

# Chi-Square Test

Alice Shizuka Hutagaol  
Angella Ananta Batubara  
Marcella Aurelia Yatijan



# Apa itu Chi-Square ( $\chi^2$ ) Test?

Chi-Squared Test adalah uji statistik non-parametrik yang digunakan untuk melihat apakah terdapat perbedaan signifikan antara data kategori (diskrit) yang diamati dan data yang diharapkan berdasarkan suatu hipotesis

Untuk mengecek:

- ? Apakah data kategori kita cocok dengan harapan
- ? Apakah dua hal kategori saling berhubungan
- ? Apakah distribusi kategori berbeda antar kelompok



# Kapan kita pakai Chi-Square Test?

Kalau kita punya data **categorical - categorical**, misalnya:

- Jenis produk (Furniture, Office Supplies, Technology)
- Tipe pelanggan (Consumer, Corporate, Home Office)
- Wilayah (US, APAC, EMEA)

Dan kamu mau cek:

- ⚖️ Apakah datanya terdistribusi seimbang?
- 🔗 Apakah 2 variabel kategori saling terkait?
- 🌐 Apakah distribusi berbeda antarwilayah?



# Rumus Utama Chi-Square

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = nilai aktual (observasi)

E = nilai harapan

Kalau O dan E jauh beda  $\rightarrow \chi^2$  besar  
 $\rightarrow$  kemungkinan nilai aktual tidak sesuai dengan harapan



# 3 Jenis Chi-Square Test

1

## Chi-Squared Goodness of Fit



Uji Chi-Squared Goodness of Fit digunakan untuk mengukur apakah distribusi frekuensi dari satu variabel kategori sesuai dengan distribusi yang diharapkan atau teoritis.

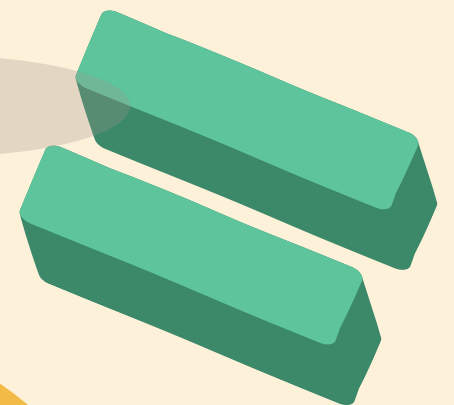


2

## Chi-Squared Test of Independence



Uji Chi-Squared Test of Independence digunakan untuk menentukan apakah terdapat hubungan atau keterkaitan yang signifikan antara dua variabel kategori dalam satu populasi.



3

## Chi-Squared Test of Homogeneity



Uji Chi-Squared Test of Homogeneity digunakan untuk menguji apakah distribusi satu variabel kategori adalah sama atau berbeda di beberapa populasi atau kelompok yang terpisah.



# Study Case

**Data :** SuperStore merupakan bisnis retail yang menjual berbagai produk dari tiga kategori utama:

**Furniture, Office Supplies, dan Technology**, yang terbagi lagi menjadi beberapa **sub-kategori**.

Dalam pengelolaan bisnis, penting untuk memastikan bahwa seluruh lini produk dapat berkontribusi secara seimbang terhadap penjualan dan profit.

Karena jika ada ketidakseimbangan, dapat memicu risiko bisnis jangka panjang seperti:

- Ketergantungan profit pada segmen tertentu
- Tidak optimalnya stok dan strategi pemasaran

# Study Case

Menggunakan metode statistik berbasis Chi-Squared, yaitu:

- ◆ **Chi-Squared Goodness of Fit**

- Untuk mengetahui apakah distribusi penjualan sudah merata atau justru didominasi oleh kategori/sub-kategori tertentu

- ◆ **Chi-Squared Test of Independence**

- Untuk mengevaluasi apakah ada hubungan signifikan antara kategori produk dan variabel lain seperti jenis pelanggan, prioritas pesanan, atau metode pengiriman

- ◆ **Chi-Squared Test of Homogeneity**

- Untuk membandingkan apakah pola pembelian antar kategori produk berbeda di berbagai wilayah atau pasar (market)

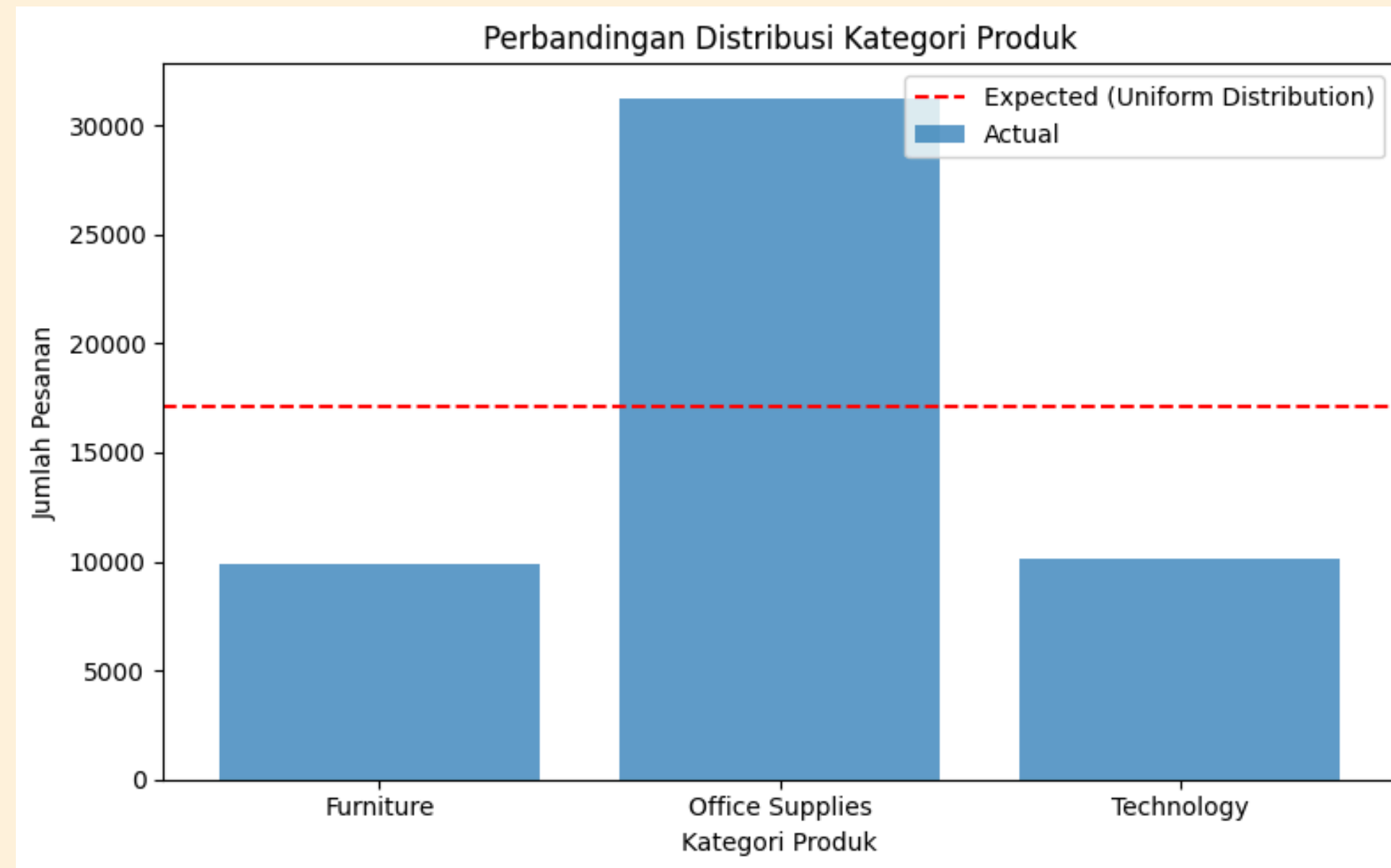
# Chi-Squared Goodness of Fit

- ✚ Hipotesis nol ( $H_0$ ) : Tidak ada perbedaan kelakuan antar kategori produk (distribusi penjualan merata)
- ✚ Hipotesis alternatif ( $H_1$ ) : Ada perbedaan kelakuan antar kategori produk (distribusi penjualan tidak merata)

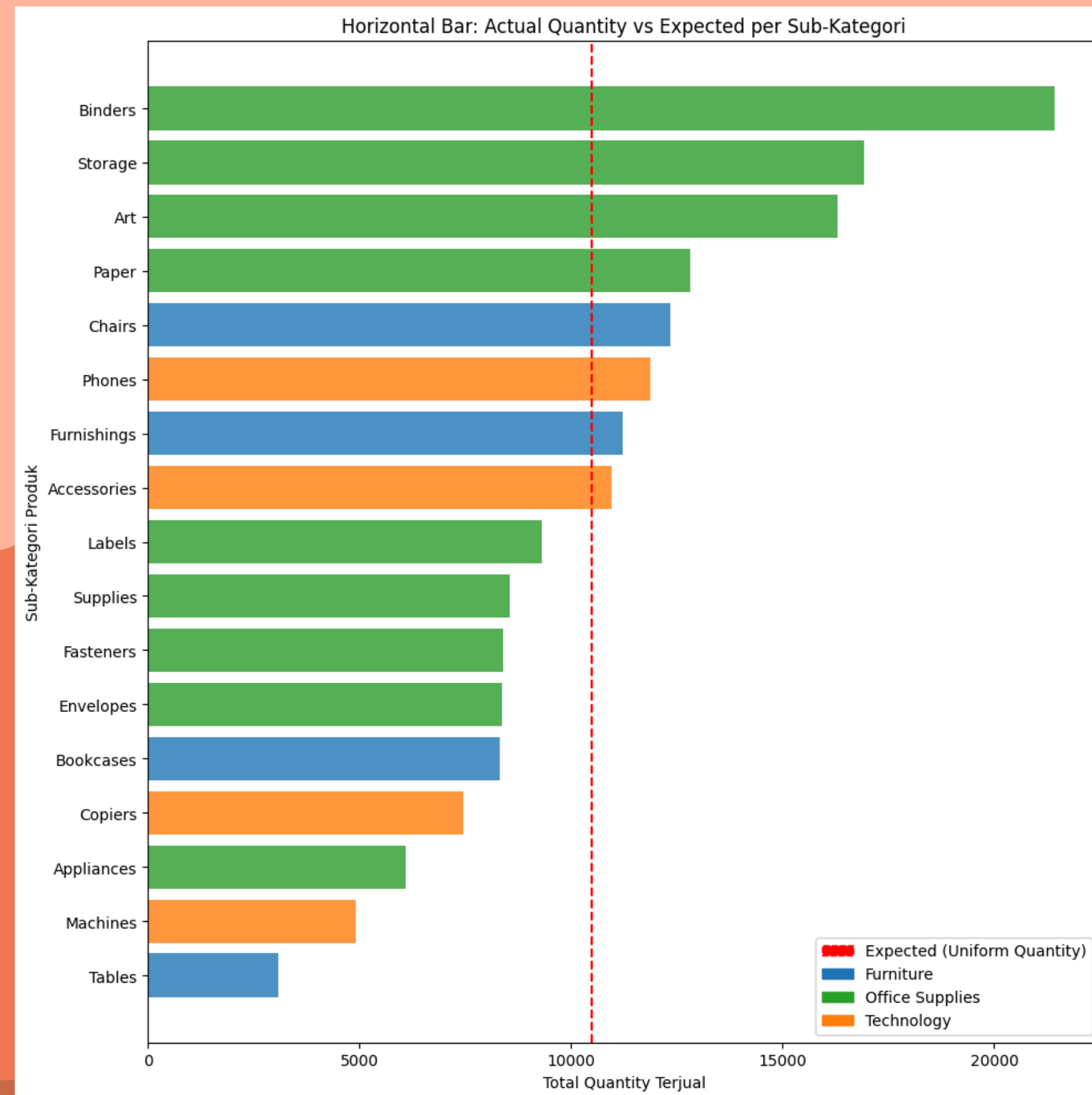
```
=== CHI-SQUARED GOODNESS OF FIT TEST ===  
Jumlah aktual kategori produk:  
category  
Furniture          9876  
Office Supplies    31273  
Technology          10141  
Name: count, dtype: int64  
  
Jumlah harapan (merata): [np.float64(17096.666666666668),  
  
Chi-Square = 17634.30  
p-value = 0.000000000000000000000000
```



# Chi-Squared Goodness of Fit



# Chi-Squared Goodness of Fit



# Chi-Squared Goodness of Fit



```
=== CHI-SQUARED GOODNESS OF FIT TEST - QUANTITY per SUB-CATEGORY ===
```

```
Quantity aktual per sub-kategori:
```

```
sub_category
```

```
Accessories      10946
```

```
Appliances       6078
```

```
Art              16301
```

```
Binders          21429
```

```
Bookcases        8310
```

```
Chairs           12336
```

```
Copiers          7454
```

```
Envelopes        8380
```

```
Fasteners        8390
```

```
Furnishings      11225
```

```
Labels           9322
```

```
Machines         4906
```

```
Paper            12822
```

```
Phones           11870
```

```
Storage          16917
```

```
Supplies         8543
```

```
Tables           3083
```

```
Name: quantity, dtype: int64
```

```
Quantity harapan (merata): [np.float64(10488.941176470587), np.float64(10488.]
```

```
Chi-Square = 32389.50
```

```
p-value = 0.000000
```



# Chi-Squared Test of Independence

📌 Hipotesis untuk Chi-Squared Test of Independence

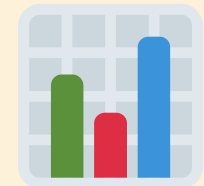
◆  $H_0$  (Hipotesis Nol):

Tidak ada hubungan (independen) antara kedua variabel kategorikal.  
Artinya, distribusi satu variabel tidak tergantung pada variabel lainnya.

◆  $H_1$  (Hipotesis Alternatif):

Ada hubungan (dependen) antara kedua variabel kategorikal.  
Artinya, distribusi satu variabel berubah tergantung pada kategori variabel lainnya.





## Tabel Ringkasan Uji Chi-Squared (category sebagai variabel utama)

| No. | Variabel Lawan | Chi <sup>2</sup> Hitung | Chi <sup>2</sup> Tabel ( $\alpha=0.05$ ) | p-value | Kesimpulan                 |
|-----|----------------|-------------------------|--|---------|----------------------------|
| 1   | segment        | 27.235                  | 94.877                                   | 0.6051  | ☑️ Gagal Tolak $H_0$       |
| 2   | order_priority | 67.773                  | 125.916                                  | 0.3419  | ☑️ Gagal Tolak $H_0$       |
| 3   | region         | 3.424.664               | 364.150                                  | 0.0000  | ❗ Tolak $H_0$ (Signifikan) |
| 4   | ship_mode      | 25.977                  | 125.916                                  | 0.8574  | ☑️ Gagal Tolak $H_0$       |
| 5   | country        | 8.529.431               | 3.328.538                                | 0.0000  | ❗ Tolak $H_0$ (Signifikan) |
| 6   | ship_date      | 29.972.012              | 30.529.553                               | 0.1757  | ☑️ Gagal Tolak $H_0$       |
| 7   | market         | 5.592.533               | 210.261                                  | 0.0000  | ❗ Tolak $H_0$ (Signifikan) |
| 8   | sub_category   | 1.025.800.000           | 461.943                                  | 0.0000  | ❗ Tolak $H_0$ (Signifikan) |

Terdapat 4 variabel yang secara signifikan mempengaruhi variabel category,  
yaitu region, country, market, dan sub category



# Tabel Ringkasan Uji Chi-Squared (sub\_category sebagai variabel utama)

| No. | Variabel Lawan | Chi <sup>2</sup> Hitung | Chi <sup>2</sup> Tabel ( $\alpha=0.05$ ) | p-value | Kesimpulan   |
|-----|----------------|-------------------------|--|---------|--|
| 1   | segment        | 47,5982                 | 46,194                                   | 0.0374  | ! Tolak H <sub>0</sub> (Signifikan)                            |
| 2   | order_priority | 49,90                   | 65,171                                   | 0.3976  | <input checked="" type="checkbox"/> Gagal Tolak H <sub>0</sub> |
| 3   | region         | 3.118,5654              | 225,329                                  | 0.0000  | ! Tolak H <sub>0</sub> (Signifikan)                            |
| 4   | ship_mode      | 65,1299                 | 65,171                                   | 0.0504  | <input checked="" type="checkbox"/> Gagal Tolak H <sub>0</sub> |
| 5   | country        | 7.366,8151              | 2.449,555                                | 0.0000  | ! Tolak H <sub>0</sub> (Signifikan)                            |
| 6   | ship_date      | 24084,7408              | 23765,0305                               | 0.0010  | ! Tolak H <sub>0</sub> (Signifikan)                            |
| 7   | market         | 5.110,1517              | 119,871                                  | 0.0000  | ! Tolak H <sub>0</sub> (Signifikan)                            |

Terdapat 5 variabel yang secara signifikan mempengaruhi variabel category, yaitu segment, region, country, ship date, dan market

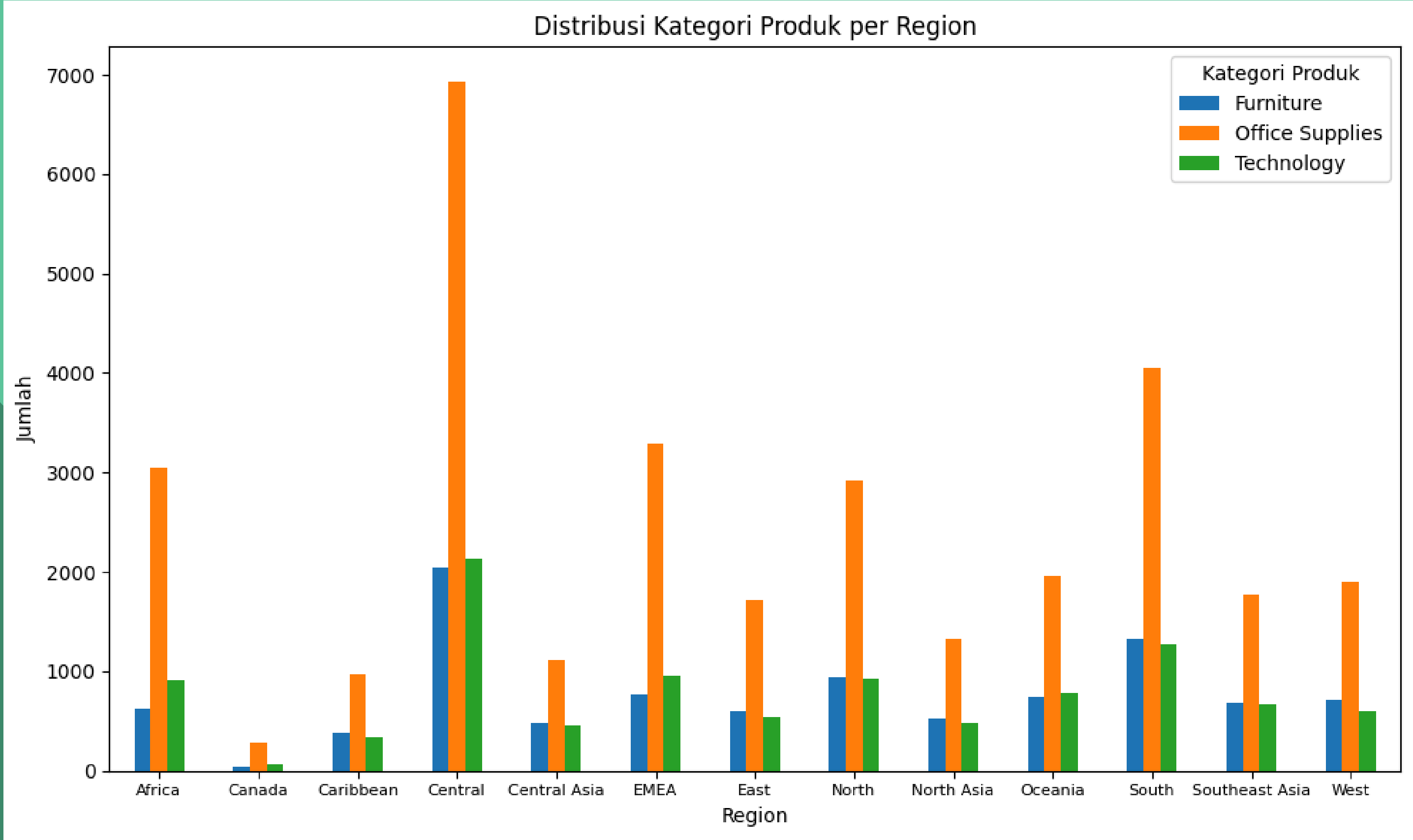
# Chi-Squared Test of Homogeneity

Hipotesis nol ( $H_0$ ) : Distribusi kategori produk sama di tiap region (homogen)

Hipotesis alternatif ( $H_1$ ) : Distribusi kategori produk tidak sama di tiap region (heterogen)



# Category & Region



Chi-Square hitung = 342.466  
(> Chi-Square tabel 36.415)

p-value = 0 (< 0.05)

Kesimpulan: Distribusi produk di setiap region heterogen ( $H_0$  ditolak)



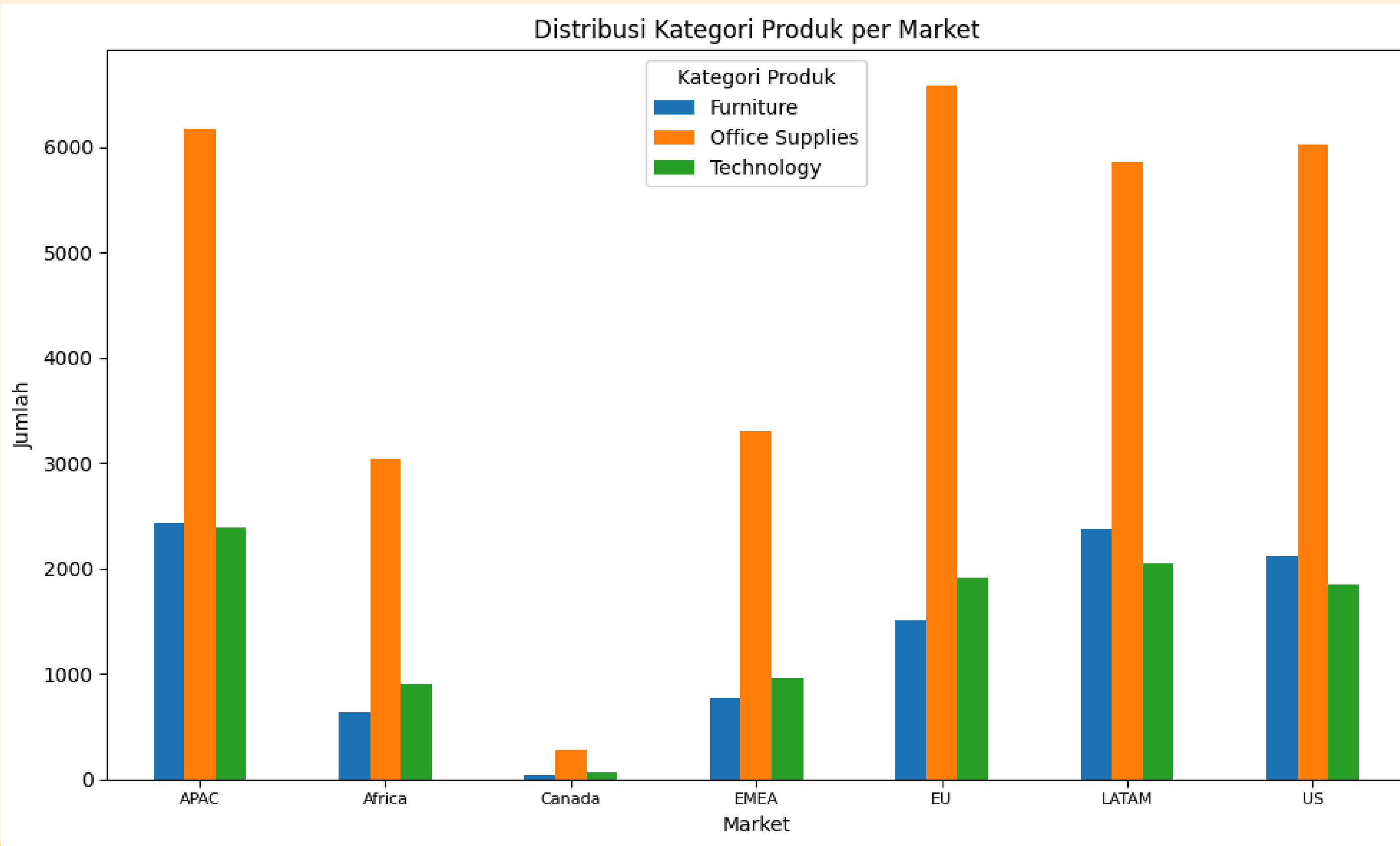
# Chi-Squared Test of Homogeneity

Hipotesis nol ( $H_0$ ) : Distribusi kategori produk sama di tiap market (homogen)

Hipotesis alternatif ( $H_1$ ) : Distribusi kategori produk tidak sama di tiap market (heterogen)



# Category & Market



Chi-Square hitung = 559.253  
(> Chi-Square tabel 21.0261)

p-value = 0 (< 0.05)

Kesimpulan: Distribusi  
kategori di setiap market  
heterogen ( $H_0$  ditolak)

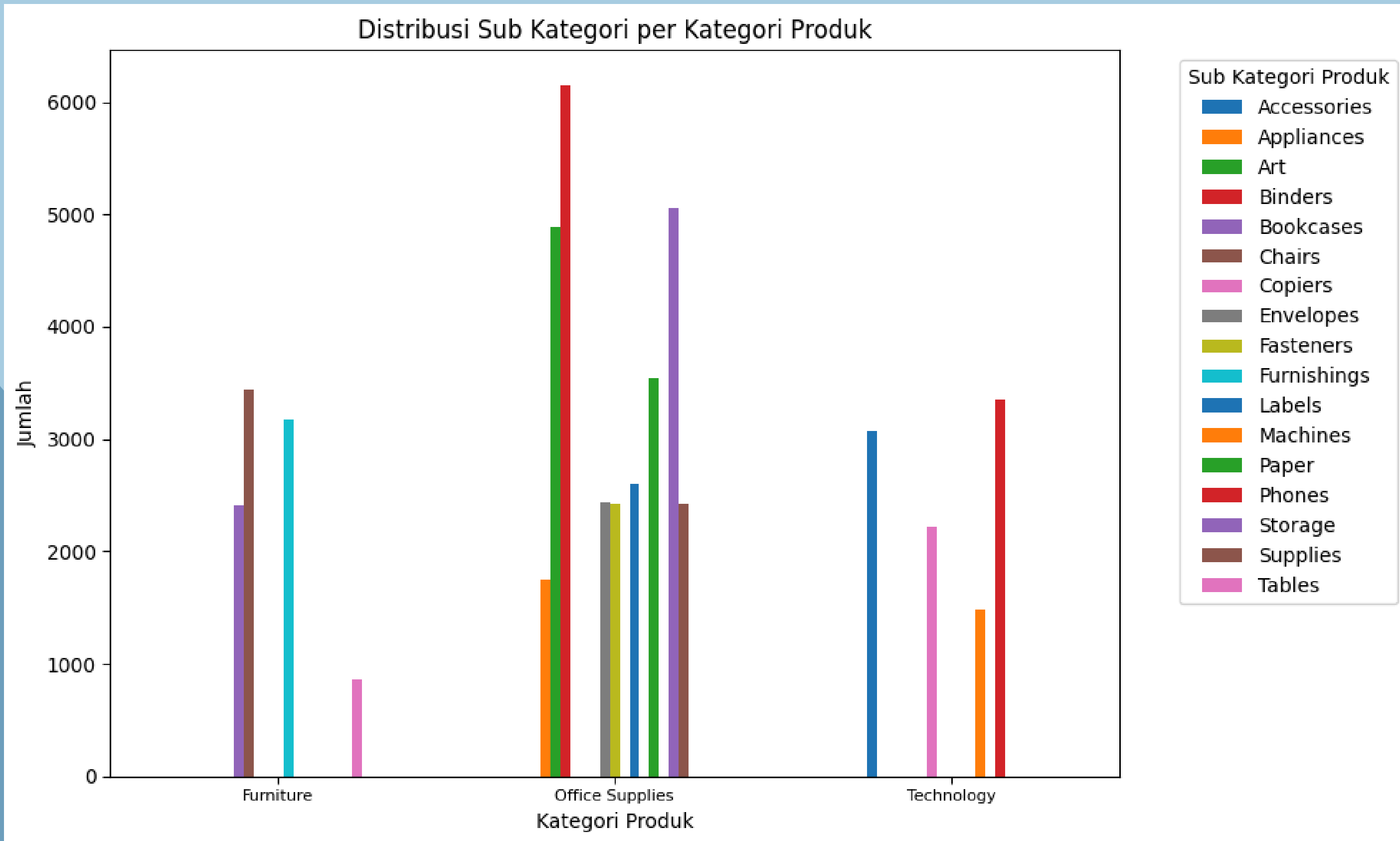
# Chi-Squared Test of Homogeneity

Hipotesis nol ( $H_0$ ) : Distribusi sub kategori produk sama di tiap kategori(homogen)

Hipotesis alternatif ( $H_1$ ) : Distribusi sub kategori produk tidak sama di tiap kategori (heterogen)



# Sub Category & Category



Chi-Square hitung = 102580  
(> Chi-Square tabel 46.194)

p-value = 0 (< 0.05)

Kesimpulan: Distribusi sub kategori di setiap produk heterogen ( $H_0$  ditolak)

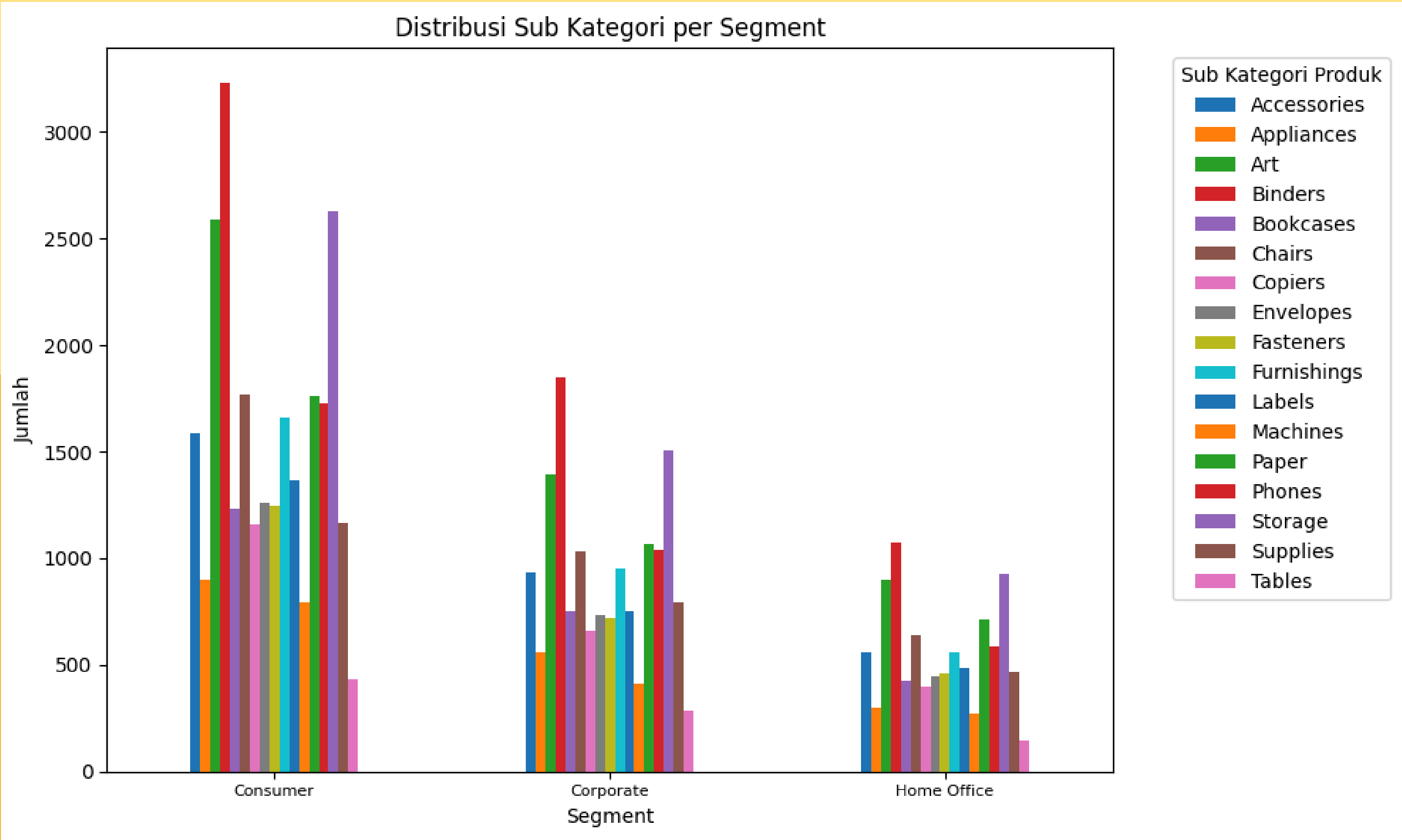
# Chi-Squared Test of Homogeneity

Hipotesis nol ( $H_0$ ) : Distribusi sub kategori produk sama di tiap segment (homogen)

Hipotesis alternatif ( $H_1$ ) : Distribusi sub kategori produk tidak sama di tiap segment (heterogen)



# Sub Category & Segment



Chi-Square hitung = 47.5982  
(> Chi-Square tabel 46.194)

p-value = 0.0374 (< 0.05)

Kesimpulan: Distribusi sub kategori di setiap segment heterogen ( $H_0$  ditolak)

# Kesimpulan dan Saran

## KESIMPULAN

- Distribusi penjualan kategori dan sub-kategori produk tidak merata, Office Supplies mendominasi, sedangkan kategori lain seperti Furniture dan beberapa sub-kategori tertinggal jauh.
  - Terdapat hubungan signifikan antara category dengan variabel seperti region, market, sub\_category, dan country.
  - Distribusi kategori produk berbeda secara signifikan antar region dan market. Contoh: Office Supplies sangat dominan di Central & EU, tapi seimbang di Asia.
- 📌 Insight: Pola pembelian kategori produk berbeda tergantung faktor demografis & operasional → Strategi produk harus disesuaikan per wilayah, bukan disamaratakan.

## SARAN

Dapat dilakukan analisis lebih lanjut terkait pengaruh ship date terhadap category untuk penjadwalan marketing

**THANK  
you**

