

10[~]



member of ASTRA

ANOVA

TEST

Ahmad Ardra Damarjati | 71486

Michael Abelard | 71487

Nabila Karin | 71478



WHAT IS ANOVA?

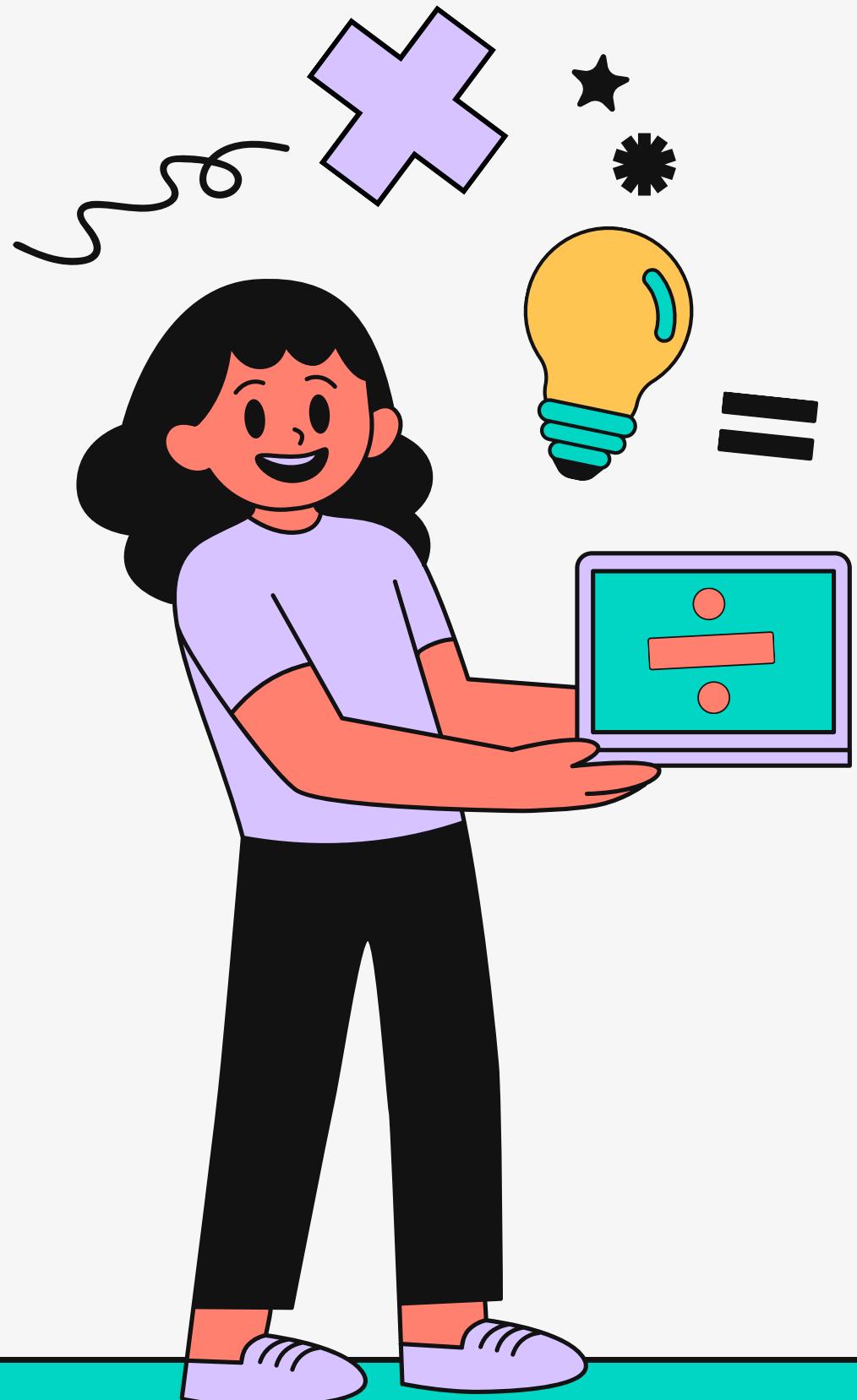


ANOVA adalah metode statistik yang digunakan untuk menguji apakah ada perbedaan rata-rata antar lebih dari dua kelompok.

Tujuan

- Menguji pengaruh dari dua atau lebih faktor
- Mendekripsi perbedaan varians antara kelompok

Core Idea of ANOVA



Variasi antar kelompok (apakah kelompoknya beda?)

Variasi dalam kelompok (apakah perbedaan ini cuma karena variasi internal?)

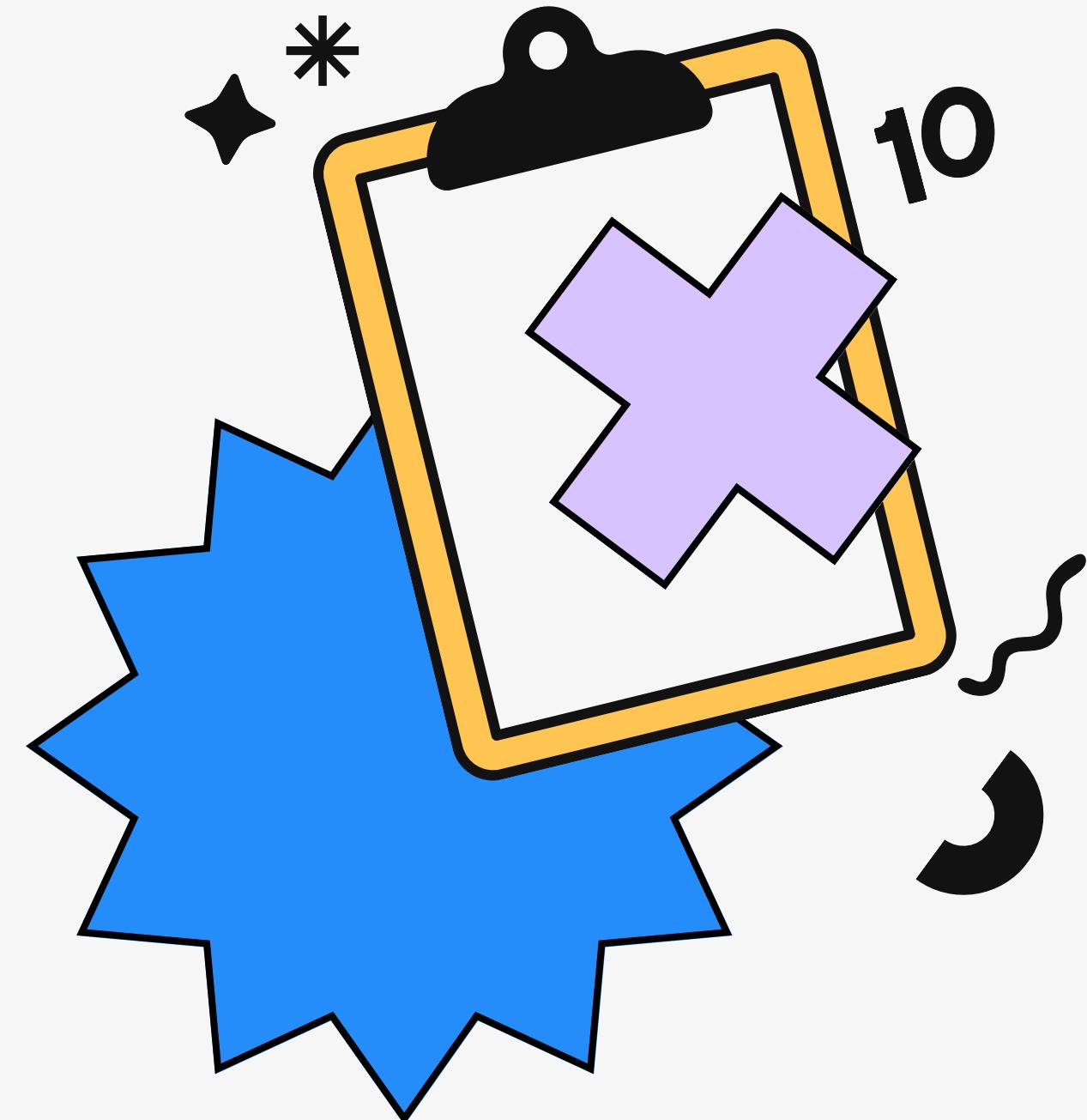
BETWEEN > WITHIN = Rata-ratanya mungkin memang benar-benar beda secara signifikan

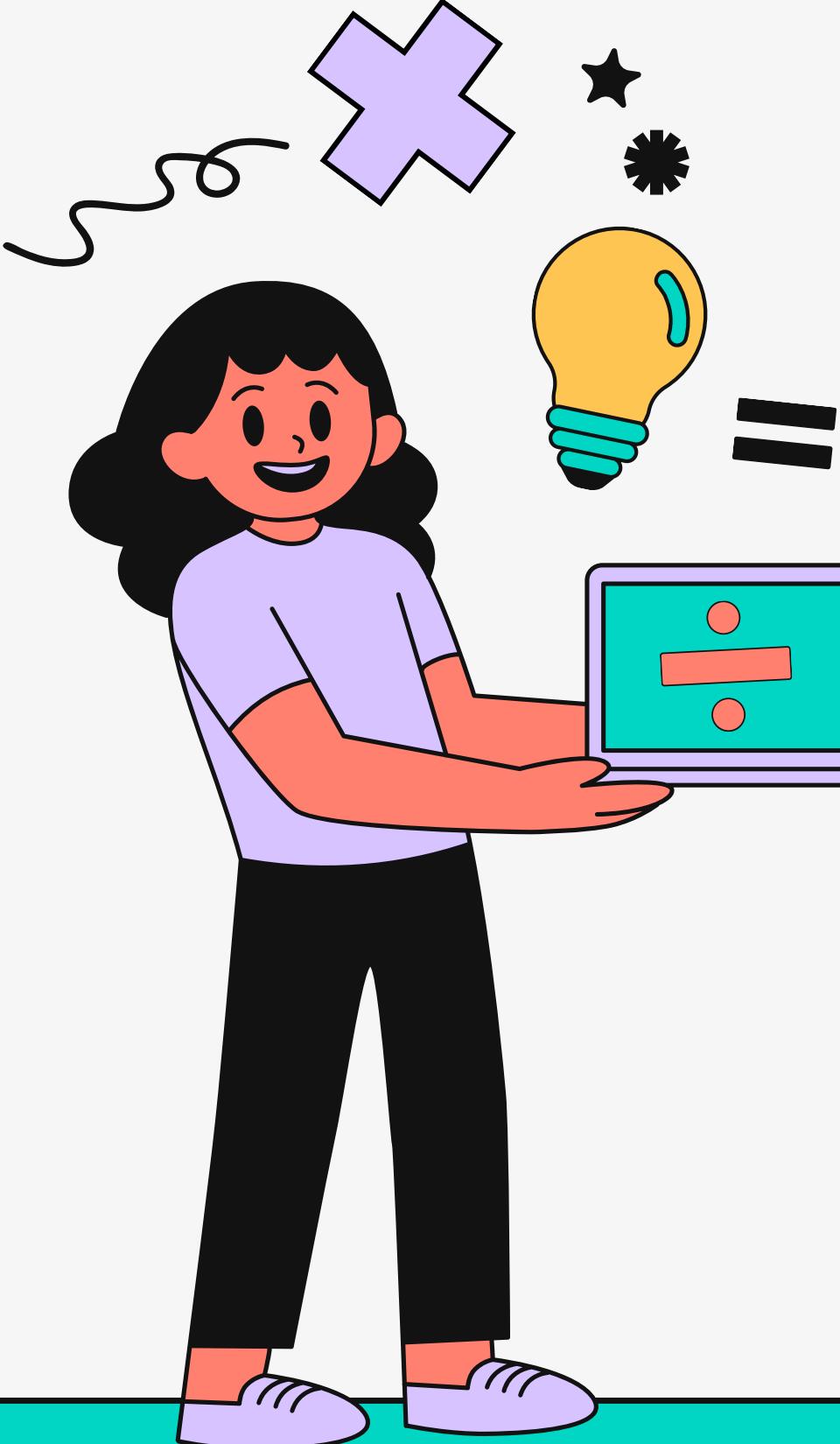
Jenis-jenis ANOVA

1	2	3
One-Way ANOVA	Two-Way ANOVA	N-Way ANOVA
Membandingkan rata-rata dari dua atau lebih kelompok berdasarkan satu faktor (variabel independen)	Menguji pengaruh dari dua faktor (dua variabel independen)	Menguji pengaruh dari tiga/lebih faktor (variabel independen) terhadap satu variabel dependen serta interaksi antar faktor-faktor tersebut.

CARA KERJA

ANOVA



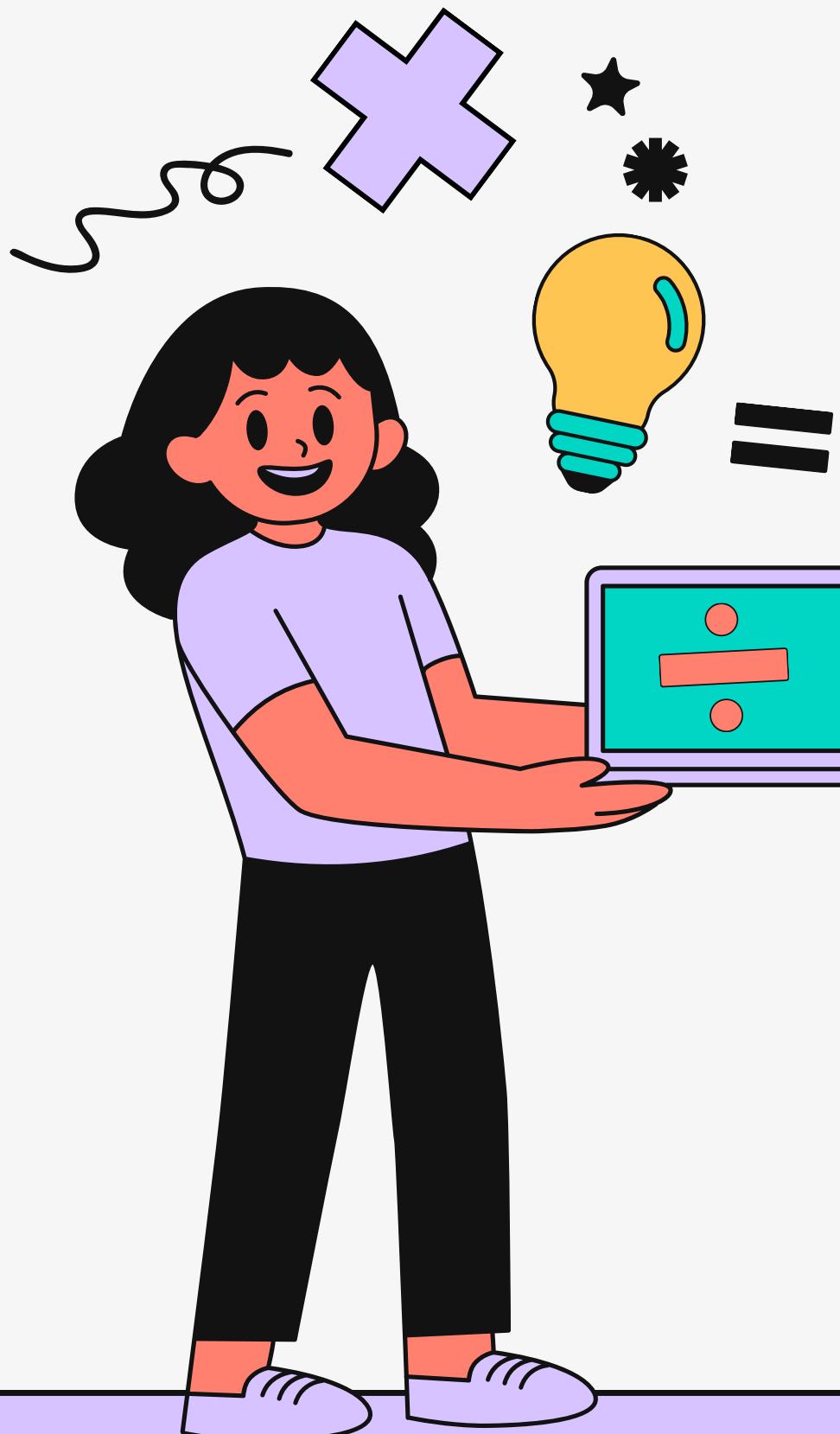


1. Formulasi HIPOTESIS

HO: Semua grup tidak memiliki perbedaan rata-rata yang signifikan

H1: Setidaknya 1 grup memiliki perbedaan rata-rata yang signifikan

2. Hitung VARIASI



SSB(Sum of Square Between Groups): Variasi karena beda antar kelompok

SSW(Sum of Square Within Group): Variasi di dalam kelompok

$$SSB = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$$

k : jumlah grup

n_i : jumlah data di grup ke- i

$N = \sum n_i$: total seluruh data

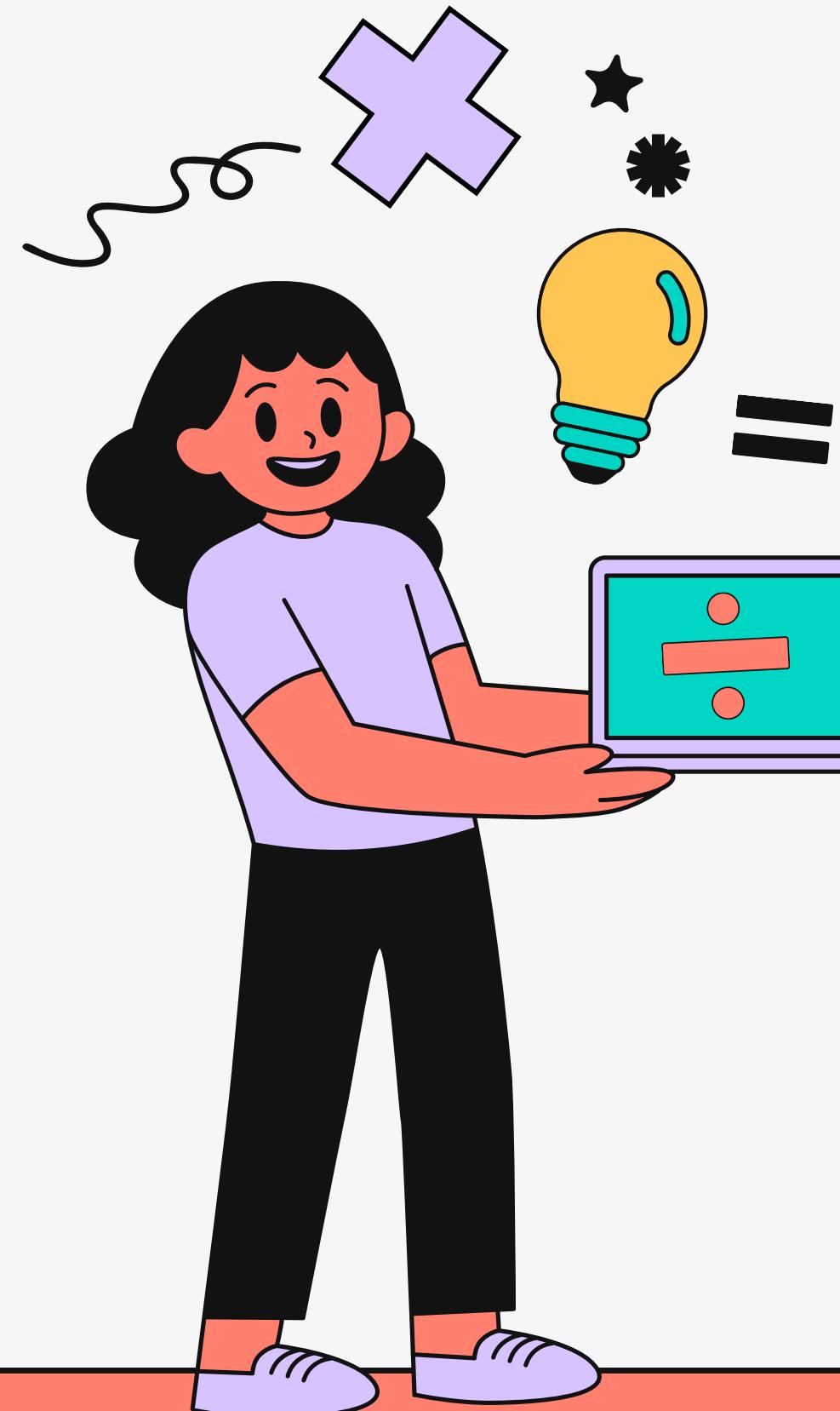
\bar{X}_i : rata-rata grup ke- i

\bar{X} : rata-rata total

$$SSW = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

X_{ij} : data ke- j dari grup ke- i

3. Hitung F-Statistics

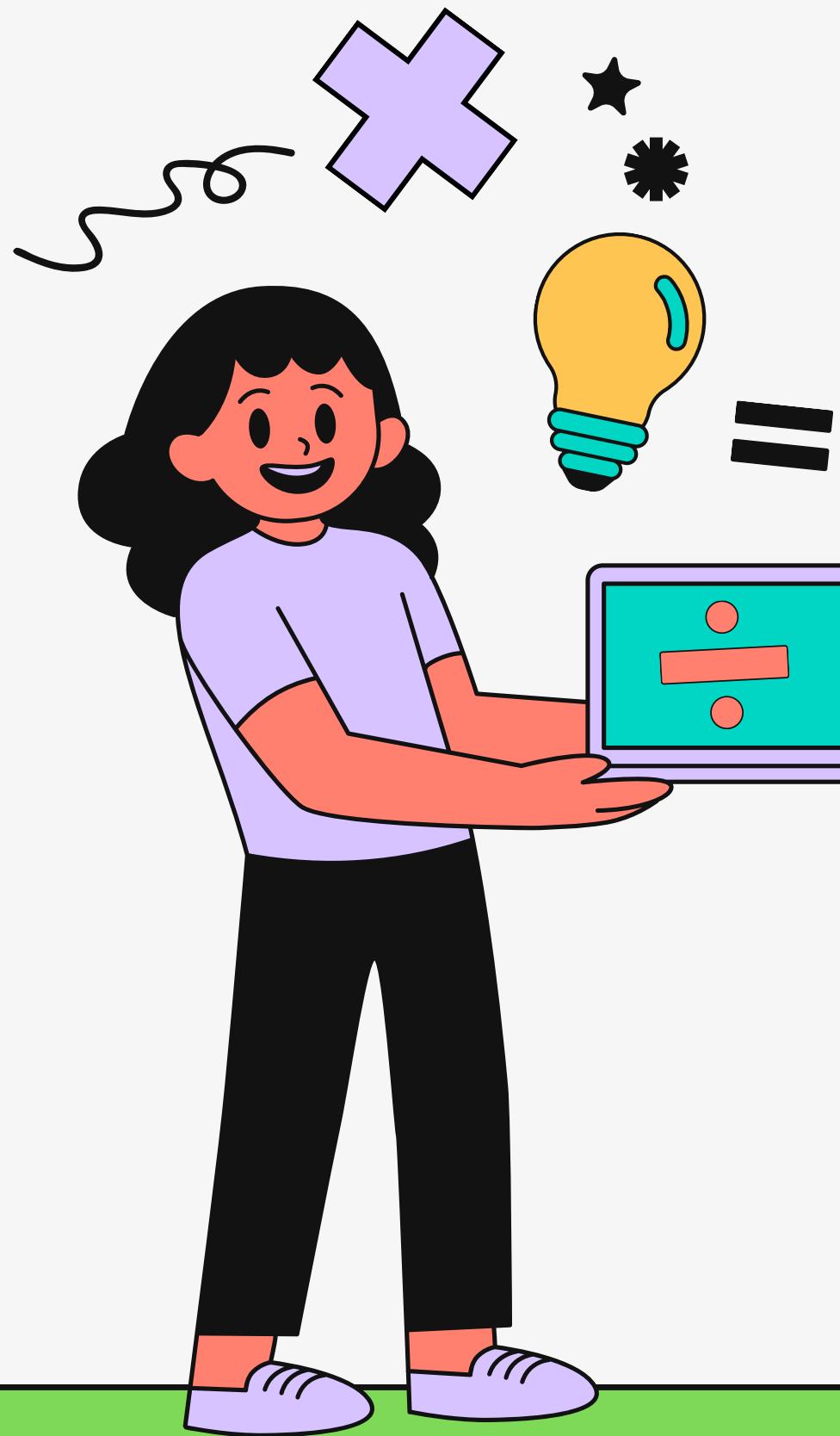


Statistik F = Variansi antar grup/ Variansi antar kelompok

$$F = \frac{SSB/(k-1)}{SSW/(N-k)}$$

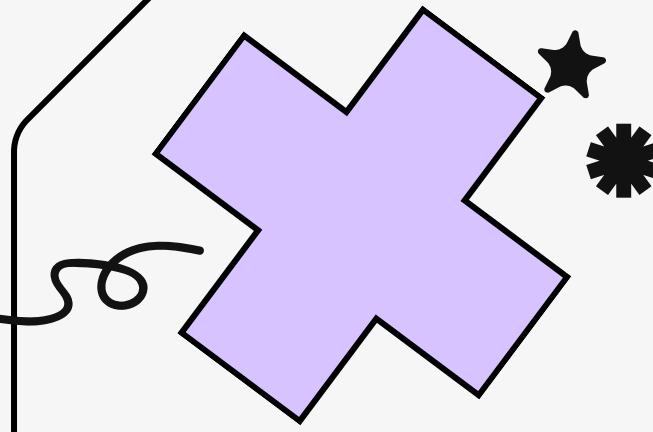
Simbol	Arti
F	Nilai statistik F
SSB	Sum of Squares Between (jumlah variasi antar grup)
SSW	Sum of Squares Within (jumlah variasi dalam grup)
k	Jumlah kelompok/grup
N	Total seluruh data
k - 1	Derajat bebas antar grup
N - k	Derajat bebas dalam grup

4 . Bandingkan Nilai



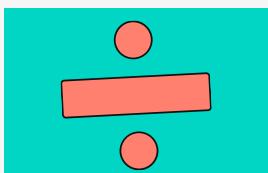
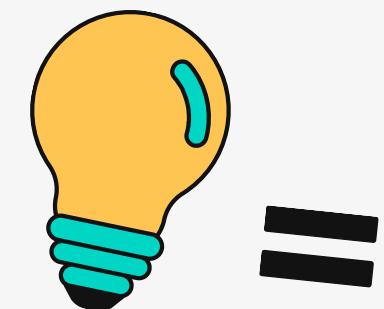
P-Value > α , Terima H_0 , Tidak terdapat perbedaan yang signifikan antar grup

P-Value < α , Tolak H_0 , terdapat perbedaan yang signifikan antar grup



Asumsi ANOVA

Normalitas	Data dari setiap grup harus berdistribusi normal
Homoskedastisitas	Variansi antar grup harus sama atau seragam
Interdependensi	Data dari satu subjek tidak boleh mempengaruhi subjek lainnya.



DATASET

SuperStore

```
● df = df[['order_priority', 'sub_category', 'region', 'sales']]  
df.head(5)
```

✓ 0.0s ━ Open 'df' in Data Wrangler

	▲ order_priority	▲ sub_category	▲ region	▲ sales
0	Medium	Storage	Africa	408
1	Medium	Supplies	Oceania	120
2	High	Storage	EMEA	66
3	High	Paper	North	45
4	Medium	Furnishings	Oceania	114



Kategorik

Numerik

ONE-WAY ANOVA

```
# One-Way ANOVA
model_1way = ols('sales ~ C(order_priority)', data=df).fit()
anova_1way = sm.stats.anova_lm(model_1way, typ=2)
# Tampilkan dengan format notasi ilmiah (sci)
pd.options.display.float_format = '{:.6e}'.format
print("One-Way ANOVA:\n", anova_1way)
```

One-Way ANOVA:

	sum_sq	df	F	PR(>F)
C(order_priority)	4.710560e+05	3.000000e+00	6.605014e-01	5.762584e-01
Residual	1.219204e+10	5.128600e+04	NaN	NaN

P-Value(0.0576) > α , menerima HO
Terdapat perbedaan yg tidak
signifikan terhadap kolom
order_priority terhadap kolom **sales**

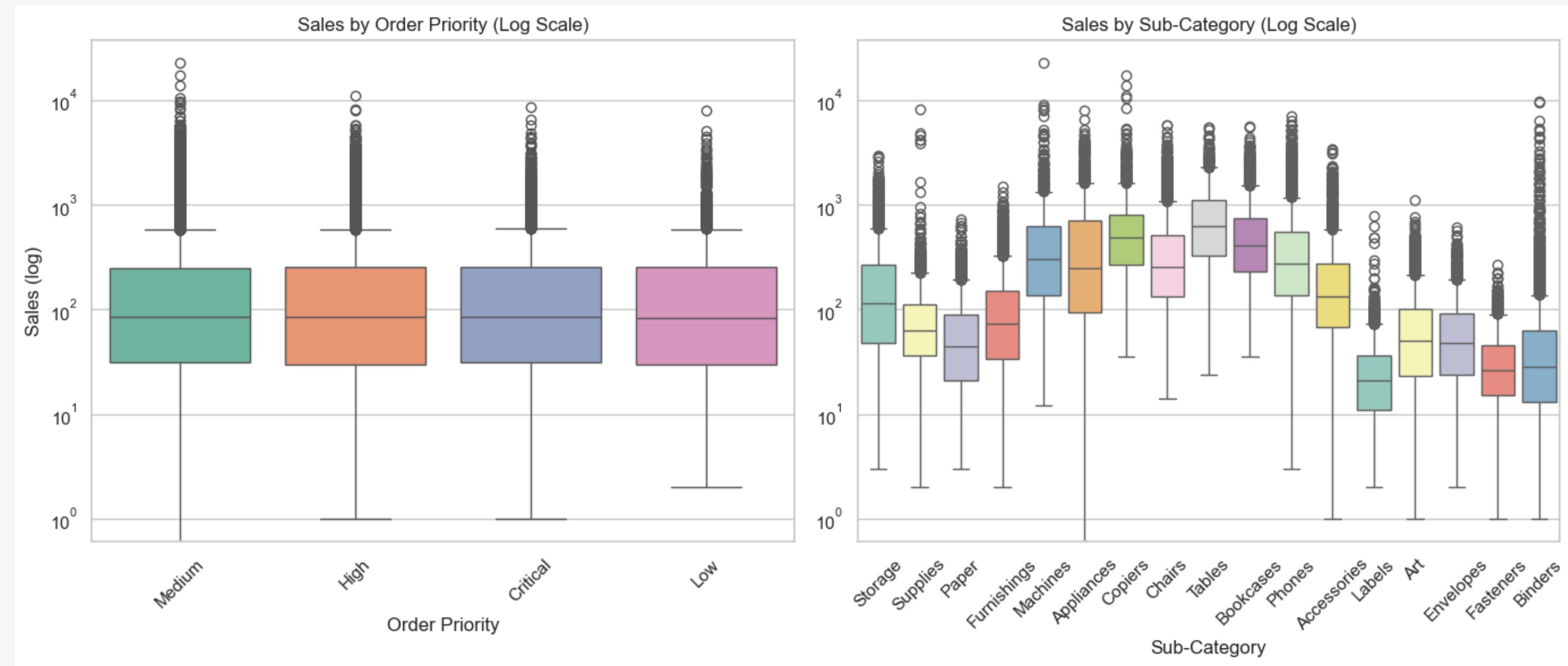
```
# One-Way ANOVA
model_1way = ols('sales ~ C(sub_category)', data=df).fit()
anova_1way = sm.stats.anova_lm(model_1way, typ=2)
# Tampilkan dengan format notasi ilmiah (sci)
pd.options.display.float_format = '{:.6e}'.format
print("One-Way ANOVA:\n", anova_1way)
```

One-Way ANOVA:

	sum_sq	df	F	PR(>F)
C(sub_category)	2.583494e+09	1.600000e+01	8.615832e+02	0.000000e+00
Residual	9.609017e+09	5.127300e+04	NaN	NaN

P-Value(0.000) < α , tolak HO
Terdapat perbedaan yg signifikan
terhadap kolom **sub_category**
terhadap kolom **sales**

ONE-WAY ANOVA



TWO-WAY ANOVA

```
model_2way = ols('sales ~ C(sub_category) * C(region)', data=df).fit()
anova_2way = sm.stats.anova_lm(model_2way, typ=2)
print("Two-Way ANOVA:\n", anova_2way)
```

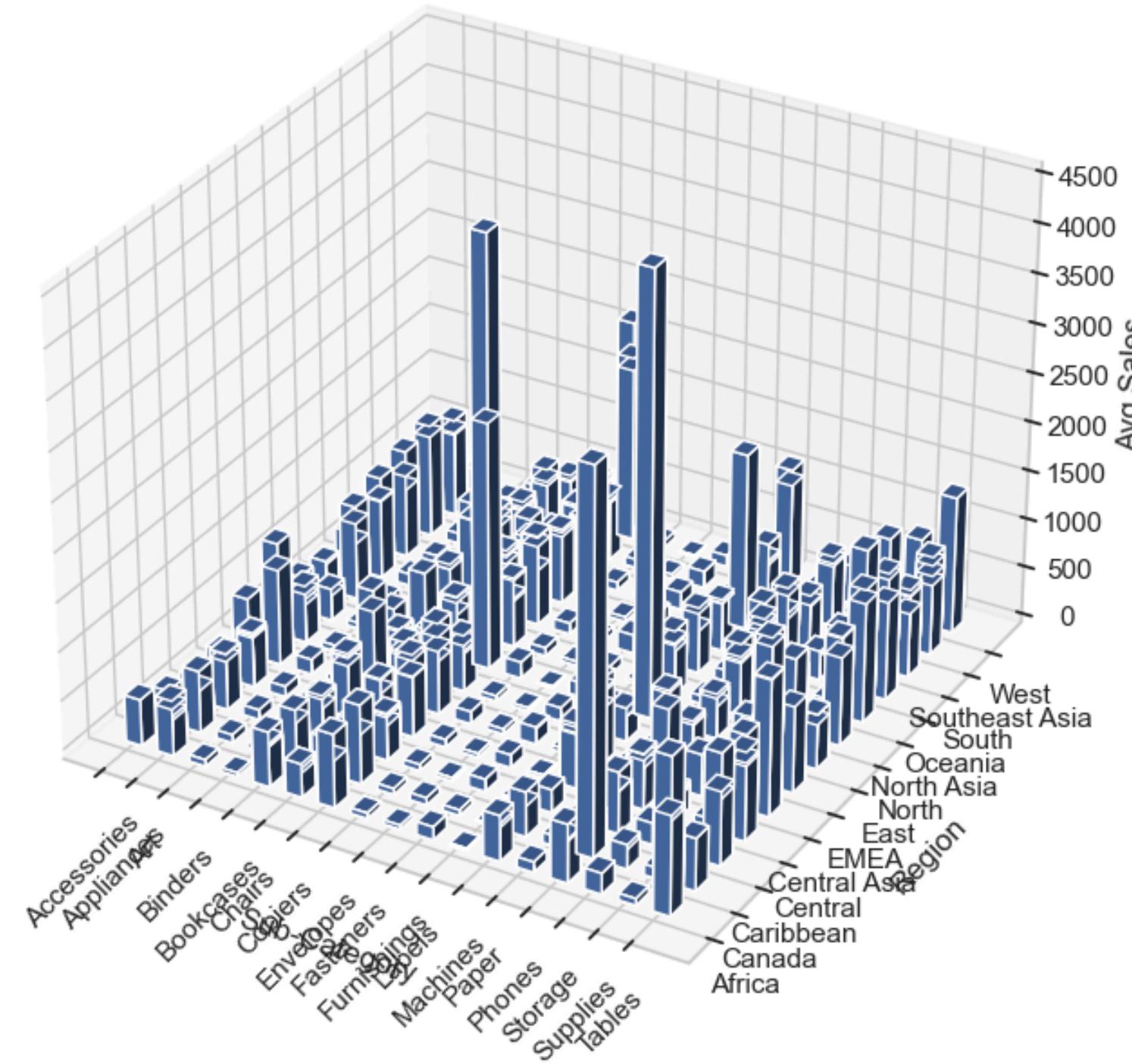
Two-Way ANOVA:

	sum_sq	df	F	PR(>F)
C(sub_category)	2.540239e+09	1.600000e+01	8.970765e+02	0.000000e+00
C(region)	1.130401e+08	1.200000e+01	5.322627e+01	3.865614e-128
C(sub_category):C(region)	4.577648e+08	1.920000e+02	1.347150e+01	0.000000e+00
Residual	9.038212e+09	5.106900e+04	NaN	NaN

P-value < α , tolak H_0 .
Terdapat perbedaan yang signifikan antara
sub_cateogery di **region** yang berbeda
terhadap kolom **sales**

TWO-WAY ANOVA

3D Bar Plot: Avg Sales by Sub-Category, Region



N-WAY ANOVA

```
# Model: N-Way ANOVA dengan 3 faktor + interaksi
model_nway = ols('sales ~ C(sub_category) * C(region) * C(order_priority)', data=df).fit()
anova_nway = sm.stats.anova_lm(model_nway, typ=2)

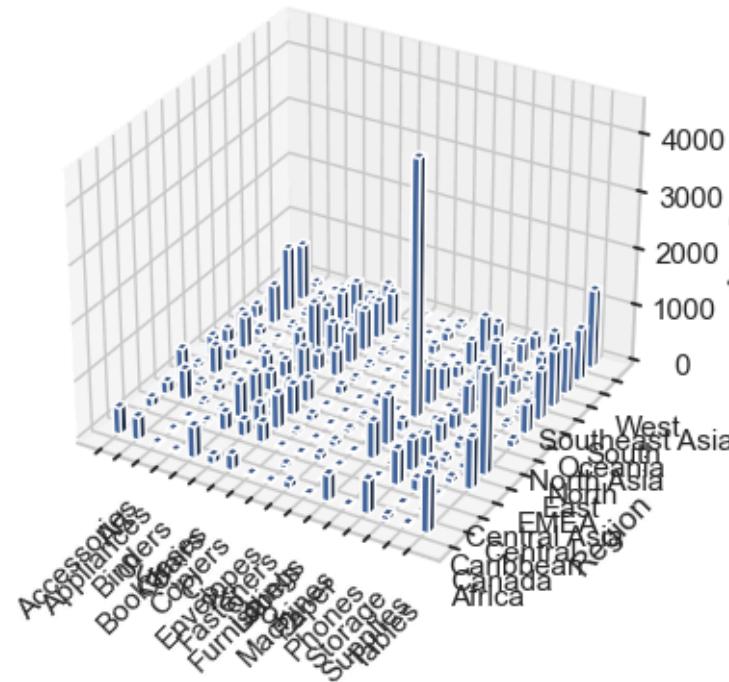
# Tampilkan hasil dengan format scientific
pd.options.display.float_format = '{:.6e}'.format
print("N-Way ANOVA (3 faktor):\n", anova_nway)
```

N-Way ANOVA (3 faktor):		
	sum_sq	df
C(sub_category)	2.430306e+09	1.600000e+01
C(region)	-1.639706e+07	1.200000e+01
C(order_priority)	2.168685e+05	3.000000e+00
C(sub_category):C(region)	4.331952e+08	1.920000e+02
C(sub_category):C(order_priority)	5.176176e+06	4.800000e+01
C(region):C(order_priority)	9.368308e+06	3.600000e+01
C(sub_category):C(region):C(order_priority)	1.588394e+08	5.760000e+02
Residual	8.883007e+09	5.042900e+04
	F	PR(>F)
C(sub_category)	8.623059e+02	0.000000e+00
C(region)	-7.757204e+00	1.000000e+00
C(order_priority)	4.103889e-01	7.455400e-01
C(sub_category):C(region)	1.280863e+01	0.000000e+00
C(sub_category):C(order_priority)	6.121927e-01	9.522743e-01
C(region):C(order_priority)	1.477335e+00	8.702869e-02
C(sub_category):C(region):C(order_priority)	1.565511e+00	4.182748e-16
Residual	NaN	NaN

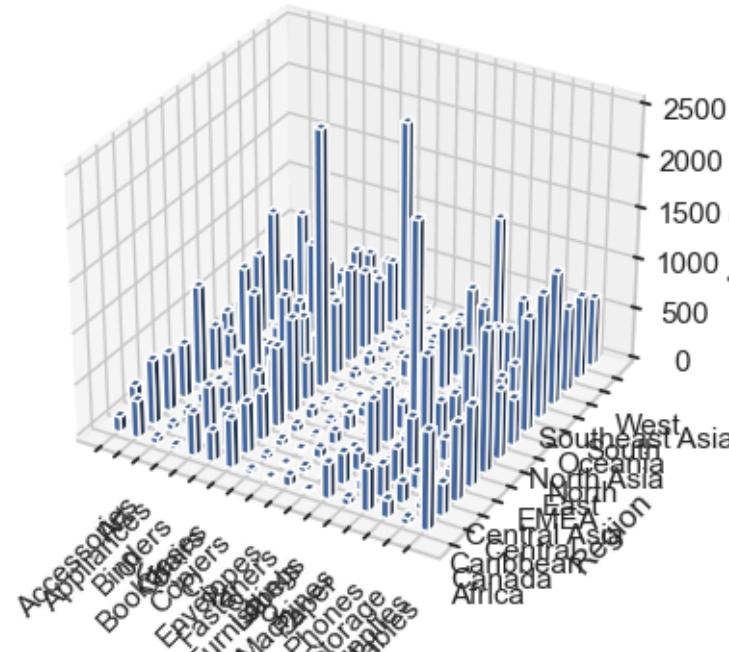
- P-value < a, tolak H₀. Terdapat perbedaan yg signifikan antara kolom **sub_category** dengan kolom **sales**.
- P-value < a, tolak H₀. Terdapat perbedaan yg signifikan antara kolom **sub_category** dengan kolom **sales** tergantung **region**-nya.
- P-value < a, tolak H₀. Terdapat perbedaan yg signifikan antara kolom **sub_category** dengan kolom **sales** tergantung **region**-nya dan **order_priority**nya
- P-value > a, menerima H₀. Terdapat perbedaan yg tidak signifikan antara kolom-kolom dengan sales

N-WAY ANOVA

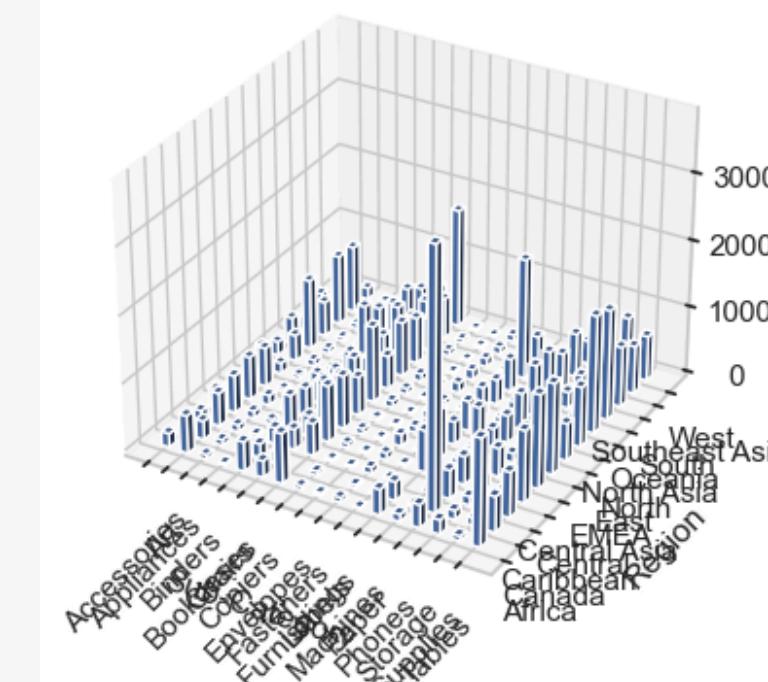
3D Bar Plot: Avg Sales (Order Priority: Low)



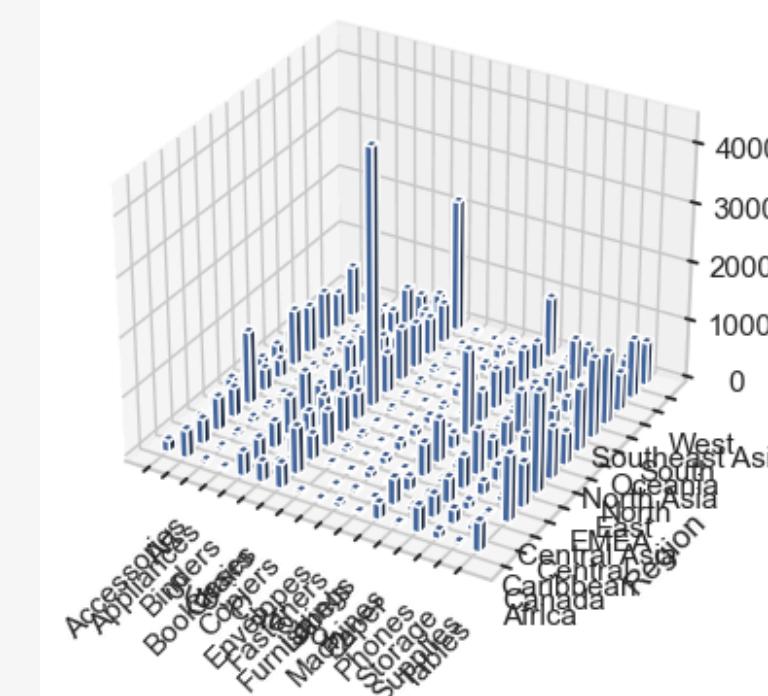
3D Bar Plot: Avg Sales (Order Priority: Medium)



3D Bar Plot: Avg Sales (Order Priority: Critical)

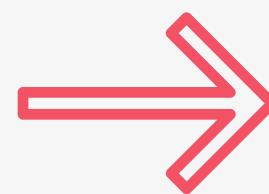


3D Bar Plot: Avg Sales (Order Priority: High)



Kesimpulan & Saran

ONE-WAY ANOVA



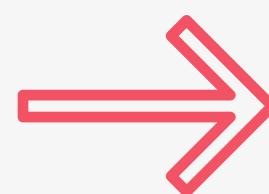
- Terdapat perbedaan yg tidak signifikan terhadap kolom **order_priority** terhadap kolom **sales**
- Terdapat perbedaan yg signifikan terhadap kolom **sub_category** terhadap kolom **sales**

TWO-WAY ANOVA

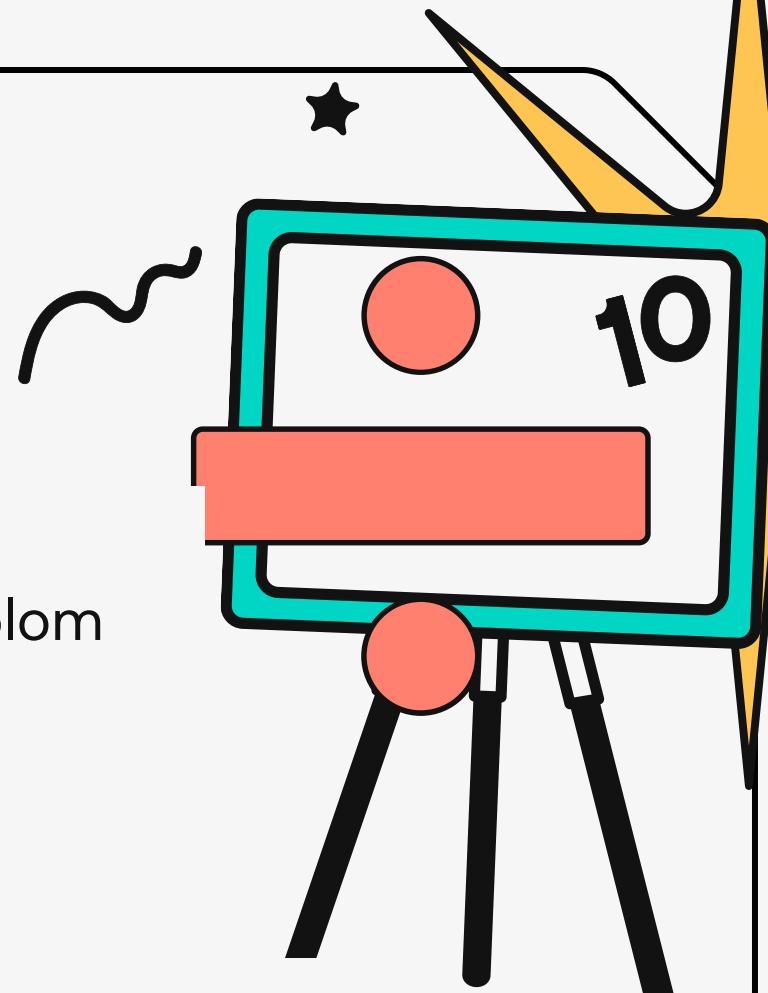


- Terdapat perbedaan yang signifikan antara **sub_categery** di **region** yang berbeda terhadap kolom **sales**

N-WAY ANOVA



- Terdapat perbedaan yg signifikan antara kolom **sub_category** dengan kolom **sales**.
- Terdapat perbedaan yg signifikan antara kolom **sub_category** dengan kolom **sales** tergantung **region**-nya.
- Terdapat perbedaan yg signifikan antara kolom **sub_category** dengan kolom **sales** tergantung **region**-nya dan **order_priority**nya
- Terdapat perbedaan yg tidak signifikan antara kolom-kolom dengan sales



Thank You

