LAPORAN
PRAKTIK DATA ENGINEERING
TEAM PROJECT KE-2

Oleh :

Dzaki Anwar	(234311037)
Ikbal Dzakwan	(234311043)
Veri Tri Anggoro	(234311055)
Widya Wulandari	(234311056)

JURUSAN TEKNIK
PROGRAM STUDI REKAYASA PERANGKAT LUNAK
POLITEKNIK NEGERI MADIUN

2025

Project Information
Project Name : Dampak Fluktuasi Nilai Tukar terhadap Kinerja Ekspor dan Import Indonesia
Created By : Data Engineering Kelompok 4
Date : 16 Februari, 2025
Version : 1.0

1. **Data Source Analysis**
   1.1 **Data Total Nilai Ekspor non-migas Indonesia**
   Source Details
   - Dataset Name : Perkembangan Ekspor Non Migas (Komoditi)
   - URL/Access Point : https://satudata.kemendag.go.id/data-informasi/perdagangan-luar-negeri/ekspor-non-migas-komoditi
   - Data Owner : Kementerian Perdagangan RI
   - Update Frequency : Tahunan

   1.2 **Perkembangan Impor Non Migas (Komoditi)**
   Source Details
   - Dataset Name : Perkembangan Impor Non Migas (Komoditi)
   - URL/Access Point : https://satudata.kemendag.go.id/data-informasi/perdagangan-luar-negeri/impor-non-migas-komoditi
   - Data Owner : Kementerian Perdagangan RI
   - Update Frequency : Tahunan

   1.3 **Data KURS Nilai Tukar Rupiah terhadap Mata Uang USD**
   Source Details
   - Dataset Name : Kurs Transaksi Bank Indonesia Mata Uang Usd(dari data center ortax.org)
   - URL/Access Point : https://datacenter.ortax.org/ortax/kursbi/show/USD?rentang_tanggal=2020-01-01,2024-12-31&show=USD&page=1
   - Data Owner : Bank Indonesia
   - Update Frequency : Harian

2. **Implementasi Teknolog**
   2.1 Python : bahasa pemrograman yang digunakan.
   2.2 Pandas : manipulasi, pembersihan, dan penyimpanan data (CSV & HTML).
   2.3 ydata-profiling : Exploratory Data Analysis (EDA).
   2.4 Web Scraping : untuk ambil data dari website menggunakan kombinasi requests dan pandas.read_html.
   2.5 Google Colab : Digunakan sebagai environment untuk menjalankan kode dan memproses data.

3. **Hasil dan Pembahasan**
   3.1 **Implementasi Pembacaan Data Dengan Python**
      1. Kode Program

```python
# Implementasi Pembacaan Data Dengan Python
import pandas as pd

file_ekspor = "/content/ekspor_non_migas.csv"
file_impor = "/content/impor_non_migas.csv"
file_kurs = "/content/kurs_usd.csv"

def read_data():
  df_ekspor = pd.read_csv(file_ekspor)
  df_impor = pd.read_csv(file_impor)
  df_kurs = pd.read_csv(file_kurs)

  return df_ekspor, df_impor, df_kurs

df_ekspor, df_impor, df_kurs = read_data()

print("\nData Ekspor:")
print(df_ekspor.head())

print("\nData Import:")
print(df_impor.head())

print("\nData Kurs:")
print(df_kurs.head())
```

2. Hasil

```
Data Ekspor:
     HS                                          URAIAN      2020  \
0    HS                                          URAIAN    2020.0
1   NaN                                      NON MIGAS  154940.8
2  27.0  Bahan bakar mineral, minyak mineral dan produk...   17255.1
3  15.0  Lemak dan minyak hewani, nabati atau mikroba s...   20716.4
4  72.0                                  Besi dan baja   10861.6

       2021      2022      2023      2024  Trend (%)  2020 -  2024  \
0    2021.0    2022.0    2023.0    2024.0  Trend (%)  2020 -  2024
1  219362.1  275906.1  242852.5  248826.6                    11.06
2   32831.2   54998.7   43572.0   39651.8                     21.5
3   32919.1   35159.5   28451.8   26821.9                     3.78
4   20925.6   27800.4   26704.6   25801.5                    21.83

   Perub (%)  2024 -  2023  Peran (%)  2024  Jan-Mar  Jan-Mar.1  \
0  Perub (%)  2024 -  2023  Peran (%)  2024   2024.0     2025.0
1                     2.46           100.0  58403.9    62982.0
2                     -9.0           15.94   9681.3     8069.6
3                    -5.73           10.78   5964.4     8109.8
4                    -3.38           10.37   6094.2     6494.4

   Perub (%)  2025/  2024  Peran (%)  2025
0  Perub (%)  2025/  2024  Peran (%)  2025
1                    7.84           100.0
2                  -16.65           12.81
3                   35.97           12.88
4                    6.57           10.31

Data Import:
     HS                                          URAIAN      2020  \
0    HS                                          URAIAN    2020.0
1   NaN                                      NON MIGAS  127312.0
2  84.0  Reaktor nuklir, ketel, mesin dan peralatan mek...   21808.5
3  85.0    Mesin dan perlengkapan elektrik serta bagiannya   19048.6
4  72.0                                  Besi dan baja    6855.2

       2021      2022      2023      2024  Trend (%)  2020 -  2024  \
0    2021.0    2022.0    2023.0    2024.0  Trend (%)  2020 -  2024
1  170660.9  197030.7  186055.8  197383.6                    10.11
2   25845.8   31571.7   32155.2   33514.3                    11.38
3   22372.2   26398.8   25782.5   27046.4                      8.8
4   11957.1   13928.2   11381.1   10664.4                      8.7

   Perub (%)  2024 -  2023  Peran (%)  2024  Jan-Mar  Jan-Mar.1  \
0  Perub (%)  2024 -  2023  Peran (%)  2024   2024.0     2025.0
1                    6.09           100.0  45891.4    47227.2
2                    4.23           16.98   7895.9     7732.0
3                     4.9            13.7   6833.5     6802.1
4                    -6.3             5.4   2603.3     2280.6

   Perub (%)  2025/  2024  Peran (%)  2025
0  Perub (%)  2025/  2024  Peran (%)  2025
1                    2.91           100.0
2                  -2.08           16.37
3                  -0.46            14.4
4                  -12.4            4.83

Data Kurs:
                    Masa Berlaku    Kurs Jual    Kurs Beli Kurs Tengah*
0  02 January 2020 - 03 January 2020  Rp13.964,48  Rp13.825,53  Rp13.895,01
1  03 January 2020 - 04 January 2020  Rp13.968,50  Rp13.829,51  Rp13.899,01
2  06 January 2020 - 07 January 2020  Rp14.030,81  Rp13.891,20  Rp13.961,01
3  07 January 2020 - 08 January 2020  Rp13.988,60  Rp13.849,41  Rp13.919,01
4  08 January 2020 - 09 January 2020  Rp14.003,67  Rp13.864,33  Rp13.934,00
```

3.2 **Analisis Struktur Data Menggunakan Pandas**

1. Kode Program

```python
# Analisis Struktur Data Menggunakan Pandas
def Struktur_Data(df, name):
  print(f"\nStruktur Data {name}:")
  print(df.info())

df_ekspor, df_impor, df_kurs = read_data()

Struktur_Data(df_ekspor, "Ekspor")
Struktur_Data(df_impor, "Import")
Struktur_Data(df_kurs, "Kurs")
```

2. Hasil

```
Struktur Data Ekspor:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   HS                      99 non-null     object
 1   URAIAN                  100 non-null    object
 2   2020                    100 non-null    float64
 3   2021                    100 non-null    float64
 4   2022                    100 non-null    float64
 5   2023                    100 non-null    float64
 6   2024                    100 non-null    float64
 7   Trend (%)  2020 -  2024  100 non-null    object
 8   Perub (%)  2024 -  2023  100 non-null    object
 9   Peran (%)  2024          100 non-null    object
 10  Jan-Mar                 100 non-null    float64
 11  Jan-Mar.1               100 non-null    float64
 12  Perub (%)  2025/  2024  100 non-null    object
 13  Peran (%)  2025          100 non-null    object
dtypes: float64(7), object(7)
memory usage: 11.1+ KB
None

Struktur Data Import:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   HS                      99 non-null     object
 1   URAIAN                  100 non-null    object
 2   2020                    100 non-null    float64
 3   2021                    100 non-null    float64
 4   2022                    100 non-null    float64
 5   2023                    100 non-null    float64
 6   2024                    100 non-null    float64
 7   Trend (%)  2020 -  2024  100 non-null    object
 8   Perub (%)  2024 -  2023  100 non-null    object
 9   Peran (%)  2024          100 non-null    object
 10  Jan-Mar                 100 non-null    float64
 11  Jan-Mar.1               100 non-null    float64
 12  Perub (%)  2025/  2024  100 non-null    object
 13  Peran (%)  2025          100 non-null    object
dtypes: float64(7), object(7)
memory usage: 11.1+ KB
None

Struktur Data Kurs:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1215 entries, 0 to 1214
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Masa Berlaku  1215 non-null   object
 1   Kurs Jual     1215 non-null   object
 2   Kurs Beli     1215 non-null   object
 3   Kurs Tengah*  1215 non-null   object
dtypes: object(4)
memory usage: 38.1+ KB
None
```

3.3 **Membuat Profiling Data Sederhana (Menggunakan ydata-profiling)**

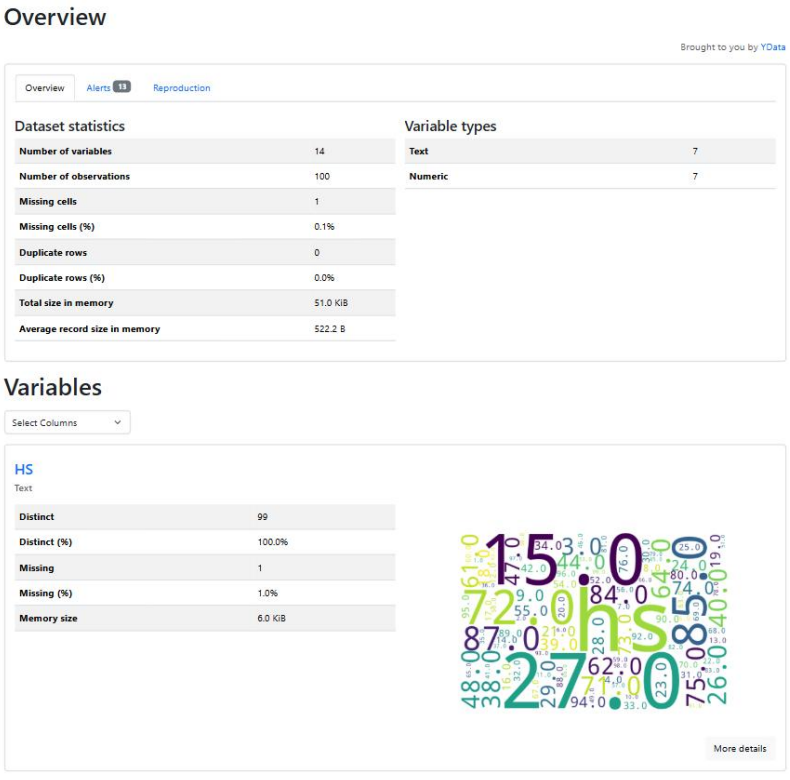
1. Kode Program

```python
#Membuat Profiling Data Sederhana (Menggunakan ydata-profiling)
import pandas as pd
from ydata_profiling import ProfileReport
import os

save_path = "/content"
os.makedirs(save_path, exist_ok=True)

df_ekspor = pd.read_csv(file_ekspor)
df_impor = pd.read_csv(file_impor)
df_kurs = pd.read_csv(file_kurs)

profile_ekspor = ProfileReport(df_ekspor, title="Profiling Data Ekspor", explorative=True)
profile_impor = ProfileReport(df_impor, title="Profiling Data Import", explorative=True)
profile_kurs = ProfileReport(df_kurs, title="Profiling Data Kurs", explorative=True)

profile_ekspor.to_file(os.path.join(save_path, "profiling_ekspor.html"))
profile_impor.to_file(os.path.join(save_path, "profiling_impor.html"))
profile_kurs.to_file(os.path.join(save_path, "profiling_kurs.html"))

print(f"Data Profiling berhasil disimpan dalam {save_path}")
```

2. Hasil

```
Upgrade to ydata-sdk
Improve your data and profiling with ydata-sdk, featuring data quality scoring, redundancy detection, outlier identification, text validation, and synthetic data generation.
Summarize dataset: 100% 72/72 [00:07<00:00, 7.50it/s, Completed]

  0%|          | 0/14 [00:00<?, ?it/s]
 50%|█████     | 7/14 [00:00<00:00, 68.32it/s]
100%|██████████| 14/14 [00:00<00:00, 55.57it/s]
Generate report structure: 100% 1/1 [00:08<00:00, 8.18s/it]
Render HTML: 100% 1/1 [00:01<00:00, 1.72s/it]
Export report to file: 100% 1/1 [00:00<00:00, 26.58it/s]
Summarize dataset: 100% 72/72 [00:06<00:00, 8.17it/s, Completed]

  0%|          | 0/14 [00:00<?, ?it/s]
 43%|████▎     | 6/14 [00:00<00:00, 59.97it/s]
100%|██████████| 14/14 [00:00<00:00, 51.06it/s]
Generate report structure: 100% 1/1 [00:08<00:00, 8.56s/it]
Render HTML: 100% 1/1 [00:01<00:00, 1.12s/it]
Export report to file: 100% 1/1 [00:00<00:00, 29.13it/s]
Summarize dataset: 100% 13/13 [00:00<00:00, 24.99it/s, Completed]

  0%|          | 0/4 [00:00<?, ?it/s]
100%|██████████| 4/4 [00:00<00:00, 30.93it/s]
Generate report structure: 100% 1/1 [00:02<00:00, 2.92s/it]
Render HTML: 100% 1/1 [00:00<00:00, 3.83it/s]
Export report to file: 100% 1/1 [00:00<00:00, 32.13it/s]
Data Profiling berhasil disimpan dalam /content
```

1. Data Ekspor



Overview

Brought to you by YData

Overview | Alerts 13 | Reproduction

Dataset statistics

Number of variables	14
Number of observations	100
Missing cells	1
Missing cells (%)	0.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	51.0 KiB
Average record size in memory	522.2 B

Variable types

Text	7
Numeric	7

Variables

Select Columns

HS

Text

Distinct	99
Distinct (%)	100.0%
Missing	1
Missing (%)	1.0%
Memory size	6.0 KiB

More details

2. Data Impor

Overview

Brought to you by YData

Overview | Alerts 14 | Reproduction

Dataset statistics

Number of variables	14
Number of observations	100
Missing cells	1
Missing cells (%)	0.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	51.0 KiB
Average record size in memory	521.9 B

Variable types

Text	7
Numeric	7

Variables

Select Columns

HS

Text

Distinct	99
Distinct (%)	100.0%
Missing	1
Missing (%)	1.0%
Memory size	6.0 KiB

More details

3. Data Kurs

Overview

Brought to you by YData

Overview | Reproduction

Dataset statistics

Number of variables	4
Number of observations	1215
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	347.1 KiB
Average record size in memory	292.5 B

Variable types

Text	4

Variables

Select Columns

Masa Berlaku

Text

Distinct	1214
Distinct (%)	99.9%
Missing	0
Missing (%)	0.0%
Memory size	105.0 KiB

More details

3.4 **Implementasi Struktur Data Dengan Python**

1. Kode Program

```
[9]  #Implementasi Struktur Data dengan Python
     import pandas as pd

     df_ekspor = pd.read_csv(file_ekspor)
     df_impor = pd.read_csv(file_impor)
     df_kurs = pd.read_csv(file_kurs)

     print("\nData Ekspor:")
     print(df_ekspor.info())

     print("\nData Import:")
     print(df_impor.info())

     print("\nData Kurs:")
     print(df_kurs.info())
```

2. Hasil

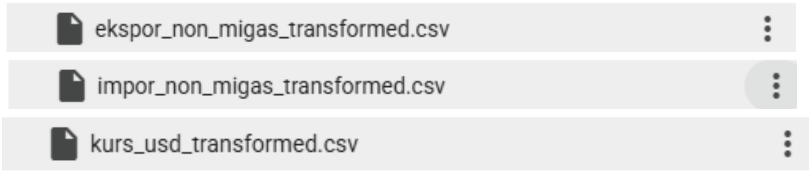
```
Data Ekspor:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   HS                        99 non-null     object
 1   URAIAN                    100 non-null    object
 2   2020                      100 non-null    float64
 3   2021                      100 non-null    float64
 4   2022                      100 non-null    float64
 5   2023                      100 non-null    float64
 6   2024                      100 non-null    float64
 7   Trend (%)  2020 -  2024   100 non-null    object
 8   Perub (%)  2024 -  2023   100 non-null    object
 9   Peran (%)  2024           100 non-null    object
 10  Jan-Apr                   100 non-null    float64
 11  Jan-Apr.1                 100 non-null    float64
 12  Perub (%)  2025/  2024    100 non-null    object
 13  Peran (%)  2025           100 non-null    object
dtypes: float64(7), object(7)
memory usage: 11.1+ KB
None

Data Import:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   HS                        99 non-null     object
 1   URAIAN                    100 non-null    object
 2   2020                      100 non-null    float64
 3   2021                      100 non-null    float64
 4   2022                      100 non-null    float64
 5   2023                      100 non-null    float64
 6   2024                      100 non-null    float64
 7   Trend (%)  2020 -  2024   100 non-null    object
 8   Perub (%)  2024 -  2023   100 non-null    object
 9   Peran (%)  2024           100 non-null    object
 10  Jan-Apr                   100 non-null    float64
 11  Jan-Apr.1                 100 non-null    float64
 12  Perub (%)  2025/  2024    100 non-null    object
 13  Peran (%)  2025           100 non-null    object
dtypes: float64(7), object(7)
memory usage: 11.1+ KB
None

Data Kurs:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1215 entries, 0 to 1214
Data columns (total 4 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Masa Berlaku  1215 non-null   object
 1   Kurs Jual     1215 non-null   object
 2   Kurs Beli     1215 non-null   object
 3   Kurs Tengah*  1215 non-null   object
dtypes: object(4)
memory usage: 38.1+ KB
None
```

3.5 **Membuat Fungsi untuk Transformasi Data**

1. Kode Program

```python
#Membuat Funsi untuk Transformasi Data
import pandas as pd

file_ekspor = "/content/ekspor_non_migas.csv"
file_impor = "/content/impor_non_migas.csv"
file_kurs = "/content/kurs_usd.csv"

df_ekspor = pd.read_csv(file_ekspor)
df_impor = pd.read_csv(file_impor)
df_kurs = pd.read_csv(file_kurs)

df_ekspor.columns = df_ekspor.columns.str.lower().str.replace(" ", "_")
df_impor.columns = df_impor.columns.str.lower().str.replace(" ", "_")
df_kurs.columns = df_kurs.columns.str.lower().str.replace(" ", "_")

if 'nilai_ekspor' in df_ekspor.columns:
    df_ekspor['nilai_ekspor_scaled'] = df_ekspor['nilai_ekspor'] * 1000

if 'nilai_impor' in df_impor.columns:
    df_impor['nilai_impor_scaled'] = df_impor['nilai_impor'] * 1000

if 'kurs' in df_kurs.columns:
    df_kurs['kurs'] = df_kurs['kurs'].astype(str).str.replace(",", "").astype(float)

df_ekspor.to_csv("/content/ekspor_non_migas_transformed.csv", index=False)
df_impor.to_csv("/content/impor_non_migas_transformed.csv", index=False)
df_kurs.to_csv("/content/kurs_usd_transformed.csv", index=False)

print("Transformasi selesai! Data tersimpan di Google Drive.")
```

2. Hasil



3.6 **Validasi Transformasi Data**

1. Kode Program

```
[11] #Validasi Transfromasi Data
     import pandas as pd

     file_ekspor_trans = "/content/ekspor_non_migas_transformed.csv"
     file_impor_trans = "/content/impor_non_migas_transformed.csv"
     file_kurs_trans = "/content/kurs_usd_transformed.csv"

     df_ekspor = pd.read_csv(file_ekspor_trans)
     df_impor = pd.read_csv(file_impor_trans)
     df_kurs = pd.read_csv(file_kurs_trans)

     print("Cek Missing Values:")
     print("Data Ekspor:", df_ekspor.isnull().sum())
     print("Data Impor:", df_impor.isnull().sum())
     print("Data Kurs:", df_kurs.isnull().sum())

     print("\nStatistik Data Ekspor:")
     print(df_ekspor.describe())

     print("\nStatistik Data Impor:")
     print(df_impor.describe())

     print("\nStatistik Data Kurs:")
     print(df_kurs.describe())

     print("\nTipe Data:")
     print("Data Ekspor:", df_ekspor.dtypes)
     print("Data Impor:", df_impor.dtypes)
     print("Data Kurs:", df_kurs.dtypes)

     print("\nValidasi Hasil Transformasi:")
     if 'nilai_ekspor_scaled' in df_ekspor.columns:
         print("Kolom nilai_ekspor_scaled berhasil ditambahkan!")

     if 'nilai_impor_scaled' in df_impor.columns:
         print("Kolom nilai_impor_scaled berhasil ditambahkan!")

     if 'kurs' in df_kurs.columns and df_kurs['kurs'].dtype == 'float64':
         print("Kolom kurs berhasil dikonversi ke float!")

     print("\nValidasi selesai! Pastikan tidak ada anomali pada data.")
```

2. Hasil

```
Cek Missing Values:
Data Ekspor: hs                          1
uraian                       0
2020                         0
2021                         0
2022                         0
2023                         0
2024                         0
trend_(%)__2020_-__2024      0
perub_(%)__2024_-__2023      0
peran_(%)__2024              0
jan-apr                      0
jan-apr.1                    0
perub_(%)__2025/__2024       0
peran_(%)__2025              0
dtype: int64
Data Impor: hs                          1
uraian                       0
2020                         0
2021                         0
2022                         0
2023                         0
2024                         0
trend_(%)__2020_-__2024      0
perub_(%)__2024_-__2023      0
peran_(%)__2024              0
jan-apr                      0
jan-apr.1                    0
perub_(%)__2025/__2024       0
peran_(%)__2025              0
dtype: int64
Data Kurs: masa_berlaku    0
kurs_jual       0
kurs_beli       0
kurs_tengah*    0
dtype: int64

Statistik Data Ekspor:
                2020           2021           2022           2023  \
count     100.00000     100.000000     100.000000     100.000000
mean     3119.01100    4407.451000    5538.341000    4877.281000
std     15669.57802   22344.079164   28257.026795   24792.151606
min         0.00000       0.000000       0.200000       0.100000
25%        60.55000      61.675000      85.625000      83.025000
50%       392.95000     507.050000     566.900000     539.300000
75%      1372.82500    1789.175000    2437.300000    2033.200000
max    154940.80000  219362.100000  275906.100000  242852.500000

                2024       jan-apr     jan-apr.1
count     100.000000    100.000000    100.000000
mean     4996.777000   1553.617000   1671.382000
std     25289.059419   7802.872263   8372.994845
min         0.100000      0.000000      0.000000
25%        81.925000     24.975000     24.775000
50%       525.950000    169.250000    169.300000
75%      2546.150000    747.225000    919.975000
max    248826.600000  76668.800000  82556.500000

Statistik Data Impor:
                2020           2021           2022           2023  \
count     100.000000     100.000000     100.000000     100.000000
mean     2566.439000    3433.425000    3960.838000    3741.344000
std     12961.375744   17305.353073   20012.544214   18934.925549
min         0.500000       0.600000       1.000000       0.900000
25%       143.575000     190.300000     197.675000     183.400000
50%       496.150000     595.900000     703.950000     630.150000
75%      1257.125000    1784.500000    1944.525000    1549.050000
max    127312.000000  170660.900000  197030.700000  186055.800000

                2024       jan-apr     jan-apr.1
count     100.000000    100.000000    100.000000
mean     3967.914000   1216.295000   1326.108000
std     20068.963836   6092.884692   6638.851911
min         0.900000      0.200000      0.200000
25%       191.850000     61.500000     62.275000
50%       672.800000    191.150000    213.600000
75%      1742.400000    510.575000    597.625000
max    197383.600000  59802.800000  65292.700000

Statistik Data Kurs:
                       masa_berlaku    kurs_jual    kurs_beli kurs_tengah*
count                          1215         1215         1215         1215
unique                         1214          928          926          933
top     23 May 2023 - 24 May 2023  Rp15.708,15  Rp14.290,19  Rp15.630,00
freq                              2            5            5            5
```

```
Tipe Data:
Data Ekspor: hs                            object
uraian                      object
2020                       float64
2021                       float64
2022                       float64
2023                       float64
2024                       float64
trend_(%)__2020_-__2024     object
perub_(%)__2024_-__2023     object
peran_(%)__2024             object
jan-apr                    float64
jan-apr.1                  float64
perub_(%)__2025/__2024      object
peran_(%)__2025             object
dtype: object
Data Impor: hs                            object
uraian                      object
2020                       float64
2021                       float64
2022                       float64
2023                       float64
2024                       float64
trend_(%)__2020_-__2024     object
perub_(%)__2024_-__2023     object
peran_(%)__2024             object
jan-apr                    float64
jan-apr.1                  float64
perub_(%)__2025/__2024      object
peran_(%)__2025             object
dtype: object
Data Kurs: masa_berlaku    object
kurs_jual       object
kurs_beli       object
kurs_tengah*    object
dtype: object

Validasi Hasil Transformasi:

Validasi selesai! Pastikan tidak ada anomali pada data.
```