Model Training

Full report of model Predict price (regression)

XGBoost (s2) - v1

Model

Model ID	**S-FINALPROJECT_MINHAL_KHURRUM-eiAJ88IP-17...**
Model type	**Regression**
Target	**price**
Backend	**Python (in memory)**
Algorithm	**Xgboost regression**
Trained on	**2025/05/27 13:26**
Columns	**19**
Train set rows	**17309**
Test set rows	**4304**
Weighting method	**No weighting**
Calibration method	**No calibration**
Code Env	**DSS builtin env**
Python version	**3.9.20**

Dataset
Model
Custom

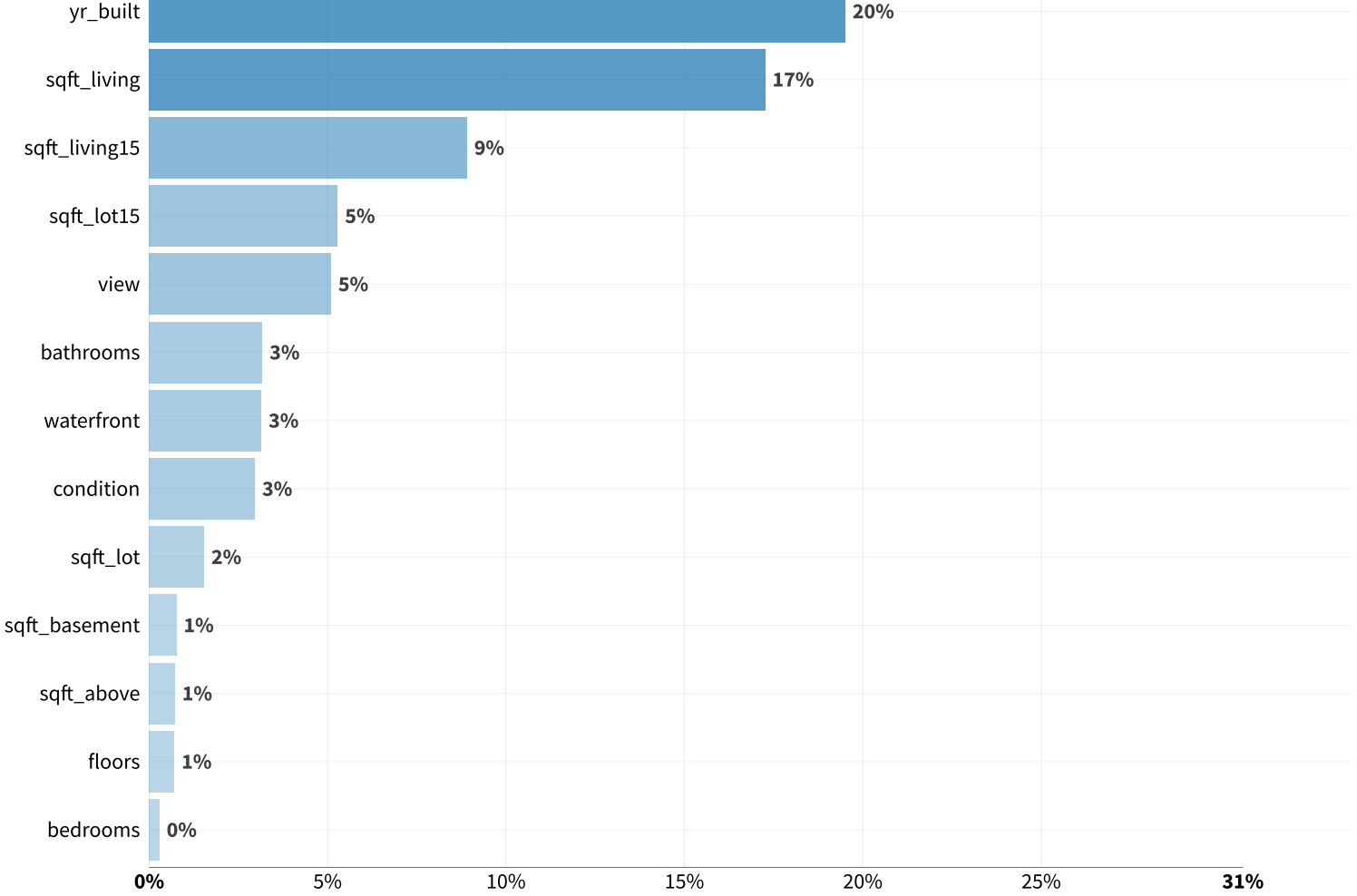
Metadata

trainDataset:dataset-name	→	kc_house_data_prepared
testDataset:dataset-name	→	kc_house_data_prepared
evaluationDataset:dataset-name	→	kc_house_data_prepared
model:algorithm	→	XGBOOST_REGRESSION
model:date	→	2025-05-27T13:27:16.344+0000
evaluation:date	→	2025-05-27T13:27:16.344+0000
model:name	→	XGBoost (s2) - v1

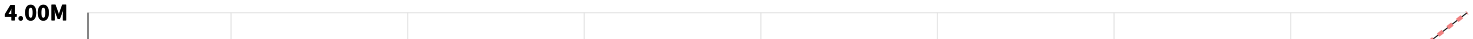
+ ADD A LABEL

Optional. Informative labels for the model. The *model:algorithm*, *model:date*, *model:name*, *trainDataset:dataset-name*, *testDataset:dataset-name* labels, *evaluation:date* and *evaluationDataset:dataset-name* are automatically added.

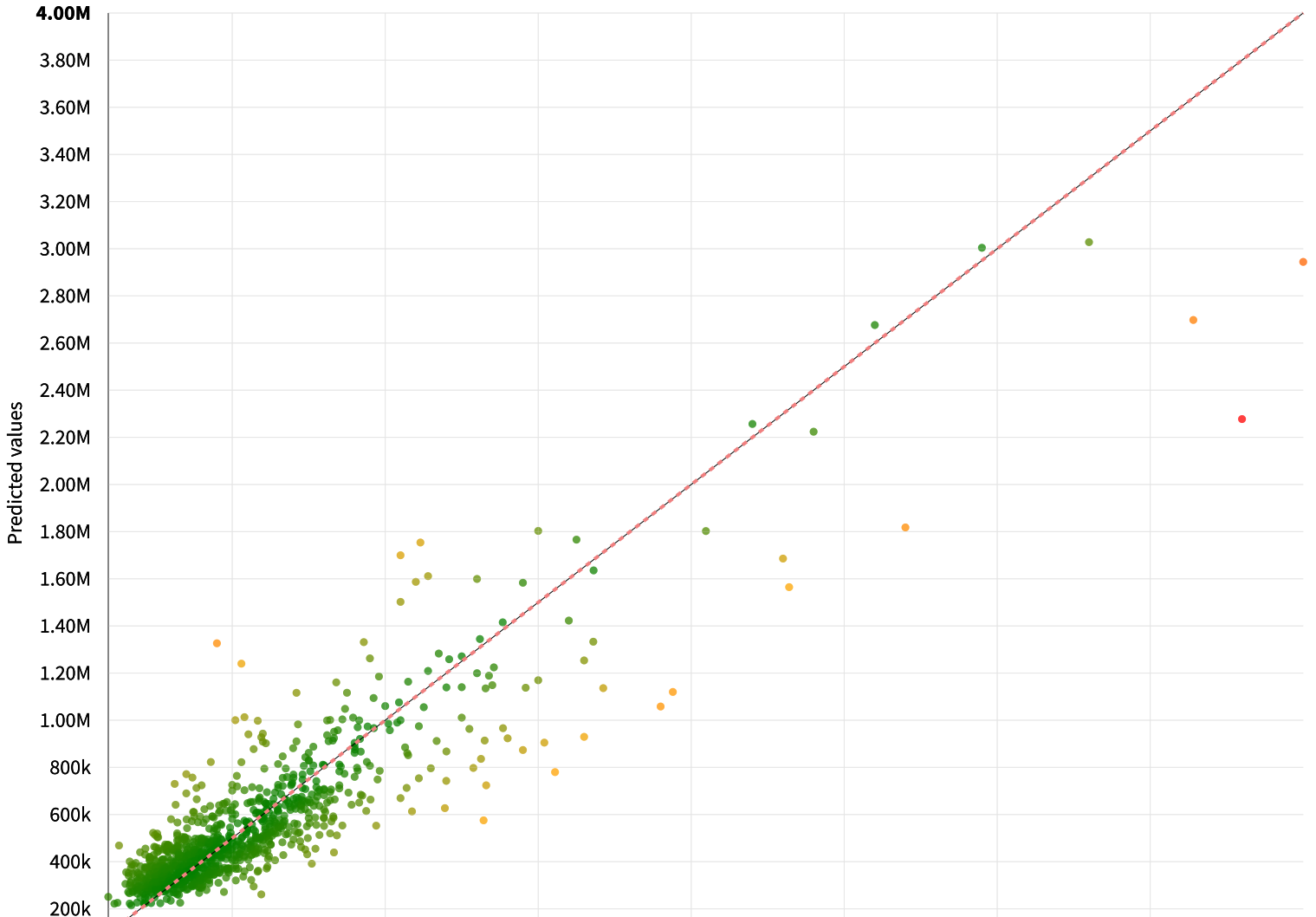
FEATURE IMPORTANCE: Predict price (regression) table



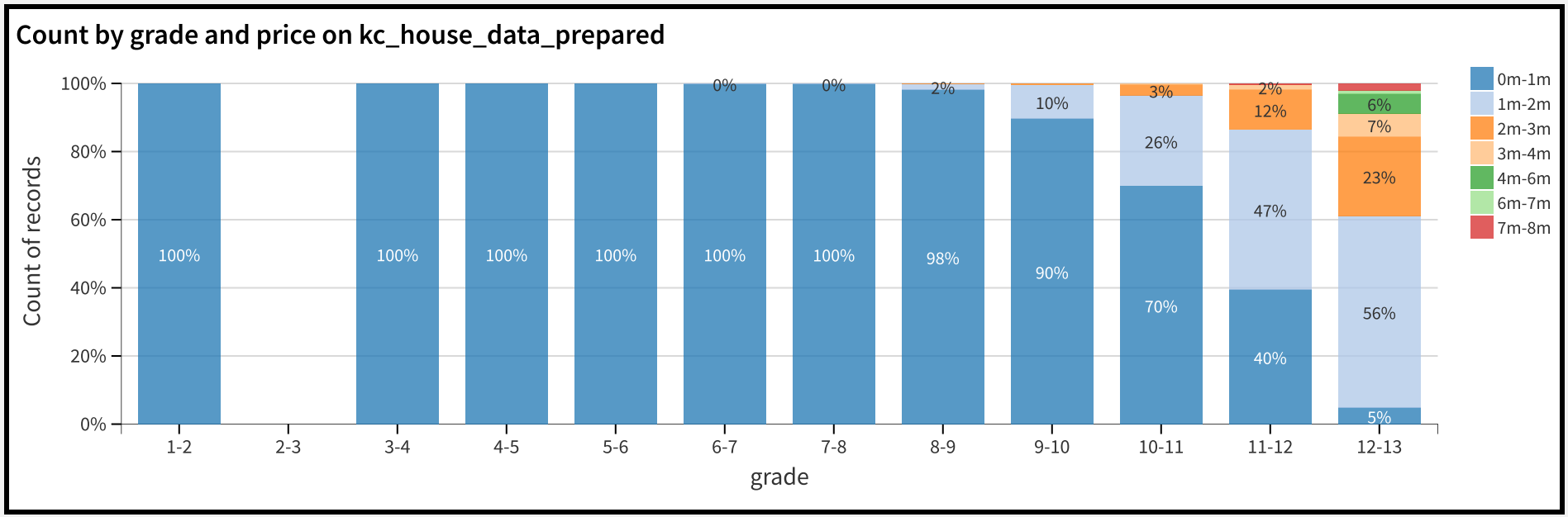
SCATTER PLOT: Predict price (regression) table



SCATTER PLOT: Predict price (regression) table



Charts by Minhal

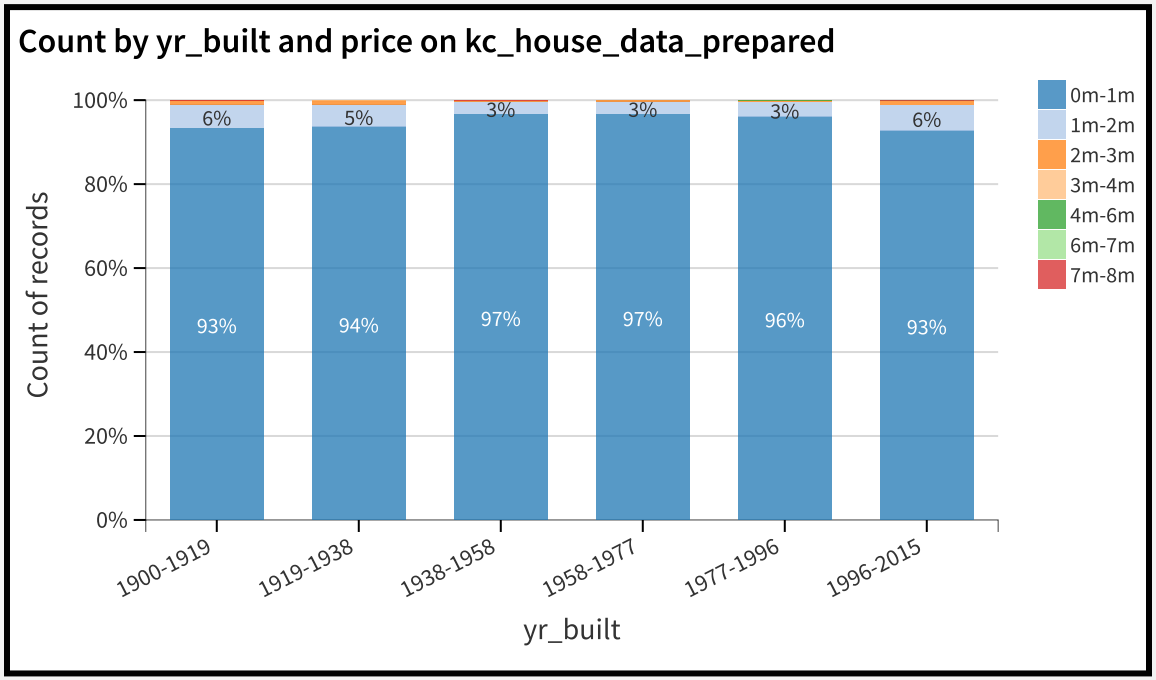


> This graph represents **count of records** by **grade** and **price** . From this graph, I analyze that most of the houses lies in between the range of 0-1 million of price range.

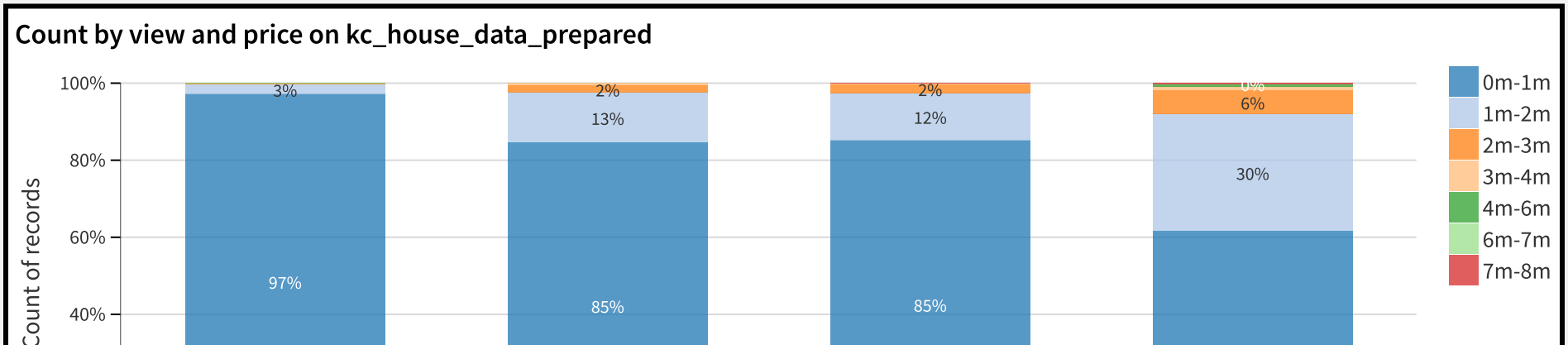
If I talk about ***grade*** column then we have values in between 1-13, so we get:

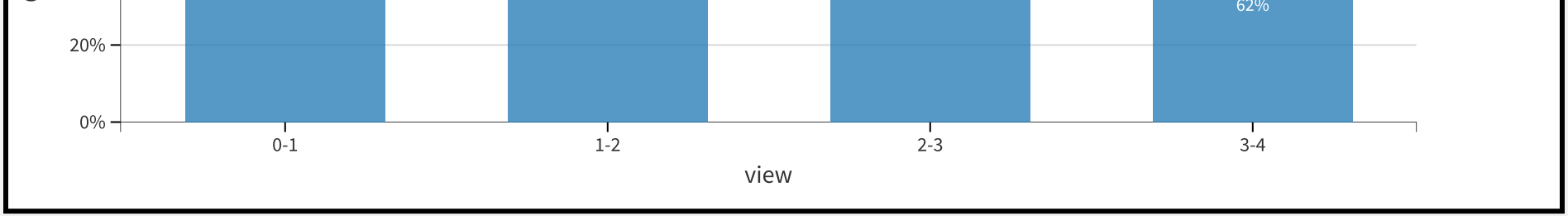
- **Grade 1-2:** 100% house pricing between 0-1M.
- **Grade 2-3:** No house available
- **Grade 3-4:** 100% house pricing between 0-1M.
- **Grade 4-5:** 100% house pricing between 0-1M.
- **Grade 5-6:** 100% house pricing between 0-1M.
- **Grade 6-7:** 100% house pricing between 0-1M.
- **Grade 7-8:** 100% house pricing between 0-1M.
- **Grade 8-9:** 98% house pricing between 0-1M & 2% between 1-2M
- **Grade 9-10:** 90% house pricing between 0-1M & 10% between 1-2M

- **Grade 10-11:** 70% house pricing between 0-1M, 26% between 1-2M & 3% between 2-3M.
- **Grade 11-12:** 40% house pricing between 0-1M, 47% between 1-2M, 12% between 2-3M & 2% between 3-4M.
- **Grade 12-13:** 5% house pricing between 0-1M, 56% between 1-2M, 23% between 2-3M, 7% between 3-4M, 6% between 4-6M & 2% between 7-8M.



This graph represents **count of records** by **year built** and **price** . From this graph, I analyze that most of the houses lies in between the range of 0-1 million of price range.

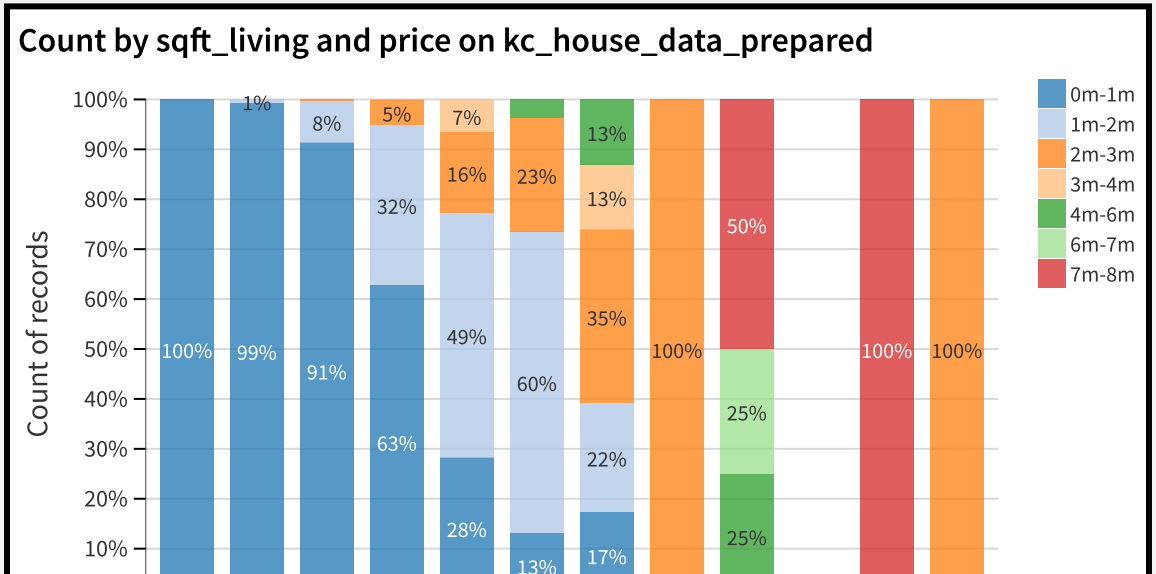




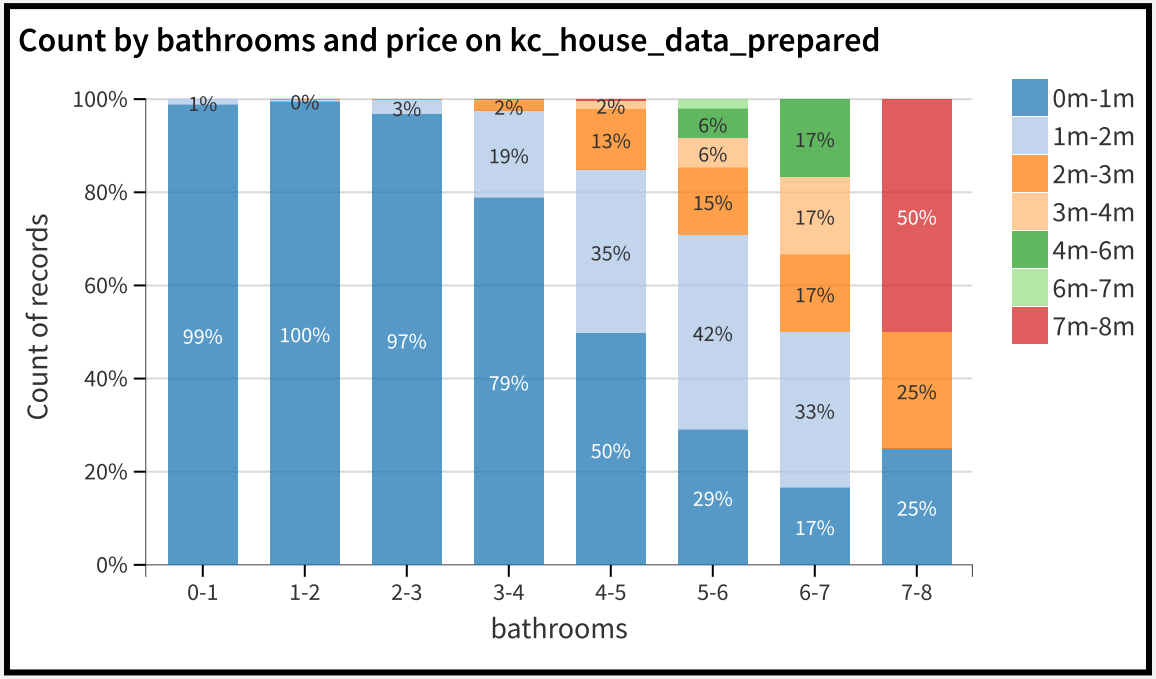
This graph represents **count of records** by **view** and **price** . From this graph, I analyze that most of the houses lies in between the range of 0-1 million of price range and some of houses lie in between 1-2M as well.

If I talk about ***view*** column then we have values in between 0-4, so we get:

- **View 0-1:** 97% house pricing between 0-1M & 3% between 1-2M
- **View 1-2:** 85% house pricing between 0-1M, 13% between 1-2M & 2% between 2-3M.
- **View 2-3:** 85% house pricing between 0-1M, 12% between 1-2M & 2% between 2-3M.
- **View 3-4:** 62% house pricing between 0-1M, 30% between 1-2M, 6% between 2-3M, 1% between 3-4M & 1% between 4-6M.

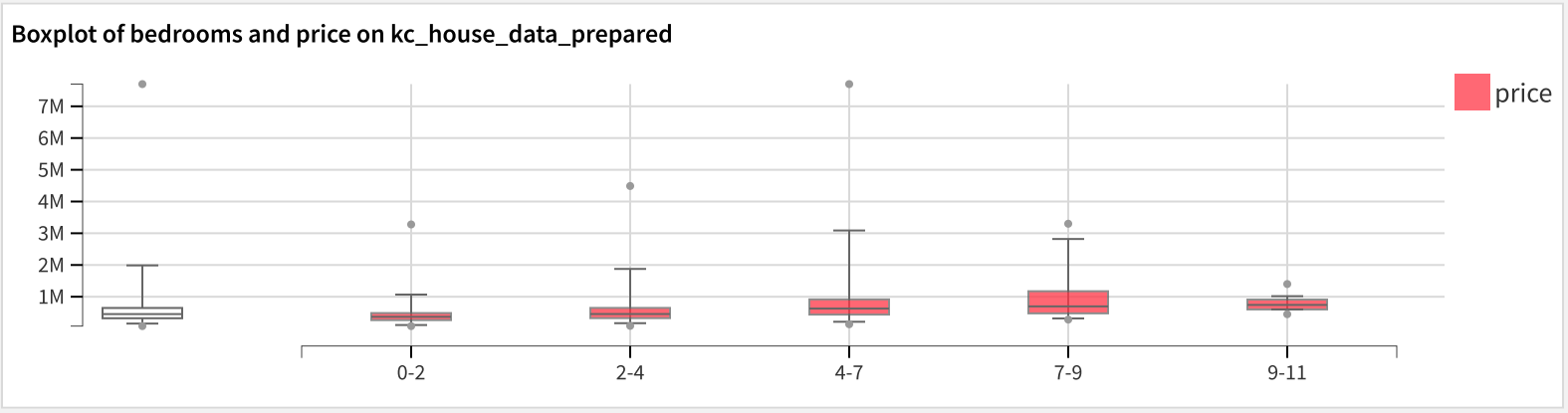


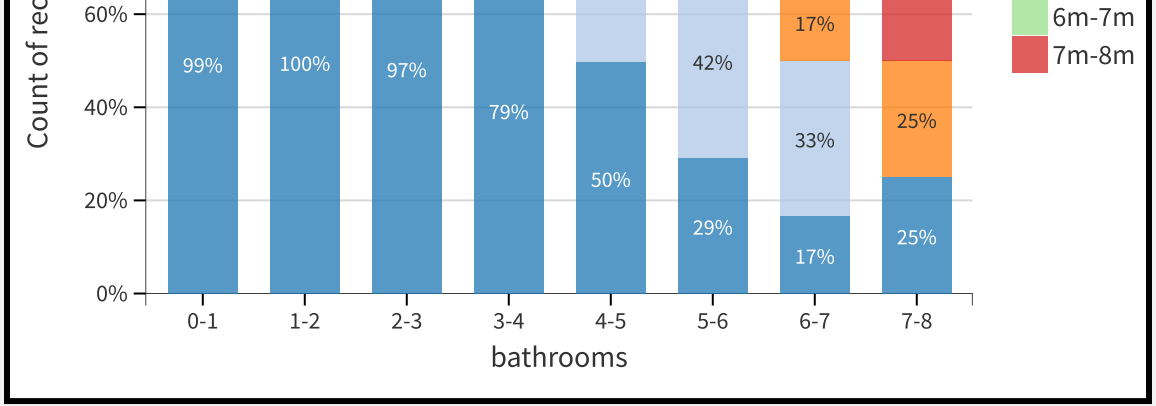
This graph represents **count of records** by **square feat. living** and **price** . From this graph, I analyze that as the number of sqft. are increasing the the pricing are increasing gradually.
Where, in the last 2 bars where range is ***11.3k- 12.4k*** the price of houses is **7-8M 100%** and where the price is ***12.4k - 13.5k*** the price lies in between **2-3M 100%.**



This graph represents **count of records** by **bathrooms** and **price** . From this graph, I analyze that:

- **From number of bathrooms (0-4),** most of the houses lies in between the range of 0-1 million of price.
- **From number of bathrooms (5-7),** there is increment of house whose price lies in between 1-2M as well.
- **For bathroom (7-8),** 50% of houses prices are lying in between 7-8M.



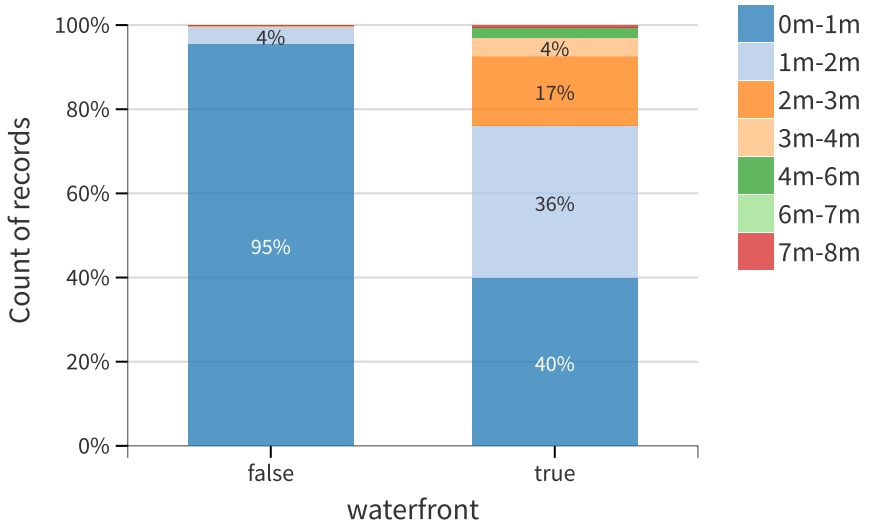


- **From number of bathrooms (5-7),** there is increment of house whose price lies in between 1-2M as well.
- **For bathroom (7-8),** 50% of houses prices are lying in between 7-8M.

Boxplot of bedrooms and price on kc_house_data_prepared

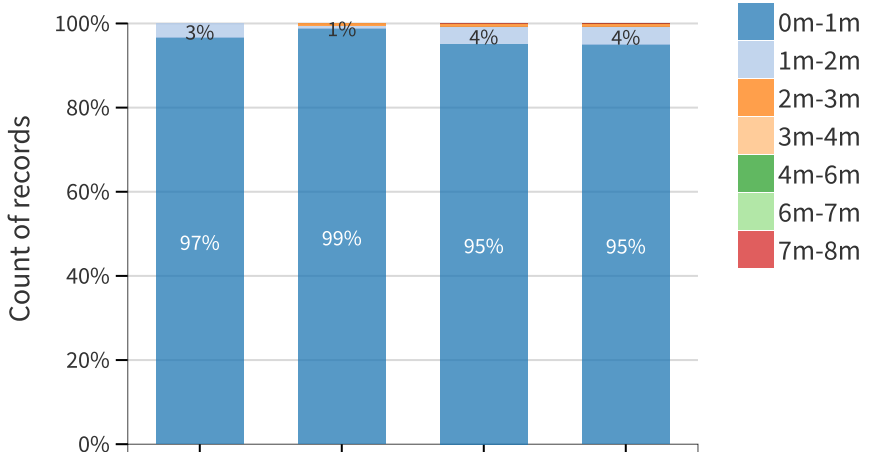
Charts by Khurrum

Count by waterfront and price on kc_house_data_prepared

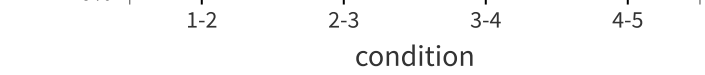


Waterfront homes are way more expensive. Only 40% of them are under $1M, while most non-waterfront homes (95%) are under $1M. About 17% of waterfront homes cost $2M–$3M. So, being near water clearly pushes the price up a lot.

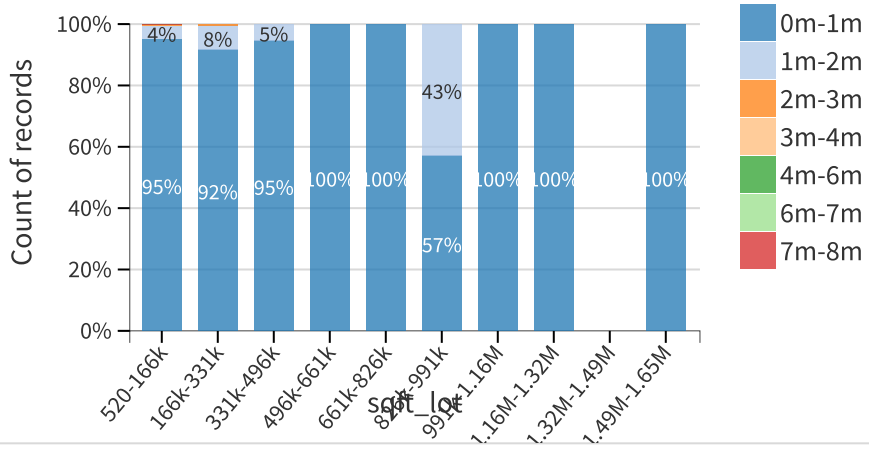
Count by conditions and price on kc_house_data_prepared



No matter the condition, nearly all homes (95–99%) are under $1M. Even top-condition homes aren't selling for much more. This means condition doesn't impact price as much as you'd think — other things matter more.

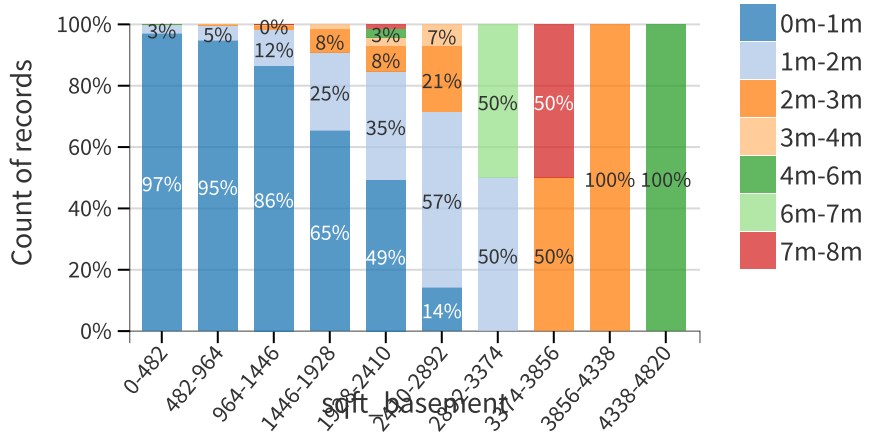


Count by sqft_lot and price on kc_house_data_prepared



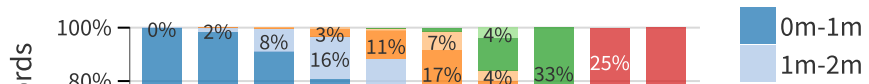
Smaller lots mostly have cheaper homes. But once the lot size goes above 991K sqft, prices jump. Homes on huge lots are always above $1M. Bigger land = higher price, simple as that.

Count by sqft_basement and price on kc_house_data_prepared

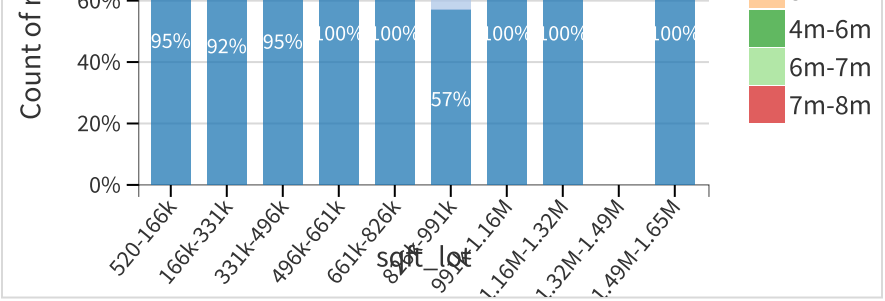


Most homes with small or no basements (under 1,000 sqft) are priced under $1M. As basement size grows, prices start climbing. Between 2,000–2,800 sqft, there's a mix of prices from $1M to $4M. Homes with basements over 3,300 sqft are all luxury — priced between $4M and $7M. Bigger basements clearly push the price higher

Count by sqft_above and price on kc_house_data_prepared



Small homes (under 3,000 sqft) are almost always under $1M. Once the size hits 4,000+ sqft, higher prices appear. Big houses (7,500+


Count of r
60%
40%
20%
0%
95%
92%
95%
100%
100%
57%
100%
100%
100%
520-166k
166k-331k
331k-496k
496k-661k
661k-826k
826k-991k
991k-1.16M
1.16M-1.32M
1.32M-1.49M
1.49M-1.65M
sqft_lo
4m-6m
6m-7m
7m-8m

Clusters

Full report of model Clustering on kc_house_data_prepared

KMeans (k=3) (s1) - v1

KMeans (k=3)

Cluster outliers

Trained in 16 seconds on 21613 records

Spacious Living Area		**3376 (15.62%)**
New Built House Year		6269 (29.01%)
Average House Condition		11568 (53.52%)
Spacious LOT		400 (1.85%)

Spacious Living Area

Enter a description for this cluster

Observations

- **sqft_living** is in average 69.13% greater : mean of 3518 against 2080 globally
- **grade** is in average 22.33% greater : mean of 9.37 against 7.66 globally
- **sqft_living15** is in average 49.90% greater : mean of 2978 against 1987 globally

Model

Model ID	**S-FINALPROJECT_MINHAL_KHURRUM-FYFZE2Zq-1748376338632**
Model type	**Clustering**
Code Env	**DSS builtin env**
Python version	**3.9.20**

Metadata

trainDataset:dataset-name	→	kc_house_data_prepared
testDataset:dataset-name	→	kc_house_data_prepared
evaluationDataset:dataset-name	→	kc_house_data_prepared
model:algorithm	→	KMEANS
model:date	→	2025-05-27T15:46:09.769+0000
evaluation:date	→	2025-05-27T15:46:09.769+0000
model:name	→	KMeans (k=3) (s1) - v1

+ ADD A LABEL

Optional. Informative labels for the model. The *model:algorithm*, *model:date*, *model:name*, *trainDataset:dataset-name*, *testDataset:dataset-name* labels, *evaluation:date* and *evaluationDataset:dataset-name* are automatically added.

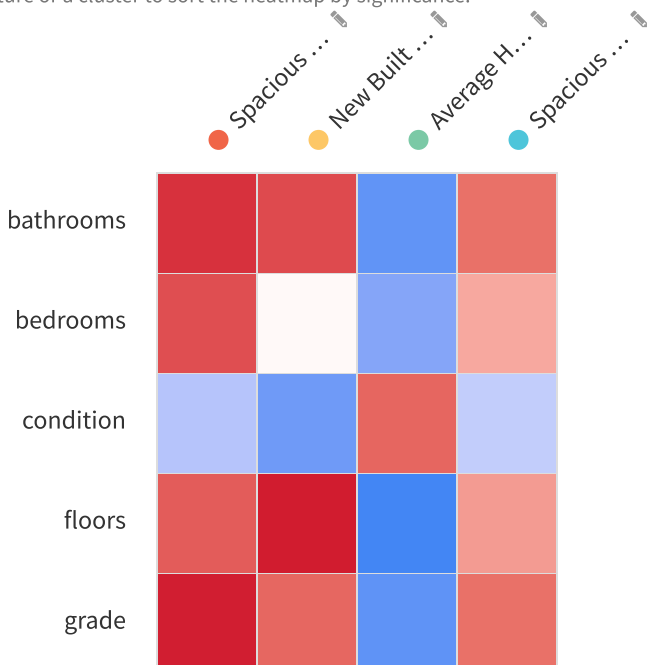
Clustering on kc_house_data_prepared table

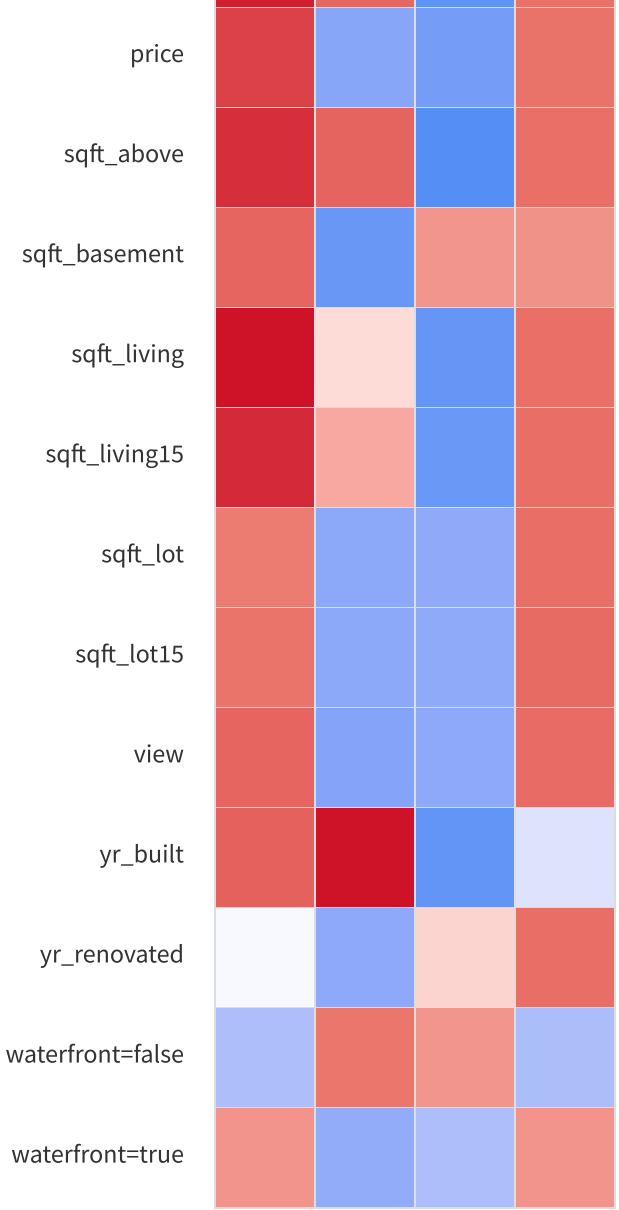
View features: ALL | NUMERIC | SINGLE CATEGORICAL

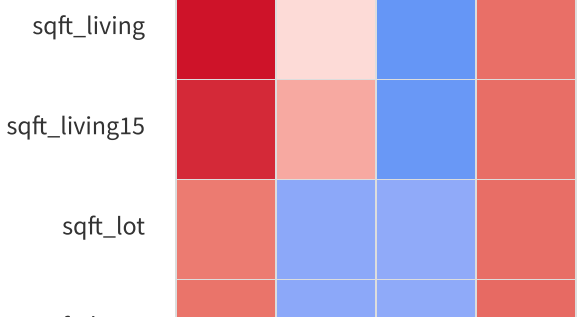
Low High

EXPORT

Click on a feature or a cluster to sort the heatmap by significance.




price
sqft_above
sqft_basement
sqft_living
sqft_living15
sqft_lot
sqft_lot15
view
yr_built
yr_renovated
waterfront=false
waterfront=true


sqft_living
sqft_living15
sqft_lot