

Power BI and Python Visualization

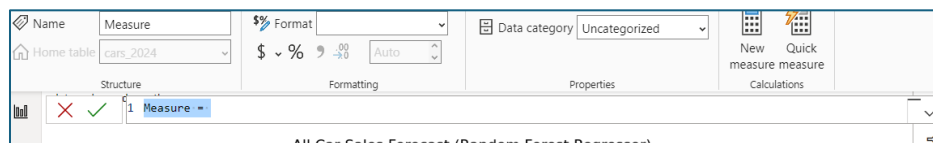
Cara untuk buat visual (Forecasting) menggunakan library python - Prophet & Scikit Learn.

1. Setup BI Environment

- Pastikan Data dh di load dalam Power BI (cars_2024)
- Pastikan Field date_reg perlu menggunakan Data Type (**Date**), sbb python nak baca timeseries.
- Set format: **yyyy-mm-dd**, dan Pilih **Don't Summarize**

2. Create New Measure pada table data cars_2024

- Di Pane Data, Click pada nama **cars_2024**
- Pada bahagian ribbon **Table tools**, click **New Measure**



- Kotak Edit 'Measure =' akan keluar.
- Masukkan code dibawah dan click tanda right hijau. Contoh, code untuk mengira total Car Sales bagi bulan Julai:

```
Total Car Sales in July 2024 =  
CALCULATE (  
    COUNT('cars_2024'[maker]),  
    MONTH('cars_2024'[date_reg]) = 7,  
    YEAR('cars_2024'[date_reg]) = 2024  
)
```

Measure ini penting untuk kita buat pengiraan yang Power BI tak boleh nk tunjukkan disebabkan oleh data kita dalam bentuk daily log dan tiada numerical values. (Especially data cars_2024 ni)

Sekiranya nak buat calculation bulan 8, gantikan value `MONTH('cars_2024'[date_reg]) = 7` kepada `MONTH('cars_2024'[date_reg]) = 8`.

```
Total Car Sales in August 2024 =  
CALCULATE (  
    COUNT('cars_2024'[maker]),  
    MONTH('cars_2024'[date_reg]) = 8,  
    YEAR('cars_2024'[date_reg]) = 2024  
)
```

- Create New Measure lagi, Copy and Paste code yang sama untuk buat bagi bulan-bulan yang lain.

- Buat untuk keseluruhan bulan, contoh:
– **Total Car Sales in January 2024, Total Car Sales in February 2024 --> Total Car Sales in October 2024**

3. Create tables (DAX Tables) untuk simpan values Sales

1. Lepas kita dah buat new measure tadi, setiap measures tu akan simpan values sales untuk bulan tertentu.
Dalam Power BI, “measures” ini tidak boleh dirujuk sebagai dataset.
Maksudnya dia tidak seperti “field data” yang boleh dipanggil oleh code python.
Python akan panggil value dari dataset dengan code: `df=dataset`. (cuba check code python)
2. Namun kita boleh buat table gunakan code DAX (Power BI) untuk create kan tables supaya code python boleh baca dataset dari table DAX. Contoh code DAX:

```
SalesTable =
UNION(
    ROW("date", DATE(2024, 1, 1), "month", 1, "sales", [Total Car Sales in January 2024]),
    ROW("date", DATE(2024, 2, 1), "month", 2, "sales", [Total Car Sales in February 2024]),
    ROW("date", DATE(2024, 3, 1), "month", 3, "sales", [Total Car Sales in March 2024]),
    ROW("date", DATE(2024, 4, 1), "month", 4, "sales", [Total Car Sales in April 2024]),
    ROW("date", DATE(2024, 5, 1), "month", 5, "sales", [Total Car Sales in May 2024]),
    ROW("date", DATE(2024, 6, 1), "month", 6, "sales", [Total Car Sales in June 2024]),
    ROW("date", DATE(2024, 7, 1), "month", 7, "sales", [Total Car Sales in July 2024]),
    ROW("date", DATE(2024, 8, 1), "month", 8, "sales", [Total Car Sales in August 2024]),
    ROW("date", DATE(2024, 9, 1), "month", 9, "sales", [Total Car Sales in September 2024]),
    ROW("date", DATE(2024, 10, 1), "month", 10, "sales", [Total Car Sales in October 2024])
)
```

Dalam code DAX atas ni, setiap nilai sales ikut bulan itu akan refer pada values yang tersimpan dalam “Measures”. contoh bulan january akan simpan nilai measures dari `[Total Car Sales in January 2024]`.

3. Jadi, jika kita nak tahu nilai sales bagi brand Toyota, kita perlu buat measures yang lain, kemudian buat tables untuk simpan nilai measures itu. Contoh DAX code untuk create measures bagi sales toyota bulan april:

```
Toyota Sales April 2024 =
CALCULATE(
    COUNT('cars_2024'[maker]),
    MONTH('cars_2024'[date_reg]) = 4,
    'cars_2024'[maker] = "Toyota"
)
```

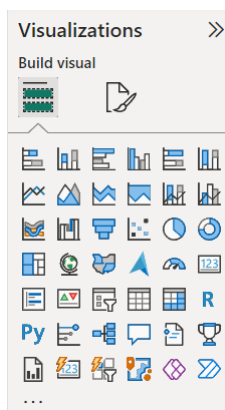
4. Perkara ni nampak macam leceh disebabkan oleh dataset asal itu tidak mempunyai sebarang nilai sales quantity tetapi hanyalah lebih kepada rekod log register jualan.

4. Guna kod python untuk baca DAX Tables

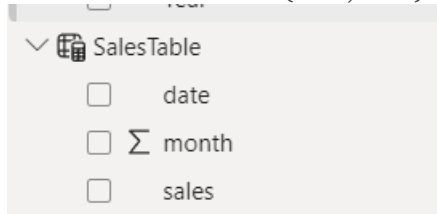
1. Setelah selesai create nilai sales dalam DAX tables seperti SalesTable, dan ToyotaSalesTable. Kita boleh gunakan code python untuk menganalisis trend dan buat forecasting.
2. Dalam contoh power BI itu, ada 3 jenis Predictive Model iaitu Prophet, Scikit-learn Linear Regression, dan juga Random Forest.
3. Perlu ingat yang kalau nak guna predictive model, kita perlu install library ini di python terlebih dahulu. Contoh yang kita pernah buat: `pip install prophet scikit-learn`. Ingat bukan?
4. Jadi barulah kita boleh gunakan library ini dalam PowerBI.

5. Cara Create Visualization Python (untuk paparkan graf forecast)

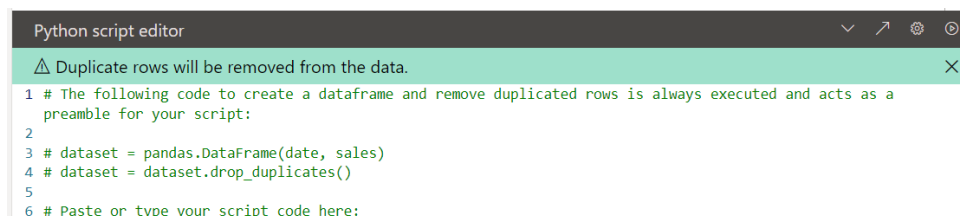
1. Klik pada pane Visualization di sebelah kanan. (sebelah pane Data)
2. Pilih ikon Py (Python visual) dan add masuk dalam report canvas.



3. Select column dataset (date,sales) dari DAX tables (SalesTables) yang dah buat tadi.



4. Bila dah select, barulah kita boleh masukkan code Python pada ruangan `Python script editor`. Jangan lupa tekan butang bulat RUN yang kecil tu kat bahagian atas-kanan Python Script editor untuk run code python.



Ini contoh code Prophet:

```
# dataset = pandas.DataFrame(date, sales)
# dataset = dataset.drop_duplicates()

# Paste or type your script code here:
# Import libraries
import pandas as pd
from prophet import Prophet
import matplotlib.pyplot as plt

# Ambil dataset dari Power BI
df = dataset # Power BI menyediakan 'dataset' secara otomatis

# Pastikan kolom nama sesuai: 'date' -> 'ds', 'sales' -> 'y'
df.rename(columns={'date': 'ds', 'sales': 'y'}, inplace=True)

# Format tarikh
df['ds'] = pd.to_datetime(df['ds'])

# Inisialisasi model Prophet
model = Prophet()
model.fit(df)

# Buat data masa depan untuk 6 bulan
future = model.make_future_dataframe(periods=6, freq='M')
forecast = model.predict(future)

# Visualisasi ramalan
model.plot(forecast)
plt.title('Car Sales Forecast (Next 6 Months)')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.show()

# Paparkan data ramalan untuk 6 bulan
print(forecast[['ds', 'yhat', 'yhat_lower', 'yhat_upper']].tail(6))
```

Ini contoh code Scikit-learn Linear Regression:

```
# Import libraries
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt

# Data: Ambil dataset dari Power BI
df = dataset # Power BI menyediakan 'dataset' secara otomatis

# Format data
df['date'] = pd.to_datetime(df['date'])
df['month'] = df['date'].dt.month # Extract bulan
df['year'] = df['date'].dt.year # Extract tahun
df['sales'] = df['sales'] # Pastikan 'sales' adalah kolom nilai jualan

# Model Training: Gunakan data bulan dan tahun sebagai feature
```

```

df['time_index'] = (df['year'] - df['year'].min()) * 12 + df['month'] # Time
index
X = df[['time_index']] # Feature: Time index (untuk menjaga urutan masa)
y = df['sales'] # Target: Jumlah penjualan

# Fit model Linear Regression
model = LinearRegression()
model.fit(X, y)

# Forecast untuk 6 bulan berikutnya dari bulan terakhir
last_time_index = df['time_index'].max()
future_time_index = np.arange(last_time_index + 1, last_time_index + 7) # Ta
mbah 6 bulan
future_months = pd.DataFrame({'time_index': future_time_index})
future_months['year'] = df['year'].min() + (future_time_index // 12)
future_months['month'] = future_time_index % 12
future_months['month'] = future_months['month'].replace(0, 12) # Betulkan bu
lan ke Desember jika modulus = 0
future_months['year'] += (future_months['month'] == 12).astype(int) # Betulk
an tahun untuk Desember

# Forecast sales
future_months['forecast_sales'] = model.predict(future_months[['time_index']]
)

# Gabungkan data sebenar dan forecast untuk visualisasi
df['type'] = 'Actual'
future_months['type'] = 'Forecast'
combined_data = pd.concat([df, future_months], ignore_index=True)

# Visualisasi
plt.figure(figsize=(10, 5))
plt.scatter(df['time_index'], df['sales'], color='blue', label='Actual Sales'
)
plt.plot(future_months['time_index'], future_months['forecast_sales'], color=
'red', linestyle='--', label='Forecast Sales')
plt.xlabel('Time Index')
plt.ylabel('Sales')
plt.title('Car Sales Forecast')
plt.legend()
plt.grid(True)
plt.show()

# Cetak ramalan
print(future_months[['year', 'month', 'forecast_sales']])

```

Ini contoh code Random Forest:

```

# The following code to create a dataframe and remove duplicated rows is alwa
ys executed and acts as a preamble for your script:

# dataset = pandas.DataFrame(date, sales)
# dataset = dataset.drop_duplicates()

# Paste or type your script code here:
# Import libraries

```

```

import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestRegressor
import seaborn as sns
import matplotlib.pyplot as plt

# Data: Ambil dataset dari Power BI
df = dataset # Power BI menyediakan 'dataset' secara otomatis

# Format data
df['date'] = pd.to_datetime(df['date'])
df['month'] = df['date'].dt.month
df['year'] = df['date'].dt.year
df['time_index'] = (df['year'] - df['year'].min()) * 12 + df['month']
df['sales'] = df['sales']

# Model Training: Random Forest Regressor
X = df[['time_index']]
y = df['sales']
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X, y)

# Forecast untuk 6 bulan berikutnya
last_time_index = df['time_index'].max()
future_time_index = np.arange(last_time_index + 1, last_time_index + 7)
future_months = pd.DataFrame({'time_index': future_time_index})
future_months['year'] = df['year'].min() + (future_time_index // 12)
future_months['month'] = future_time_index % 12
future_months['month'] = future_months['month'].replace(0, 12)
future_months['year'] += (future_months['month'] == 12).astype(int)

# Predict future sales
future_months['forecast_sales'] = model.predict(future_months[['time_index']])

# Gabungkan data sebenar dan forecast untuk visualisasi
df['type'] = 'Actual'
future_months['type'] = 'Forecast'
combined_data = pd.concat([df, future_months], ignore_index=True)

# Meniru gaya Seaborn dengan matplotlib
plt.style.use('ggplot')
plt.figure(figsize=(12, 6))
plt.plot(df['time_index'], df['sales'], marker='o', label='Actual Sales', color='blue')
plt.plot(future_months['time_index'], future_months['forecast_sales'], marker='o', linestyle='--', label='Forecast Sales', color='red')
plt.title('All Car Sales Forecast (Random Forest Regressor)', fontsize=16)
plt.xlabel('Time Index', fontsize=12)
plt.ylabel('Sales', fontsize=12)
plt.legend(title='Type')
plt.grid(True)
plt.show()

# Cetak ramalan
print(future_months[['year', 'month', 'forecast_sales']])

```

Secara asasnya, ini lah kaedah nak run code forecast dalam Power BI.
untuk kes data ***cars_2024*** ini memang perlu create banyak **Measures** supaya dapat
hasilkan lebih banyak lagi data untuk tujuan forecast dan juga visual dashboard di Power BI.

WISH U ALL THE BEST SAYANG!