

Nama : Rizqi Rohmatul Huda
Kelas : 2G-TI
Nomor Absen : 26
NIM : 2141720264
Link Github : https://github.com/RizqiRHuda/uas_statistik

1. Tentukan deskripsi dari data

- a. Berdasarkan data yang diberikan, kolom mana yang dapat digunakan untuk permodelan regresi dan tidak

Jawaban :

- Kolom yang dapat digunakan permodelan regresi
 - Weekly_Sales
 - Fuel_Price
 - CPI
 - Unemployment
 - Holiday_Flag
 - Temperature
- Kolom yang tidak dapat digunakan permodelan regresi
 - Store
 - Date

- b. Carilah nilai-nilai deskriptif mean, median, simpangan baku, varians pada kolom yang dapat digunakan untuk regresi sesuai nomor 1a hanya **untuk store**

Jawaban :

```
print('1.b')
store_id = 4
df_filtered = df[df['Store'] == store_id]

weekly_sale_stats = df_filtered['Weekly_Sales'].describe()
holiday_flag_stats = df_filtered['Holiday_Flag'].describe()
temperature_stats = df_filtered['Temperature'].describe()
fuel_price_stats = df_filtered['Fuel_Price'].describe()
customer_price_index_stats = df_filtered['CPI'].describe()
unemployment_stats = df_filtered['Unemployment'].describe()

print("Statistics for Store ID = 4:")
print("Weekly Sale:")
print(weekly_sale_stats)

print("Holiday Flag:")
print(holiday_flag_stats)

print("Temperature:")
print(temperature_stats)

print("Fuel Price:")
print(fuel_price_stats)

print("Customer Price Index:")
print(customer_price_index_stats)

print("Unemployment:")
print(unemployment_stats)
```

```

PS E:\Local Disc A\semester4\statistika & komput
statistika & komputasi/uas_statistik/uas.py"
Nomor 1

1.b
Statistics for Store ID = 4:
Weekly Sale:
count    1.430000e+02
mean     2.094713e+06
std      2.662014e+05
min      1.762539e+06
25%      1.929611e+06
50%      2.073951e+06
75%      2.175039e+06
max      3.676389e+06
Name: Weekly_Sales, dtype: float64
Holiday Flag:
count    143.000000
mean     0.069930
std      0.255926
min      0.000000
25%      0.000000
50%      0.000000
75%      0.000000
max      1.000000
Name: Holiday_Flag, dtype: float64
Temperature:
count    143.000000
mean     62.253357
std      16.180023
min      28.840000
25%      48.470000
50%      64.220000
75%      77.440000
max      86.090000
Name: Temperature, dtype: float64
Fuel Price:
count    143.000000
mean     3.216972
std      0.416967
min      2.540000
25%      2.764500
50%      3.290000
75%      3.586500
max      3.881000
Name: Fuel_Price, dtype: float64
Customer Price Index:
count    143.000000
mean     128.679669
std      1.858300
min      126.064000
25%      126.590094
50%      129.075677
75%      130.502414
max      131.193097

```

```

Name: CPI, dtype: float64
Unemployment:
count    143.000000
mean     5.964692
std      1.421267
min      3.879000
25%      4.607000
50%      5.946000
75%      7.127000
max      8.623000
Name: Unemployment, dtype: float64

```

- c. Tentukan nilai dari QQ1,Q2,Q3 dan IQR untuk 'Fuel_Price', 'CPI', dan 'Unemployment' **HANYA UNTUK STORE**

Jawaban :

```

print('1.c')
fuel_price_q1 = df_filtered['Fuel_Price'].quantile(0.25)
fuel_price_q2 = df_filtered['Fuel_Price'].quantile(0.50)
fuel_price_q3 = df_filtered['Fuel_Price'].quantile(0.75)
fuel_price_iqr = fuel_price_q3 - fuel_price_q1

cpi_q1 = df_filtered['CPI'].quantile(0.25)
cpi_q2 = df_filtered['CPI'].quantile(0.50)
cpi_q3 = df_filtered['CPI'].quantile(0.75)
cpi_iqr = cpi_q3 - cpi_q1

unemployment_q1 = df_filtered['Unemployment'].quantile(0.25)
unemployment_q2 = df_filtered['Unemployment'].quantile(0.50)
unemployment_q3 = df_filtered['Unemployment'].quantile(0.75)
unemployment_iqr = unemployment_q3 - unemployment_q1

print("Statistics for Store ID = 4:")
print("Fuel Price:")
print("Q1:", fuel_price_q1)
print("Q2:", fuel_price_q2)
print("Q3:", fuel_price_q3)
print("IQR:", fuel_price_iqr)

print("Customer Price Index:")
print("Q1:", cpi_q1)
print("Q2:", cpi_q2)
print("Q3:", cpi_q3)
print("IQR:", cpi_iqr)

print("Unemployment:")
print("Q1:", unemployment_q1)
print("Q2:", unemployment_q2)
print("Q3:", unemployment_q3)
print("IQR:", unemployment_iqr)

```

```

1.c
Statistics for Store ID = 4:
Fuel Price:
Q1: 2.7645
Q2: 3.29
Q3: 3.5865
IQR: 0.8220000000000001
Customer Price Index:
Q1: 126.59009445
Q2: 129.0756774
Q3: 130.5024138
IQR: 3.9123193500000004
Unemployment:
Q1: 4.607
Q2: 5.946
Q3: 7.127
IQR: 2.5199999999999996

```

- d. Variansi dari 1-holiday week dan 0-non holiday week

Jawaban :

```

print('1d')
grouped_data = df.groupby('Holiday_Flag')['Weekly_Sales'].var()
print("Variance Description:")
for flag, variance in grouped_data.items():
    if flag == 1:
        print("Holiday Week:")
    else:
        print("Non-Holiday Week:")
    print("Variance:", variance)
    print()

```

```

1d
Variance Description:
Non-Holiday Week:
Variance: 312433415424.3816

Holiday Week:
Variance: 393988373714.1925

```

- e. Apakah rata-rata weekly sales di setiap toko(store) sama?

Jawaban: Tidak, rata-rata Weekly Sales untuk setiap storenya adalah berbeda.
Berikut adalah datanya

```
print('1e')
average_sales_by_store = df.groupby('Store')['Weekly_Sales'].mean()
is_average_sales_equal = average_sales_by_store.nunique() == 1
if is_average_sales_equal:
    print("Rata-rata Weekly Sales di setiap toko sama.")
else:
    print("Rata-rata Weekly Sales di setiap toko berbeda.")
```

```
1e
Rata-rata Weekly Sales di setiap toko adalah berbeda.
```

- f. Dari setiap toko(store), CPI mana yang lebih tinggi

Jawaban :

```
93     print('1f')
94     max_cpi_by_store = df.groupby('Store')['CPI'].max()
95     higher_cpi_by_store = max_cpi_by_store.idxmax()
96     higher_cpi_value = max_cpi_by_store.max()
97
98     print("CPI yang lebih tinggi di setiap toko:")
99     for store_id in max_cpi_by_store.index:
100         cpi_value = max_cpi_by_store.loc[store_id]
101         print("Store ID:", store_id)
102         print("CPI:", cpi_value)
103         print()
```

```
1f
CPI yang lebih tinggi di setiap toko:
Store ID: 1
CPI: 223.4442513
```

- g. CPI mana yang lebih tinggi, holiday week atau non holiday week?

Jawaban :

```
05     print('1g')
06     average_cpi_holiday = df[df['Holiday_Flag'] == 1]['CPI'].mean()
07     average_cpi_non_holiday = df[df['Holiday_Flag'] == 0]['CPI'].mean()
08     if average_cpi_holiday > average_cpi_non_holiday:
09         print("Rata-rata CPI pada holiday week lebih tinggi.")
10     elif average_cpi_holiday < average_cpi_non_holiday:
11         print("Rata-rata CPI pada non-holiday week lebih tinggi.")
12     else:
13         print("Rata-rata CPI pada holiday week dan non-holiday week sama.")
```

```
1g
Rata-rata CPI pada non-holiday week lebih tinggi.
```

2. Berdasarkan data, tentukan

- a. Lakukan uji normalitas pada 'Weekly_Sales' dan 'Fuel_Price' menggunakan uji KS dengan $\alpha=0.05$. Apakah kedua nilai tersebut berasal dari distribusi normal? Kemukakan jawaban Anda dengan bukti hasil analisis

Jawaban :

```
7 class Nomor2() :
8     weekly_sales = df['Weekly_Sales']
9     fuel_price = df['Fuel_Price']
10    alpha = 0.05
11
12    print('Nomor 2\n')
13
14    statistic, p_value = kstest(weekly_sales, norm.fit(weekly_sales))
15    print("Uji Normalitas Weekly Sales:")
16    print(f"Statistic: {statistic}")
17    print(f"P-value: {p_value}")
18    if p_value > alpha:
19        print("Weekly Sales didistribusikan secara normal")
20    else:
21        print("Weekly Sales tidak didistribusikan secara normal")
22
23    statistic, p_value = kstest(fuel_price, norm.fit(fuel_price))
24    print("Uji Normalitas Fuel Price:")
25    print(f"Statistic: {statistic}")
26    print(f"P-value: {p_value}")
27    if p_value > alpha:
28        print("Fuel Price didistribusikan secara normal")
29    else:
30        print("Fuel Price tidak didistribusikan secara normal")
31
```

```
Uji Normalitas Weekly Sales:
Statistic: 0.44693084693084695
P-value: 0.6900128635984475
Weekly Sales didistribusikan secara normal
Uji Normalitas Fuel Price:
Statistic: 0.5383061383061383
P-value: 0.42655414389291124
Fuel Price didistribusikan secara normal
```

b. Berdasarkan soal 1a, tentukan variable independent dan variable dependen

➤ Variabel Independen:

- Holiday Flag: suatu variabel yang digunakan untuk menunjukkan apakah saat ini adalah hari libur atau tidak
- Temperature: variabel untuk mengukur suhu.
- Fuel Price: Variabel ini digunakan untuk mencatat atau memperoleh informasi terkini tentang harga bahan bakar yang dapat digunakan dalam perhitungan, pemantauan, atau tindakan lain yang berkaitan dengan harga bahan bakar.

- Customer Price Index: Variabel yang mengukur indeks harga konsumen.
- Unemployment: Variabel yang mengukur tingkat pengangguran.

➤ Variabel Dependen:

- Weekly Sale: Variabel target yang ingin diprediksi atau dijelaskan menggunakan variabel-variabel independen di atas

3. Berdasarkan data, tentukan

- Lakukan uji korelasi masing-masing variable indepent dengan variable dependen. Berapa nilai korelasinya? (Contoh pasangan korelasi:Weekly_Sales-CPI)
- Apakah ada pasangan variable independent dan dependen yang memiliki korelasi negative? Jika ya, sebutkan pasangan variabelnya

Jawaban :

```
32 class Nomor3() :
33     print()
34     print('Nomer 3')
35     print('3a')
36     correlation = df[['Holiday_Flag', 'Temperature', 'Fuel_Price', 'CPI', 'Unemployment', 'Weekly_Sales']].corr()
37     print("Nilai korelasi antara variabel independen dan variabel dependen:")
38     print(correlation['Weekly_Sales'])
39
40     print('3b')
41     correlation = df[['Holiday_Flag', 'Temperature', 'Fuel_Price', 'CPI', 'Unemployment', 'Weekly_Sales']].corr()
42     negative_correlations = correlation[correlation['Weekly_Sales'] < 0]
43     negative_correlations = negative_correlations['Weekly_Sales'].drop('Weekly_Sales', errors='ignore')
44     if negative_correlations.empty:
45         print("Tidak ada pasangan variabel independen dan dependen dengan korelasi negatif.")
46     else:
47         print("Pasangan variabel independen dan dependen dengan korelasi negatif:")
48         print(negative_correlations)
```

3a.

```
Nomer 3
3a
Nilai korelasi antara variabel independen dan variabel dependen:
Holiday_Flag    0.036891
Temperature     -0.063810
Fuel_Price      0.009464
CPI             -0.072634
Unemployment    -0.106176
Weekly_Sales    1.000000
Name: Weekly_Sales, dtype: float64
```

3b.

```
3b
Pasangan variabel independen dan dependen dengan korelasi negatif:
Temperature     -0.063810
CPI             -0.072634
Unemployment    -0.106176
Name: Weekly_Sales, dtype: float64
```

4. Buatlah model regresi berdasarkan variable independent 'Fuel_Price'. Nyatakan model Anda dalam bentuk $y=a+bx$

Jawaban :

```
51 class Nomor4() :
52     print()
53     print('Nomer 4')
54     data = df[['Fuel_Price', 'Weekly_Sales']]
55
56     x = data[['Fuel_Price']]
57     y = data[['Weekly_Sales']]
58
59     model = LinearRegression()
60     model.fit(x, y)
61
62     a = model.intercept_
63     b = model.coef_[0]
64
65     print("Model regresi: y = {} + {}x".format(a, b))
66     data = df[['Fuel_Price', 'Weekly_Sales']]
67
68     x = data[['Fuel_Price']]
69     y = data[['Weekly_Sales']]
70
71     model = LinearRegression()
72
73     model.fit(x, y)
74     y_pred = model.predict(x)
75     plt.scatter(x, y, color='blue', label='Data')
76     plt.plot(x, y_pred, color='red', linewidth=2, label='Regression Line')
77
78     plt.xlabel('Fuel_Price')
79     plt.ylabel('Weekly_Sales')
80     plt.title('Linear Regression')
81
82     plt.legend()
83
84     plt.show()
```

Nomer 4

Model regresi: $y = 1007884.9200801749 + 11635.764282979584x$

