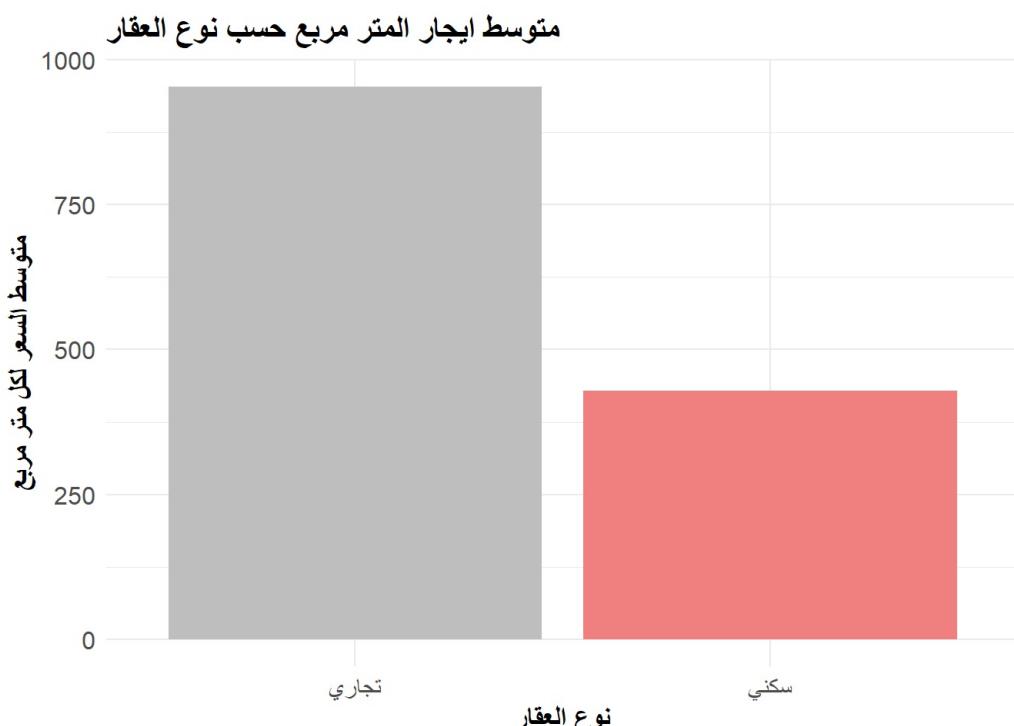
```
df3$type <- ifelse(df3$category == 1, 'سكنى', ifelse(df3$category %in% c(8, 14), 'تجاري', 'غير مفروض'))  
skeny_data <- df3[df3$type == 'سكنى', ]  
  
Q1_skeny <- quantile(skeny_data$price, 0.25)  
Q3_skeny <- quantile(skeny_data$price, 0.75)  
IQR_skeny <- Q3_skeny - Q1_skeny  
lower_bound_skeny <- Q1_skeny - 1.5 * IQR_skeny  
upper_bound_skeny <- Q3_skeny + 1.5 * IQR_skeny  
skeny_data_filtered <- skeny_data[skeny_data$price >= lower_bound_skeny & skeny_data$price <= upper_bound_skeny, ]  
tagari_data <- df3[df3$type == 'تجاري', ]  
Q1_tagari <- quantile(tagari_data$price, 0.25)  
Q3_tagari <- quantile(tagari_data$price, 0.75)  
IQR_tagari <- Q3_tagari - Q1_tagari  
lower_bound_tagari <- Q1_tagari - 1.5 * IQR_tagari  
upper_bound_tagari <- Q3_tagari + 1.5 * IQR_tagari  
tagari_data_filtered <- tagari_data[tagari_data$price >= lower_bound_tagari & tagari_data$price <= upper_bound_tagari, ]  
mean_skeny <- mean(skeny_data_filtered$price)  
mean_tagari <- mean(tagari_data_filtered$price)  
mean_values <- c(mean_skeny, mean_tagari)  
types <- c("سكنى", "تجاري")  
  
library(ggplot2)  
  
ggplot(data.frame(type = types, mean_price = mean_values), aes(x = type, y = mean_price, fill = type)) +  
  geom_bar(stat = "identity", show.legend = FALSE) +  
  scale_fill_manual(values = c("سكنى" = "lightcoral", "تجاري" = "gray")) +  
  labs(  
    title = "المتوسط الحسابي للأسعار العقارية",  
    x = "نوع العقار",  
    y = "المتوسط الحسابي للسعر")  
  ) +  
  theme_minimal() +  
  theme(  
    plot.title = element_text(size = 16, face = "bold"),  
    axis.title = element_text(size = 14, face = "bold"),  
    axis.text.x = element_text(size = 12),  
    axis.text.y = element_text(size = 12)  
  )
```

المتوسط الحسابي لأسعار العقارات



```
df3$price_per_sqm <- df3$price / df3$area
df3$type <- ifelse(df3$category == 1, 'سكنى', ifelse(df3$category %in% c(8, 14), 'تجاري', 'غير مفروض'))
mean_price_per_sqm <- aggregate(price_per_sqm ~ type, data = df3, FUN = mean)

ggplot(mean_price_per_sqm, aes(x = type, y = price_per_sqm, fill = type)) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  scale_fill_manual(values = c("سكنى" = "lightcoral", "تجاري" = "gray")) +
  labs(
    title = "متوسط ايجار المتر مربع حسب نوع العقار",
    x = "نوع العقار",
    y = "متوسط السعر لكل متر مربع"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14, face = "bold"),
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12)
  )
```



How do the prices of offices, institutions, and real estate companies in Riyadh compare? This analysis compares the prices of offices, institutions, and real estate companies in Riyadh to understand pricing trends across different property types. By examining these price variations, we can identify which types of properties are more expensive or affordable in the city, helping investors and buyers make informed decisions.

```

akar2_r=filter(akar2_no_outliers,city == "الرياض")
akar2_rr <- remove_outliers(akar2_r, c("price"))
akar2_rr <- akar2_rr %>%
  mutate(company_type = case_when(
    grepl("مؤسسة", user.name, ignore.case = TRUE) ~ "مؤسسة",
    grepl("شركة", user.name, ignore.case = TRUE) ~ "شركة",
    grepl("مكتب", user.name, ignore.case = TRUE) ~ "مكتب",
    TRUE ~ NA_character_
  ))
#View(akar2_rr)
akar2_rr <- akar2_rr %>%
  filter(rent_period == "3" & category %in% c("1", "4","5"))

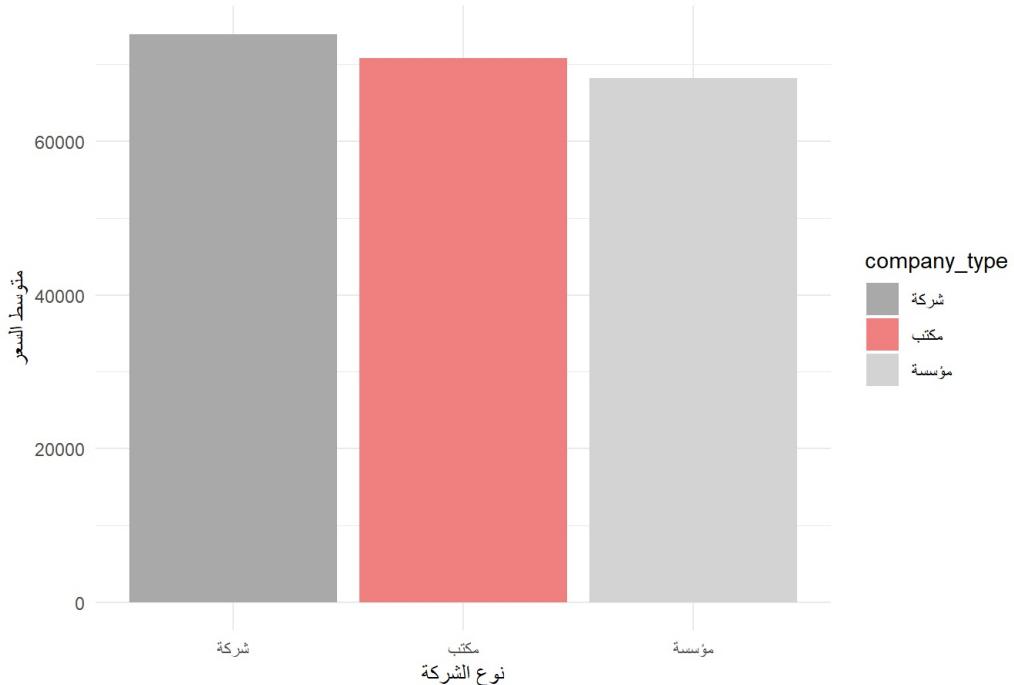
companies_with_type <- akar2_rr %>%
  filter(!is.na(company_type))
#View(companies_with_type)

average_price <- companies_with_type %>%
  group_by(company_type) %>%
  summarise(average_price = mean(price, na.rm = TRUE))

ggplot(average_price, aes(x = company_type, y = average_price, fill = company_type)) +
  geom_bar(stat = "identity") +
  labs(
    title = "مقارنة متوسط الأسعار للمؤسسة، الشركة، والمكتب",
    x = "نوع الشركة",
    y = "متوسط السعر"
  ) +
  theme_minimal() +
  theme(text = element_text(family = "IBM")) +
  scale_fill_manual(values = c("مؤسسة" = "lightgray", "شركة" = "darkgray", "مكتب" = "lightcoral"))

```

مقارنة متوسط الأسعار للمؤسسة، الشركة، والمكتب



What are the differences between leasing through a corporation and leasing from an individual? This analysis explores the differences between leasing through a corporation and leasing from an individual landlord. It examines factors such as lease terms, rental rates, property management, and flexibility. The goal is to understand how leasing practices vary based on the type of landlord and how these differences impact tenants in terms of costs, management, and overall leasing experience. By comparing both options, the analysis helps potential tenants make informed decisions about their leasing choices.

```

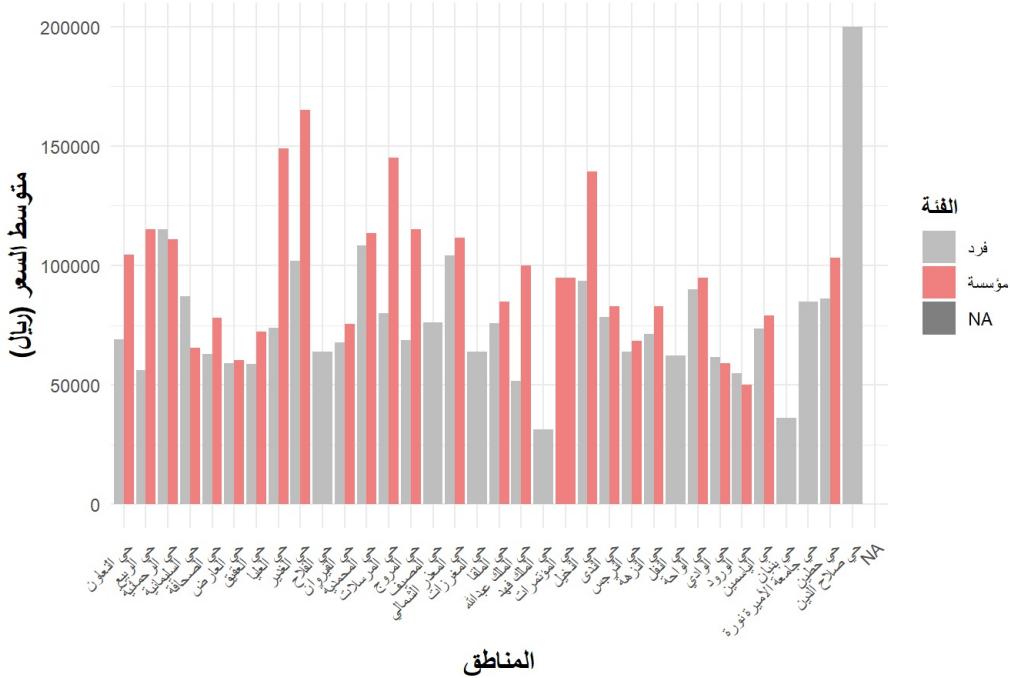
df7=read_csv("akar2024.csv")
type1 <- df7[df7$city == 'الرياض', ]
type1 <- type1[type1$rent_period == 3.0, ]
type1 <- type1[type1$category == 1, ]
type1 <- type1[type1$rooms == 3, ]
type1$user_type <- ifelse(grepl("فرد", "شركة|مؤسسة", type1$user.name), "مؤسسة", "فرد")
type12 <- type1[type1$direction == 'شمال الرياض', ]

avg_price_north <- type12 %>%
  group_by(district, user_type) %>%
  summarise(avg_price = mean(price)) %>%
  ungroup()

ggplot(avg_price_north, aes(x = district, y = avg_price, fill = user_type)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    x = "المناطق",
    y = "(متوسط السعر (ريال",
    title = "متوسط أسعار الإيجارات في شمال الرياض حسب المنطقة"
  ) +
  scale_fill_manual(values = c("gray", "lightcoral")) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, family = "Arial"),
    axis.text.y = element_text(family = "Arial"),
    plot.title = element_text(family = "Arial", face = "bold", size = 16),
    axis.title = element_text(family = "Arial", face = "bold", size = 14),
    legend.title = element_text(family = "Arial", face = "bold", size = 12)
  ) +
  guides(fill = guide_legend(title = "الفئة"))

```

متوسط أسعار الإيجارات في شمال الرياض حسب المنطقة



```

type123 <- subset(type1, direction == 'جنوب الرياض')
type123$user_type <- ifelse(grepl("شركة|مؤسسة", type123$user.name), 'مؤسسة', 'فرد')
avg_price1 <- aggregate(price ~ district + user_type, data = type123, FUN = mean)
avg_price_pivot <- reshape(avg_price1, idvar = "district", timevar = "user_type", direction = "wide")
library(ggplot2)

ggplot(avg_price_pivot, aes(x = district)) +
  geom_bar(aes(y = `price.فرد`, fill = "فرد"), stat = "identity", position = "dodge") +
  geom_bar(aes(y = `price.مؤسسة`, fill = "مؤسسة"), stat = "identity", position = "dodge") +
  labs(
    title = "متوسط أسعار الإيجارات في جنوب الرياض حسب المنطقة",
    x = "المناطق",
    y = "متوسط السعر (ريال)",
    fill = "الفئة"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14, face = "bold")
  ) +
  scale_fill_manual(values = c("فرد" = "lightcoral", "مؤسسة" = "gray")) +
  theme(legend.position = "top") +
  theme(axis.text = element_text(size = 12, family = "Arial"))

```

متوسط أسعار الإيجارات في جنوب الرياض حسب المنطقة



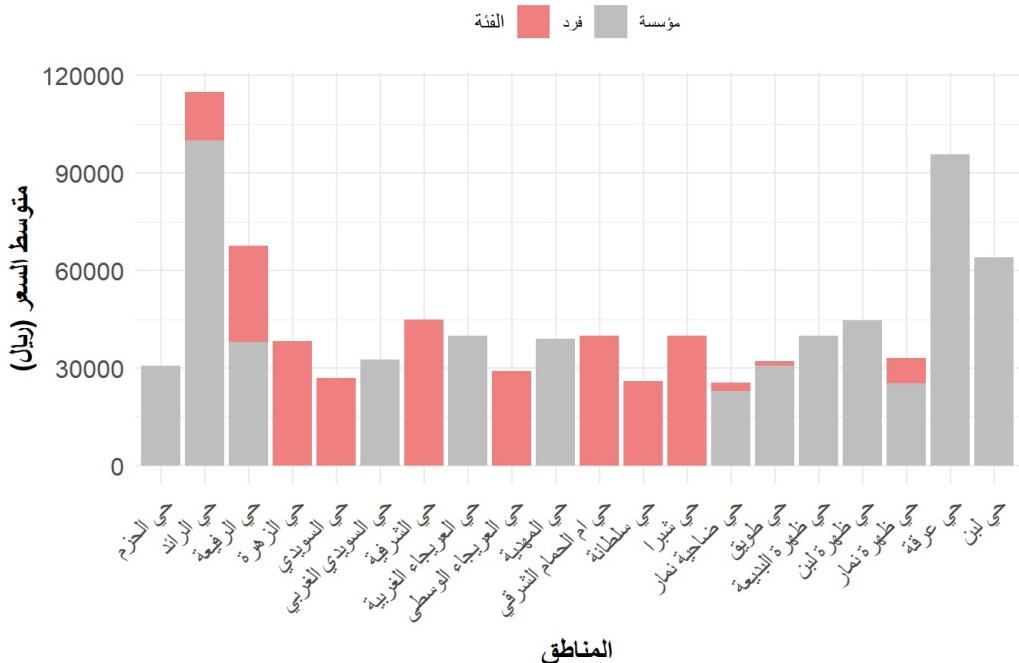
```

type122 <- subset(type1, direction == 'غرب الرياض')
type122$user_type <- ifelse(grepl("شركة|مؤسسة", type122$user.name), 'مؤسسة', 'فرد')
avg_price1 <- aggregate(price ~ district + user_type, data = type122, FUN = mean)
avg_price_pivot <- reshape(avg_price1, idvar = "district", timevar = "user_type", direction = "wide")
library(ggplot2)

ggplot(avg_price_pivot, aes(x = district)) +
  geom_bar(aes(y = `price.فرد`, fill = "فرد"), stat = "identity", position = "dodge") +
  geom_bar(aes(y = `price.مؤسسة`, fill = "مؤسسة"), stat = "identity", position = "dodge") +
  labs(
    title = "متوسط أسعار الإيجارات في غرب الرياض حسب المنطقة",
    x = "المناطق",
    y = "متوسط السعر (ريال)",
    fill = "الفئة"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14, face = "bold")
  ) +
  scale_fill_manual(values = c("فرد" = "lightcoral", "مؤسسة" = "gray")) +
  theme(legend.position = "top") +
  theme(axis.text = element_text(size = 12, family = "Arial"))

```

متوسط أسعار الإيجارات في غرب الرياض حسب المنطقة

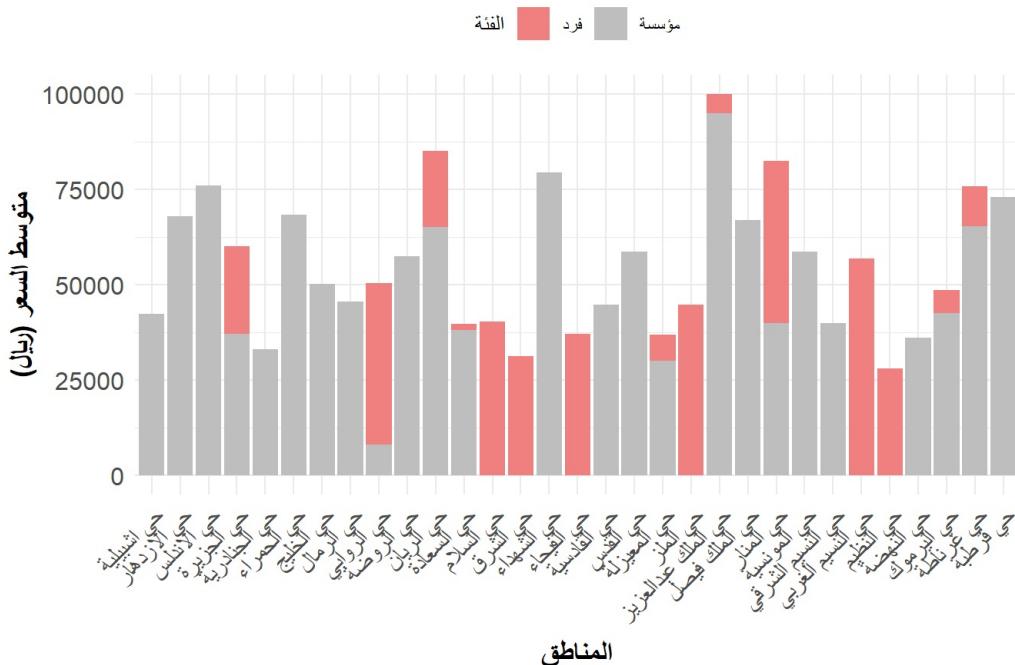


```

type124 <- subset(type1, direction == 'شرق الرياض')
type124$user_type <- ifelse(grepl("شركة|مؤسسة", type124$user.name), 'مؤسسة|فرد', 'شركة|فرد')
avg_price1 <- aggregate(price ~ district + user_type, data = type124, FUN = mean)
avg_price_pivot <- reshape(avg_price1, idvar = "district", timevar = "user_type", direction = "wide")
library(ggplot2)

ggplot(avg_price_pivot, aes(x = district)) +
  geom_bar(aes(y = `price.فرد`, fill = "فرد"), stat = "identity", position = "dodge") +
  geom_bar(aes(y = `price.مؤسسة`, fill = "مؤسسة"), stat = "identity", position = "dodge") +
  labs(
    title = "متوسط أسعار الإيجارات في شرق الرياض حسب المنطقة",
    x = "المناطق",
    y = "متوسط السعر (ريل)",
    fill = "الفئة"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14, face = "bold")
  ) +
  scale_fill_manual(values = c("فرد" = "lightcoral", "مؤسسة" = "gray")) +
  theme(legend.position = "top") +
  theme(axis.text = element_text(size = 12, family = "Arial"))
  
```

متوسط أسعار الإيجارات في شرق الرياض حسب المنطقة

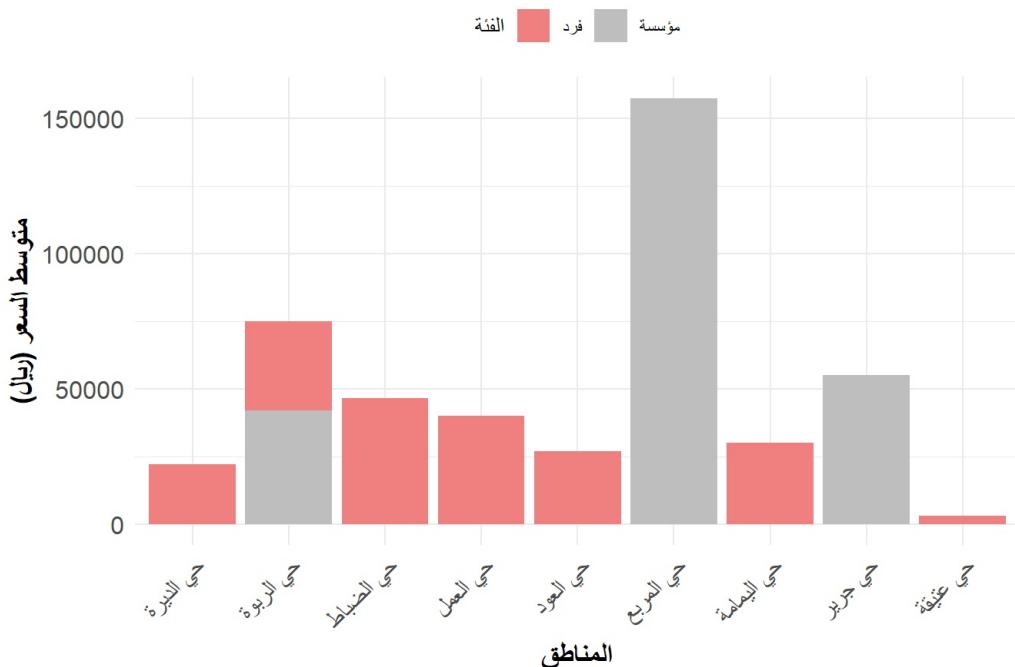


```

type125 <- subset(type1, direction == 'وسط الرياض')
type125$user_type <- ifelse(grepl("شركة|مؤسسة", type125$user.name), "شركة|مؤسسة", "فرد")
avg_price1 <- aggregate(price ~ district + user_type, data = type125, FUN = mean)
avg_price_pivot <- reshape(avg_price1, idvar = "district", timevar = "user_type", direction = "wide")
library(ggplot2)

ggplot(avg_price_pivot, aes(x = district)) +
  geom_bar(aes(y = `price.فرد`, fill = "فرد"), stat = "identity", position = "dodge") +
  geom_bar(aes(y = `price.مؤسسة`, fill = "مؤسسة"), stat = "identity", position = "dodge") +
  labs(
    title = "متوسط أسعار الإيجارات في وسط الرياض حسب المنطقة",
    x = "المناطق",
    y = "متوسط السعر (ريال)",
    fill = "الفئة"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14, face = "bold")
  ) +
  scale_fill_manual(values = c("فرد" = "lightcoral", "مؤسسة" = "gray")) +
  theme(legend.position = "top") +
  theme(axis.text = element_text(size = 12, family = "Arial"))
  
```

متوسط أسعار الإيجارات في وسط الرياض حسب المنطقة



part three

What are the most important attributes affecting the price other than area? This analysis indicates which attributes are most valuable in Riyadh's property market. For instance, attributes like "pool" or "playground" might have higher associated average prices, suggesting they add significant value to properties. Property developers and sellers can use these insights to enhance property features that attract buyers or renters.

```

data_riyadh <- data %>%
  filter(city == "الرِّيَاضُ")

# Filter and group data for categories 6, 3, and 2
furnish_summary_riyadh <- data_riyadh %>%
  filter(category %in% c(6, 3, 2)) %>%
  group_by(furnished, category) %>%
  summarise(avg_price = mean(price, na.rm = TRUE), .groups = "drop")
# Function to remove outliers based on IQR
remove_outliers <- function(data, column) {
  Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
  Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
  IQR_value <- Q3 - Q1
  lower_bound <- Q1 - 1.5 * IQR_value
  upper_bound <- Q3 + 1.5 * IQR_value
  data %>% filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
}

# Remove outliers in the price column
data_no_outliers <- remove_outliers(data_riyadh, "price")

# List of attributes to analyze
attributes <- c("beds", "two_entrances", "special_entrance", "kitchen", "livings",
               "playground", "pool", "rooms", "private_roof", "wc", "lift")

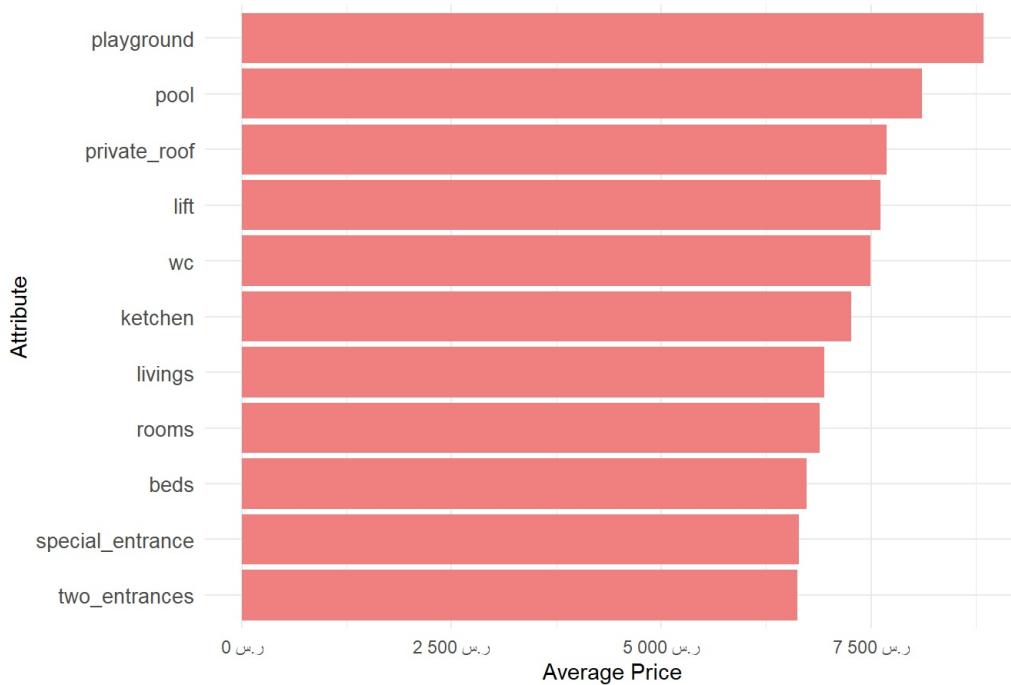
# Calculate average price for each attribute only for category 3
avg_price_attributes <- lapply(attributes, function(attr) {
  data_no_outliers %>%
    filter(category == 3) %>% # Filter for category 3
    mutate(price_per_area = price / area) %>%
    remove_outliers("price_per_area") %>%
    group_by(.data[[attr]]) %>%
    summarise(avg_price_per_area = mean(price_per_area, na.rm = TRUE)) %>%
    mutate(attribute = attr)
}) %>% bind_rows()

# Arrange attributes from highest to lowest price effect
avg_price_summary <- avg_price_attributes %>%
  group_by(attribute) %>%
  summarise(avg_price = mean(avg_price_per_area, na.rm = TRUE)) %>%
  arrange(desc(avg_price))

# Plot the results
ggplot(avg_price_summary, aes(x = reorder(attribute, avg_price), y = avg_price)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "السمات المؤثرة على أسعار العقارات في الرياض (فيلا للبيع",
       x = "Attribute",
       y = "Average Price") +
  coord_flip() +
  scale_y_continuous(labels = label_number(scale = 1, suffix = " ₡.", decimal.mark = "."))
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10))

```

السمات المؤثرة على أسعار العقارات في الرياض (فيلا للبيع)



Does the presence of features like a private entrance, parking, rooftop, or external additions (e.g., yard) increase property value? We will check if there are differences and how big they are.

```

data=read_csv("akar2024.csv")
data <- data %>% distinct(id, .keep_all = TRUE)%>% filter(city=="الرياض")
cat=c(1, 6)
# Function to remove outliers
remove_outliers <- function(data, column) {
  Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
  Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
  IQR_value <- IQR(data[[column]], na.rm = TRUE)
  lower_bound <- Q1 - 1.5 * IQR_value
  upper_bound <- Q3 + 1.5 * IQR_value
  filtered_data <- data %>% filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
  return(filtered_data)
}

p_three_df=data %>%
  filter(category %in% c(1, 6)) %>%
  # For category 6, remove rows where rent_period is not NA
  filter(!(category == 6 & !is.na(rent_period))) %>%
  # For category 1, remove rows where rent_period is NA
  filter(!(category == 1 & is.na(rent_period)))
# Define the function
p_three <- function(data, group_col, rent_per = NA, title_suffix = "") {

  # Determine title based on rent_period
  title_text <- switch(as.character(rent_per),
    "3" = paste("متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار السنوي", title_suffix),
    "2" = paste("متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار الشهري", title_suffix),
    "NA" = paste("متوسط سعر المتر المربع حسب عدد الغرف في الرياض - للبيع", title_suffix))

  # Filter and summarize data
  avg_price <- data %>%
    filter(city == "الرياض",
      (is.na(rent_per) & is.na(rent_period)) | rent_period == rent_per) %>%
    mutate(price_per_area = price / area) %>%
    remove_outliers("price_per_area") %>%
    group_by({{ group_col }}) %>%
    summarise(avg_price_per_area = mean(price_per_area, na.rm = TRUE)) %>%
    arrange({{ group_col }})

  # Calculate percentage difference between consecutive bars
  avg_price <- avg_price %>%
    mutate(percentage_change = c(NA, diff(avg_price_per_area) / head(avg_price_per_area, -1) * 100)) # Use head() to align lengths

  # Print the percentage differences
  print(avg_price) # This will display the calculated percentage change

  # Create plot
  plot <- ggplot(avg_price, aes(x = as.factor({{ group_col }}), y = avg_price_per_area)) +
    geom_bar(stat = "identity", fill = "lightcoral") +
    labs(title = title_text,
      x = paste("عدد", title_suffix),
      y = "متوسط سعر المتر المربع") +
    scale_y_continuous(labels = scales::comma) +
    theme_minimal()

  return(plot)
}

```

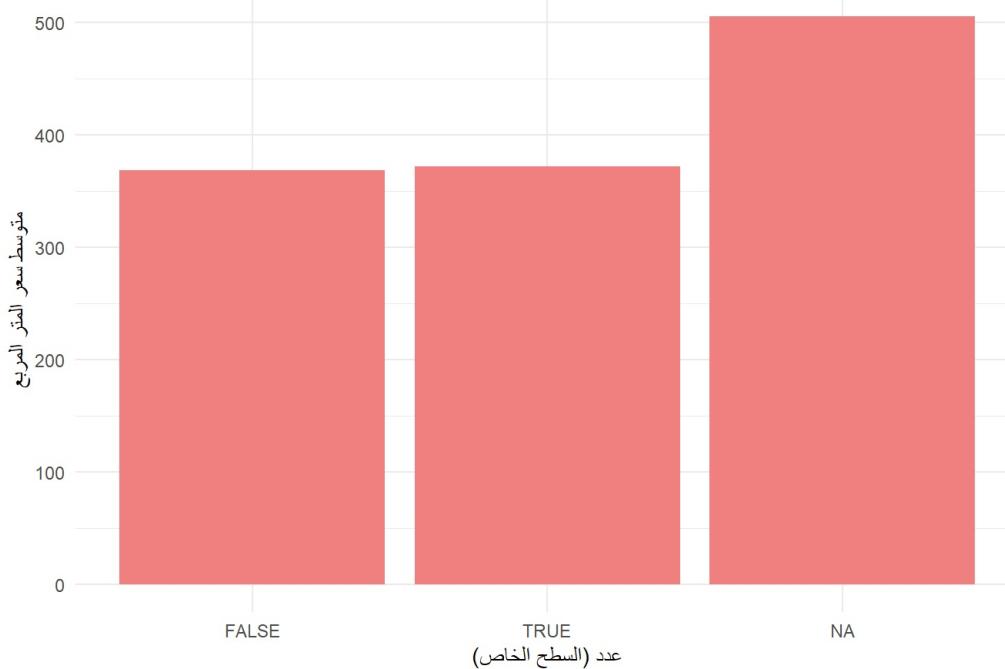
p_three(p_three_df, private_roof, rent_per = 3, title_suffix = "(السطح الخام)")

```

## # A tibble: 3 × 3
##   private_roof avg_price_per_area percentage_change
##   <lgl>           <dbl>            <dbl>
## 1 FALSE          369.             NA
## 2 TRUE           372.             0.921
## 3 NA             506.             36.0

```

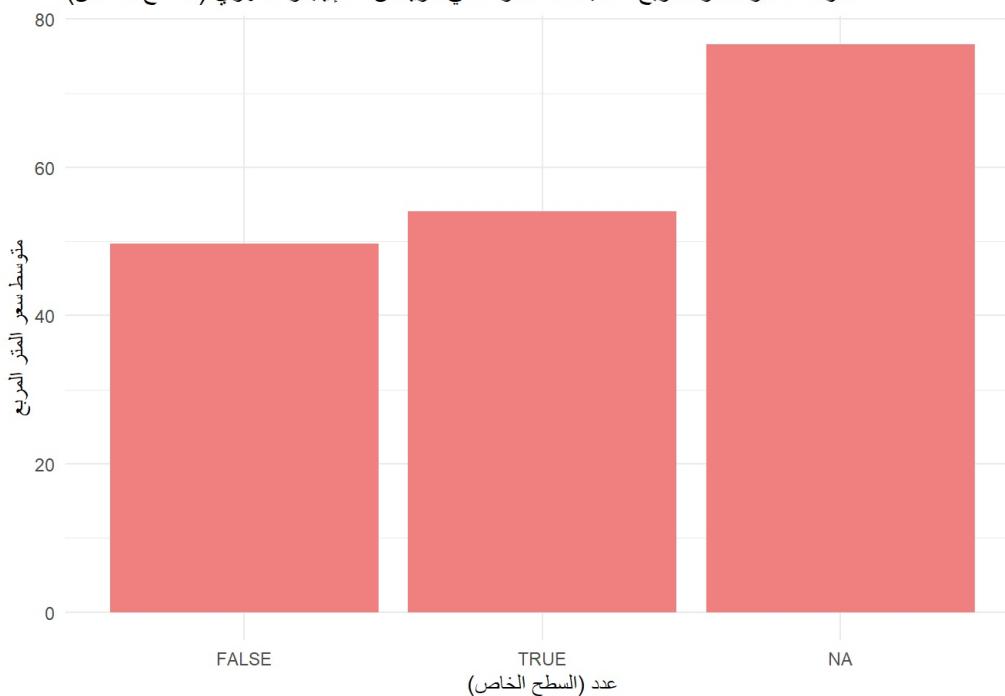
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار السنوي (السطح الخاص)



```
p_three(p_three_df, private_roof, rent_per = 2, title_suffix = "(السطح الخاص)")
```

```
## # A tibble: 3 × 3
##   private_roof avg_price_per_area percentage_change
##   <lgl>           <dbl>            <dbl>
## 1 FALSE          49.7             NA
## 2 TRUE           54.0             8.69
## 3 NA              76.6            41.9
```

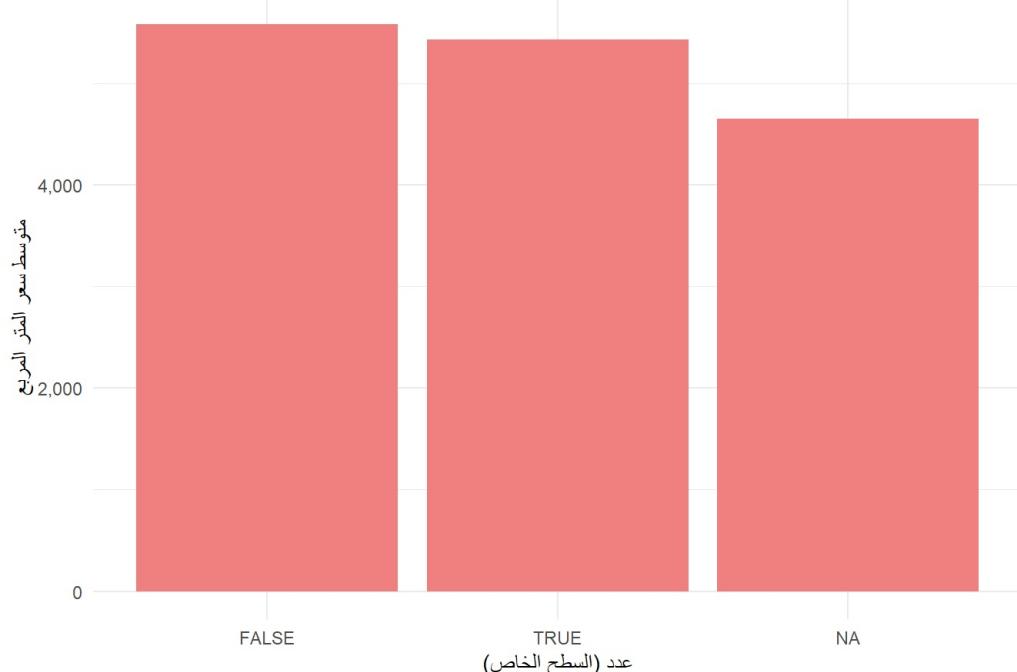
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار الشهري (السطح الخاص)



```
p_three(p_three_df, private_roof, rent_per = NA, title_suffix = "(السطح الخاص)")
```

```
## # A tibble: 3 × 3
##   private_roof avg_price_per_area percentage_change
##   <lgl>           <dbl>            <dbl>
## 1 FALSE          5585.             NA
## 2 TRUE           5434.            -2.71
## 3 NA              4655.            -14.3
```

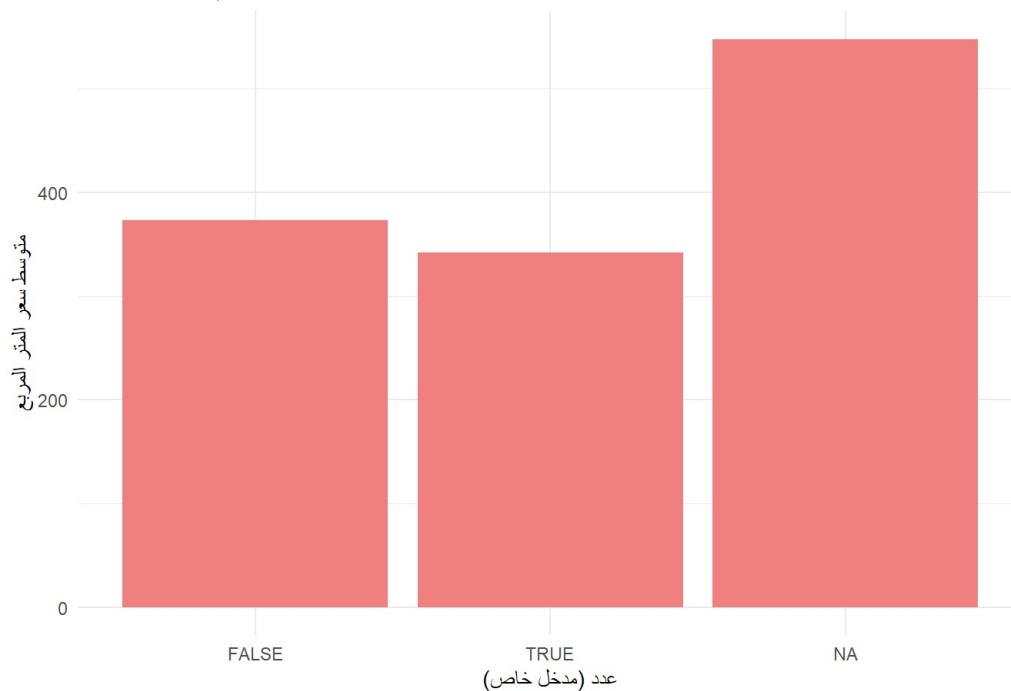
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - للبيع (السطح الخاص)



```
p_three(p_three_df, special_entrance, rent_per = 3, title_suffix = "(مدخل خاص)")
```

```
## # A tibble: 3 × 3
##   special_entrance avg_price_per_area percentage_change
##   <lgl>              <dbl>            <dbl>
## 1 FALSE             373.             NA
## 2 TRUE              342.            -8.27
## 3 NA                548.            60.2
```

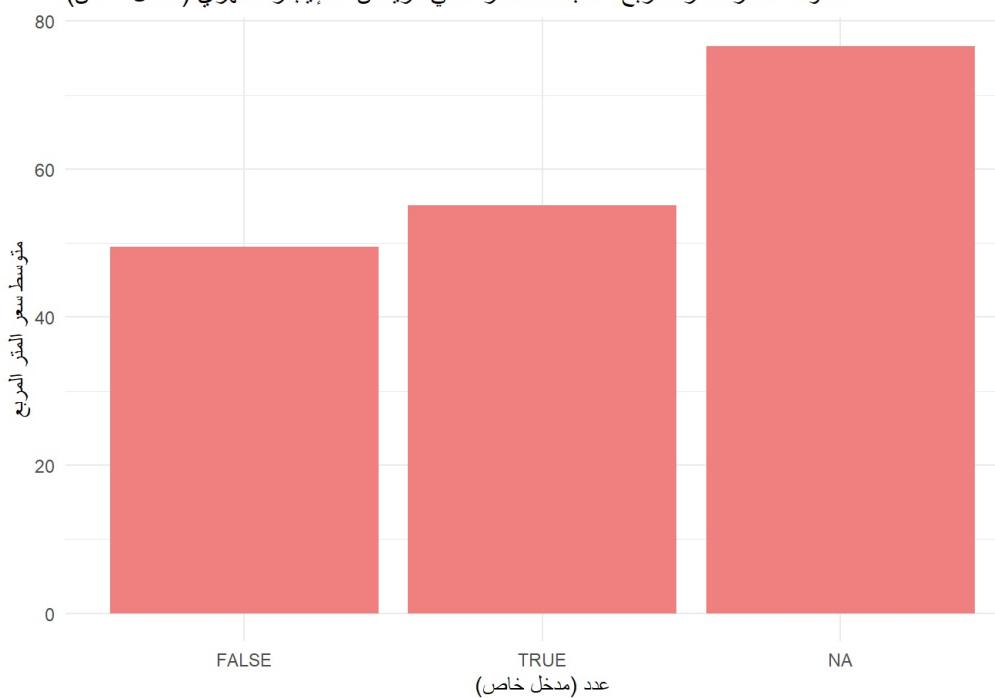
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار السنوي (مدخل خاص)



```
p_three(p_three_df, special_entrance, rent_per = 2, title_suffix = "(مدخل خاص)")
```

```
## # A tibble: 3 × 3
##   special_entrance avg_price_per_area percentage_change
##   <lgl>              <dbl>            <dbl>
## 1 FALSE             49.5             NA
## 2 TRUE              55.1            11.4
## 3 NA                76.6            39.1
```

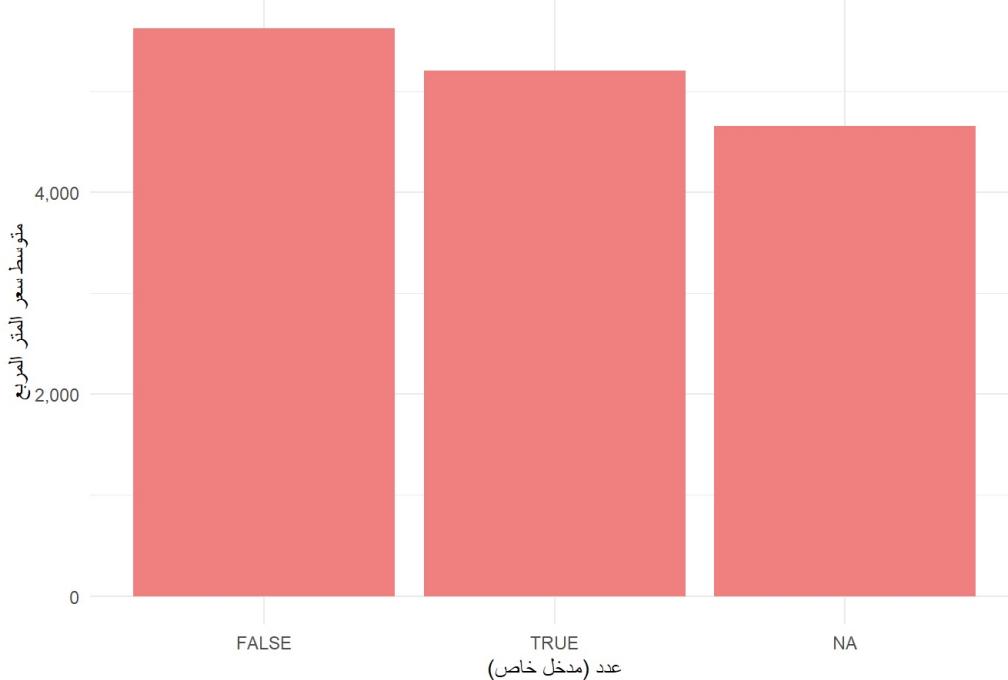
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار الشهري (مدخل خاص)



```
p_three(p_three_df, special_entrance, rent_per = NA, title_suffix = "(مدخل خاص)")
```

```
## # A tibble: 3 × 3
##   special_entrance avg_price_per_area percentage_change
##   <lgl>              <dbl>            <dbl>
## 1 FALSE             5624.             NA
## 2 TRUE              5205.             -7.46
## 3 NA                4656.             -10.6
```

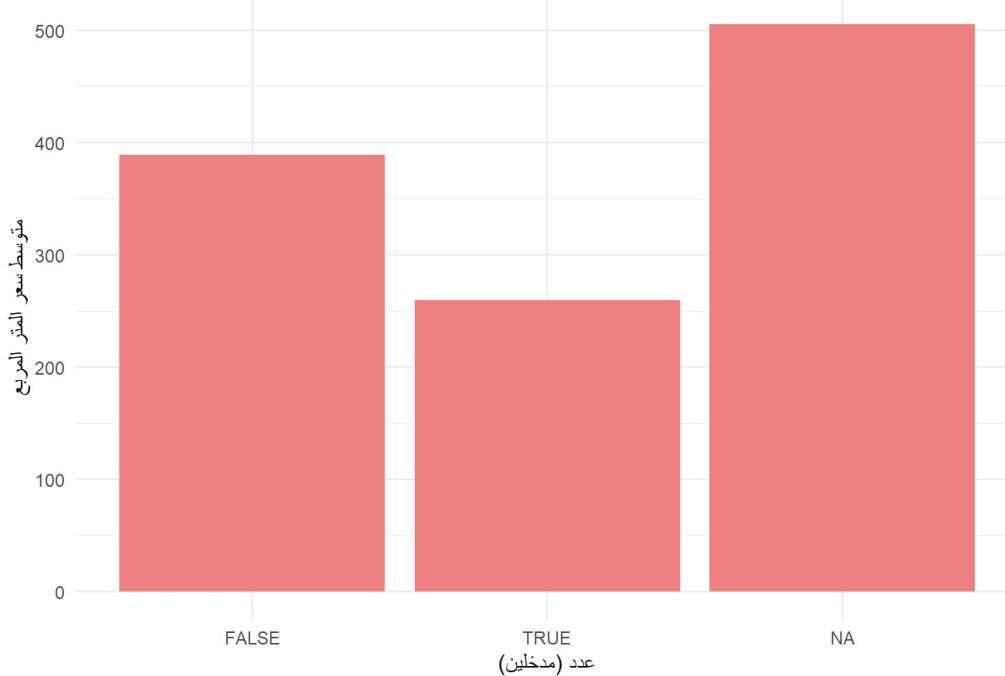
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - للبيع (مدخل خاص)



```
p_three(p_three_df, two_entances, rent_per = 3, title_suffix = "(مدخلين)")
```

```
## # A tibble: 3 × 3
##   two_entances avg_price_per_area percentage_change
##   <lgl>          <dbl>            <dbl>
## 1 FALSE         389.             NA
## 2 TRUE          259.             -33.4
## 3 NA            506.             95.2
```

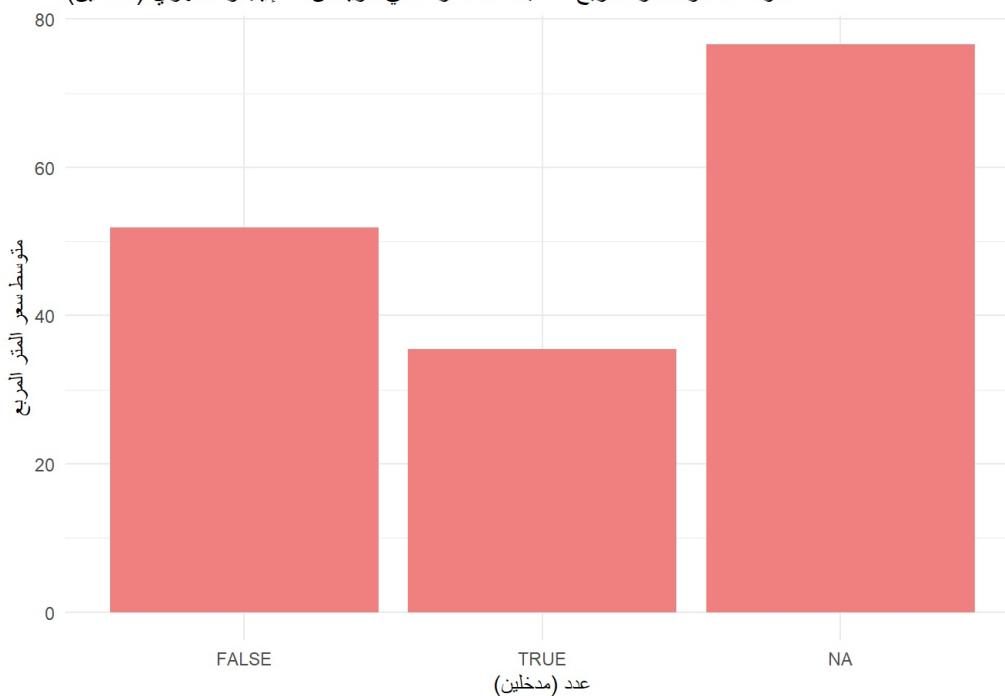
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار السنوي (مدخلين)



```
p_three(p_three_df, two_entrances, rent_per = 2, title_suffix = "(مدخلين)")
```

```
## # A tibble: 3 × 3
##   two_entrances avg_price_per_area percentage_change
##   <lgl>           <dbl>            <dbl>
## 1 FALSE          51.9             NA
## 2 TRUE           35.4            -31.7
## 3 NA              76.6            116.
```

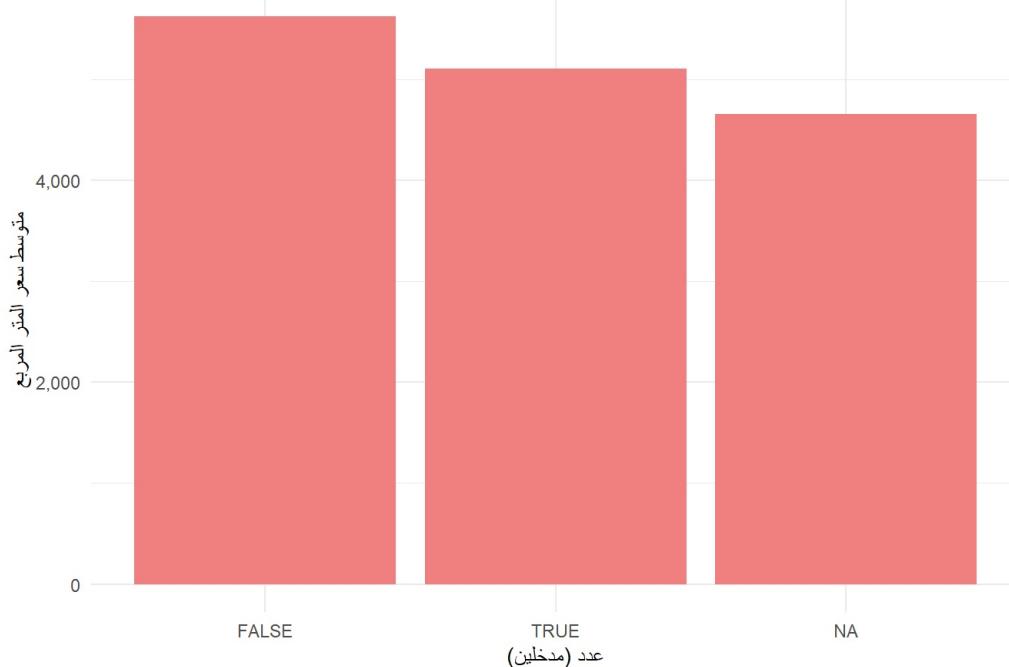
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار الشهري (مدخلين)



```
p_three(p_three_df, two_entrances, rent_per = NA, title_suffix = "(مدخلين)")
```

```
## # A tibble: 3 × 3
##   two_entrances avg_price_per_area percentage_change
##   <lgl>           <dbl>            <dbl>
## 1 FALSE          5626.             NA
## 2 TRUE           5107.            -9.22
## 3 NA              4657.            -8.81
```

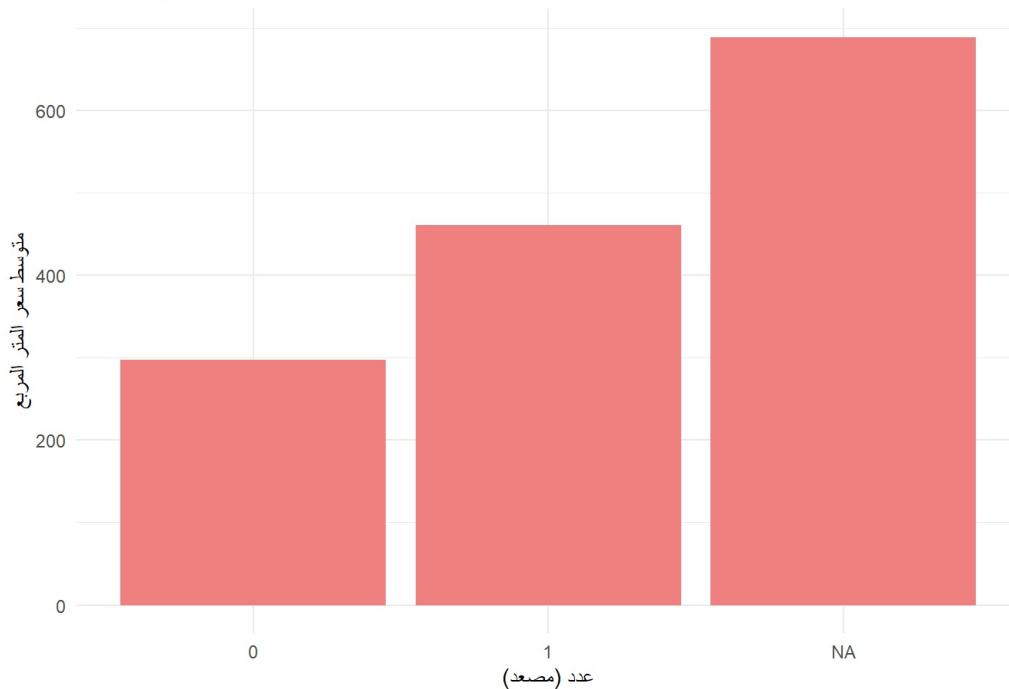
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - للبيع (مدخلين)



```
p_three(p_three_df, lift, rent_per = 3, title_suffix = "(مصد)"")
```

```
## # A tibble: 3 × 3
##   lift avg_price_per_area percentage_change
##   <dbl>      <dbl>            <dbl>
## 1     0       297.             NA
## 2     1       461.            55.1
## 3    NA       689.            49.5
```

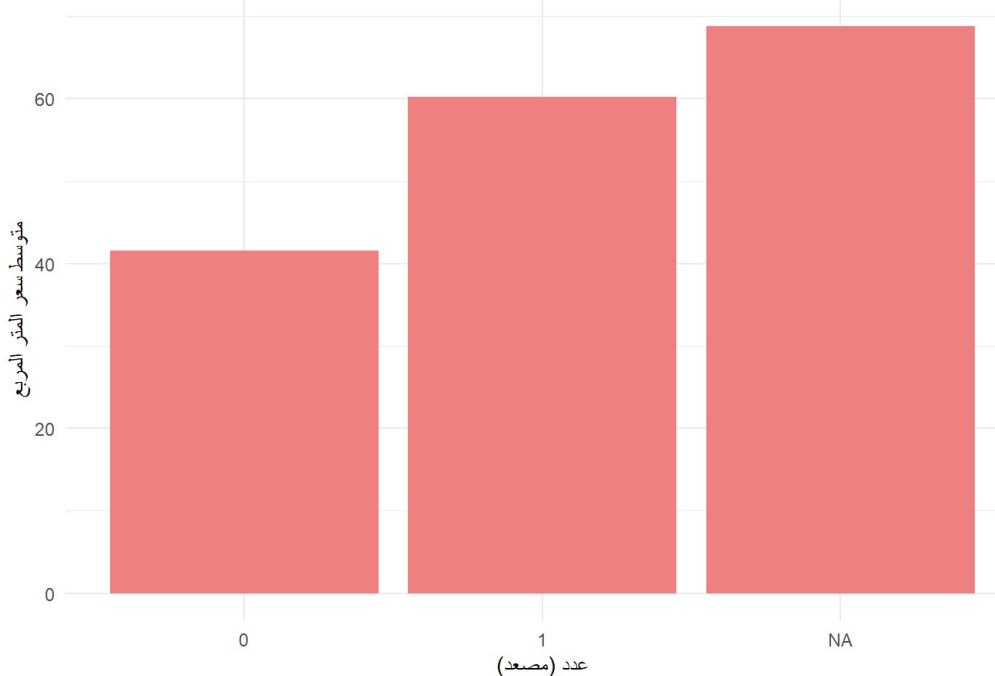
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار السنوي (مصد)



```
p_three(p_three_df, lift, rent_per = 2, title_suffix = "(مصد)"")
```

```
## # A tibble: 3 × 3
##   lift avg_price_per_area percentage_change
##   <dbl>      <dbl>            <dbl>
## 1     0       41.6             NA
## 2     1       60.3            44.9
## 3    NA       68.9            14.2
```

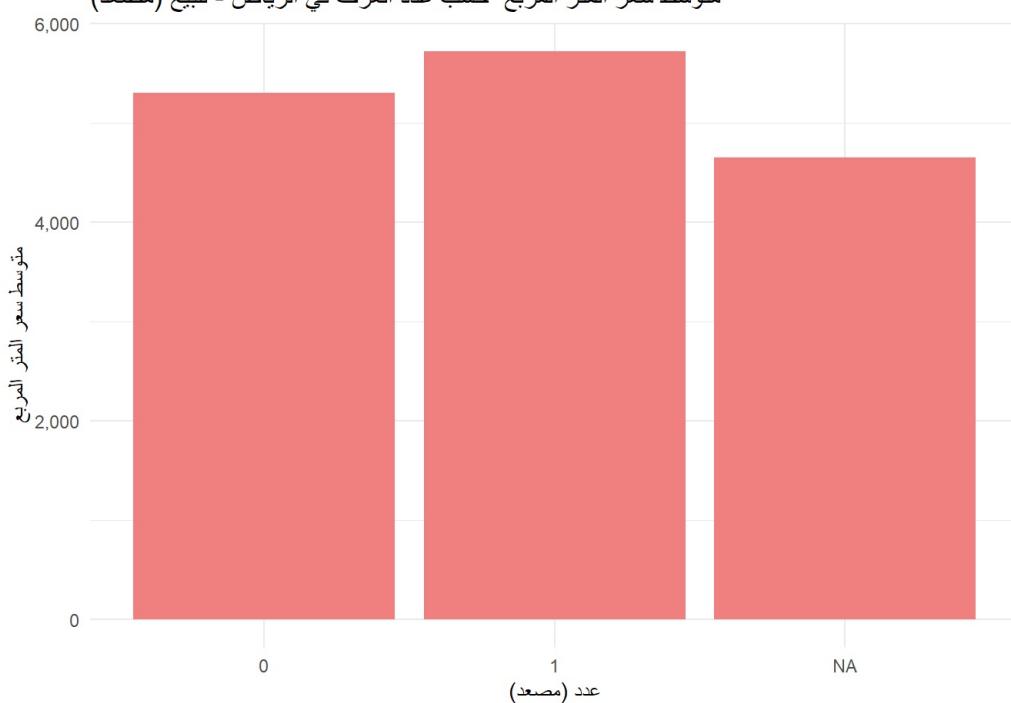
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار الشهري (مصد



```
p_three(p_three_df, lift, rent_per = NA, title_suffix = "(مصد)" )
```

```
## # A tibble: 3 × 3
##   lift avg_price_per_area percentage_change
##   <dbl>      <dbl>            <dbl>
## 1     0       5305.           NA
## 2     1       5723.            7.88
## 3    NA       4648.           -18.8
```

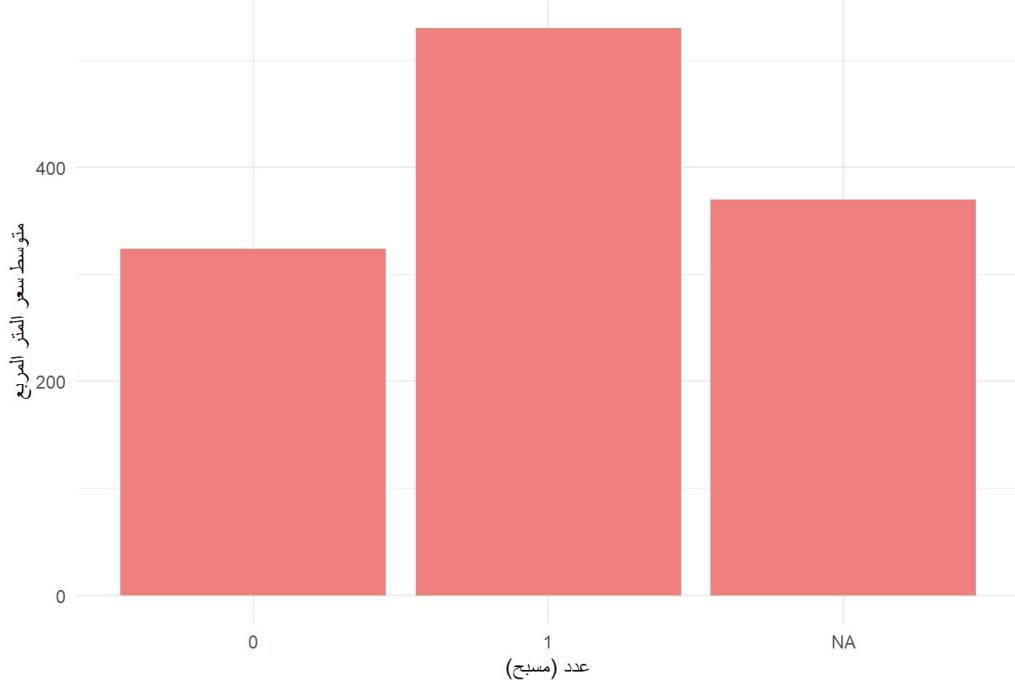
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - للبيع (مصد)



```
p_three(p_three_df, pool, rent_per = 3, title_suffix = "(مبين)" )
```

```
## # A tibble: 3 × 3
##   pool avg_price_per_area percentage_change
##   <dbl>      <dbl>            <dbl>
## 1     0       324.           NA
## 2     1       530.            63.6
## 3    NA       370.           -30.2
```

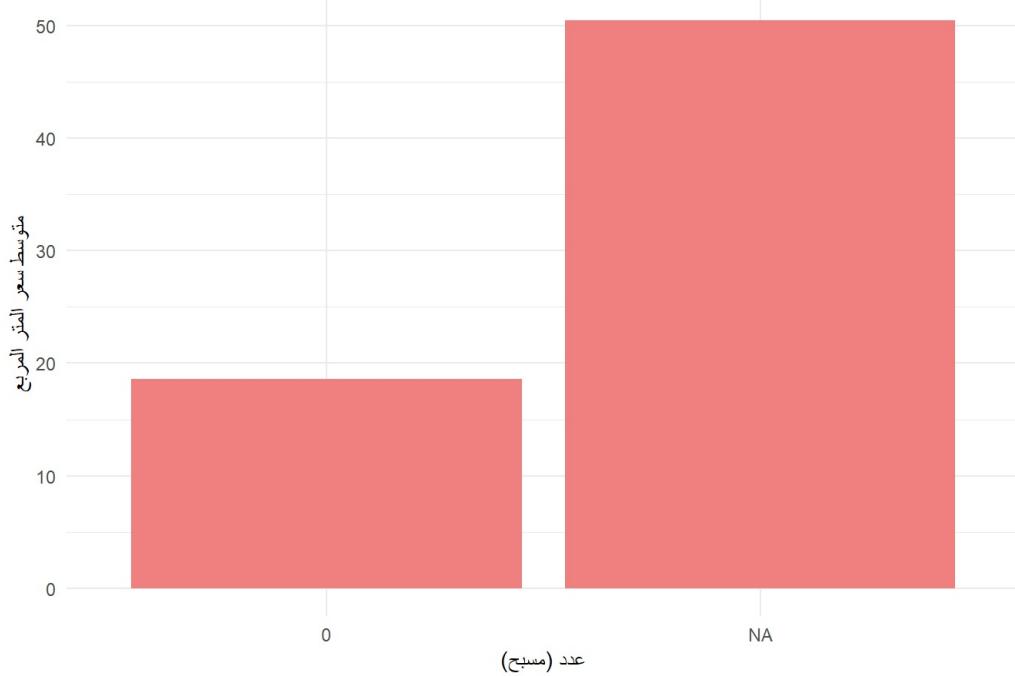
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار السنوي (مسبح)



```
p_three(p_three_df, pool, rent_per = 2, title_suffix = "(مسبح)")
```

```
## # A tibble: 2 × 3
##   pool avg_price_per_area percentage_change
##   <dbl>          <dbl>            <dbl>
## 1     0           18.6             NA
## 2    NA           50.5            172.
```

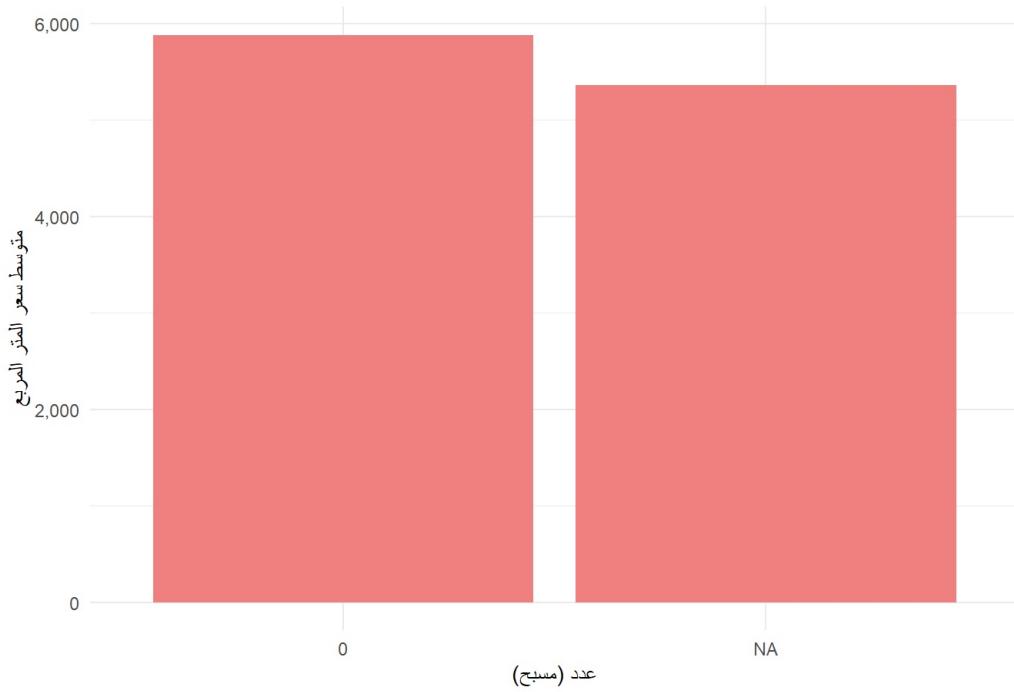
متوسط سعر المتر المربع حسب عدد الغرف في الرياض - الإيجار الشهري (مسبح)



```
p_three(p_three_df, pool, rent_per = NA, title_suffix = "(مسبح)")
```

```
## # A tibble: 2 × 3
##   pool avg_price_per_area percentage_change
##   <dbl>          <dbl>            <dbl>
## 1     0           5884.            NA
## 2    NA           5363.           -8.86
```

متوسط سعر المتر المربع حسب عدد الغرف في الرياض - للبيع (مسبح)



Does furnished vs. unfurnished apartments, as well as the presence of water, electrical, drainage systems, impact the price? Determining whether furnished properties and the availability of water, electrical, and drainage systems impact prices helps identify key factors influencing property value. This insight can refine predictive models, improve pricing strategies, and guide buyers and sellers in making informed decisions. Additionally, it highlights market trends and supports modernization efforts by emphasizing the importance of essential utilities in property valuation.

```

data=read_csv("akar2024.csv")
data <- data %>% distinct(id, .keep_all = TRUE)%>% filter(city=="الرياض")

# Function to plot average price per area with outliers removed based on IQR
plot_price_per_area <- function(data, feature, categories = c(3, 2, 6)) {

  # Check if the specified feature exists in the dataset
  if (!(feature %in% colnames(data))) {
    stop("The specified feature does not exist in the dataset.")
  }

  # Function to remove outliers based on IQR
  remove_outliers <- function(data, column) {
    Q1 <- quantile(data[[column]], 0.25, na.rm = TRUE)
    Q3 <- quantile(data[[column]], 0.75, na.rm = TRUE)
    IQR_value <- IQR(data[[column]], na.rm = TRUE)
    lower_bound <- Q1 - 1.5 * IQR_value
    upper_bound <- Q3 + 1.5 * IQR_value
    filtered_data <- data %>% filter(data[[column]] >= lower_bound & data[[column]] <= upper_bound)
    return(filtered_data)
  }

  # Remove outliers for the price and area columns
  data_no_outliers <- remove_outliers(data, "price")
  data_no_outliers <- remove_outliers(data_no_outliers, "area")

  # Create a new column 'price_per_area' that calculates price per area
  feature_summary <- data_no_outliers %>%
    filter(category %in% categories) %>% # Filter by the given categories
    mutate(price_per_area = price / area) %>% # Create new column for price per area
    group_by(!!(sym(feature)), category) %>%
    summarise(avg_price_per_area = mean(price_per_area, na.rm = TRUE), .groups = "drop")

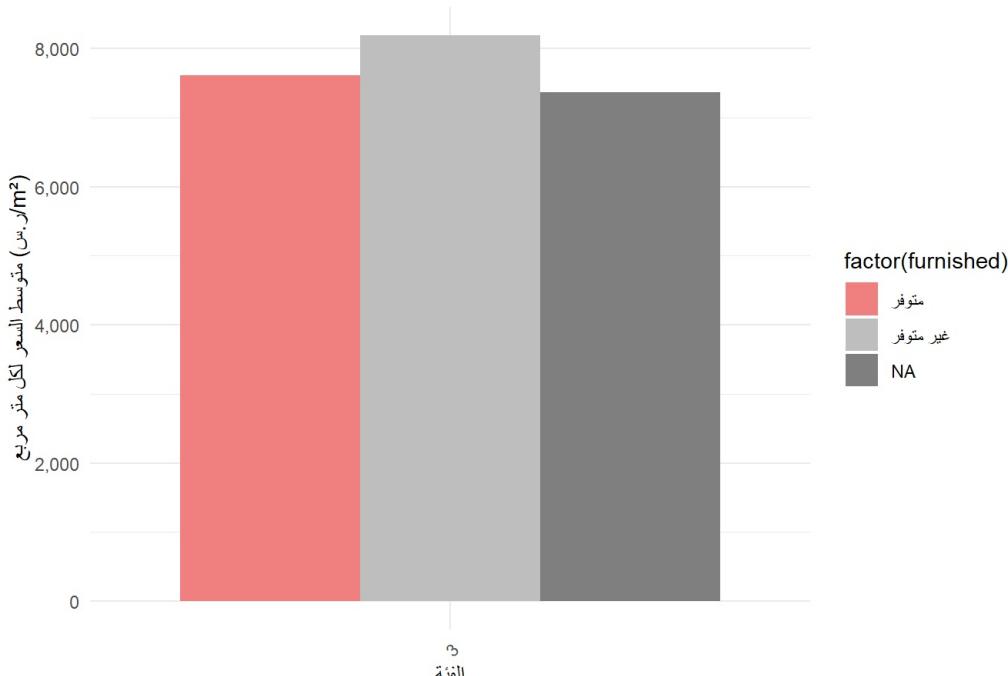
  # Plot the average price per area for the specified feature and categories
  ggplot(feature_summary, aes(x = factor(category), y = avg_price_per_area, fill = factor(!!(sym(feature))))) +
    geom_bar(stat = "identity", position = "dodge") +
    labs(title = paste("متوسط السعر لكل متر مربع حسب", feature, "للفئة", paste(categories)),
        x = "الفئة",
        y = "متوسط السعر لكل متر مربع (ر.س./م²)") +
    scale_fill_manual(values = c("lightcoral", "gray"),
                      labels = c("متوفر", "غير متوفر"))
    scale_y_continuous(labels = label_comma()) +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate x labels for readability
}

# Example usage:

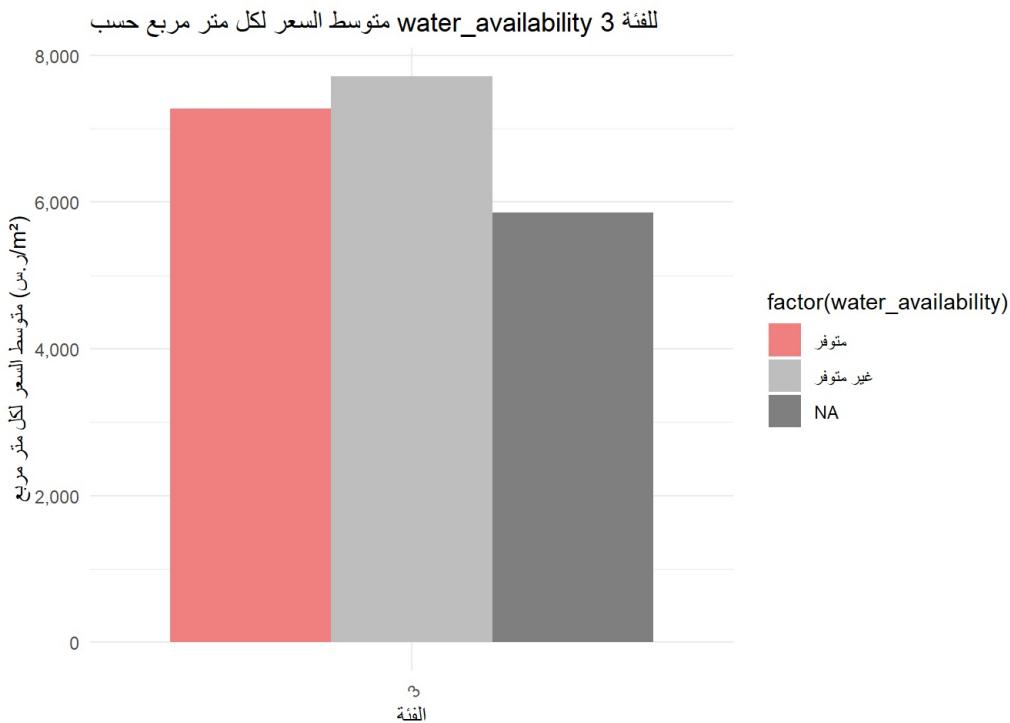
# Plot for 'furnished' feature
plot_price_per_area(data_riyadh, "furnished", 3)

```

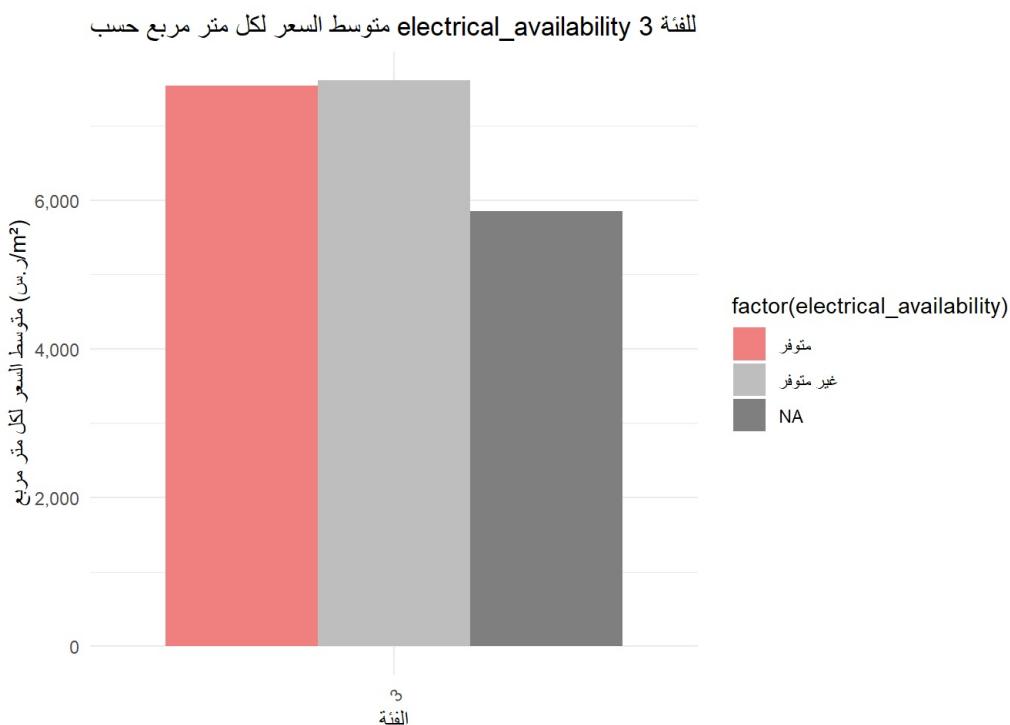
للفة 3 متوسط السعر لكل متر مربع حسب furnished



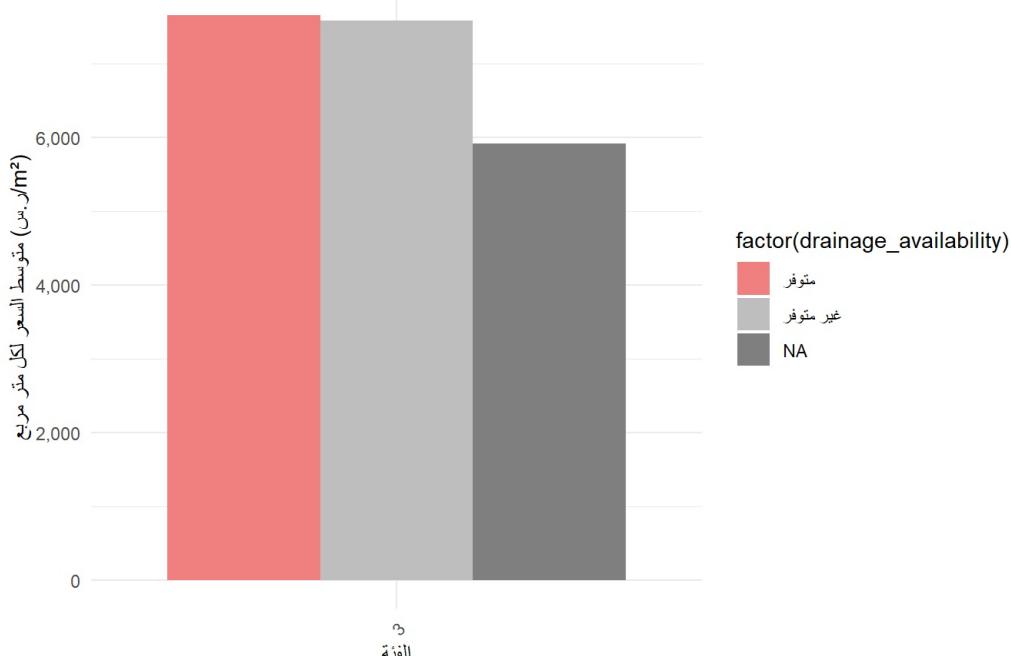
```
# Plot for another feature, say 'water_availability'  
plot_price_per_area(data_riyadh, "water_availability", 3)
```



```
# Plot for another feature, say 'electrical_availability'  
plot_price_per_area(data_riyadh, "electrical_availability", 3)
```



```
# Plot for another feature, say 'drainage_availability'  
plot_price_per_area(data_riyadh, "drainage_availability", 3)
```



Is the annual rent cheaper than monthly rent?, and what is the difference?

```

yearr <- data %>%
  filter(city == "الرياض", rent_period == 3) %>%
  mutate(price_per_month = price / 12)

monthh <- data %>%
  filter(city == "الرياض", rent_period == 2)

# Remove outliers for both datasets
yearr_no_outliers <- remove_outliers(yearr, "price_per_month")
monthh_no_outliers <- remove_outliers(monthh, "price")

mean_yearr <- mean(yearr_no_outliers$price_per_month, na.rm = TRUE)
mean_monthh <- mean(monthh_no_outliers$price, na.rm = TRUE)

# Create a data frame for plotting
mean_values <- data.frame(
  Dataset = c("الإيجار السنوي", "إيجار شهري"),
  Mean_Price = c(mean_yearr, mean_monthh)
)

ggplot(mean_values, aes(x = Dataset, y = Mean_Price, fill = Dataset)) +
  geom_bar(stat = "identity") +
  labs(title = "مقارنة بين متوسط اسعار الشقق",
       x = "Dataset",
       y = "Mean Price") +
  theme_minimal()

```

مقارنة بين متوسط اسعار الشقق



```
# Filter the data for yearly and monthly rent periods, and compute price per square meter
yearr <- data %>%
  filter(city == "الرياض", rent_period == 3) %>%
  mutate(price = price / 12) %>% # Price per square meter
  mutate(price_per_square_meter = price / area)
monthh <- data %>%
  filter(city == "الرياض", rent_period == 2) %>%
  mutate(price_per_square_meter = price / area) # Price per square meter

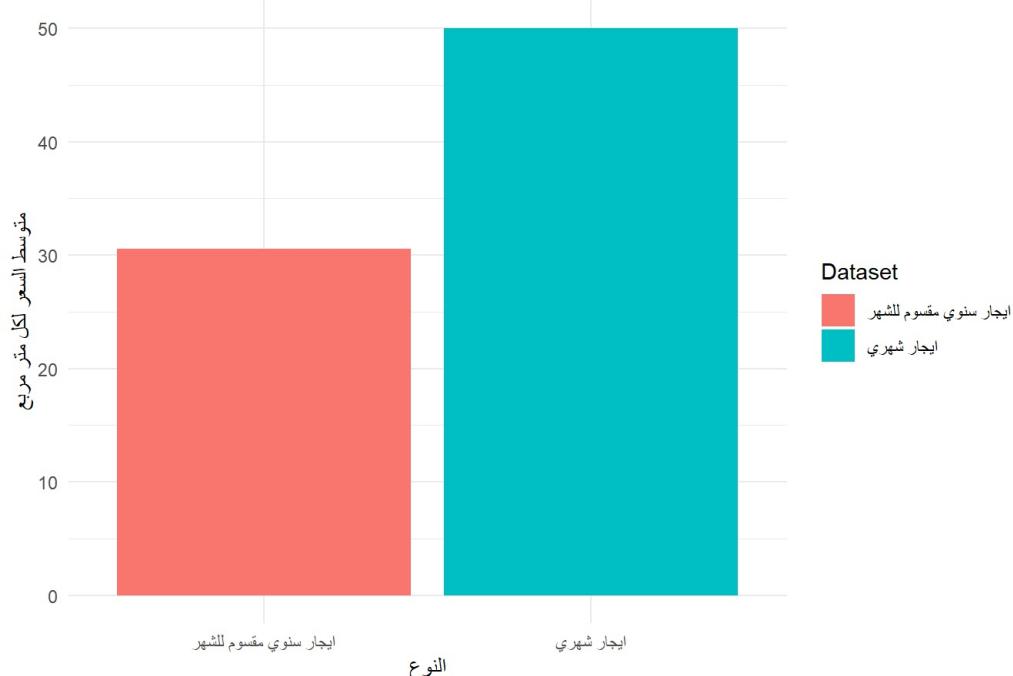
# Remove outliers for both datasets
yearr_no_outliers <- remove_outliers(yearr, "price_per_square_meter")
monthh_no_outliers <- remove_outliers(monthh, "price_per_square_meter")

# Calculate the mean price per square meter for each dataset (after removing outliers)
mean_yearr <- mean(yearr_no_outliers$price_per_square_meter, na.rm = TRUE)
mean_monthh <- mean(monthh_no_outliers$price_per_square_meter, na.rm = TRUE)

# Create a data frame for plotting
mean_values <- data.frame(
  Dataset = c("الإيجار السنوي مقسم إلى شهري", "الإيجار الشهري"),
  Mean_Price = c(mean_yearr, mean_monthh)
)

# Create a bar plot comparing the mean prices
ggplot(mean_values, aes(x = Dataset, y = Mean_Price, fill = Dataset)) +
  geom_bar(stat = "identity") +
  labs(title = "مقارنة بين متوسط اسعار الشقق ل كل متر مربع",
       x = "النوع",
       y = "متوسط السعر لكل متر مربع") +
  theme_minimal()
```

مقارنة بين متوسط اسعار الشقق ل كل مترا مربع



Does the area (in square meters) affect the price?

```
Westriyadhmeans <- Westriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

Centralriyadhmeans <- Centralriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

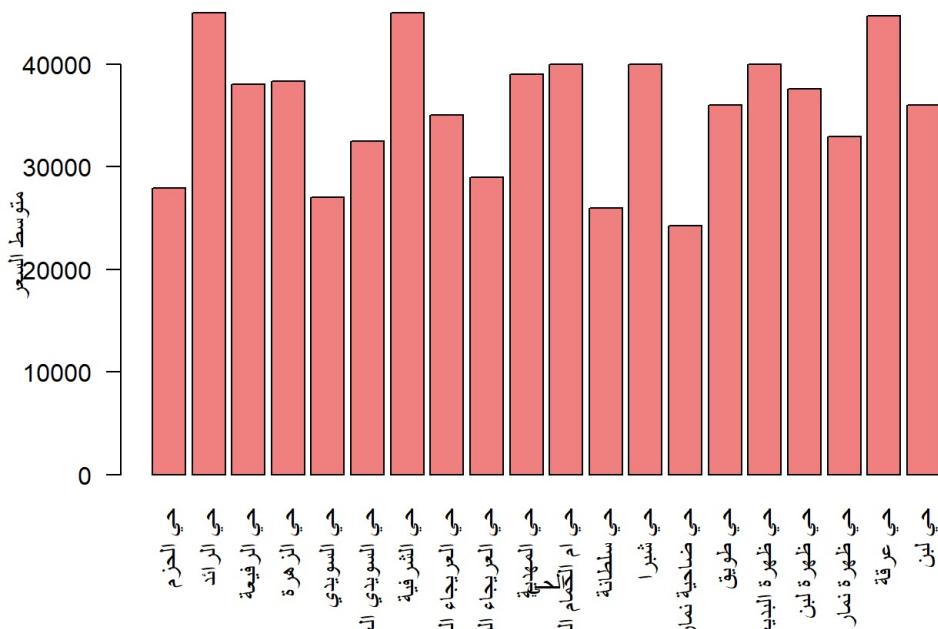
eastriyadhmeans <- eastriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

Southriyadhmeans <- Southriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

northriyadhmeans <- northriyadh %>%
  group_by(district) %>%
  summarise(mean_price = mean(price))

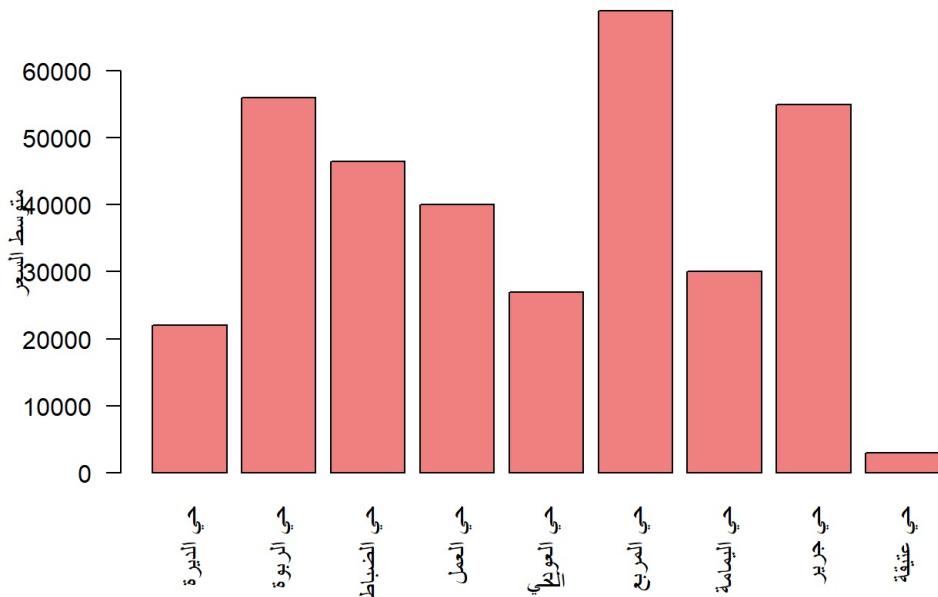
districts_west <- Westriyadhmeans$district
mean_prices_west <- Westriyadhmeans$mean_price
barplot(mean_prices_west, names.arg = districts_west, col = "lightcoral", main = "توسط السعر حسب الحي في غرب الرياض", lab = "متوسط السعر", las = 2)
```

متوسط السعر حسب الحي في غرب الرياض



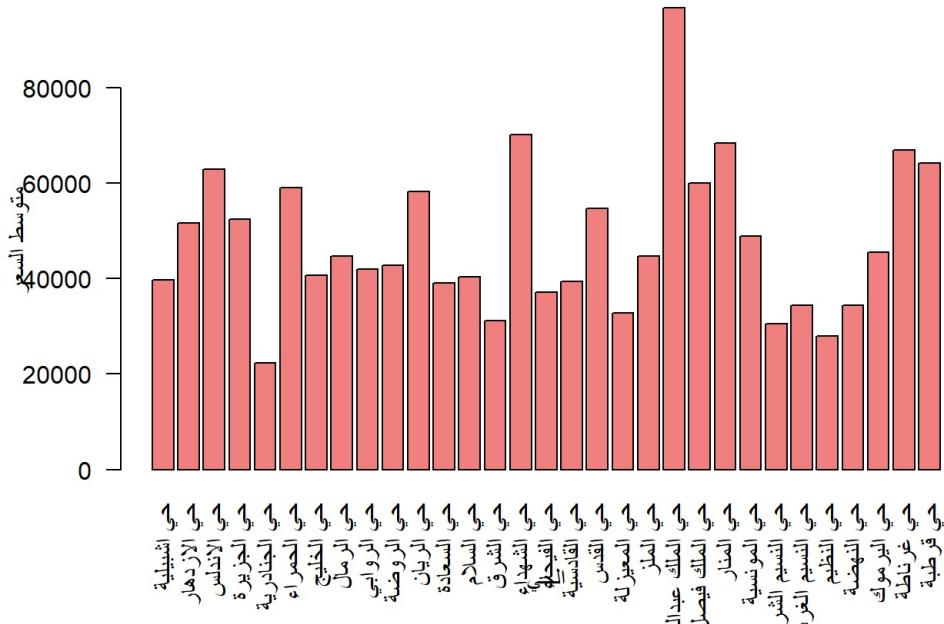
```
districts_central <- Centralriyadhmeans$district
mean_prices_central <- Centralriyadhmeans$mean_price
barplot(mean_prices_central, names.arg = districts_central, col = "lightcoral", main = "لسعر حسب الحي في وسط الرياض", ylab = "متوسط السعر", las = 2)
```

متوسط السعر حسب الحي في وسط الرياض



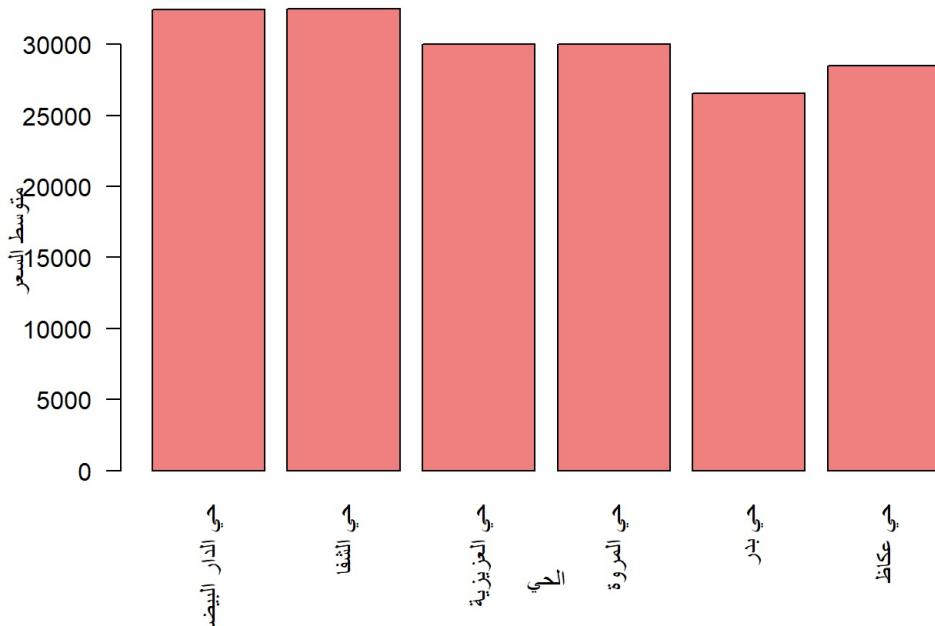
```
districts_east <- eastriyadhmeans$district
mean_prices_east <- eastriyadhmeans$mean_price
barplot(mean_prices_east, names.arg = districts_east, col = "lightcoral", main = "متوسط السعر حسب الحي في شرق الرياض", ylab = "متوسط السعر", las = 2)
```

متوسط السعر حسب الحي في شرق الرياض



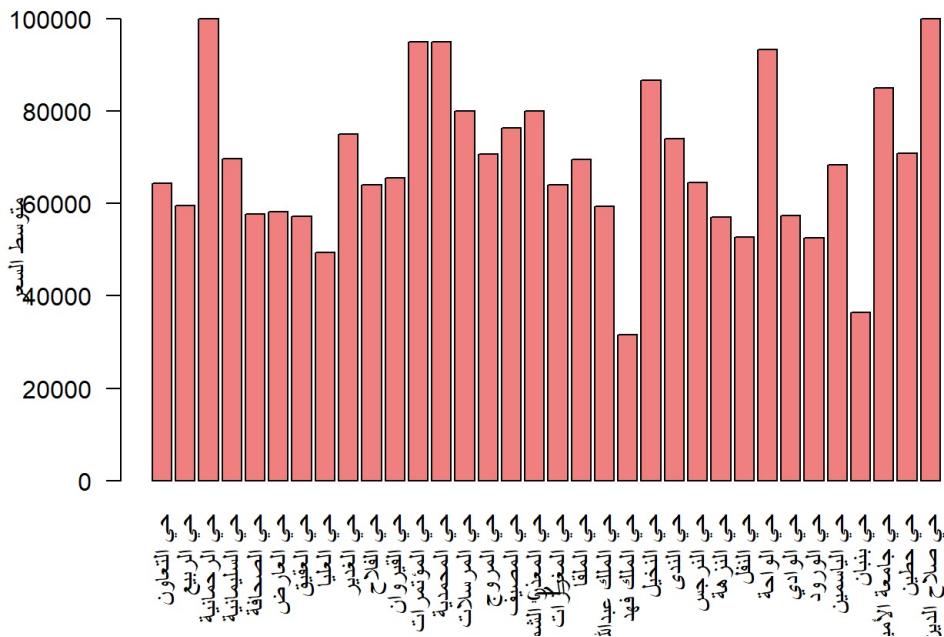
```
districts_south <- Southriyadhmeans$district
mean_prices_south <- Southriyadhmeans$mean_price
barplot(mean_prices_south, names.arg = districts_south, col = "lightcoral", main = "متوسط السعر حسب الحي في جنوب الرياض", ylab = "متوسط السعر", las = 2)
```

متوسط السعر حسب الحي في جنوب الرياض



```
districts_north <- northriyadhmeans$district
mean_prices_north <- northriyadhmeans$mean_price
barplot(mean_prices_north, names.arg = districts_north, col = "lightcoral", main = "متوسط السعر حسب الحي في شمال الرياض", ylab = "متوسط السعر", las = 2)
```

متوسط السعر حسب الحي في شمال الرياض

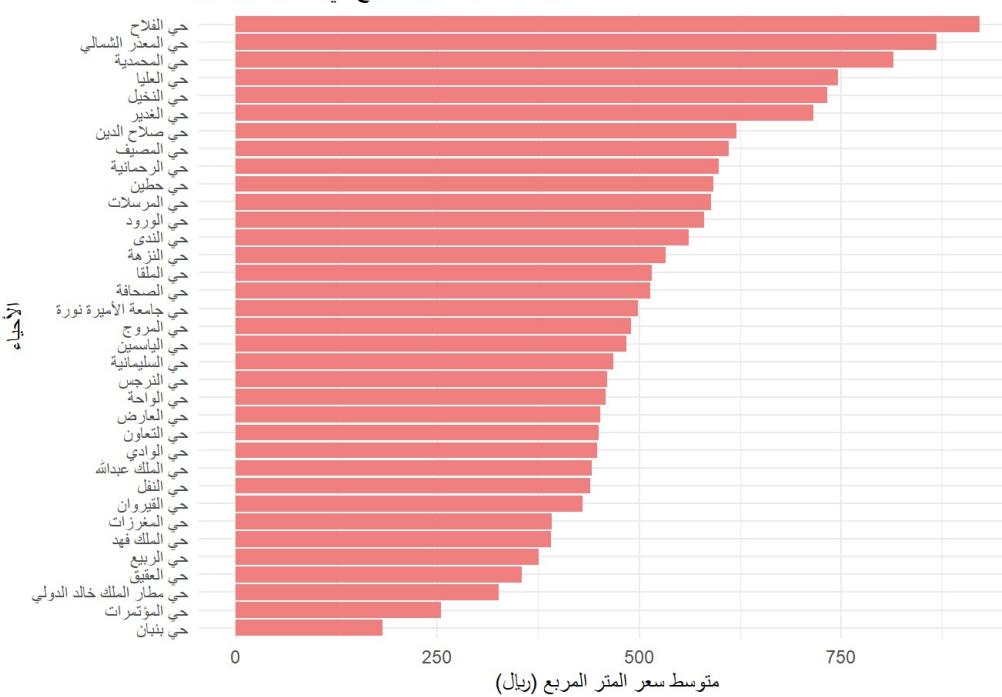


```
df=read_csv("akar2024.csv")
```

```
area <- subset(df, city == "الرياض" & rent_period == 3.0 & category == 1)
area <- area[, c("price", "area", "district", "direction")]
riyadharean <- subset(area, direction == "شمال الرياض")
riyadharean$price_per_area <- riyadharean$price / riyadharean$area
avedis <- aggregate(price_per_area ~ district, data = riyadharean, FUN = mean)
```

```
library(ggplot2)
ggplot(avedis, aes(x = reorder(district, price_per_area), y = price_per_area)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "متوسط سعر المتر المربع في شمال الرياض",
       x = "الأحياء",
       y = "(متوسط سعر المتر المربع (ريال))") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme_minimal() +
  coord_flip() +
  theme(axis.title.x = element_text(hjust = 0.5))
```

متوسط سعر المتر المربع في شمال الرياض



```
average <- mean(riyadharean$price_per_area, na.rm = TRUE)
cat("متوسط سعر المتر المربع في شمال الرياض هو:", average, " ريال")
```

```
east_riyadh <- subset(area, direction == "شرق الرياض")
east_riyadh$price_per_area <- east_riyadh$price / east_riyadh$area
average_east <- mean(east_riyadh$price_per_area, na.rm = TRUE)
cat("متوسط سعر المتر المربع في شرق الرياض هو:", average_east, " ريال")
```

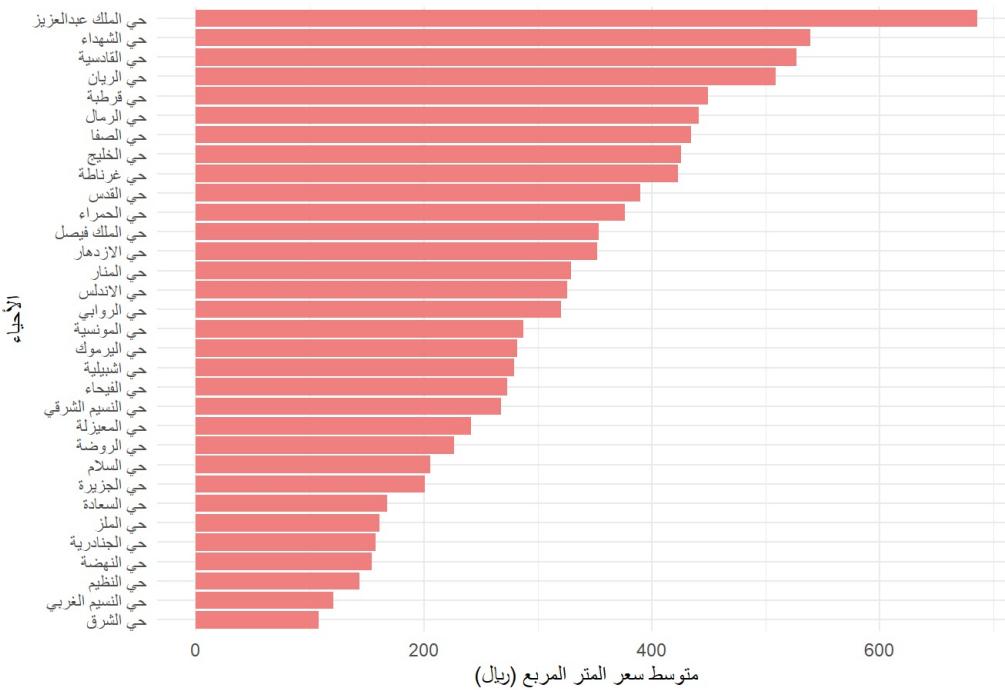
متوسط سعر المتر المربع في شرق الرياض هو: 3402.34 لـ ##

```
avedis_east <- aggregate(price_per_area ~ district, data = east_riyadh, mean)

library(ggplot2)

ggplot(avedis_east, aes(x = price_per_area, y = reorder(district, price_per_area))) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    x = "متوسط سعر المتر المربع (ريال)",
    y = "الأحياء",
    title = "متوسط سعر المتر المربع في شرق الرياض"
  ) +
  theme_minimal() +
  theme(axis.text.y = element_text(angle = 0, hjust = 1))
```

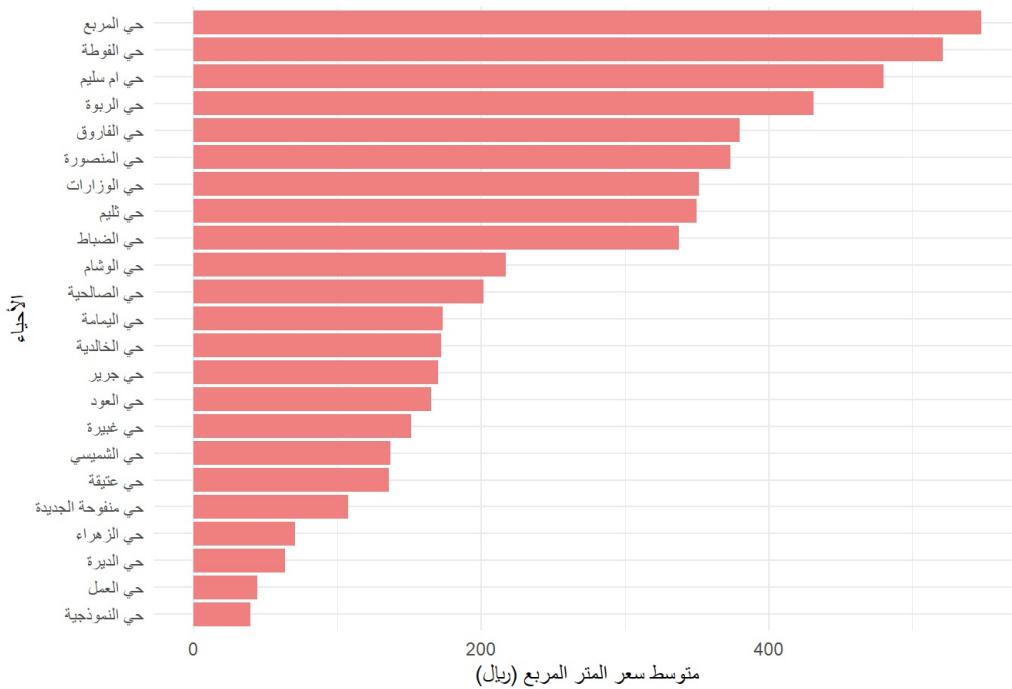
متوسط سعر المتر المربع في شرق الرياض



```
Central_riyadh <- subset(area, direction == "وسط الرياض")
Central_riyadh <- Central_riyadh %>%
  mutate(price_per_area = price / area)
Central_riyadh <- Central_riyadh %>%
  filter(district != "حي منفوجة")
avedis5 <- Central_riyadh %>%
  group_by(district) %>%
  summarise(price_per_area = mean(price_per_area, na.rm = TRUE))

ggplot(avedis5, aes(x = price_per_area, y = reorder(district, price_per_area))) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    x = "متوسط سعر المتر المربع (ريال)",
    y = "الأحياء",
    title = "متوسط سعر المتر المربع في وسط الرياض"
  ) +
  theme_minimal() +
  theme(axis.text.y = element_text(angle = 0, hjust = 1))
```

متوسط سعر المتر المربع في وسط الرياض



```
average_price <- mean(avedis5$price_per_area)
cat("متوسط سعر المتر المربع في وسط الرياض هو:", average_price, " ريال")\n
```

متوسط سعر المتر المربع في وسط الرياض هو: 244.5479 ##

```
west_riyadh <- subset(area, direction == "غرب الرياض")
west_riyadh$price_per_area <- west_riyadh$price / west_riyadh$area

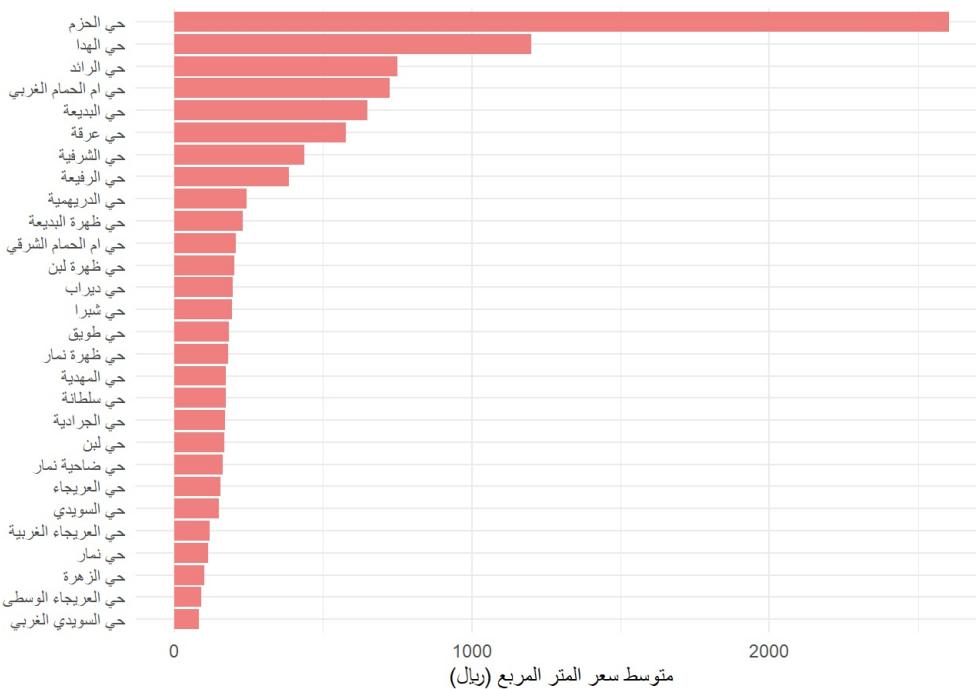
average_west <- mean(west_riyadh$price_per_area, na.rm = TRUE)
cat("متوسط سعر المتر المربع في غرب الرياض هو:", average_west, " ريال")\n
```

متوسط سعر المتر المربع في غرب الرياض هو: 116.0823 ##

```
avedis_west <- west_riyadh %>%
  group_by(district) %>%
  summarise(price_per_area = mean(price_per_area, na.rm = TRUE))

ggplot(avedis_west, aes(x = price_per_area, y = reorder(district, price_per_area))) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    x = "متوسط سعر المتر المربع (ريال)",
    y = "الأحياء",
    title = "متوسط سعر المتر المربع في غرب الرياض"
  ) +
  theme_minimal() +
  theme(axis.text.y = element_text(angle = 0, hjust = 1))
```

متوسط سعر المتر المربع في غرب الرياض



```
south_riyadh <- subset(area, direction == "جنوب")
south_riyadh$price_per_area <- south_riyadh$price / south_riyadh$area
```

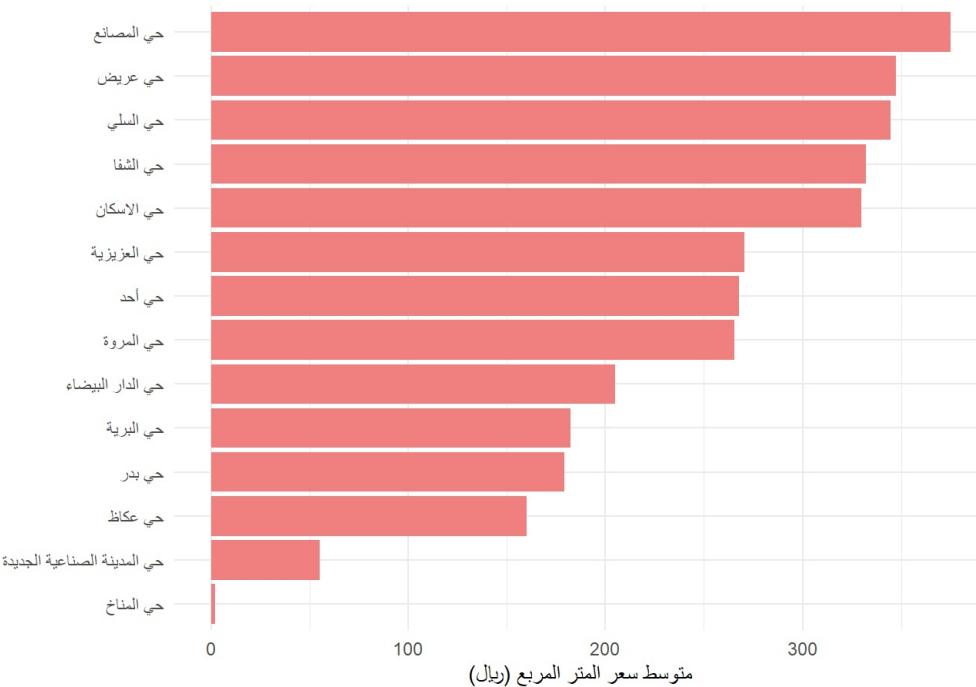
```
average_south <- mean(south_riyadh$price_per_area, na.rm = TRUE)
cat("متوسط سعر المتر المربع في جنوب الرياض هو:", average_south, "\n")
```

متوسط سعر المتر المربع في جنوب الرياض هو: 215.0021 ##

```
avedis_south <- south_riyadh %>%
  group_by(district) %>%
  summarise(price_per_area = mean(price_per_area, na.rm = TRUE))

ggplot(avedis_south, aes(x = price_per_area, y = reorder(district, price_per_area))) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    x = "متوسط سعر المتر المربع (ريال)",
    y = "الأحياء",
    title = "متوسط سعر المتر المربع في جنوب الرياض = "
  ) +
  theme_minimal() +
  theme(axis.text.y = element_text(angle = 0, hjust = 1))
```

متوسط سعر المتر المربع في جنوب الرياض



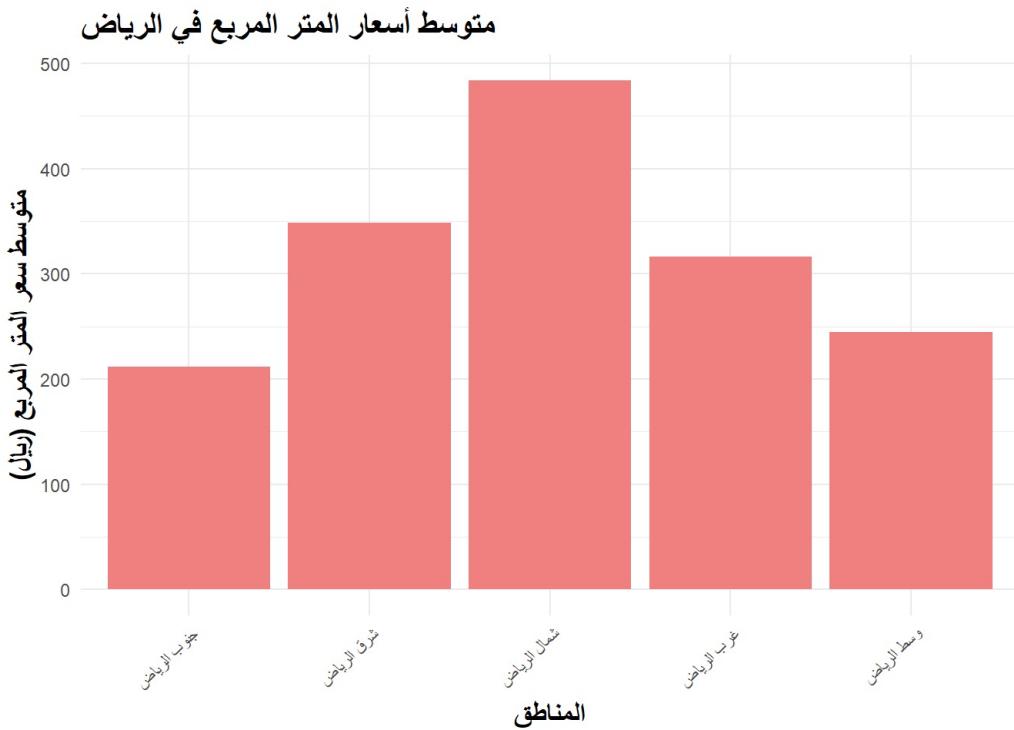
```

average_prices <- data.frame(
  شمال_الرياض = "شمال الرياض", "شرق_الرياض", "وسط_الرياض", "غرب_جليوبالهريفن") = المنطقه
  c = average, average_east, average_price, average_west, average_south)
)

library(ggplot2)

ggplot(average_prices, aes(x = المنطقه, y = متوسط_سعر_المتر_الربع)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(
    x = "المناطق",
    y = "(متوسط سعر المتر المربع (ريال",
    title = "متوسط أسعار المتر المربع في الرياض"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, family = "Arial"),
    axis.text.y = element_text(family = "Arial"),
    plot.title = element_text(family = "Arial", face = "bold", size = 16),
    axis.title = element_text(family = "Arial", face = "bold", size = 14)
  )
)

```



part four

Who are the sellers with the best reviews? Identifying the sellers with the best reviews provides valuable insights into the relationship between seller reputation and property prices.

```

# Function to compare sellers by average review score for a specific category
compare_top_sellers <- function(data, category_value, top_n = 10) {

  # Filter the data for the specific category
  filtered_data <- data %>% filter(category == category_value)

  # Remove outliers in the 'review' column
  filtered_data_no_outliers <- remove_outliers(filtered_data, "user.review")

  # Summarize the data by user name and category, after removing outliers
  review_summary <- filtered_data_no_outliers %>%
    group_by(user.name) %>%
    summarise(avg_review = mean(user.review, na.rm = TRUE), count = n()) %>%
    filter(count >= 10) %>% # Filter out sellers with less than 10 records
    arrange(desc(avg_review))

  # Get the top N districts by average review score
  top_sellers <- head(review_summary, top_n)
  #View(top_sellers)

  # Plot the top N districts by average review score
  ggplot(top_sellers, aes(x = reorder(user.name, avg_review), y = avg_review)) +
    geom_bar(stat = "identity", fill = "lightcoral") +
    labs(title = paste("أفضل 20 بائع حسب متوسط التقييم في ( فليفلة ٢٠١٩ )", x = "Sellers", y = "Average Review Score") +
    coord_flip() #+
    #theme(axis.text.y = element_text(family = "amiri", size = 10))
}

# Example usage:
compare_top_sellers(data_riyadh, category_value = 3, top_n = 20)

```



```
unique(compare_top_sellers(data_riyadh, category_value = 3, top_n = 20))
```

```

## [[1]]
## # A tibble: 20 × 3
##   user.name          avg_review count
##   <chr>              <dbl> <int>
## 1 55                 5     1
## 2 27                 5     1
## 3 11                 5     1
## 4 23                 5     1
## 5 14                 5     1
## 6 10                 5     1
## 7 14                 5     1
## 8 30                 5     1
## 9 17                 5     1
## 10 11                5     1
## 11 37                5     1
## 12 29                5     1
## 13 51                 5     1
## 14 45                 5     1
## 15 32                 5     1
## 16 62                 5     1
## 17 48                 5     1
## 18 19                 5     1
## 19 42                 5     1
## 20 46                 5     1

```

```

## 13 41      5          مكتب دار المستفيد للعقارات
## 14 14      5          مكتب عبيد للخدمات العقارية
## 15 21      5          مكتب مرسى الشمال للخدمات العقارية
## 16 11     4.85        عبدالله محمد
## 17 66     4.82        خالد فيصل
## 18 14     4.79        تركي
## 19 59     4.75        عبدالله
## 20 93     4.74        خالد مفرح
##                               مؤسسة بنيان العاصمة العقارية
##
## [[2]]
## [[2]][[1]]
## geom_bar: just = 0.5, width = NULL, na.rm = FALSE, orientation = NA
## stat_identity: na.rm = FALSE
## position_stack
##
##
## [[3]]
## <ggproto object: Class ScalesList, gg>
##   add: function
##   add_defaults: function
##   add_missing: function
##   backtransform_df: function
##   clone: function
##   find: function
##   get_scales: function
##   has_scale: function
##   input: function
##   map_df: function
##   n: function
##   non_position_scales: function
##   scales: list
##   train_df: function
##   transform_df: function
##   super: <ggproto object: Class ScalesList, gg>
##
## [[4]]
## <Guides[0] ggproto object>
##
## <empty>
##
## [[5]]
## Aesthetic mapping:
## * `x` -> `reorder(user.name, avg_review)`
## * `y` -> `avg_review`
##
## [[6]]
## list()
##
## [[7]]
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##   aspect: function
##   backtransform_range: function
##   clip: on
##   default: FALSE
##   distance: function
##   expand: TRUE
##   is_free: function
##   is_linear: function
##   labels: function
##   limits: list
##   modify_scales: function
##   range: function
##   render_axis_h: function
##   render_axis_v: function
##   render_bg: function
##   render_fg: function
##   setup_data: function
##   setup_layout: function
##   setup_panel_guides: function
##   setup_panel_params: function
##   setup_params: function
##   train_panel_guides: function
##   transform: function
##   super: <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##
## [[8]]
## <ggproto object: Class FacetNull, Facet, gg>
##   compute_layout: function
##   draw_back: function

```

```

##   draw_front: function
##   draw_labels: function
##   draw_panels: function
##   finish_data: function
##   init_scales: function
##   map_data: function
##   params: list
##   setup_data: function
##   setup_params: function
##   shrink: TRUE
##   train_scales: function
##   vars: function
##   super: <ggproto object: Class FacetNull, Facet, gg>
##
## [[9]]
## <environment: 0x000001a59cf8af0>
##
## [[10]]
## <ggproto object: Class Layout, gg>
##   coord: NULL
##   coord_params: list
##   facet: NULL
##   facet_params: list
##   finish_data: function
##   get_scales: function
##   layout: NULL
##   map_position: function
##   panel_params: NULL
##   panel_scales_x: NULL
##   panel_scales_y: NULL
##   render: function
##   render_labels: function
##   reset_scales: function
##   resolve_label: function
##   setup: function
##   setup_panel_guides: function
##   setup_panel_params: function
##   train_position: function
##   super: <ggproto object: Class Layout, gg>
##
## [[11]]
## [[11]]$x
## [1] "Sellers"
##
## [[11]]$y
## [1] "Average Review Score"
##
## [[11]]$title
## [1] "نوع حسب متوسط التقييم في ( فالييرا 2014 )"

```

Interactive Maps

1- Leaflet library This code creates an interactive map to visualize real estate data using the Leaflet package. It places markers at property locations and latitude and clusters nearby markers for better clarity. Each marker displays detailed property information in a popup when clicked, including the title, price, area, number of rooms, and available amenities like a pool . This allows users to explore and understand property details directly on the map in an intuitive and visually engaging way.

2- Shiny App We built an interactive web app that enables users to filter real estate listings based on criteria like price, bedrooms, and amenities. The filtered data is shown on a map using Leaflet, with color-coded markers representing different price categories. Users can click on the markers to view detailed property information. The app updates in real time according to the user's selections, providing an easy and intuitive way to explore the listings.

```

library(readr)
library(leaflet)
library(dplyr)
library(shiny)
df=read_csv("akar2024.csv")

# 1. Remove Outliers Using IQR
remove_outliers <- function(df, columns) {
  for (col in columns) {
    # Calculate the IQR (Interquartile Range)
    Q1 <- quantile(df[[col]], 0.25, na.rm = TRUE)
    Q3 <- quantile(df[[col]], 0.75, na.rm = TRUE)
    IQR <- Q3 - Q1

    # Define outlier bounds
    lower_bound <- Q1 - 1.5 * IQR
    upper_bound <- Q3 + 1.5 * IQR

    # Filter rows based on the outlier condition
    df <- df[df[[col]] >= lower_bound & df[[col]] <= upper_bound, ]
  }
  return(df)
}

# List of numerical columns to check for outliers
numerical_columns <- c('price', 'area', 'beds', 'wc', 'livings', 'rooms', 'age')

# Apply the outlier removal function
df <- remove_outliers(df, numerical_columns)

# 2. Remove districts with fewer than 5 records
district_counts <- table(df$district)
valid_districts <- names(district_counts[district_counts >= 5])

# Filter the dataframe to keep only valid districts
df <- df[df$district %in% valid_districts, ]

leaflet(data = df) %>%
  addTiles() %>%
  addMarkers(
    lng = ~location.lng, lat = ~location.lat,
    clusterOptions = markerClusterOptions(),
    popup = ~paste(
      "<b>Title:</b>", title, "<br>",
      "<b>Price:</b>", format(price, scientific = FALSE), "<br>",
      "<b>Area:</b>", area, "sqm<br>",
      "<b>Rooms:</b>", rooms, "<br>",
      "<b>Bedrooms:</b>", beds, "<br>",
      "<b>City:</b>", city, "<br>",
      "<b>District:</b>", district, "<br>",
      "<b>Furnished:</b>", ifelse(furnished == 1, "Yes", "No"), "<br>",
      "<b>Pool:</b>", ifelse(pool == 1, "Yes", "No"), "<br>",
      "<b>Playground:</b>", ifelse(playground == 1, "Yes", "No"), "<br>",
      "<b>Kitchen:</b>", ifelse(kitchen == 1, "Yes", "No"), "<br>",
      "<b>Lift:</b>", ifelse(lift == 1, "Yes", "No"), "<br>",
      "<b>Water Availability:</b>", ifelse(water_availability == TRUE, "Yes", "No"), "<br>",
      "<b>Electrical Availability:</b>", ifelse(electrical_availability == TRUE, "Yes", "No"), "<br>",
      "<b>Drainage Availability:</b>", ifelse(drainage_availability == TRUE, "Yes", "No")
    )
  )
)

```





```

# Load and preprocess the data
df <- df %>%
  filter(!is.na(location.lat) & !is.na(location.lng)) %>% # Filter out rows with missing coordinates
  mutate(price_category = cut(price, breaks = quantile(price, probs = c(0, 0.25, 0.5, 0.75, 1), na.rm = TRUE),
    labels = c("Low", "Medium", "High", "Very High"), include.lowest = TRUE))

  # Sample Data (replace df with actual data)
  # df <- your_data_frame

  # UI
  ui <- fluidPage(
    titlePanel("Real Estate Listings"),
    sidebarLayout(
      sidebarPanel(
        sliderInput("priceRange", "Price Range:", min = min(df$price, na.rm = TRUE),
        max = max(df$price, na.rm = TRUE), value = c(min(df$price, na.rm = TRUE), max(df$price, na.rm = TRUE))),
        sliderInput("bedrooms", "Number of Bedrooms:", min = min(df$beds, na.rm = TRUE),
        max = max(df$beds, na.rm = TRUE), value = c(min(df$beds, na.rm = TRUE), max(df$beds, na.rm = TRUE))),
        sliderInput("rooms", "Number of Rooms:", min = min(df$rooms, na.rm = TRUE),
        max = max(df$rooms, na.rm = TRUE), value = c(min(df$rooms, na.rm = TRUE), max(df$rooms, na.rm = TRUE))),
        sliderInput("livingRooms", "Number of Living Rooms:", min = min(df$livings, na.rm = TRUE),
        max = max(df$livings, na.rm = TRUE), value = c(min(df$livings, na.rm = TRUE), max(df$livings, na.rm = TRUE))),
        sliderInput("areaRange", "Area (sqm):", min = min(df$area, na.rm = TRUE),
        max = max(df$area, na.rm = TRUE), value = c(min(df$area, na.rm = TRUE), max(df$area, na.rm = TRUE))),
        checkboxInput("furnished", "Furnished", value = FALSE),
        checkboxGroupInput("amenities", "Amenities:", choices = c("pool", "playground",
        "kitchen", "lift", "water", "electricity", "drainage")),
        ),
        mainPanel(
          leafletOutput("mapOutput")
        )
      )
    )
  )

  # Server
  server <- function(input, output, session) {
    # Filter data based on input
    filteredData <- reactive({
      df %>%
        filter(
          price >= input$priceRange[1] & price <= input$priceRange[2],
          beds >= input$bedrooms[1] & beds <= input$bedrooms[2],
          rooms >= input$rooms[1] & rooms <= input$rooms[2],
          livings >= input$livingRooms[1] & livings <= input$livingRooms[2],
          area >= input$areaRange[1] & area <= input$areaRange[2],
          if (input$furnished) furnished == 1 else TRUE,
          if ("pool" %in% input$amenities) pool == 1 else TRUE,
          if ("playground" %in% input$amenities) playground == 1 else TRUE,
          if ("kitchen" %in% input$amenities) kitchen == 1 else TRUE,
          if ("lift" %in% input$amenities) lift == 1 else TRUE,
          if ("water" %in% input$amenities) water_availability == TRUE else TRUE,
          if ("electricity" %in% input$amenities) electrical_availability == TRUE else TRUE,
          if ("drainage" %in% input$amenities) drainage_availability == TRUE else TRUE
        )
    })
  }

  # Render the Leaflet map
  output$mapOutput <- renderLeaflet({
    leaflet(data = filteredData()) %>%
      addTiles() %>%
      addCircleMarkers(
        lng = ~location.lng, lat = ~location.lat,
        color = ~case_when(
          price_category == "Low" ~ "green",
          price_category == "Medium" ~ "orange",

```

```

    price_category == "High" ~ "red",
    TRUE ~ "purple"
),
radius = 5, fillOpacity = 0.8,
popup = ~paste(
  "<b>Title:</b>", title, "<br>",
  "<b>Price:</b>", scales::dollar(price), "<br>",
  "<b>Area:</b>", area, "sqm<br>",
  "<b>Rooms:</b>", rooms, "<br>",
  "<b>Living Rooms:</b>", livings, "<br>",
  "<b>Bedrooms:</b>", beds, "<br>",
  "<b>City:</b>", city, "<br>",
  "<b>District:</b>", district, "<br>",
  "<b>Furnished:</b>", ifelse(furnished == 1, "Yes", "No"), "<br>",
  "<b>Pool:</b>", ifelse(pool == 1, "Yes", "No"), "<br>",
  "<b>Playground:</b>", ifelse(playground == 1, "Yes", "No"), "<br>",
  "<b>Kitchen:</b>", ifelse(kitchen == 1, "Yes", "No"), "<br>",
  "<b>Lift:</b>", ifelse(lift == 1, "Yes", "No"), "<br>",
  "<b>Water Availability:</b>", ifelse(water_availability == TRUE, "Yes", "No"),
  "<b>Electrical Availability:</b>", ifelse(electrical_availability == TRUE,
  "<b>Drainage Availability:</b>", ifelse(drainage_availability == TRUE, "Yes",
  "No"))
)
) %>%
addLegend("bottomright", colors = c("green", "orange", "red", "purple"),
          labels = c("Low", "Medium", "High", "Very High"), title = "Price Category")
})
}

# Run the Shiny app
shinyApp(ui, server)

```

Shiny applications not supported in static R Markdown documents