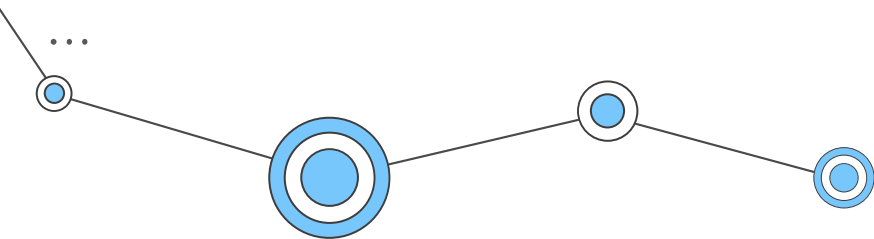
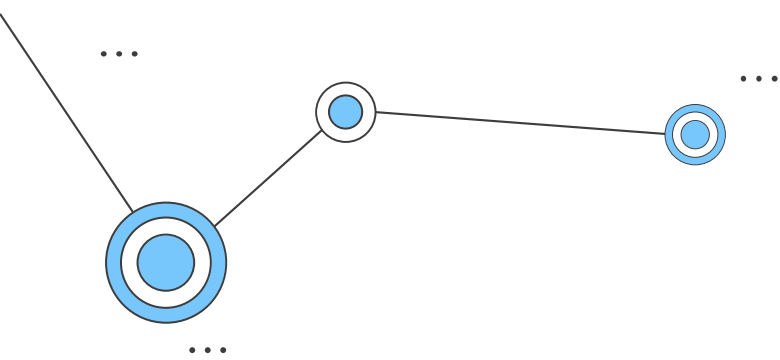


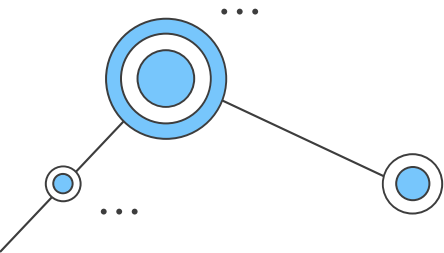
## Analisis Segmentasi Pelanggan Menggunakan Algoritma Clustering






## Kelompok 2 :

- **Hafiizh Taufiqul Hakim  
(2012500720)**
- **Pandu Aryo Utomo  
(2012500498)**






# Tujuan

- Menghasilkan sebuah segmentasi pelanggan yang dapat membantu perusahaan dalam menemukan karakteristik dari setiap pelanggan agar perusahaan mampu memprioritaskan sumber daya dan tenaga untuk pelanggan tertentu.
  - Menganalisis kluster yang sudah terbentuk yang akan digunakan menjadi bahan untuk menarik kesimpulan dari penelitian ini sehingga bisa memberikan solusi kepada perusahaan.
- 




# Permasalahan

Pelanggan memiliki karakteristik yang berbeda sehingga menyulitkan bisnis dalam melakukan pemasaran produk.





# Metode

- Data Cleaning
  - Data Clustering dengan menggunakan K-means
    - \*K-means merupakan metode pengelompokan yang dapat dilakukan untuk melakukan segmentasi.
  - Data Visualisasi
- 

# Library

Library yang kami gunakan untuk melakukan proses segmentasi pelanggan sebagai berikut :

- **Import Numpy** : Berfungsi yang siap pakai untuk memudahkan kita melakukan perhitungan saintifik seperti matriks, aljabar, statistik, dan sebagainya.
- **Import Pandas** : sebuah library open source yang ada pada bahasa pemrograman Python yang sering digunakan untuk memproses data, mulai pembersihan data, manipulasi data, hingga melakukan analisis data.
- **Import Seaborn** : library untuk membuat grafik dan statistik dengan menggunakan Python.
- **Import Matplotlib** : library Python yang fokus pada visualisasi data seperti plot grafik.
- **Import K-Means** : Melakukan perhitungan cluster dengan menggunakan metode k-means

# Data Set

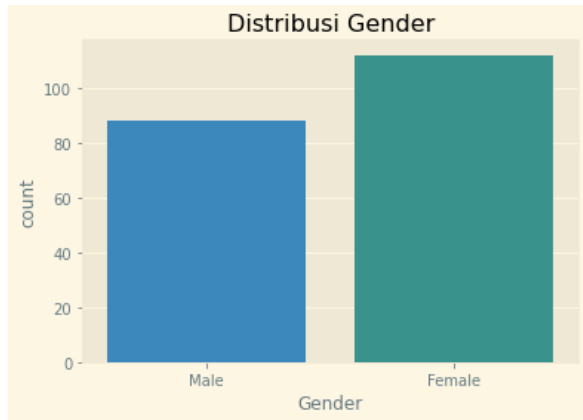
## Mall\_Customer.csv

Sumber : <https://www.kaggle.com/code/agusmuhdiaji/customers-segmentation-indonesian-language>

| 1  | CustomerID | Gender | Age | Annual Income (k\$) | Spending Score (1-100) |
|----|------------|--------|-----|---------------------|------------------------|
| 2  | 1          | Male   | 19  | 15                  | 39                     |
| 3  | 2          | Male   | 21  | 15                  | 81                     |
| 4  | 3          | Female | 20  | 16                  | 6                      |
| 5  | 4          | Female | 23  | 16                  | 77                     |
| 6  | 5          | Female | 31  | 17                  | 40                     |
| 7  | 6          | Female | 22  | 17                  | 76                     |
| 8  | 7          | Female | 35  | 18                  | 6                      |
| 9  | 8          | Female | 23  | 18                  | 94                     |
| 10 | 9          | Male   | 64  | 19                  | 3                      |
| 11 | 10         | Female | 30  | 19                  | 72                     |

# Hasil Pembahasan

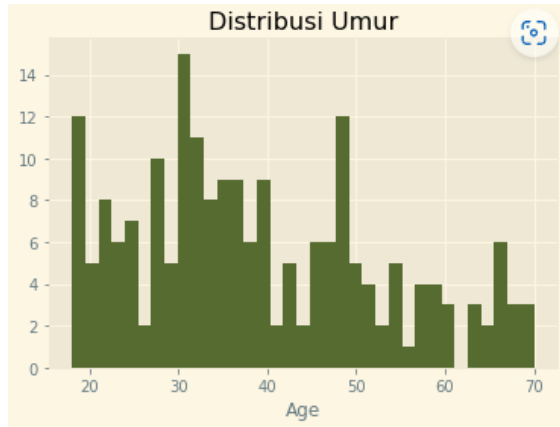
```
sns.countplot(x='Gender', data=customers)
plt.style.use('Solarize_Light2') #memilih style plot
plt.title('Distribusi Gender')
```



1. Visualisasi data dengan melihat distribusi gender, dapat terlihat bahwa gender female atau perempuan lebih mendominasi dalam pengunjung mall.

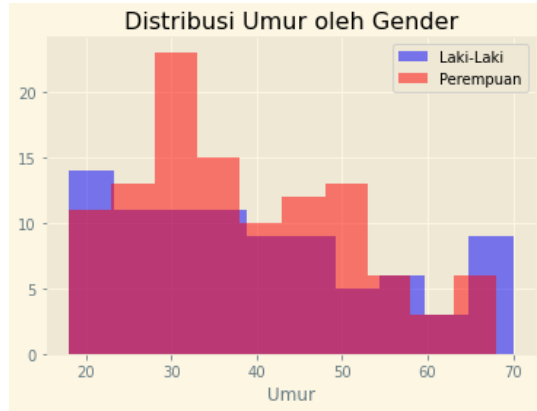


```
customers.hist('Age', bins=35, color = 'darkolivegreen')  
plt.title('Distribusi Umur')  
plt.xlabel('Age')
```



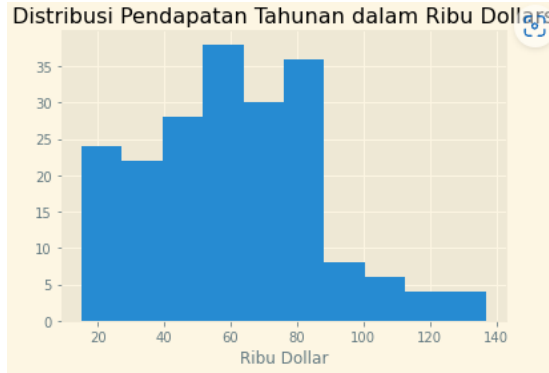
2. Visualisasi data untuk distribusi umur dengan bins 35 dikarenakan mengambil nilai tengah 0-70. berdasarkan grafik dapat kita lihat bahwa pengunjung mall terbanyak berumur diantara 30 – 33.

```
plt.hist('Age', data=customers[customers['Gender'] == 'Male'], alpha=0.5, label='Laki-Laki', color = 'blue')
plt.hist('Age', data=customers[customers['Gender'] == 'Female'], alpha=0.5, label='Perempuan', color = 'red')
plt.title('Distribusi Umur oleh Gender')
plt.xlabel('Umur')
plt.legend()
```



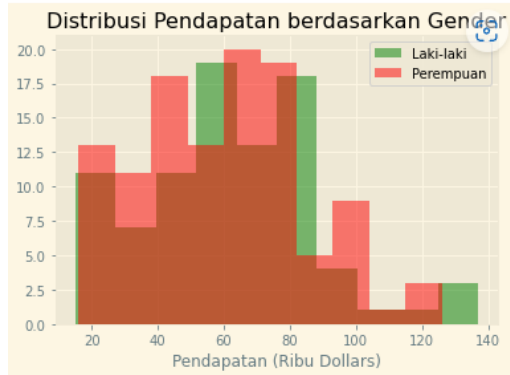
3. Visualisasi data dengan histogram melihat distribusi umur pengunjung mall berdasarkan gendernya, dapat dilihat bahwa perempuan dengan umur diatas 30 mendominasi sebagai pengunjung terbanyak ,sedangkan pengunjung yang sedikit adalah laki-laki dan perempuan berumur diantara 59-65 tahun

```
customers.hist('Annual Income (k$)')  
plt.title('Distribusi Pendapatan Tahunan dalam Ribu Dollars')  
plt.xlabel('Ribu Dollar')
```



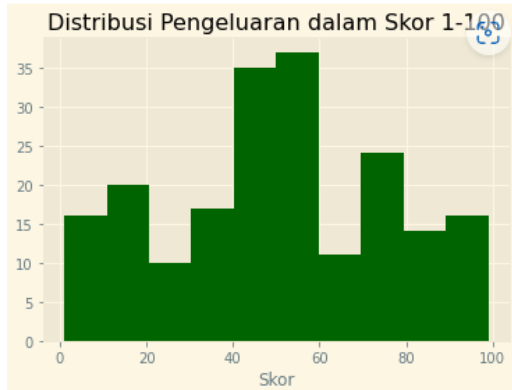
4. Visualisasi data untuk distribusi pendapatan tahunan distribusi pendapatan terbesar berada di kisaran 60 hingga 80 ribu dollar.

```
plt.hist('Annual Income (k$)', data=customers[customers['Gender'] == 'Male'], alpha=0.5, label='Laki-laki', color = 'green')
plt.hist('Annual Income (k$)', data=customers[customers['Gender'] == 'Female'], alpha=0.5, label='Perempuan', color = 'red')
plt.title('Distribusi Pendapatan berdasarkan Gender')
plt.xlabel('Pendapatan (Ribu Dollars)')
plt.legend()
```



5. apakah gender mempengaruhi pendapatan? maka dari itu kita mencari distribusinya berdasarkan tabel di samping dapat terlihat bahwa pendapatan lebih banyak oleh laki-laki dibandingkan perempuan Pendapatan perempuan terbesar berada di kisaran 60-80 ribu dollar

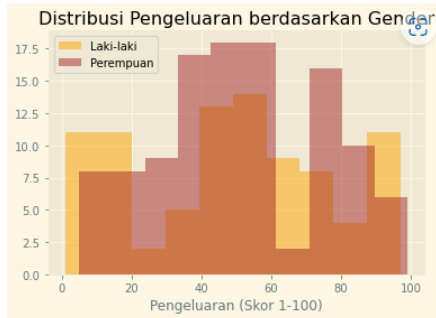
```
customers.hist('Spending Score (1-100)', color = 'darkgreen')  
plt.title('Distribusi Pengeluaran dalam Skor 1-100')  
plt.xlabel('Skor')
```



6. visualisasi data untuk distribusi pengeluaran dalam skor 1 sampai 100

distribusi pendapatan terbesar berada di kisaran skor 40-60

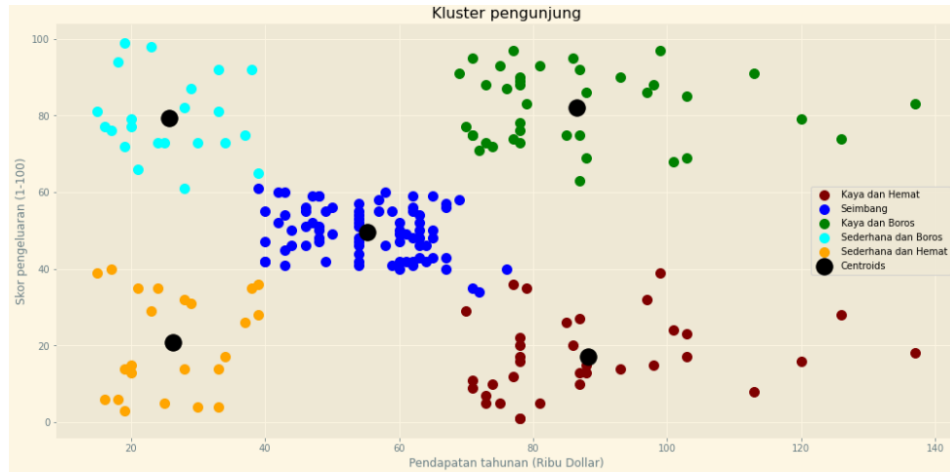
```
plt.hist('Spending Score (1-100)', data=customers[customers['Gender'] == 'Male'], alpha=0.5, label='Laki-laki', color = 'orange')
plt.hist('Spending Score (1-100)', data=customers[customers['Gender'] == 'Female'], alpha=0.5, label='Perempuan', color = 'brown')
plt.title('Distribusi Pengeluaran berdasarkan Gender')
plt.xlabel('Pengeluaran (Skor 1-100)')
plt.legend()
```



7. Apakah gender mempengaruhi pengeluaran? maka dari itu kita mencari distribusinya

berdasarkan tabel di bawah dapat terlihat bahwa pengeluaran lebih banyak oleh perempuan dibandingkan laki-laki. Pendapatan perempuan terbesar berada di kisaran skor 60-80

```
plt.figure(1, figsize = (17, 8))
plt.scatter(Seg[y_kmeans == 0, 0], Seg[y_kmeans == 0, 1], s = 100, c = 'maroon', label = 'Kaya dan Hemat')
plt.scatter(Seg[y_kmeans == 1, 0], Seg[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Seimbang')
plt.scatter(Seg[y_kmeans == 2, 0], Seg[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Kaya dan Boros')
plt.scatter(Seg[y_kmeans == 3, 0], Seg[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Sederhana dan Boros')
plt.scatter(Seg[y_kmeans == 4, 0], Seg[y_kmeans == 4, 1], s = 100, c = 'orange', label = 'Sederhana dan Hemat')
plt.scatter(kmeans.cluster_centers[:, 0], kmeans.cluster_centers[:, 1], s = 300, c = 'black', label = 'Centroids')
plt.title('Kluster pengunjung')
plt.xlabel('Pendapatan tahunan (Ribuan Dollar)')
plt.ylabel('Skor pengeluaran (1-100)')
plt.legend()
plt.show()
```

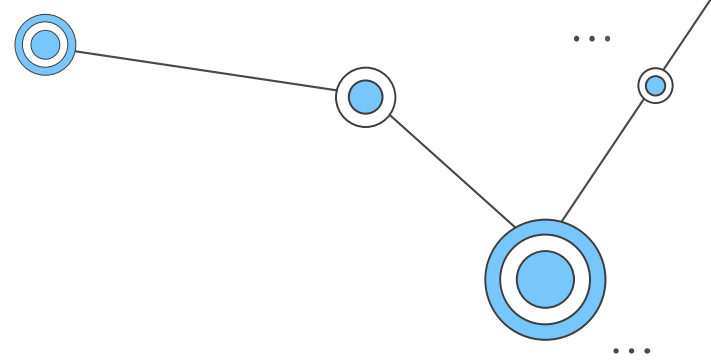


Visualisasi hasil clusters

# Kesimpulan

- Segmentasi pengunjung dari mall yang akan dijadikan target marketing yaitu kluster 3 adalah orang yang kaya dengan pendapatan tinggi dan pengeluaran yang tinggi atau boros dan juga kluster 4 adalah orang yang sederhana dengan pendapatan yang rendah dan pengeluaran yang tinggi atau boros.





**TERIMA KASIH....**

