



HOUSIE

MariBisnis Capstone Project

Nafisan Nauroti Salsa Bila

DAI-005

Data and Artificial Intelligence
Microsoft



≡ About this Project

MariBisnis ingin mengetahui trend bisnis dari penjualan rumah yang ada di Amerika Serikat. Harapannya, MariBisnis dapat untuk memprediksi harga suatu rumah dan melakukan pemetaan terhadap sebaran data yang ada.



≡ D a t a s e t → ←

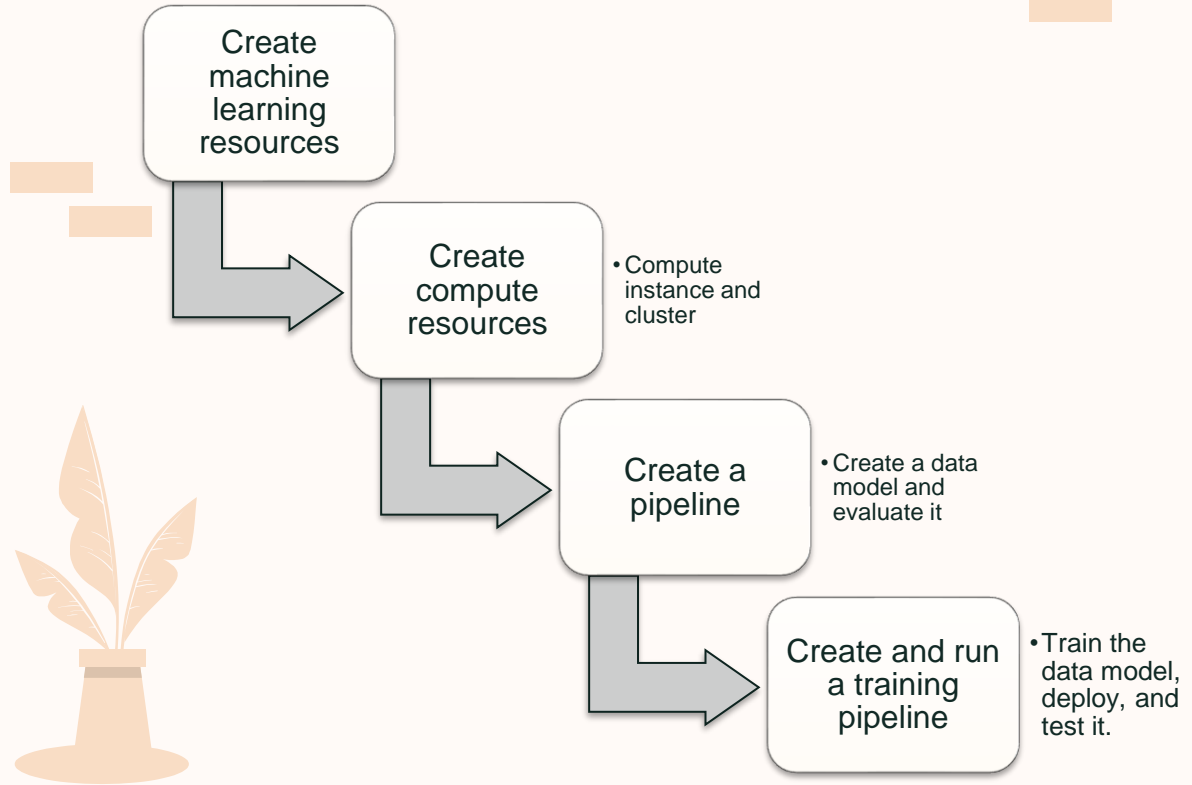
21 features that need to be explained and processed:

Column name	Description	Column name	Description
id	A notation for a house	grade	The overall grade is given to the housing unit, based on the King County
date	Date house was sold		
price	Price in the prediction target	sqft_above	Sqft of house apart from the basement
bedrooms	Number of bedrooms/house	sqft_basement	Sqft of the basement
bathrooms	Number of bathrooms/bedrooms	yr_built	The year when the house was built
sqft_living	Square footage of the house	yr_renovated	The year when the house was renovated
sqft_lot	Square footage of the lot	zipcode	Zip code
floors	Total floors (levels) in the house	lat	Latitude coordinate
waterfront	House which has a view to a waterfront	long	Longitude coordinate
view	Has been viewed	sqft_living15	Square footage living in 2015
condition	How good the condition is overall	sqft_lot15	Square footage lot in 2015

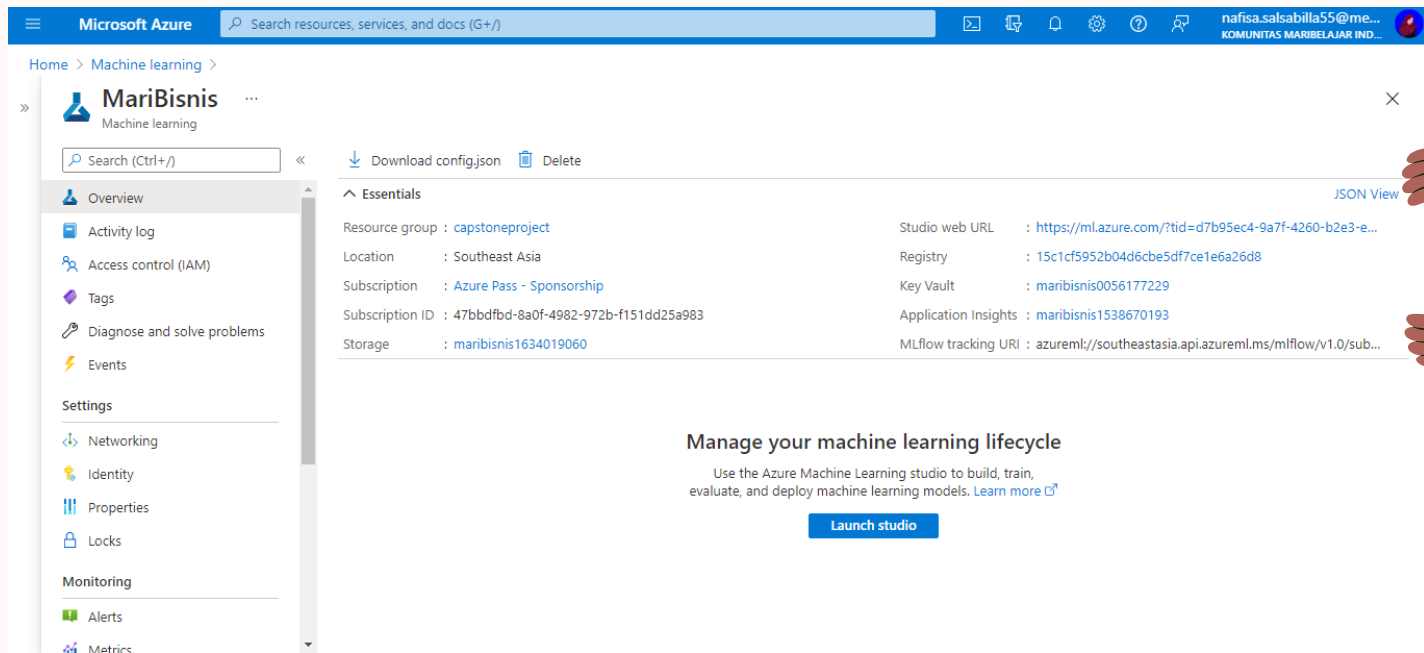
Link download:

<https://onedrive.live.com/?authkey=%21AKqjhQKiqhcKPho&id=1A1C328B89C1FDC0%21455863&cid=1A1C328B89C1FDC0>

≡ Pre-processing and Processing



≡ Create Machine Learning Resources



Microsoft Azure Search resources, services, and docs (G+/)

Home > Machine learning >

MariBisnis Machine learning

Search (Ctrl+/)

Download config.json Delete

JSON View

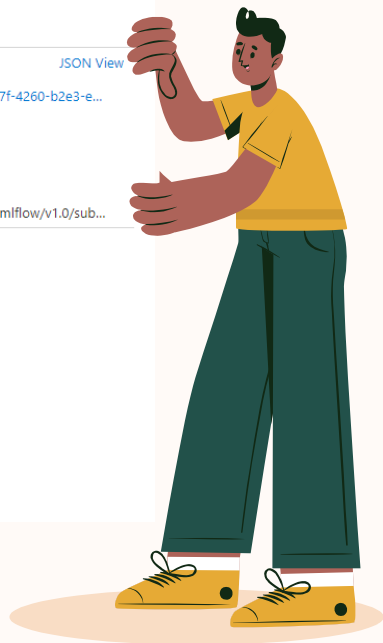
Essentials

Resource group	: capstoneproject	Studio web URL	: https://ml.azure.com/?tid=d7b95ec4-9a7f-4260-b2e3-e...
Location	: Southeast Asia	Registry	: 15c1cf5952b04d6cbe5df7ce1e6a26d8
Subscription	: Azure Pass - Sponsorship	Key Vault	: maribisnis0056177229
Subscription ID	: 47bbdbfd-8a0f-4982-972b-f151dd25a983	Application Insights	: maribisnis1538670193
Storage	: maribisnis1634019060	MLflow tracking URI	: azureml://southeastasia.api.azureml.ms/mlflow/v1.0/sub...

Manage your machine learning lifecycle

Use the Azure Machine Learning studio to build, train, evaluate, and deploy machine learning models. [Learn more](#)

Launch studio



≡ Create Compute Resources

Compute instances

Resource properties

Status

● Stopped

Last operation

Stopped at Dec 2, 2021 5:02 PM: Succeeded

Virtual machine size

Standard_DS11_v2 (2 cores, 14 GB RAM, 28 GB disk)

Processing unit

CPU - Memory optimized

Additional data storage

--

Applications

JupyterLab Jupyter VS Code RStudio Terminal

Created on

11/24/2021, 10:08:35 AM

SSH access

Disabled

Attributes

Compute name

Project

Compute type

Compute instance

Subscription ID

47bbdfbd-8a0f-4982-972b-f151dd25a983

Resource group

capstoneproject

Workspace

MariBisnis

Region

southeastasia

Created by

NAFISAN NAUROTİ SALSA BILA

Compute clusters

Resource properties

Virtual machine size

Standard_DS11_v2 (2 cores, 14 GB RAM, 28 GB disk)

Processing unit

CPU - Memory optimized

OS Type

Linux

Virtual machine priority

Dedicated

Minimum number of nodes

0

Maximum number of nodes

2

Idle seconds before scale down

120

Virtual network/subnet

--

Attributes

Compute name

Capstone

Resource ID

--

Compute type

Machine Learning compute

Subscription ID

47bbdfbd-8a0f-4982-972b-f151dd25a983

Resource group

capstoneproject

Workspace

MariBisnis

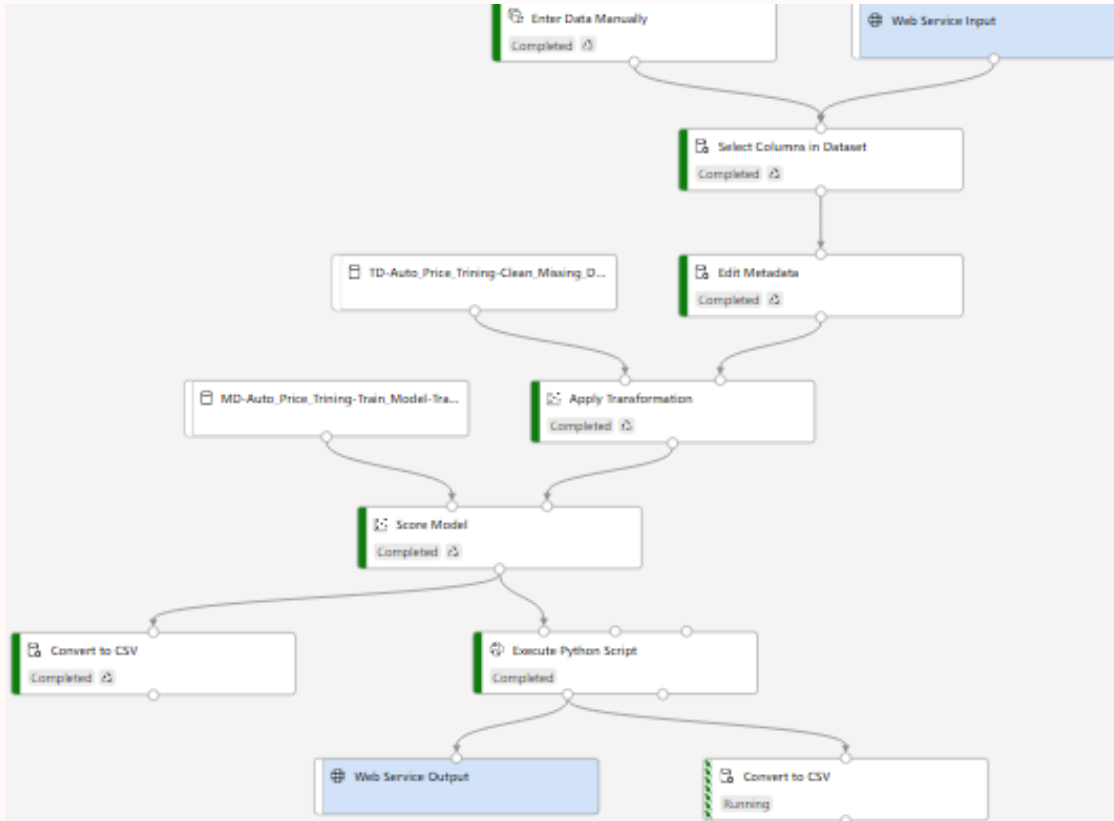
Region

southeastasia


≡ Create a Pipeline



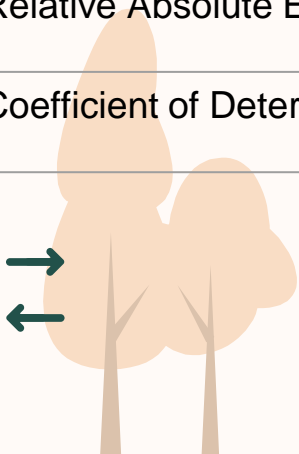
≡ Create and Run a Training Pipeline





≡ Evaluate Model



Mean Absolute Error	90077.066624
Root Mean Squared Error	161582.656996
Relative Squared Error	0.202203
Relative Absolute Error	0.386073
Coefficient of Determination (r^2)	0.797797



- represents the mean absolute error between the predicted result and the original result.
 - represents that the variation in the value produced by a forecast model is close to the variation in the observed value.
 - a relative metric between 0 and 1 based on the square of the differences between predicted and true values.
 - a relative metric between 0 and 1 based on the absolute differences between predicted and true values.
 - represents summarizes how much of the variance between predicted and true values is explained by the model.
- 
- 



Create Visualization





Correlation

By
grade

Condition

Price by
grade

Price by
bathroom

Price by
sqft



```
corr_matrix = df_housing.corr()  
corr_matrix['price'].sort_values(ascending=False)
```

```
price          1.000000  
sqft_living    0.702035  
grade          0.667434  
sqft_above     0.605567  
sqft_living15  0.585379  
bathrooms      0.525138  
view           0.397293  
sqft_basement  0.323816  
bedrooms       0.308350  
lat            0.307003  
waterfront     0.266369  
floors         0.256794  
yr_renovated   0.126434  
sqft_lot       0.089661  
sqft_lot15     0.082447  
yr_built       0.054012  
condition      0.036362  
long           0.021626  
id             -0.016762  
zipcode        -0.053203  
Name: price, dtype: float64
```



Correlation

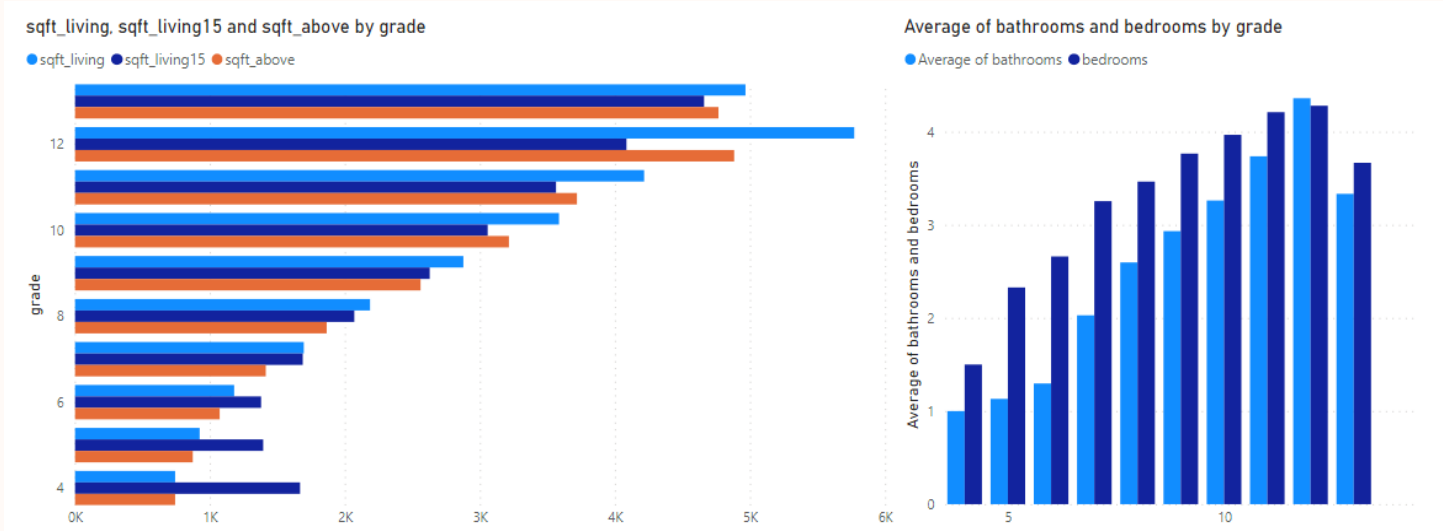
By grade

Condition

Price by
grade

Price by
bathroom

Price by
sqft





Correlation

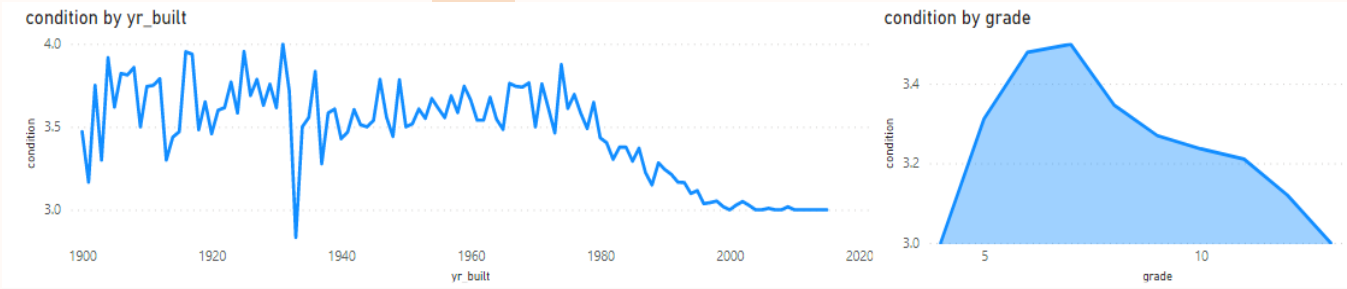
By
grade

Condition

Price by
grade

Price by
bathroom

Price by
sqft





Correlation

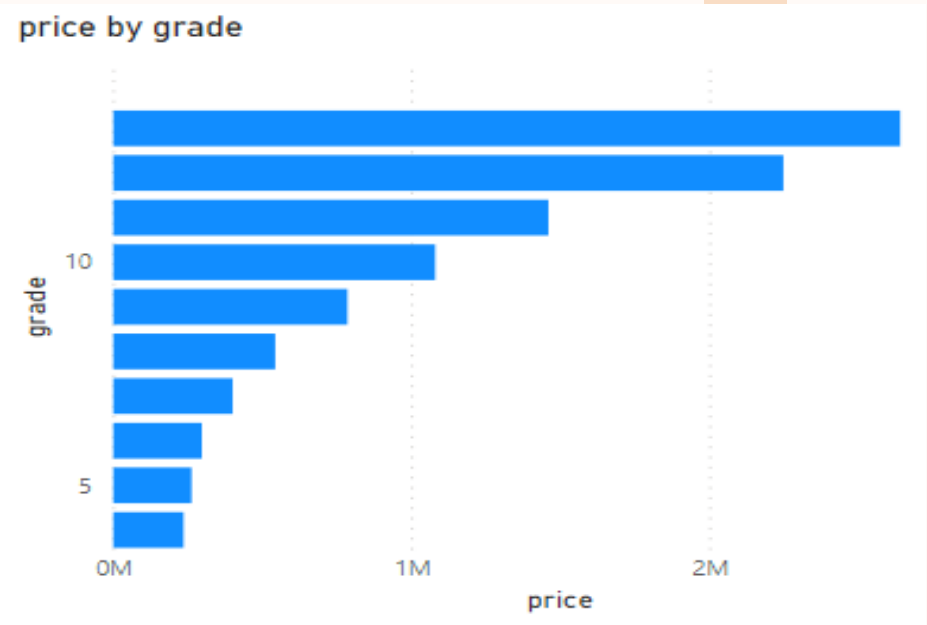
By
grade

Condition

Price by grade

Price by
bathroom

Price by
sqft





Correlation

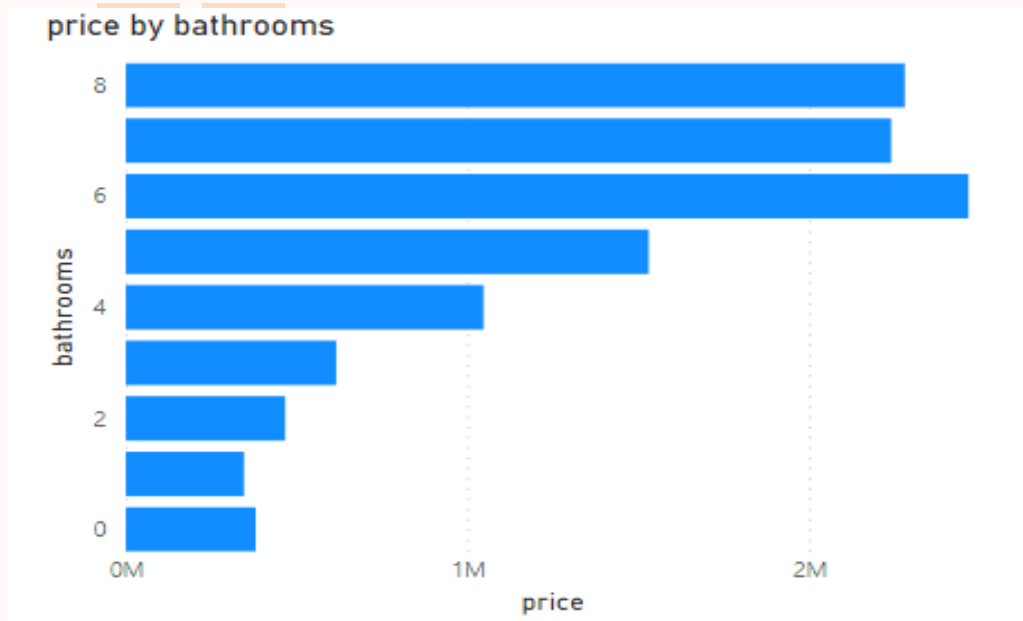
By
grade

Condition

Price by
grade

Price by
bathrooms

Price by
sqft





Correlation

By
grade

Condition

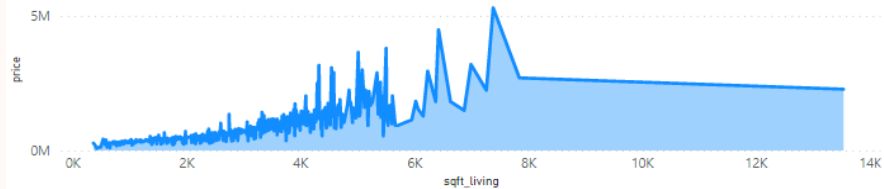
Price by
grade

Price by
bathroom

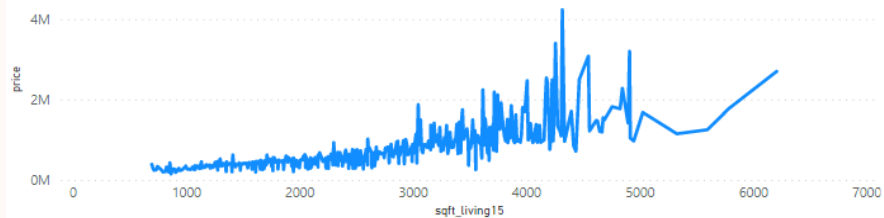
Price by sqft



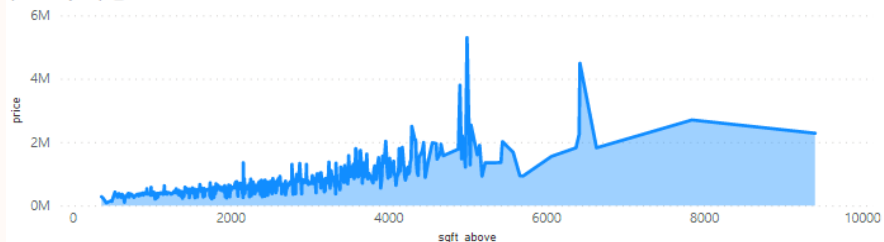
price by sqft_living



price by sqft_living15



price by sqft_above



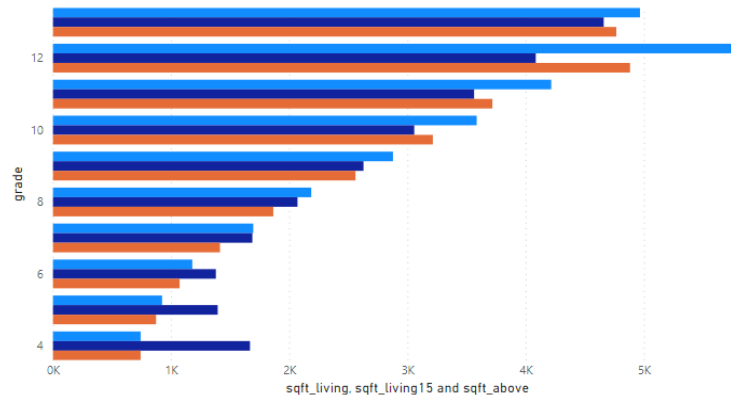


House Condition

Housie
CONDITION

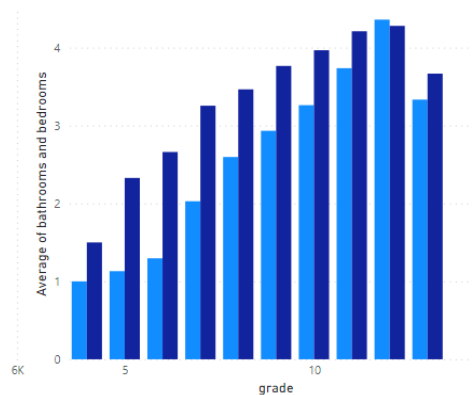
sqft_living, sqft_living15 and sqft_above by grade

● sqft_living ● sqft_living15 ● sqft_above

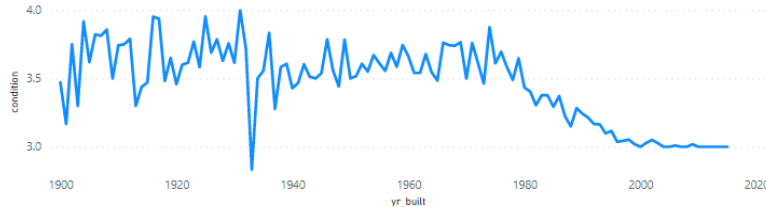


Average of bathrooms and bedrooms by grade

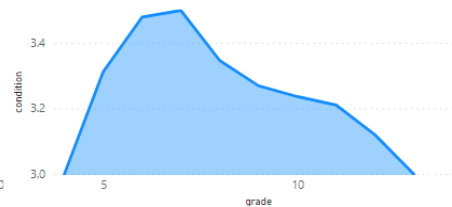
● Average of bathrooms ● bedrooms



condition by yr_built



condition by grade

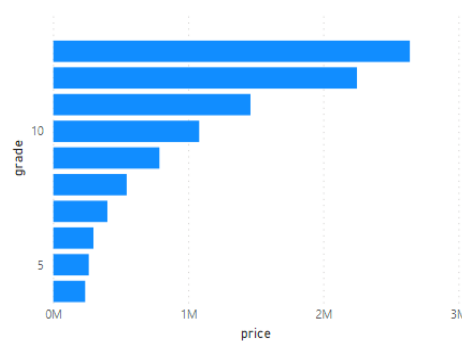




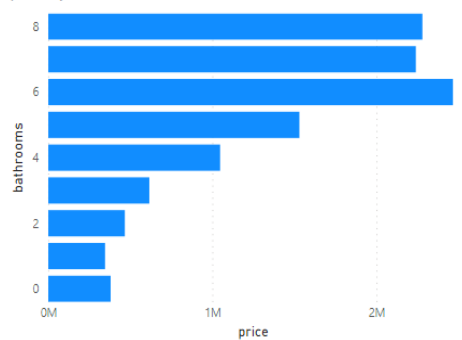
House Pricing

Housie
PRICING

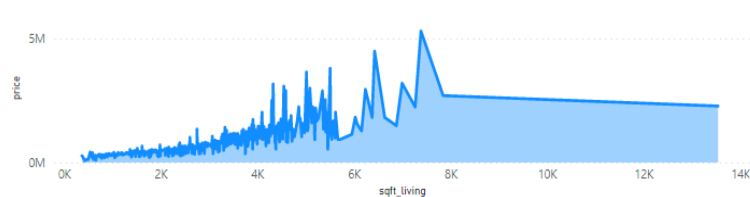
price by grade



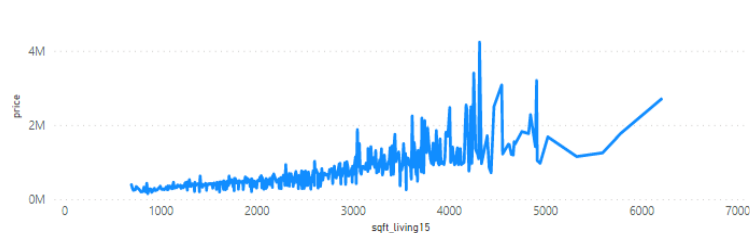
price by bathrooms



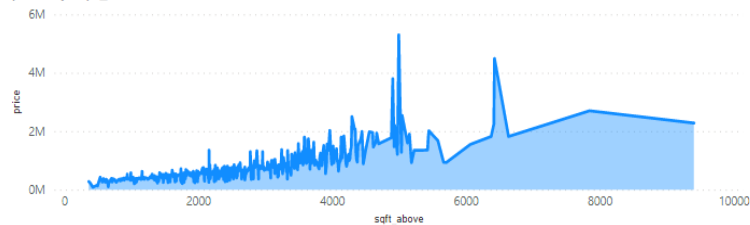
price by sqft_living



price by sqft_living15



price by sqft_above





Thank you!

HOUSIE

MariBisnis
Capstone Project

By: Nafisan Nauroti Salsa Bila

DAI-005

Data and Artificial Intelligence
Microsoft

