

Assignment: Case Study Customer Segmentation with Power BI

Link Drive: [Evan_Assignment_Day26](#)

Link PBIX: [Evan_Assignment_Day26.pbix](#)

Objectives

1. Student mampu melakukan data understanding
2. Student mampu melakukan preprocessing data
3. Student mampu melakukan visualisasi data
4. Student mampu mengimplementasikan Customer segmentation
5. Student mampu melakukan analisis hasil untuk customer segmentation

Data Understanding

Data yang diberikan untuk tugas ini adalah "superstore_dataset.xlsx". Data ini memiliki beberapa kolom. Tabel *description* adalah deskripsi dari kolom untuk tabel *superstore*. Perlu diketahui bahwa kolom-kolom tersebut memberikan pengertian akan data yang akan dikelola. Secara umum, dapat dikatakan bahwa kita akan mengelola data tentang order, shipping, customer, market segment (kategori pelanggan: konsumen, distributor, etc), tempat melakukan order dan detailnya, produk, kategori dan subkategori dari produk, sales, jumlah unit terjual, diskon, keuntungan.

Data Preprocessing

Kita kelola data yang sudah kita peroleh dengan metode RFM (recency, frequency dan monetary). Berikut adalah definisi tabel RFM:

```
rfm_table =  
SUMMARIZE(  
    superstore,  
    superstore[customer_name],  
    "last_transaction", MAX(superstore[order_date]),  
    "frequency", DISTINCTCOUNT(superstore[order_id]),  
    "monetary", SUM(superstore[sales])  
)
```

Last transaction digunakan untuk mendapatkan kolom *recency* nantinya. Setelah itu, set date today sebagai 1 Januari 2018 karena data terakhir adalah Desember 2017.

```
recency = DATEDIFF([last_transaction], DATE(2018, 1, 1), DAY)
```

Dari sini kita definisikan *r_score*, *f_score*, *m_score* dan *rfm_score*:

1. RScore:

```
r_score = SWITCH(
    TRUE(),
    [recency] <= PERCENTILEX.INC(rfm_table, [recency], 0.2), "5",
    [recency] <= PERCENTILEX.INC(rfm_table, [recency], 0.4), "4",
    [recency] <= PERCENTILEX.INC(rfm_table, [recency], 0.6), "3",
    [recency] <= PERCENTILEX.INC(rfm_table, [recency], 0.8), "2",
    "1"
)
```

2. FScore:

```
f_score = SWITCH(
    TRUE(),
    [frequency] >= PERCENTILEX.INC(rfm_table, [frequency], 0.8), "5",
    [frequency] >= PERCENTILEX.INC(rfm_table, [frequency], 0.6), "4",
    [frequency] >= PERCENTILEX.INC(rfm_table, [frequency], 0.4), "3",
    [frequency] >= PERCENTILEX.INC(rfm_table, [frequency], 0.2), "2",
    "1"
)
```

3. MScore:

```
m_score = SWITCH(
    TRUE(),
    [monetary] >= PERCENTILEX.INC(rfm_table, [monetary], 0.8), "5",
    [monetary] >= PERCENTILEX.INC(rfm_table, [monetary], 0.6), "4",
    [monetary] >= PERCENTILEX.INC(rfm_table, [monetary], 0.4), "3",
    [monetary] >= PERCENTILEX.INC(rfm_table, [monetary], 0.2), "2",
    "1"
)
```

4. RFM Score: rfm_score = [r_score] & [f_score] & [m_score]

Terakhir, kita mengelompokkan customer ke dalam kategori segmen berikut: champion, loyal, potential, at risk, dan uncategorized. Tidak seperti segmen dalam superstore yaitu corporate, home office, consumer, etc.

```
customer_segment = SWITCH(
    TRUE(),
    [rfm_score] >= "511" && [rfm_score] <= "555", "Champions",
    [rfm_score] >= "451" && [rfm_score] <= "510", "Loyal",
    [rfm_score] >= "351" && [rfm_score] <= "450", "Potential",
    [rfm_score] >= "151" && [rfm_score] <= "350", "At Risk",
    "Uncategorized"
)
```

Selain itu, kita juga dapat melakukan encoding dengan mengidentifikasi segmen tersebut ke dalam grup yang direpresentasikan dengan angka (bilangan bulat 1–5).

segment_group =

```
SWITCH(  
    TRUE(),  
    [customer_segment] = "Champions", 1,  
    [customer_segment] = "Loyal", 2,  
    [customer_segment] = "Potential", 3,  
    [customer_segment] = "At Risk", 4,  
    5 // Uncategorized  
)
```

Kita juga melakukan pengelompokan periode resensi seperti berikut:

recency_bucket =

```
SWITCH(  
    TRUE(),  
    [recency] <= 30, "0-30 days",  
    [recency] <= 60, "31-60 days",  
    [recency] <= 90, "61-90 days",  
    "90+ days"  
)
```

Selain kolom-kolom RFM table di atas, kita juga menambahkan *measures* pada superstore dataset:

1. Orders per customer

```
orders_per_customer =  
DIVIDE(  
    COUNTROWS(superstore),  
    DISTINCTCOUNT(superstore[customer_id])  
)
```

2. Profit ratio:

```
profit_ratio =  
DIVIDE(  
    SUM(superstore[profit]),  
    SUM(superstore[sales]),  
    0  
)
```

3. Total customers: total_customers = `distinctcount(rfm_table[customer_name])`

4. Total orders: total_orders = `DISTINCTCOUNT(superstore[order_id])`

5. Total profit: total_profit = `sum(superstore[profit])`

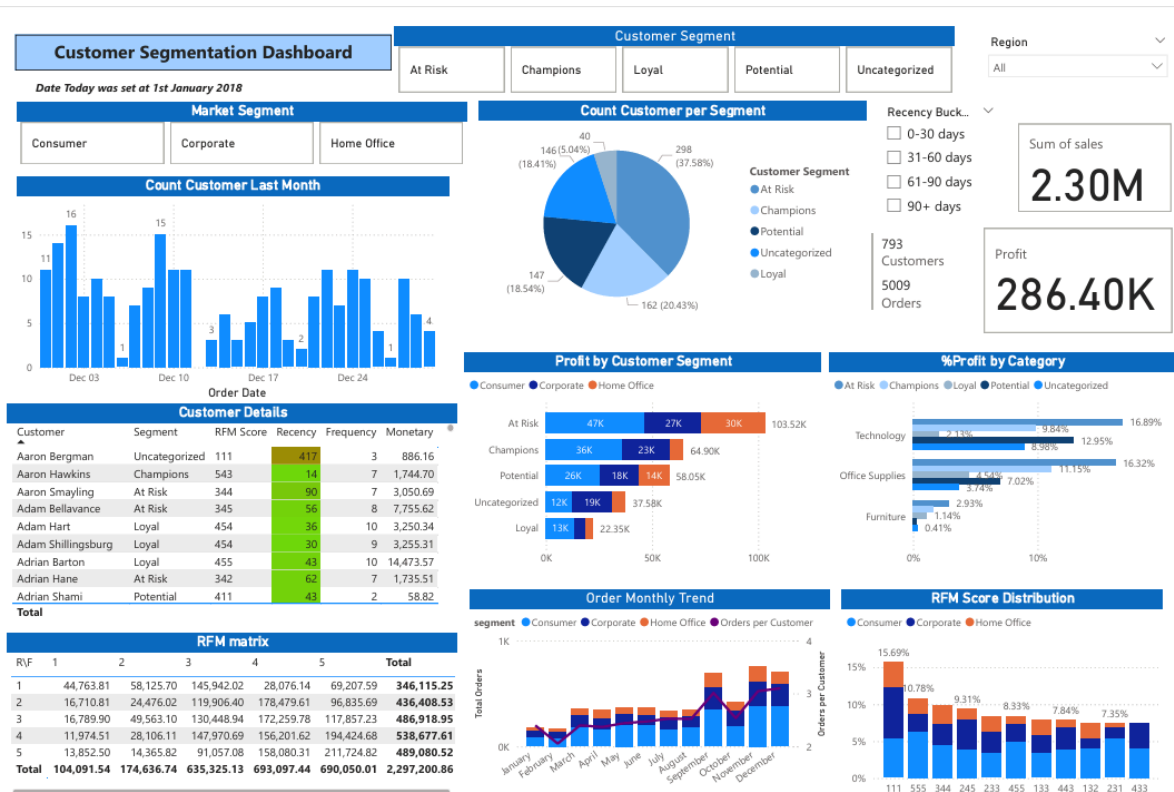
6. Total revenue (sales): total_revenue = `sum(superstore[sales])`

Visualisasi Data, Insight dan Analisis

Link untuk visualisasi dashboard:

[PDF Evan_AssDay26_Dashboard_Screenshot.pdf](#)

[PDF Evan_AssDay26_Dashboard_Screenshot_2.pdf](#)



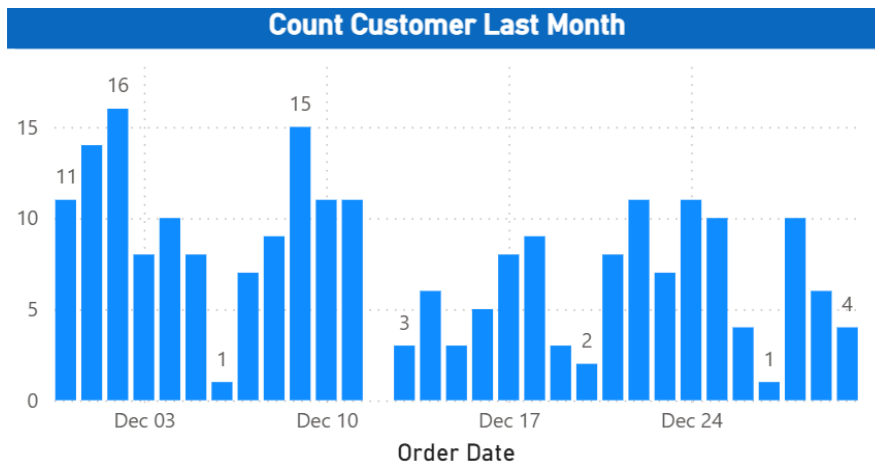
Ini adalah screenshot dashboard versi terakhir.

Card:

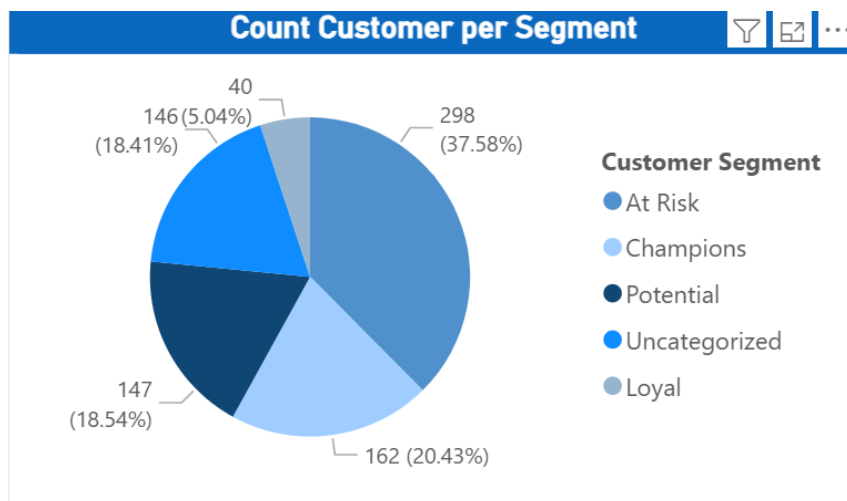
- Total sales: 2.3 M
- Profit: 286.4 K

Perlu diketahui bahwa kita memiliki button slicer sebagai berikut:

- Customer Segment: at risk, champion, loyal, potential, uncategorized
- Market Segment: Home Office, Corporate, Consumer
- Recency Bucket: 0-30 days, 30-60 days, 60-90 days, 90+ days
- Region (drop down): central, west, east, south

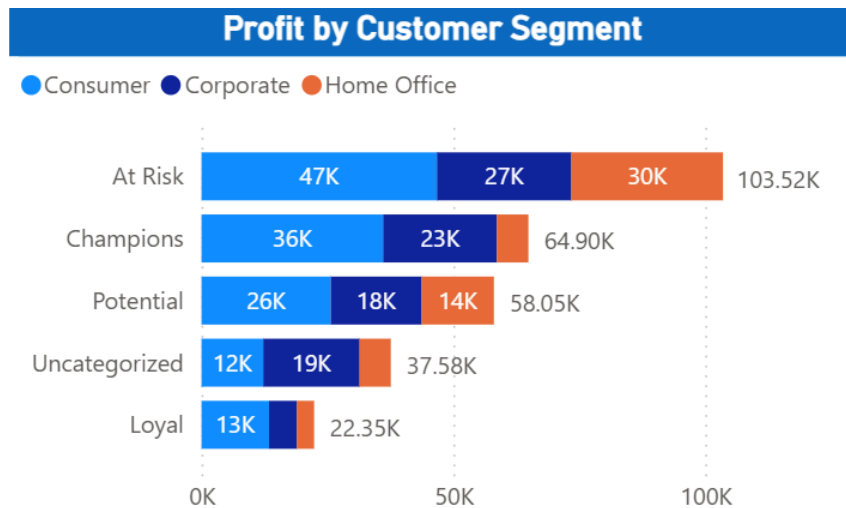


Pelanggan-pelanggan yang melakukan transaksi dalam bulan terakhir (30 hari terakhir) dihitung dalam grafik ini.



Di sini dapat terlihat bahwa ada:

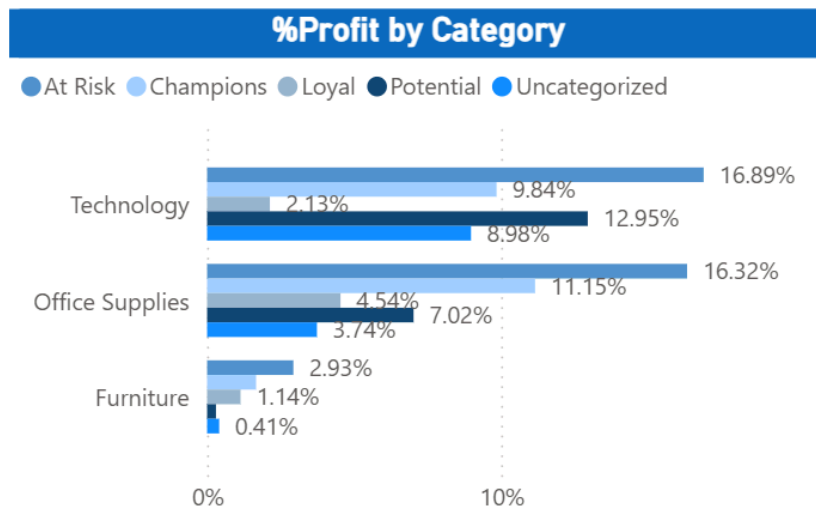
- Loyal: 40 orang
- At Risk: 298 orang
- Champion: 162 orang
- Potential: 147 orang
- Uncategorized: 146 orang



Grafik ini menunjukkan pembagian keuntungan berdasarkan segmen pelanggan dan segmen market

- at risk: 104K – consumer: 47K, corporate: 27K, home office: 30 K.
- champions: 65K – consumer: 36K, corporate: 23K, home office: 6K
- potential: 58K – consumer: 26K, corporate: 18K, home office: 14K
- uncategorized: 38K – consumer: 12K, corporate: 19K, home office: 11K
- loyal: 22.35K – consumer: 13K, corporate: N/A, home office: N/A – both corporate and home office: 9K

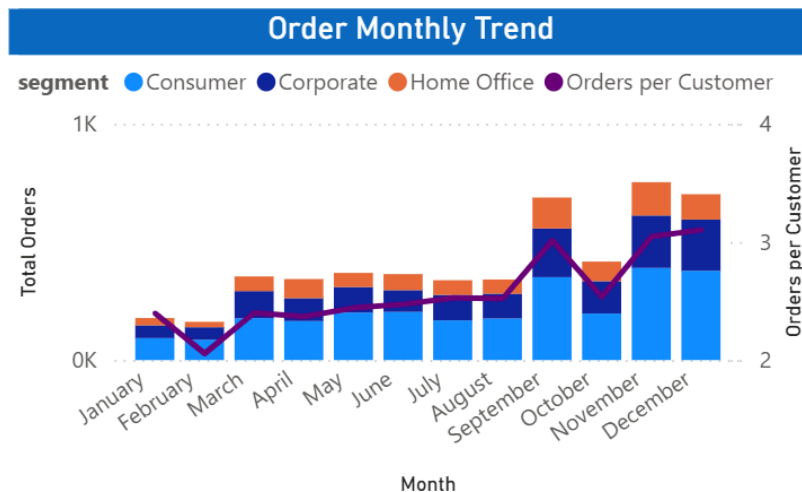
Dari sini bisa dilihat bahwa pelanggan yang RFM skornya paling rendah, segmen “At Risk” ternyata memberikan keuntungan paling tinggi, terutama dari sektor konsumen. Selain itu, dapat dilihat dari setiap segmen pelanggan hampir semuanya didominasi oleh konsumen (Consumer) dari market segment, kecuali kategori “Uncategorized”. Sementara itu “Corporate” atau corporate market segment selalu berada di urutan kedua sebagai sumber keuntungan kecuali kategori “Uncategorized” di mana corporate menempati urutan pertama. Perlu diingat bahwa berdasarkan RFM score, segmen pelanggan dibagi jadi champion (511–555), loyal (451–510), potential (351–450), at risk (151–350) dan uncategorized.



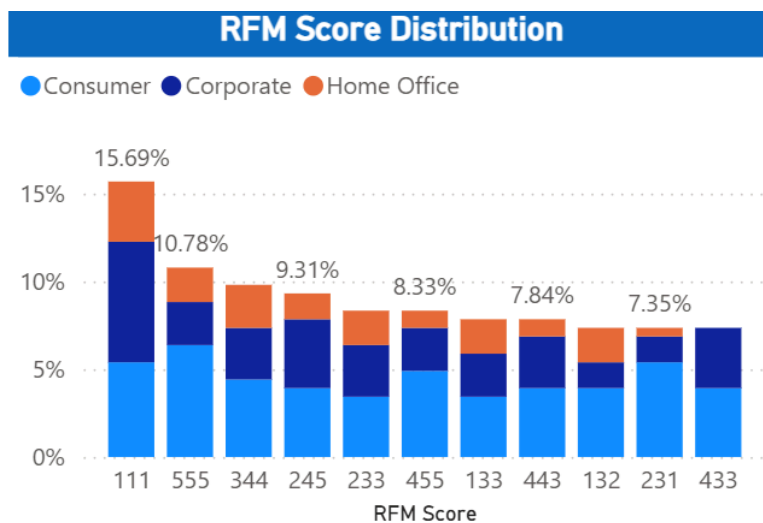
Pelanggan kategori “At Risk”, “Champions”, dan “Potential” merupakan kategori dari mayoritas pelanggan. Berdasarkan persentase,

- **Champions** adalah sumber profit kedua terbesar di *Technology* dan *Furniture* setelah segmen **At Risk**. **Potential** juga berkontribusi di urutan kedua untuk kategori *Technology*.
- Kategori **Furniture** perlu perhatian lebih karena profit rendah di semua segmen.
- Kategori *Office Supplies* memiliki segmen **Champion** terbanyak.
- Segmen **Uncategorized** relatif besar, khususnya di *Office Supplies*, yang berarti banyak pelanggan belum tersegmentasi. Biasanya, ini peluang untuk customer profiling lebih lanjut. Akan tetapi, menurut RFM score, biasanya pelanggan di segmen ini dapat diabaikan.

Berdasarkan visualisasi **%Profit by Category**, strategi utama yang dapat diambil adalah memaksimalkan kontribusi pelanggan segmen **Champions** yang terbukti memberikan profit tertinggi di kategori *Technology* dan *Office Supplies*, melalui program loyalitas eksklusif dan penawaran khusus. Segmen **Potential** di kategori *Technology* juga memiliki kontribusi besar sehingga layak dijadikan prioritas untuk diubah menjadi pelanggan **Loyal** atau **Champions** melalui pendekatan personalisasi dan promosi yang tepat sasaran. Pelanggan **Uncategorized** di kategori *Office Supplies* merupakan pelanggan dengan nilai RFM di bawah 151, yang mencerminkan perilaku kurang aktif dan nilai pembelian rendah, sehingga dapat dianggap sebagai pelanggan berisiko rendah yang perlu ditinjau ulang untuk melihat apakah masih layak dipertahankan atau dialihkan fokusnya ke segmen yang lebih potensial. Sementara itu, kategori *Furniture* menunjukkan profit yang sangat rendah di semua segmen sehingga perlu evaluasi strategi, baik dari sisi produk, pemasaran, maupun target pasar. Terakhir, pelanggan **At Risk** dapat diaktivasi kembali dengan strategi *win-back* seperti penawaran diskon personal atau kampanye retensi berdasarkan histori pembelian sebelumnya.



Bulan **Oktober dan Desember** merupakan periode performa penjualan terbaik, didorong oleh peningkatan jumlah pesanan dan frekuensi pemesanan (order) per pelanggan. Segmen **Consumer** menjadi kontributor utama terhadap volume pesanan, sehingga strategi pemasaran dan promosi bisa difokuskan pada segmen ini, terutama menjelang kuartal akhir. Sementara segmen **Home Office** dan **Corporate** juga menunjukkan peningkatan di bulan-bulan puncak, yang berarti potensi untuk dikembangkan melalui pendekatan yang lebih targeted.



Visualisasi RFM Score Distribution menunjukkan bahwa sebagian besar pelanggan berada pada RFM score 111 dengan persentase tertinggi sebesar 15.69%, yang mencerminkan pelanggan dengan aktivitas rendah, jarang melakukan pembelian, dan nilai transaksi yang kecil. Hal ini menunjukkan adanya konsentrasi pelanggan yang kurang bernilai dan perlu pendekatan khusus seperti kampanye reaktivasi atau segmentasi ulang. Di sisi lain, RFM score 555 yang merepresentasikan segmen Champions memiliki proporsi sebesar 10.78%, didominasi oleh pelanggan dari segmen Consumer, dan merupakan kelompok yang sangat aktif, loyal, serta memberikan kontribusi tinggi terhadap profit, sehingga layak diberikan perlakuan khusus seperti program loyalitas dan penghargaan eksklusif. Distribusi skor lainnya seperti 344, 245, 233, dan seterusnya menunjukkan pelanggan dengan performa menengah yang

tersebar relatif merata di semua segmen, membuka peluang untuk peningkatan nilai pelanggan melalui strategi upsell dan engagement yang lebih personal. Secara keseluruhan, segmen Consumer mendominasi hampir seluruh rentang skor, sementara Corporate dan Home Office muncul dalam proporsi lebih kecil namun tetap konsisten, sehingga strategi pemasaran dan retensi perlu disesuaikan dengan karakteristik dominan dari masing-masing kelompok.

Mayoritas pelanggan berada di skor **111**, yang merupakan kelompok paling tidak aktif dan bernilai rendah, sehingga perlu strategi khusus seperti kampanye reaktivasi atau filtering untuk efisiensi pemasaran. Sebaliknya, kelompok dengan skor **555** merupakan aset paling berharga (Champions) dan harus dijaga melalui program loyalitas dan penawaran eksklusif. Pelanggan dengan skor menengah dapat menjadi target potensial untuk dinaikkan ke level lebih tinggi melalui pendekatan personalisasi dan promosi intensif. Distribusi segmen yang merata menunjukkan bahwa strategi retensi dan konversi perlu mempertimbangkan karakteristik masing-masing segmen, khususnya Consumer yang paling mendominasi.

Customer Details					
Customer	Segment	RFM Score	Recency	Frequency	Monetary
Aaron Bergman	Uncategorized	111	417	3	886.16
Aaron Hawkins	Champions	543	14	7	1,744.70
Aaron Smayling	At Risk	344	90	7	3,050.69
Adam Bellavance	At Risk	345	56	8	7,755.62
Adam Hart	Loyal	454	36	10	3,250.34
Adam Shillingsburg	Loyal	454	30	9	3,255.31
Adrian Barton	Loyal	455	43	10	14,473.57
Adrian Hane	At Risk	342	62	7	1,735.51
Adrian Shami	Potential	411	43	2	58.82
Total					

Berdasarkan logika perhitungan skor RFM pada gambar **Customer Details**, setiap pelanggan diberikan skor Recency, Frequency, dan Monetary dari 1 hingga 5, yang kemudian digabung menjadi satu kode tiga digit (RFM Score). Skor yang lebih tinggi menunjukkan pelanggan yang lebih baru bertransaksi, lebih sering membeli, dan memiliki nilai pembelian lebih tinggi. Misalnya, *Aaron Hawkins* memiliki RFM Score 543 karena ia baru saja bertransaksi (recency rendah), sering melakukan pembelian, dan memiliki nilai transaksi cukup tinggi, sehingga masuk ke segmen **Champions**. Sebaliknya, *Aaron Bergman* memiliki RFM Score 111, yang menunjukkan sudah lama tidak melakukan transaksi, jarang membeli, dan nilai pembeliannya rendah, sehingga dimasukkan ke segmen **Uncategorized**. Pelanggan seperti *Adam Hart* dan *Adrian Barton* memiliki skor yang tinggi pada frequency dan monetary, tetapi recency mereka sedikit lebih lama, sehingga masuk ke segmen **Loyal**. Sementara pelanggan seperti *Adrian Shami*, yang memiliki recency dan frequency cukup baik namun monetary sangat

rendah, masuk ke segmen **Potential**. Skor-skor ini mencerminkan perilaku pelanggan dan digunakan untuk mengelompokkan mereka ke dalam segmen yang relevan, sehingga perusahaan dapat menentukan strategi pemasaran dan retensi yang lebih efektif berdasarkan karakteristik setiap pelanggan.

RFM matrix						
R\F	1	2	3	4	5	Total
1	44,763.81	58,125.70	145,942.02	28,076.14	69,207.59	346,115.25
2	16,710.81	24,476.02	119,906.40	178,479.61	96,835.69	436,408.53
3	16,789.90	49,563.10	130,448.94	172,259.78	117,857.23	486,918.95
4	11,974.51	28,106.11	147,970.69	156,201.62	194,424.68	538,677.61
5	13,852.50	14,365.82	91,057.08	158,080.31	211,724.82	489,080.52
Total	104,091.54	174,636.74	635,325.13	693,097.44	690,050.01	2,297,200.86

Penjelasan matrix adalah sebagai berikut. **Baris (R)**: Skor recency, dari 1 hingga 5. **Kolom (F)**: Skor frequency, dari 1 hingga 5. **Isi Matriks**: Jumlah total pembelanjaan (monetary) dari pelanggan dengan skor R dan F tertentu.

Kombinasi R=5 dan F=5 menghasilkan nilai transaksi **tertinggi** sebesar **211,724.82**, artinya pelanggan yang *baru saja melakukan transaksi dan sangat sering membeli* adalah yang paling menguntungkan. Kombinasi R=4, F=5 juga sangat tinggi (**194,424.68**) dan R=5, F=4 (**158,080.31**), memperkuat bahwa pelanggan dengan **recency dan frequency tinggi menghasilkan profit paling besar**.

Kombinasi R=1 dan F=1 menghasilkan total transaksi hanya **44,763.81**, menandakan bahwa pelanggan yang sudah lama tidak bertransaksi dan jarang membeli memiliki nilai transaksi terendah. Ini adalah area pelanggan yang berisiko churn dan kurang bernilai untuk saat ini.

Secara horizontal, kolom F=4 dan F=5 menyumbang nilai transaksi tertinggi (masing-masing **693,097.44** dan **690,050.01**) dibandingkan kolom lainnya. Artinya, **semakin tinggi frekuensi pembelian, semakin besar kontribusi terhadap total penjualan**. Secara vertikal, R=4 dan R=5 juga menyumbang total tertinggi (masing-masing **538,677.61** dan **489,080.52**), menunjukkan bahwa **pelanggan yang masih aktif (recency tinggi)** adalah kontributor utama profit.

Matriks ini menunjukkan bahwa pelanggan dengan **recency dan frequency tinggi adalah sumber profit utama**, sehingga harus dijaga dengan strategi loyalitas, penghargaan, dan penawaran eksklusif. Sebaliknya, pelanggan dengan recency rendah dan frequency rendah menghasilkan nilai transaksi yang kecil, dan perlu ditinjau apakah layak diaktifkan kembali atau tidak. Strategi upsell dan engagement harus difokuskan pada pelanggan di segmen tengah (R dan F = 3 atau 4) yang memiliki potensi naik kelas jika dikelola dengan baik.