1. **Biến Floors và Price**

A graph of different colored bars

Description automatically generated

Dùng Stacked Column Chart

**Xu hướng tăng giá theo số tầng**:

* Các căn nhà có **1 tầng** và **1.5 tầng** có giá trung bình thấp hơn so với các căn nhà có nhiều tầng hơn.
* Khi số tầng tăng lên (từ **2 tầng** đến **2.5 tầng**), giá trung bình tăng đáng kể, đạt mức cao nhất ở **2.5 tầng**.

**Giá giảm ở một số tầng cao hơn**:

* Số tầng **3.0** có giá trung bình thấp hơn so với **2.5 tầng**, điều này có thể do các yếu tố khác như diện tích, vị trí, hoặc thiết kế nhà.

**Số tầng 3.5 vượt trội**:

* Các căn nhà **3.5 tầng** có giá trung bình cao gần bằng **2.5 tầng**, có thể là do thiết kế sang trọng hoặc vị trí đặc biệt.

1. **Biến Waterfront và Price**

A screenshot of a graph

Description automatically generated

Dùng Clusterd Column Chart

**Giá nhà ở gần mặt nước (waterfront = 1) cao hơn nhiều**:

* Các căn nhà gần mặt nước (**waterfront = 1**) có giá trung bình cao hơn rõ rệt, gần **2 triệu USD**.
* Trong khi đó, các căn nhà không gần mặt nước (**waterfront = 0**) chỉ có giá trung bình khoảng **0.5 triệu USD**.

**Waterfront là yếu tố có ảnh hưởng mạnh đến giá nhà**:

* Sự khác biệt lớn giữa hai nhóm cho thấy vị trí gần mặt nước là yếu tố quan trọng làm tăng giá trị bất động sản.

1. **Biến Condition và Price**

A graph with different colored bars

Description automatically generated

Dùng Clusterd Column Chart

**Xu hướng giá tăng theo tình trạng nhà tốt hơn**:

* Những ngôi nhà có tình trạng tốt hơn (**condition = 5**) có giá trung bình cao nhất, khoảng **0.6 triệu USD**.
* Trong khi đó, nhà ở tình trạng kém hơn (**condition = 1**) có giá trung bình thấp nhất, khoảng **0.2 triệu USD**.

**Tăng giá đáng kể từ condition = 3 trở lên**:

* Giá trung bình tăng mạnh khi tình trạng từ **condition = 3** trở lên, cho thấy chất lượng nhà là một yếu tố ảnh hưởng quan trọng đến giá trị bất động sản.

**Phân phối giá ổn định trong các điều kiện trung bình**:

* Từ **condition = 2 đến condition = 4**, sự khác biệt về giá không quá lớn

1. 4 biến và price

A screenshot of a graph

Description automatically generated

**1. sqft\_living vs Price (Biểu đồ đầu tiên bên trái trên):**

* Có mối tương quan dương rõ ràng giữa diện tích sử dụng trong nhà (sqft\_living) và giá bán (price).
* Đường hồi quy màu đỏ cho thấy rằng khi diện tích sử dụng tăng, giá bán có xu hướng tăng.
* Tuy nhiên, một số điểm dữ liệu có giá trị rất cao (outliers), điều này có thể là do những ngôi nhà sang trọng hoặc đặc biệt.

**2. sqft\_lot vs Price (Biểu đồ bên phải trên):**

* Tương quan giữa diện tích lô đất (sqft\_lot) và giá bán ít rõ ràng hơn so với sqft\_living.
* Mặc dù đường hồi quy màu xanh cho thấy xu hướng tăng giá theo diện tích lô đất, nhưng dữ liệu khá phân tán, cho thấy diện tích lô đất không phải là yếu tố chính quyết định giá.

**3. sqft\_living15 vs Price (Biểu đồ bên trái dưới):**

* sqft\_living15 (diện tích sử dụng trung bình của 15 nhà xung quanh) có mối quan hệ tích cực với giá bán, tương tự như sqft\_living.
* Đường hồi quy màu xanh lá chỉ ra xu hướng giá bán tăng khi giá trị sqft\_living15 tăng. Điều này cho thấy giá trị của một ngôi nhà có thể bị ảnh hưởng bởi khu vực lân cận.

**4. sqft\_lot15 vs Price (Biểu đồ bên phải dưới):**

* sqft\_lot15 (diện tích lô đất trung bình của 15 nhà xung quanh) có mối quan hệ yếu với giá bán, tương tự sqft\_lot.
* Dữ liệu khá phân tán và không thể hiện mối tương quan chặt chẽ.
  + **sqft\_living** và **sqft\_living15** có tác động mạnh mẽ hơn đến giá bán so với **sqft\_lot** và **sqft\_lot15**.

A screen shot of a graph

Description automatically generated

**Mối quan hệ với Price:**

* **sqft\_living** và **sqft\_living15** (các điểm xanh và xanh lá cây) vẫn cho thấy mối tương quan tích cực và rõ ràng hơn so với **sqft\_lot** và **sqft\_lot15** (các điểm màu cam và đỏ).
* Đường hồi quy màu xanh dương đại diện cho **sqft\_living** và màu tím cho **sqft\_lot** cho thấy xu hướng giá tăng khi diện tích tăng, nhưng mức độ dốc của các đường khác nhau.

**Sự phân tán dữ liệu:**

* Dữ liệu của **sqft\_lot** và **sqft\_lot15** rất phân tán ở giá trị thấp, đặc biệt ở lô đất lớn hơn nhưng giá không tăng rõ ràng. Điều này chứng minh diện tích lô đất không phải yếu tố chính ảnh hưởng đến giá bất động sản.
* Ngược lại, **sqft\_living** và **sqft\_living15** tập trung và có xu hướng mạnh mẽ hơn trong việc dự đoán giá nhà.

**Outliers:**

* Một số điểm bất thường (các bất động sản giá trị rất cao) xuất hiện trong cả bốn biến, đặc biệt ở những giá trị lớn của diện tích. Điều này có thể đại diện cho các bất động sản cao cấp hoặc có đặc điểm đặc biệt.

1. 2 biến sqft\_living và sqft\_living15

A screen shot of a graph

Description automatically generated

**Mối tương quan giữa sqft\_living, sqft\_living15 và Price:**

* **sqft\_living** (điểm xanh) có xu hướng tăng rõ ràng theo giá bán, điều này cho thấy diện tích sử dụng thực tế của ngôi nhà là một yếu tố quan trọng ảnh hưởng đến giá.
* **sqft\_living15** (điểm cam) cũng cho thấy mối tương quan tích cực với giá, nhưng mức độ dốc có vẻ thấp hơn một chút so với **sqft\_living**, gợi ý rằng giá trị nhà xung quanh có ảnh hưởng nhưng không mạnh bằng diện tích sử dụng của ngôi nhà đang xét.

**Đường hồi quy:**

* Đường hồi quy màu đỏ dường như phù hợp tốt với dữ liệu, chỉ ra rằng cả hai yếu tố đều góp phần giải thích giá bán.

**Sự phân tán dữ liệu:**

* Dữ liệu của cả hai yếu tố tương đối tập trung, nhưng vẫn có một số điểm ngoại lai (outliers) ở giá trị cao hơn. Những ngôi nhà này có thể là nhà cao cấp hoặc có các đặc điểm độc đáo khác ngoài diện tích sử dụng.

**Phạm vi dữ liệu:**

* **sqft\_living** và **sqft\_living15** đều trải dài từ khoảng 0 đến hơn 12,000 sqft, nhưng phần lớn tập trung dưới 4,000 sqft. Điều này cho thấy hầu hết các ngôi nhà trong tập dữ liệu thuộc dạng nhà ở tiêu chuẩn, không quá lớn.

1. 2 biến sqft\_lot và sqft\_lot15

A screen shot of a graph

Description automatically generated

**Mối quan hệ giữa diện tích đất (sqft\_lot, sqft\_lot15) và giá nhà (Price):**

* Có sự phân tán khá lớn của dữ liệu, đặc biệt ở những giá trị diện tích nhỏ. Điều này cho thấy diện tích đất không phải là yếu tố duy nhất ảnh hưởng mạnh đến giá nhà.
* Giá nhà tăng lên nhẹ khi diện tích đất lớn hơn, nhưng mối quan hệ này có vẻ không tuyến tính rõ ràng.

**Đường hồi quy:**

* Đường hồi quy với bóng mờ (confidence interval) cho thấy mối quan hệ giữa diện tích đất và giá nhà là dương, tức là khi diện tích đất tăng, giá nhà có xu hướng tăng.
* Tuy nhiên, độ nghiêng của đường hồi quy khá nhỏ, nghĩa là tác động của diện tích đất lên giá nhà có thể không quá mạnh.

**So sánh giữa sqft\_lot và sqft\_lot15:**

* Hai biến diện tích này (trước và sau cải tạo) có xu hướng tương tự nhau trong việc giải thích giá nhà.
* Không có sự khác biệt rõ ràng giữa ảnh hưởng của sqft\_lot (diện tích ban đầu) và sqft\_lot15 (diện tích sau cải tạo).

Bổ sung:

A screenshot of a computer screen

Description automatically generated

**1. Biểu đồ trên (Bedrooms và Price):**

* **Mối quan hệ:**
  + Số lượng phòng ngủ (Bedrooms) và giá nhà (Price) có một mối quan hệ tuyến tính yếu. Khi số phòng ngủ tăng, giá nhà có xu hướng tăng, nhưng không quá mạnh.
  + Biểu đồ cho thấy phần lớn các căn nhà có từ 2 đến 6 phòng ngủ, với giá nhà chủ yếu nằm trong khoảng dưới 1 triệu USD.
* **Ngoại lệ (Outliers):**
  + Một số điểm dữ liệu có giá trị phòng ngủ rất cao (trên 10 và thậm chí hơn 30), nhưng đây là các trường hợp hiếm gặp. Điều này có thể đại diện cho những bất động sản rất đặc biệt hoặc dữ liệu sai lệch cần kiểm tra thêm.
  + Một số căn nhà có giá rất cao (>6 triệu USD) nhưng không tỷ lệ thuận với số lượng phòng ngủ, cho thấy rằng các yếu tố khác (như vị trí, diện tích sống) quan trọng hơn.

**2. Biểu đồ dưới (Bathrooms, Bedrooms, Sqft Living và Price):**

* **Mối quan hệ tổng quan:**
  + Biểu đồ thể hiện mối quan hệ giữa nhiều yếu tố (số phòng ngủ, số phòng tắm, diện tích sống) và giá nhà.
  + Dữ liệu cho thấy diện tích sống (Sqft Living) có mối tương quan mạnh nhất với giá nhà, biểu hiện qua sự phân bố dọc theo đường hồi quy của diện tích sống.
  + Số phòng ngủ và phòng tắm có tác động đến giá nhà, nhưng mức độ ảnh hưởng không cao bằng diện tích sống.
* **Phân bố dữ liệu:**
  + Phần lớn các căn nhà nằm trong phạm vi diện tích sống từ 1000 đến 4000 sqft, giá dưới 1 triệu USD.
  + Một số bất động sản với diện tích lớn (>10.000 sqft) và giá trị cao là ngoại lệ.