

Technical Document of MariBisnis

Capstone Project

Nov-Des 2021

HOUSIE

ISSUED BY

MICROSOFT

REPRESENTATIVE

NAFISAN NAUROTU SALSABILA

DAI-005

Data and Artificial Intelligence

Microsoft

Outlines

Introduction

Dataset

Resources

- Membuat Machine Learning Resource

- Membuat Compute Resources

- Membuat Dataset Resources

Pipeline

- Membuat Pipeline

- Membuat Training House Pricing

- Membuat Training House Pricing - real time inference

- Percobaan Endpoint

Evaluasi

Power BI

- Mengapa Power BI Service?

- Pengimporan Data MariBisnis

- Visualisasi House Condition

- Visualisasi House Pricing Correlated

References

Introduction

Tujuan dari *capstone project* dalam bidang bisnis ini yaitu untuk mengetahui *trend* bisnis dari penjualan rumah yang ada di Amerika Serikat. Harapannya, keluaran *project* ini dapat memprediksi harga suatu rumah dan melakukan pemetaan terhadap sebaran data yang ada.

Dataset

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

Sebelum menuju langkah pembuatan *training* dan *deployment* model, dataset MariBisnis akan diolah lebih dulu di Excel. Pengolahan ini bertujuan untuk mengubah *value* pada kolom *bathrooms* dan *floors* yang awalnya *decimal* atau *float* menjadi *integer*. Pada Excel, kolom *bathrooms* dan *floors* diubah *value*-nya menjadi integer dengan menggunakan *decrease decimal*. Hal ini hanya diimplementasikan ke kolom *bathrooms* dan *floors* karena memiliki *value* desimal, yang mana akan memunculkan ambiguitas.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	id	date	price	bedrooms	bathroom	sqft_living	sqft_lot	floors	waterfront	view	condition	grade	sqft_abov	sqft_base	yr_built	yr_renov	zipcode	lat	long	sqft_living	sqft_l
2	7.13E+09	20141013T	221900	3	1	1180	5650	1	0	0	3	7	1180	0	1955	0	98178	47.5112	-122.257	1340	5
3	6.41E+09	20141209T	538000	3	2	2570	7242	2	0	0	3	7	2170	400	1951	1991	98125	47.721	-122.319	1690	7
4	5.63E+09	20150225T	180000	2	1	770	10000	1	0	0	3	6	770	0	1933	0	98028	47.7379	-122.233	2720	8
5	2.49E+09	20141209T	604000	4	3	1960	5000	1	0	0	5	7	1050	910	1965	0	98136	47.5208	-122.393	1360	5
6	1.95E+09	20150218T	510000	3	2	1680	8080	1	0	0	3	8	1680	0	1987	0	98074	47.6168	-122.045	1800	7
7	7.24E+09	20140512T	1230000	4	5	5420	101930	1	0	0	3	11	3890	1530	2001	0	98053	47.6561	-122.005	4760	101
8	1.32E+09	20140627T	257500	3	2	1715	6819	2	0	0	3	7	1715	0	1955	0	98003	47.3097	-122.327	2238	6
9	2.01E+09	20150115T	291850	3	2	1060	9711	1	0	0	3	7	1060	0	1963	0	98198	47.4095	-122.315	1650	9
10	2.41E+09	20150415T	229500	3	1	1780	7470	1	0	0	3	7	1050	730	1960	0	98146	47.5123	-122.337	1780	8
11	3.79E+09	20150312T	323000	3	3	1890	6560	2	0	0	3	7	1890	0	2003	0	98038	47.3684	-122.031	2390	7
12	1.74E+09	20150403T	662500	3	3	3560	9796	1	0	0	3	8	1860	1700	1965	0	98007	47.6007	-122.145	2210	8
13	9.21E+09	20140527T	468000	2	1	1160	6000	1	0	0	4	7	860	300	1942	0	98115	47.69	-122.292	1330	6
14	1.14E+08	20140528T	310000	3	1	1430	19901	2	0	0	4	7	1430	0	1927	0	98028	47.7558	-122.229	1780	12
15	6.05E+09	20141007T	400000	3	2	1370	9680	1	0	0	4	7	1370	0	1977	0	98074	47.6127	-122.045	1370	10
16	1.18E+09	20150312T	530000	5	2	1810	4850	2	0	0	3	7	1810	0	1900	0	98107	47.67	-122.394	1360	4
17	9.3E+09	20150124T	650000	4	3	2950	5000	2	0	3	3	9	1980	970	1979	0	98126	47.5714	-122.375	2140	4
18	1.88E+09	20140731T	395000	3	2	1890	14040	2	0	0	3	7	1890	0	1994	0	98019	47.7277	-121.962	1890	14
19	6.87E+09	20140529T	485000	4	1	1600	4300	2	0	0	4	7	1600	0	1916	0	98103	47.6648	-122.343	1610	4
20	16000397	20141205T	189000	2	1	1200	9850	1	0	0	4	7	1200	0	1921	0	98002	47.3089	-122.21	1060	5
21	7.98E+09	20150424T	230000	3	1	1250	9774	1	0	0	4	7	1250	0	1969	0	98003	47.3343	-122.306	1280	8
22	6.3E+09	20140514T	385000	4	2	1620	4980	1	0	0	4	7	860	760	1947	0	98133	47.7025	-122.341	1400	4
23	2.52E+09	20140826T	2000000	3	3	3050	44867	1	0	4	3	9	2330	720	1968	0	98040	47.5316	-122.233	4110	20

Resources

Membuat Machine Learning Resource

Azure Machine Learning adalah platform berbasis *cloud* untuk membuat dan mengoperasikan solusi *machine learning* di Azure. Ini termasuk berbagai fitur dan kemampuan yang membantu ilmuwan data menyiapkan data, melatih model, menerbitkan layanan prediktif, dan memantau penggunaannya. Salah satu fitur ini adalah antarmuka visual yang disebut *designer*, yang dapat digunakan untuk melatih, menguji, dan menyebarkan model *machine learning* tanpa menulis kode apa pun.

Untuk menggunakan Azure Machine Learning, Anda membuat workspace di *Azure subscription*. Kemudian, Anda dapat menggunakan *workspace* ini untuk mengelola data, *compute resources*, kode, model, dan artefak lainnya yang terkait dengan beban kerja *machine learning*.

Berikut langkah-langkah membuat *machine learning workspace*:

1. Masuk ke [portal Azure](#) menggunakan info masuk Microsoft Anda.
2. Pilih **+Create a Resource**, cari **Machine Learning**, dan buat *Machine Learning resource* baru dengan pengaturan berikut:

Essentials		JSON View	
Resource group :	capstoneproject	Studio web URL :	https://ml.azure.com/?tid=d7b95ec4-9a7f-4260-b2e3-eb5...
Location :	Southeast Asia	Registry :	15c1cf5952b04d6cbe5df7ce1e6a26d8
Subscription :	Azure Pass - Sponsorship	Key Vault :	maribisnis0056177229
Subscription ID :	47bbdfbd-8a0f-4982-972b-f151dd25a983	Application Insights :	maribisnis1538670193
Storage :	maribisnis1634019060	MLflow tracking URI :	azureml://southeastasia.api.azureml.ms/mlflow/v1.0/subscr...

3. Pada laman Overview, buka Azure Machine Learning Studio menggunakan akun Microsoft yang terdaftar Azure.

Anda dapat mengelola *workspace* menggunakan portal Azure, tetapi untuk ilmuwan data dan teknisi operasi Machine Learning, studio Azure Machine Learning menyediakan antarmuka pengguna yang lebih terfokus untuk mengelola *workspace resources*.

Membuat Compute Resources

Untuk melatih dan menyebarkan model menggunakan designer Azure Machine Learning, Anda perlu melakukan komputasi untuk menjalankan proses pelatihan, dan untuk menguji model terlatih setelah menyebarkannya.

Di studio Azure Machine Learning, tampilkan halaman **Compute** (di bawah **Manage**). Di sinilah Anda mengelola target komputasi untuk aktivitas ilmu data Anda. Ada empat jenis sumber daya komputasi yang dapat dibuat:

- **Compute Instances:** *Development workstations* yang dapat digunakan oleh *data scientist* untuk bekerja dengan data dan model.
- **Compute Clusters:** *Scalable clusters of virtual machines* untuk pemrosesan kode eksperimen sesuai permintaan.
- **Inference Clusters:** *Deployment targets* untuk layanan prediktif yang menggunakan model terlatih.
- **Attached Compute:** *link ke Azure compute resources* yang ada, seperti Virtual Machines atau Azure Databricks clusters.

Langkah-langkah membuat Compute Instances dan Compute Clusters:

1. Di tab **Compute Instances**, tambahkan *new Compute Instances*.

Resource properties	Attributes
Status Stopped	Compute name Project
Last operation Stopped at Dec 2, 2021 5:02 PM: Succeeded	Compute type Compute instance
Virtual machine size Standard_DS11_v2 (2 cores, 14 GB RAM, 28 GB disk)	Subscription ID 47bbdfbd-8a0f-4982-972b-f151dd25a983
Processing unit CPU - Memory optimized	Resource group capstoneproject
Additional data storage --	Workspace MariBisnis
Applications JupyterLab Jupyter VS Code RStudio Terminal	Region southeastasia
Created on 11/24/2021, 10:08:35 AM	Created by NAFISAN NAUROT SALSA BILA
SSH access Disabled	

2. Saat Compute Instances sedang dibuat, beralihlah ke tab **Compute Clusters**, dan tambahkan compute clusters baru.

Attributes	Resource properties
Properties Tabular	Virtual machine size Standard_DS11_v2 (2 cores, 14 GB RAM, 28 GB disk)
Created by NAFISAN NAUROTOI SALSA BILA	Processing unit CPU - Memory optimized
Profile No profile generated	OS Type Linux
Files in dataset 1	Virtual machine priority Dedicated
Total size of files in dataset ⓘ 2.21 MiB	Minimum number of nodes 0
Current version 1	Maximum number of nodes 2
Latest version 1	Idle seconds before scale down 120
Created time Dec 1, 2021 11:11 AM	Virtual network/subnet --
Modified time Dec 1, 2021 11:11 AM	

Membuat Dataset Resources

Sebagai resource dalam pelatihan dan penerapan model nantinya di Machine Learning Azure, diperlukan akses dataset untuk sebagai sumber datanya.

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

Pada Azure Machine Learning Studio, lihat laman Dataset. Di sinilah nantinya dataset MariBisnis akan diolah menjadi lebih compact dan lebih bermakna.

1. Pada laman **Dataset**, pilih **+Create dataset import from local files** untuk menambahkan dataset MariBisnis.
2. Pada **Basic info**, beri nama untuk dataset, kemudian pilih Tabular untuk tipe dataset, dan beri deskripsi jika berkenan.
3. Pada **Datastore and files selection**, upload file dataset.
4. Pada **Settings and preview** dan **Schema**, klik next.
5. Pada **Confirm Details**, klik create.

Pipeline

Untuk melatih model regresi, diperlukan himpunan data yang menyertakan fitur historis (karakteristik entitas yang prediksinya ingin Anda buat) dan nilai label yang diketahui (nilai numerik yang modelnya ingin Anda latih untuk memprediksi).

Membuat Pipeline

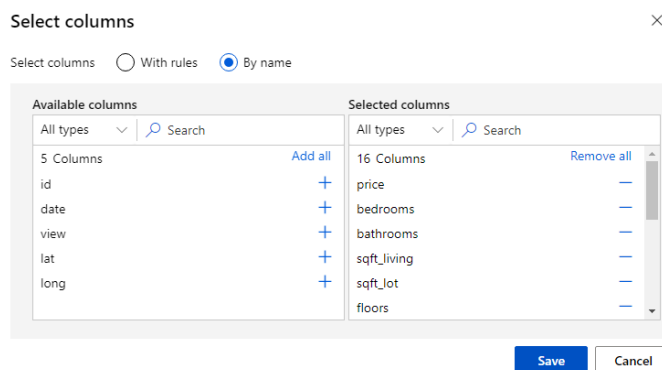
Untuk menggunakan Azure Machine Learning Designer, dapat membuat pipeline yang akan digunakan untuk melatih model *machine learning*. Pipeline ini dimulai dengan himpunan data yang ingin dilatih menjadi model.

1. Di studio Azure Machine Learning, lihat halaman **Designer**, dan pilih + untuk membuat pipeline baru.
2. Tentukan target komputasi untuk menjalankan pipeline. Di panel **Setting**, klik **Select compute target** untuk memilih Compute Clusters yang telah dibuat sebelumnya.

Membuat Training House Pricing

Tujuan pada pipeline Training House Pricing adalah untuk melatih model regresi yang akan memprediksi harga unit rumah.

1. Di samping nama pipeline di sebelah kiri, pilih tombol >> untuk meluaskan panel. Temukan bagian **Dataset**, dan seret dataset Maribisnis ke kanvas.
2. Pada **Data Transformation**, Seret modul **Select Columns in Dataset** ke kanvas, di bawah modul Maribisnis. Selanjutnya sambungkan output di bagian bawah modul Dataset ke input di bagian atas modul **Select Columns in Dataset**.
3. Pilih modul **Select Columns in Dataset**, dan di panel settingnya di sebelah kanan, pilih **Edit kolom**. Selanjutnya di jendela **Pilih kolom**, dan masukkan data apa saja yang akan digunakan (disini saya memasukkan seluruh data kecuali id, date, view, lat, dan long).



4. Pilih **Data Transformation** > **Edit Metadata**, drag and drop ke kanvas. Hubungkan **Select Columns in Dataset** dan **Edit Metadata**. Masukkan waterfront dan zipcode pada column dan pada **categorical** pilih categorical.

Edit Metadata

Column * Edit column

Column names: waterfront, zipcode

Data type *

Unchanged

Categorical *

Categorical

Fields *

Unchanged

5. **Data Transformation > Clean missing data**, hubungkan **Clean missing data** dengan **Edit metadata**.

Columns to be cleaned

Select columns ☐ With rules ☒ By name

Available columns

All types Search

6 Columns Add all

id +
date +
price +
view +
lat +
long +

Selected columns

All types Search

15 Columns Remove all

bedrooms -
bathrooms -
sqft_living -
sqft_lot -
floors -
waterfront -

Clean Missing Data

Minimum missing value ratio 0.0

Maximum missing value ratio 1.0

Cleaning mode Custom substitution value

Replacement value 0

Generate missing value indicator column False

Save Cancel

6. **Data Transformation > Split data**, hubungkan dengan **Clean missing data**.

Split Data

Splitting mode Split Rows

Fraction of rows in the first output data... 0.75

Randomized split True

Random seed 123

Stratified split False

7. **Model Training > Train Model**, hubungkan node kanan dengan **Split data**.

Train Model

Label column price

Model explanations False

8. **Machine learning algorithms > Decision Forest Regression**, hubungkan dengan node kiri **Train Model**.
9. **Model Scoring and Evaluation > Score Model**, node kiri hubungkan dengan **Train model**, sedangkan node kanan dengan **Split Data**.
10. **Model Scoring and Evaluation > Evaluate Model**, hubungkan node kiri dengan **Score Model**.
11. **Convert to CSV**, hubungkan dengan output **Score Model**.
12. Submit.



Setelah ter-submit, berikut merupakan hasil dari evaluasi model:

Mean Absolute Error (MAE)	90077.066624
Root Mean Squared Error (RMSE)	161582.656996
Relative Squared Error (RSE)	0.202203
Relative Absolute Error (RAE)	0.386073
Coefficient of Determination (R2)	0.797797

Membuat Training House Pricing - real time inference

Setelah membuat dan menjalankan pipeline untuk melatih model, diperlukan pipeline kedua yang menjalankan transformasi data yang sama untuk data baru, lalu menggunakan model terlatih untuk menyimpulkan (dengan kata lain, memprediksi) nilai label berdasarkan fiturnya. Ini akan membentuk dasar untuk layanan prediktif yang dapat diterbitkan untuk digunakan aplikasi.

1. Kanvas pada Training House Pricing - real time inference men-generate pipeline dari Training House Pricing yang telah dibuat sebelumnya.
2. Ganti dataset MariBisnis dengan **Enter Data Manually**. Pada kolom data, masukkan kode sebagai berikut:

Data

×

```
1 id,date,bedrooms,bathrooms,sqft_living,sqft_lot,floors,waterfront,view,condition,grade,sqft_above,sqft_basement,yn_
2 7129300520,20141013T000000,3,1,1180,5650,1,0,0,3,7,1180,0,1955,0,98178,47.5112,-122.257,1340,5650
3 6414100192,20141209T000000,3,2,2570,7242,2,0,0,3,7,2170,400,1951,98125,47.721,-122.319,1690,7639
4 5631500400,20150225T000000,2,1,770,10000,1,0,0,3,6,770,0,1933,0,98028,47.7379,-122.233,2720,8062
```

save

cancel

3. Pada **Select column in dataset**, hapus data price pada select column.
4. Sisipkan **Execute Python Script** diantar **Score Model** dan **web service output**. Masukkan script pada python script.

Python script

×

```
1 import pandas as pd
2
3 def azureml_main(dataframe1 = None, dataframe2 = None):
4
5     scored_results = dataframe1[['Scored Labels']]
6     scored_results.rename(columns={'Scored Labels':'predicted_price'},
7                               inplace=True)
8     return scored_results
```

save

cancel

5. Submit

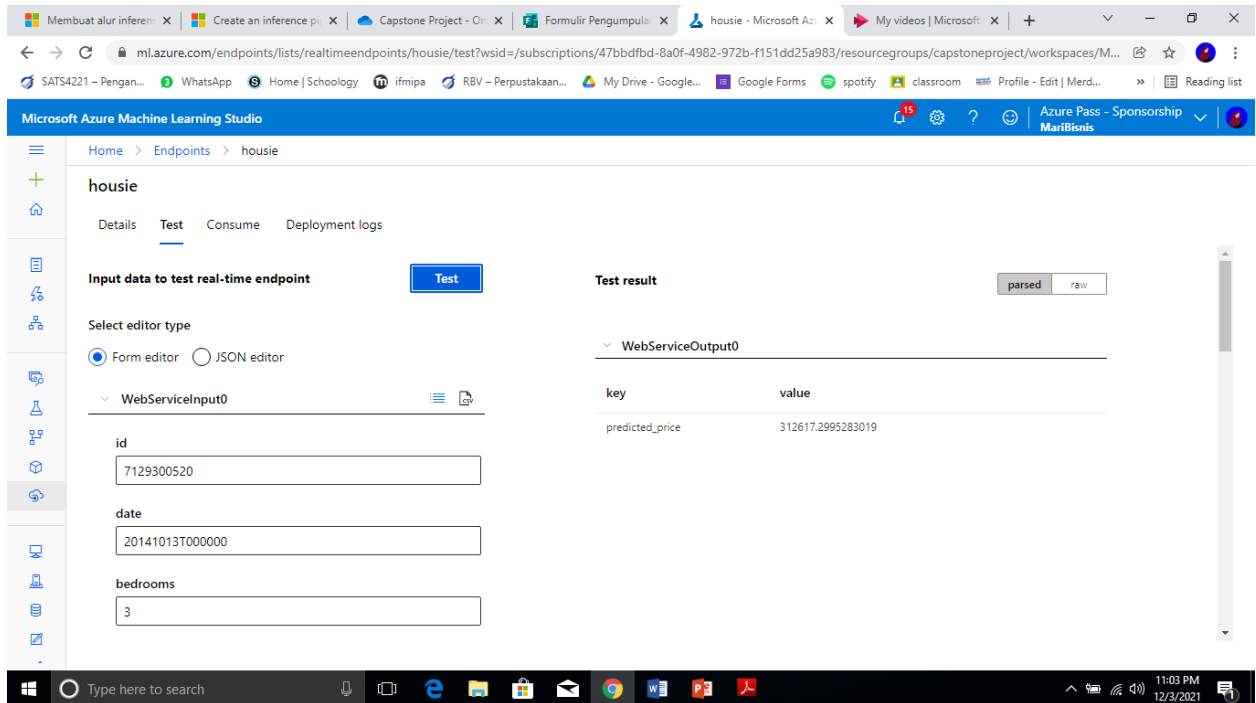


6. Deploy

Percobaan Endpoint

Setelah model dibuat dan diuji untuk inferensi real-time, selanjutnya adalah publikasi sebagai layanan untuk digunakan. Pada Endpoint predict-auto-price, buka Consume tab dan perhatikan informasi mengenai REST endpoint dan Primary Key.

Percobaan dapat dilakukan juga dengan cara membuka tab Test, tanpa perlu memasukkan REST endpoint dan Primary Key sebagai input. Dengan mudah dapat memasukkan input sesuai dengan dataset MariBisnis, tentunya dengan input yang berbeda. Hal ini dibantu oleh Web Service Input yang ada pada pipeline Training House Pricing - real time inference. Web Service Output juga berperan untuk menampilkan output dari input yang dimasukkan sebelumnya.



Evaluasi

Ketika merancang pipeline Training House Pricing dan Training House Pricing - real time inference, tidak lupa pada keduanya ditambahkan langkah Evaluate Model. Langkah ini memberikan informasi mengenai matriks evaluasi yang berjumlah lima. Berikut adalah penjelasannya:

Mean Absolute Error (MAE)	Perbedaan rata-rata antara nilai prediksi dan nilai sebenarnya. Nilai ini didasarkan pada unit yang sama dengan label, dalam hal ini dolar. Semakin rendah nilai ini, semakin baik model memprediksi.
Root Mean Squared Error (RMSE)	Akar kuadrat dari perbedaan kuadrat rata-rata antara nilai yang diprediksi dan benar. Hasilnya adalah metrik berdasarkan unit yang sama dengan label (dolar). Jika dibandingkan dengan MAE (di atas), perbedaan yang lebih besar menunjukkan varians yang lebih besar dalam kesalahan individu (misalnya, dengan beberapa kesalahan sangat kecil, sementara yang lain besar).
Relative Squared Error (RSE)	Metrik relatif antara 0 dan 1 berdasarkan kuadrat selisih antara nilai prediksi dan nilai sebenarnya. Semakin mendekati 0 metrik ini, semakin baik performa model. Karena metrik ini relatif, metrik ini dapat digunakan untuk membandingkan model yang labelnya berada dalam unit yang berbeda.
Relative Absolute Error (RAE)	Metrik relatif antara 0 dan 1 berdasarkan perbedaan absolut antara nilai prediksi dan nilai sebenarnya. Semakin mendekati 0 metrik ini, semakin baik performa model. Seperti RSE, metrik ini dapat digunakan untuk

	membandingkan model yang labelnya berada di unit yang berbeda.
Coefficient of Determination (R2)	Metrik ini lebih sering disebut sebagai R-Squared , dan merangkum berapa banyak varians antara nilai prediksi dan nilai sebenarnya yang dijelaskan oleh model. Semakin mendekati 1 nilai ini, semakin baik kinerja model.

Power BI

Power BI adalah kumpulan layanan perangkat lunak, aplikasi, dan konektor yang bekerja bersama untuk membantu, membuat, berbagi, dan menggunakan wawasan bisnis dengan cara yang paling efektif untuk layanan bisnis. Microsoft Power BI service, terkadang disebut sebagai Power BI online, adalah bagian SaaS (Software as a Service) dari Power BI. Di layanan Power BI, dasbor membantu untuk memantau perkembangan bisnis. Dasbor menampilkan tiles, yang dapat dipilih untuk membuka laporan guna menjelajahi lebih lanjut. Dasbor dan laporan terhubung ke kumpulan data yang menyatukan semua data yang relevan di satu tempat.

Mengapa Power BI Service?

Karena adanya keterbatasan dalam operating system yang digunakan, berupa Linux Elementary OS based on Ubuntu 18.04 LTS (Bionic Beaver), maka digunakanlah layanan Power BI Service yang dapat dijalankan secara online lewat browser.

Pengimporan Data MariBisnis

Tidak jauh berbeda dengan Power BI Desktop, pengambilan data pada Power BI Service cukup mudah, yakni dengan memilih Get data dan memilih pengimporan melalui file *.csv yang sebelumnya sudah diolah. Pengambilan file dilakukan melalui Local File dan memilih MariBisnis dataset, yang kemudian akan diproses dan diletakkan ke dalam My Workspace.

Setelah pengimporan file dataset selesai, Power BI Service akan men-generate sebuah Reports dan sebuah Dashboard yang nantinya akan membantu dalam visualisasi. Visualisasi nantinya akan dilakukan pada Report file, sedangkan kumpulan visualisasi yang hendak ditampilkan sebagai final akan diletakkan di Dashboard.

Visualisasi yang akan dibuat pada MariBisnis ialah mengenai kondisi rumah - House Condition, dan korelasi price terhadap kolom lainnya - Correlation. Sebelum mengarah pada visualisasi di Power BI, perlu diketahui mengenai korelasi price. Pemrosesan untuk mengetahui korelasi ini dilakukan dengan scripting python code.

```
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
```

Dari scripting python code diatas, dapat diketahui bahwa kolom price sangat berkorelasi dengan sqft_living, grade, sqft_above, sqft_living15, dan bathrooms. Hal ini dapat dilihat dari hasil sort_values() yang menunjukkan kolom-kolom yang nilainya mendekati angka 1. Nilai-nilai berikut nantinya akan membantu dalam memvisualisasikan prediksi harga sebuah unit rumah yang menjadi tujuan dari MariBisnis.

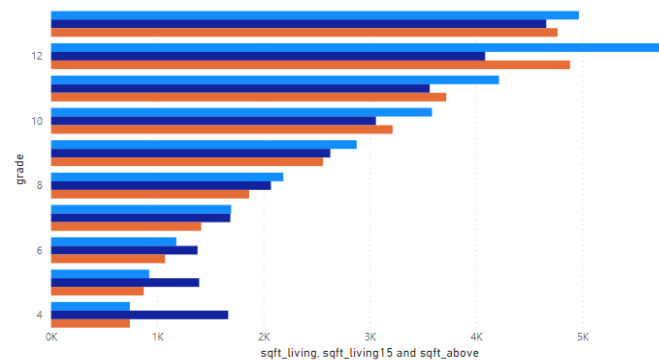
Visualisasi House Condition

Pada visualisasi yang pertama, bertujuan untuk mengetahui kondisi sebaran unit rumah yang terjual. Visualisasi ini menggunakan 3 (tiga) buah chart yang menjelaskan mengenai kondisi rumah dan grade rumah.

House CONDITION

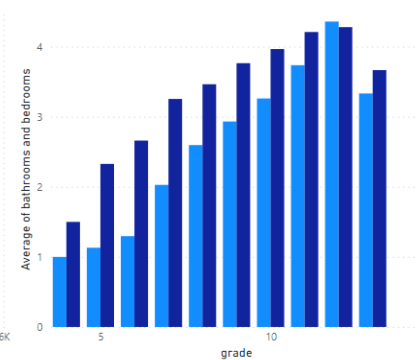
sqft_living, sqft_living15 and sqft_above by grade

● sqft_living ● sqft_living15 ● sqft_above

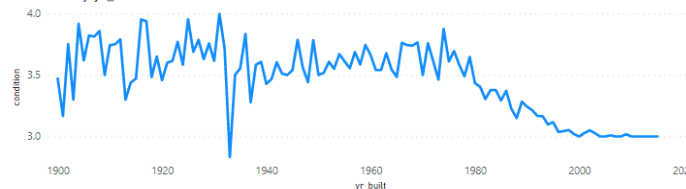


Average of bathrooms and bedrooms by grade

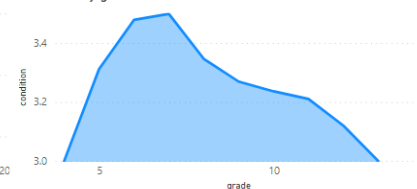
● Average of bathrooms ● bedrooms



condition by yr_built



condition by grade



- Visualisasi grade unit rumah berdasarkan square footage; sqft_living by grade.

Visualisasi ini dimaksudkan untuk mengetahui sebaran area tinggal yang dimiliki unit rumah terhadap grade. Chart yang digunakan adalah stacked bar chart. Hasil dari analisisnya dapat dilihat dari bar yang semakin naik, hal ini juga berbanding lurus dengan grade dengan angka yang semakin besar dan square footage yang menunjukkan luas semakin besar pula. Dapat diambil simpulan bahwa, ketika grade suatu unit rumah besar, maka square footage yang dimiliki rumah tersebut juga luas.
- Visualisasi grade unit rumah berdasarkan rerata jumlah kamar tidur dan kamar mandi; average of bedrooms and bathrooms by grade.

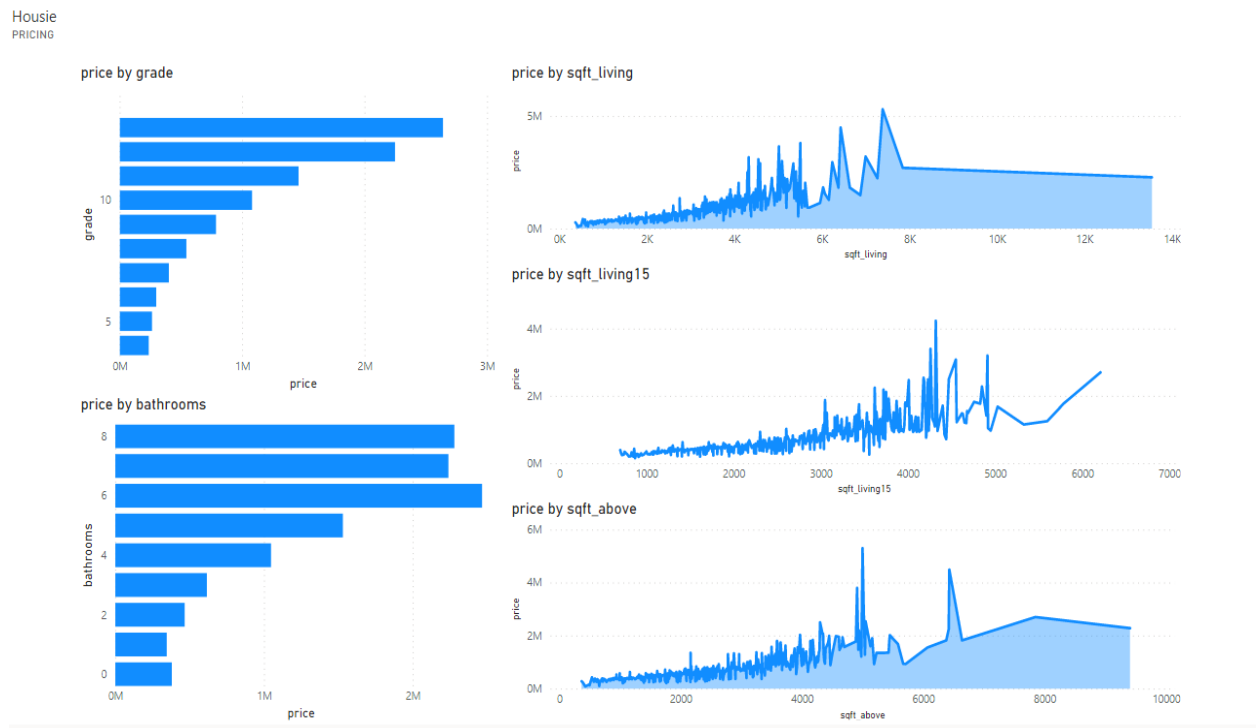
Visualisasi ini dimaksudkan untuk mengetahui sebaran kamar tidur dan kamar mandi berdasarkan grade sebuah unit rumah. Chart yang digunakan adalah stacked bar chart. Hasil dari analisisnya ialah semakin banyak kamar tidur dan kamar mandi pada sebuah unit rumah, maka unit rumah tersebut memiliki grade yang tinggi, berdasarkan grading system King Country.
- Visualisasi kondisi unit rumah berdasarkan tahun dibangunnya rumah; condition by yr_built.

Visualisasi ini dimaksudkan untuk mengetahui sebaran kondisi sebuah rumah berdasarkan tahun rumah tersebut dibangun. Chart yang digunakan adalah line chart. Hasil dari analisisnya ialah kondisi rumah yang dibangun pada tahun 1900-an ternyata dinilai lebih baik dibanding dengan kondisi rumah yang dibangun pada tahun 2000-an. Dapat dinyatakan bahwa, kondisi rumah lama tidak kalah bagusnya dengan unit rumah yang belum lama dibangun.
- Visualisasi kondisi unit rumah berdasarkan grade; condition by grade.

Visualisasi ini dimaksudkan untuk mengetahui sebaran kondisi sebuah rumah berdasarkan grade yang diberikan kepada rumah. Hasil dari analisisnya ialah rumah dengan kondisi baik tidak berbanding lurus dengan grade yang diberikan.

Visualisasi House Pricing Correlated

Pada visualisasi yang kedua, bertujuan untuk mengetahui apa saja yang menyebabkan atau berkorelasi dengan price sebuah unit rumah. Visualisasi ini menggunakan 5 (lima) buah chart yang menjelaskan mengenai hal yang berkorelasi terhadap prediksi rumah.



- Visualisasi rerata harga unit rumah berdasarkan grade; average of price by grade.
Visualisasi ini dimaksudkan untuk mengetahui hal yang berkorelasi dengan harga sebuah unit rumah. Chart yang digunakan adalah stacked column chart. Hasil dari analisisnya ialah semakin besar grade yang dimiliki sebuah unit rumah, maka hal tersebut berbanding lurus dengan harganya. Dimaksudkan, bila sebuah rumah memiliki grade tinggi, maka harga dari rumah tersebut juga tinggi. Grade ini dinilai berdasarkan grading system King Country.
- Visualisasi rerata harga unit rumah berdasarkan bathrooms; average of price by bathrooms.
Visualisasi ini dimaksudkan untuk mengetahui hal yang berkorelasi dengan harga sebuah unit rumah. Chart yang digunakan adalah stacked column chart. Hasil dari analisisnya ialah semakin banyak kamar mandi yang dimiliki sebuah unit rumah, maka hal tersebut berbanding lurus dengan harganya. Dimaksudkan, bila sebuah rumah memiliki banyak kamar mandi, maka harga dari rumah tersebut juga tinggi.
- Visualisasi price unit rumah berdasarkan rerata square footage rumah; average of price by sqft_above.
Visualisasi ini dimaksudkan untuk mengetahui hal yang berkorelasi dengan harga sebuah unit rumah. Chart yang digunakan adalah line chart. Hasil dari analisisnya ialah semakin besar square footage semua area tinggal (tidak termasuk area basement) yang dimiliki sebuah unit rumah, maka hal tersebut berbanding lurus dengan harganya. Dimaksudkan, bila sebuah rumah memiliki square footage semua area tinggal yang luas, maka harga dari rumah tersebut juga tinggi.
- Visualisasi price unit rumah berdasarkan rerata square footage area tinggal; average of price by sqft_living.

Visualisasi ini dimaksudkan untuk mengetahui hal yang berkorelasi dengan harga sebuah unit rumah. Chart yang digunakan adalah line chart. Hasil dari analisisnya ialah semakin besar square footage semua area tinggal (termasuk area basement) yang dimiliki sebuah unit rumah, maka hal tersebut berbanding lurus dengan harganya. Dimaksudkan, bila sebuah rumah memiliki square footage semua area tinggal yang luas, maka harga dari rumah tersebut juga tinggi.

- Visualisasi price unit rumah berdasarkan rerata square footage area tinggal pada tahun 2015, baik direnovasi maupun tidak; average of price by sqft_living15.

Visualisasi ini dimaksudkan untuk mengetahui hal yang berkorelasi dengan harga sebuah unit rumah. Chart yang digunakan adalah line chart. Hasil dari analisisnya ialah semakin besar square footage semua area tinggal pada tahun 2015 (termasuk area basement, baik unit rumah mengalami renovasi ataupun tidak) yang dimiliki sebuah unit rumah, maka hal tersebut berbanding lurus dengan harganya. Dimaksudkan, bila sebuah rumah memiliki square footage semua area tinggal yang luas pada tahun 2015, maka harga dari rumah tersebut juga tinggi.

References

Create a Regression Model with Azure Machine Learning designer (<https://docs.microsoft.com/en-us/learn/modules/create-regression-model-azure-machine-learning-designer/?ns-enrollment-type=LearningPath&ns-enrollment-id=learn.www.create-no-code-predictive-models-with-azure-machine-learning>)

Tutorial: Get started creating in the Power BI service (<https://docs.microsoft.com/en-us/power-bi/fundamentals/service-get-started>)