



**Dibimbing**



# Marketing Analysis and Customer Segmentation

Final Project  
Business Intelligence Bootcamp  
Batch V



**Dartwin**

Passionate, Data-Driven, and Life-Long  
Learner





# Daftar Isi

- Business Understanding
- Business Objectives
- Data Understanding
- Exploratory Data Analysis
- Marketing Analysis
- Customer Segmentation RFM
- Conclusion
- Business Recommendation

Source:  
<https://healthyeatingresearch.org>



# Business Understanding



Source:  
<https://www.nytimes.com>  
<https://www.meatpoultry.com>  
<https://lilyvolt.com>  
<https://www.seafoodsource.com>  
<https://www.givaudan.com>

- iFood adalah perusahaan yang bergerak di bidang ritel makanan
- Produk utamanya berupa:
  - Anggur (wines)
  - Produk daging (rare meat products)
  - Buah-buahan (exotic fruits)
  - Ikan olahan khusus (specially prepared fish products)
  - Produk manis (sweet products)
- Untuk meningkatkan pertumbuhan laba, iFood memutuskan meningkatkan kinerja kegiatan pemasaran melalui **kampanye pemasaran**





# Business Objectives



## Marketing Analysis

1. Melakukan exploratory data analysis terhadap data kampanye marketing yang sudah ada
2. Menentukan marketing metrik conversion rate dan retention rate kampanye marketing
3. Menganalisa pengaruh komplain customer terhadap retention rate kampanye marketing



## Customer Segmentation RFM

1. Melakukan segmentasi customer perusahaan menjadi tipe-tipe customer berupa best customer, loyal customer, big spender, dan almost lost and lost good customer
2. Menentukan jumlah masing-masing tipe customer yang ada
3. Menganalisa customer transaction berdasarkan tipe customer yang ada





# Data Understanding

#	Column	Non-Null Count	Dtype
0	Income	2021 non-null	float64
1	Kidhome	2021 non-null	int64
2	Teenhome	2021 non-null	int64
3	Recency	2021 non-null	int64
4	MntWines	2021 non-null	int64
5	MntFruits	2021 non-null	int64
6	MntMeatProducts	2021 non-null	int64
7	MntFishProducts	2021 non-null	int64
8	MntSweetProducts	2021 non-null	int64
9	MntGoldProds	2021 non-null	int64
10	NumDealsPurchases	2021 non-null	int64
11	NumWebPurchases	2021 non-null	int64
12	NumCatalogPurchase	2021 non-null	int64
13	NumStorePurchases	2021 non-null	int64
14	NumWebVisitsMonth	2021 non-null	int64
15	AcceptedCmp3	2021 non-null	int64
16	AcceptedCmp4	2021 non-null	int64
17	AcceptedCmp5	2021 non-null	int64
18	AcceptedCmp1	2021 non-null	int64
19	AcceptedCmp2	2021 non-null	int64
20	Complain	2021 non-null	int64

#	Column	Non-Null Count	Dtype
21	Z_CostContact	2021 non-null	int64
22	Z_Revenue	2021 non-null	int64
23	Response	2021 non-null	int64
24	Age	2021 non-null	int64
25	Customer_Days	2021 non-null	int64
26	marital_Divorced	2021 non-null	int64
27	marital_Married	2021 non-null	int64
28	marital_Single	2021 non-null	int64
29	marital_Together	2021 non-null	int64
30	marital_Widow	2021 non-null	int64
31	education_2nCycle	2021 non-null	int64
32	education_Basic	2021 non-null	int64
33	education_Graduati	2021 non-null	int64
34	education_Master	2021 non-null	int64
35	education_PhD	2021 non-null	int64
36	MntTotal	2021 non-null	int64
37	MntRegularProds	2021 non-null	int64
38	AcceptedCmpOverall	2021 non-null	int64

- Dataset diperoleh dari Kaggle
- Dataset ifood\_df.csv terdiri atas:
  - **38 kolom**, dimana semua kolom merupakan kolom data numerical
  - **2205 baris**

Source:

<https://www.kaggle.com/datasets/jackdaoud/marketing-data>



# Data Preparation



```
[2] # Cek data missing values
## Mengetahui berapa banyak data yang missing value
df_marketing_copy.isna().sum()

Income      0
Kidhome     0
Teenhome    0
Recency     0
MntWines    0
MntFruits   0
MntMeatProducts 0
MntFishProducts 0
MntSweetProducts 0
MntGoldProds 0
NumDealsPurchases 0
NumWebPurchases 0
NumCatalogPurchases 0
NumStorePurchases 0
NumWebVisitsMonth 0
AcceptedCmp3 0
AcceptedCmp4 0
AcceptedCmp5 0
```

Missing  
Values Check  
and Handling



```
[63] # Cek data yang duplikat
## Mengetahui berapa banyak data yang duplikat
df_marketing_copy.duplicated().sum()

184

Karena terdapat data duplicated values sebanyak 184 baris,
terhadap data tsb. sehingga hal tsb dapat diperbaiki dan ke
terhadap data ybh, maka baru dilakukan tindak lebih lanjut

Dalam kasus ini, data duplicated values akan di-drop.

[64] # Men-drop duplicated values dari data
df_marketing_copy.drop_duplicates(inplace=True)

# Menampilkan jumlah data setelah data duplicated
df_marketing_copy
```

Duplications  
Check and  
Handling



```
ted

dan if untuk menentukan nilai data kolom IsRetented
rall + Response > 1, maka IsRetented = 1
rall + Response <= 1, maka Is Retented = 0
, Response in zip(df_marketing_copy['AcceptedCmpOverall'], d
rall + Response) > 1:
end(1)

end(0)

etented ke dalam data df_marketing copy
etented'] = IsRetented

asil penambahan data kolom IsRetented
ceptedCmpOverall", "Response", "IsRetented"].sample(10)

all Response IsRetented
0 0 0
0 0 0
```

New Data  
Columns  
Creation



Google Colab:

<https://colab.research.google.com/drive/17NrzywJPqOq6raqmT3McoR1ZrULwfDrp?usp=sharing>





# Data Preparation

No.	Column	Non-Null Count	Dtype
0	Income	2021 non-null	float64
1	Kidhome	2021 non-null	int64
2	Teenhome	2021 non-null	int64
3	Recency	2021 non-null	int64
4	MntWines	2021 non-null	int64
5	MntFruits	2021 non-null	int64
6	MntMeatProducts	2021 non-null	int64
7	MntFishProducts	2021 non-null	int64
8	MntSweetProducts	2021 non-null	int64
9	MntGoldProds	2021 non-null	int64
10	NumDealsPurchases	2021 non-null	int64
11	NumWebPurchases	2021 non-null	int64
12	NumCatalogPurchase	2021 non-null	int64
13	NumStorePurchases	2021 non-null	int64
14	NumWebVisitsMonth	2021 non-null	int64
15	AcceptedCmp3	2021 non-null	int64
16	AcceptedCmp4	2021 non-null	int64
17	AcceptedCmp5	2021 non-null	int64
18	AcceptedCmp1	2021 non-null	int64
19	AcceptedCmp2	2021 non-null	int64
20	Complain	2021 non-null	int64

No.	Column	Non-Null Count	Dtype
21	Z_CostContact	2021 non-null	int64
22	Z_Revenue	2021 non-null	int64
23	Response	2021 non-null	int64
24	Age	2021 non-null	int64
25	Customer_Days	2021 non-null	int64
26	marital_Divorced	2021 non-null	int64
27	marital_Married	2021 non-null	int64
28	marital_Single	2021 non-null	int64
29	marital_Together	2021 non-null	int64
30	marital_Widow	2021 non-null	int64
31	education_2nCycle	2021 non-null	int64
32	education_Basic	2021 non-null	int64
33	education_Graduati	2021 non-null	int64
34	education_Master	2021 non-null	int64
35	education_PhD	2021 non-null	int64
36	MntTotal	2021 non-null	int64
37	MntRegularProds	2021 non-null	int64
38	AcceptedCmpOverall	2021 non-null	int64
39	IsRetented	2021 non-null	int64
40	NumPurchasesTotal	2021 non-null	int64

- Dataset ifood\_df.csv setelah dilakukan preparation terdiri atas:
  - 41 kolom, dimana 39 kolom di awal dan 2 kolom yang ditambahkan
  - 2021 baris, setelah dilakukan data preparation (missing values dan duplicated data handling)

Source:

<https://www.kaggle.com/datasets/jackdaoud/marketing-data>



# Data Preparation

No.	Column	Description
1	Recency	Jumlah hari sejak purchase terakhir customer
2	MntWines	Jumlah produk anggur yang dibelanjakan customer
3	MntFruits	Jumlah produk buah-buahan yang dibelanjakan customer
4	MntMeatProducts	Jumlah produk daging yang dibelanjakan customer
5	MntFishProducts	Jumlah produk ikan olahan khusus yang dibelanjakan customer
6	MntSweetProducts	Jumlah produk manis yang dibelanjakan customer
7	MntGoldProds	Jumlah produk emas yang dibelanjakan customer
8	MntTotal	Jumlah produk total yang dibelanjakan customer
9	NumDealsPurchases	Jumlah purchases yang dibuat melalui diskon
10	NumWebPurchases	Jumlah purchases yang dibuat melalui website
11	NumCatalogPurchase	Jumlah purchases yang dibuat melalui katalog
12	NumStorePurchases	Jumlah purchases yang dibuat melalui toko
13	AcceptedCmp1	1 jika customer menerima tawaran pada kampanye pertama, 0 jika tidak'
14	AcceptedCmp2	1 jika customer menerima tawaran pada kampanye kedua, 0 jika tidak'
15	AcceptedCmp3	1 jika customer menerima tawaran pada kampanye ketiga, 0 jika tidak'
16	AcceptedCmp4	1 jika customer menerima tawaran pada kampanye keempat, 0 jika tidak'
17	AcceptedCmp5	1 jika customer menerima tawaran pada kampanye kelima, 0 jika tidak'
18	Response	1 jika customer menerima tawaran pada kampanye terakhir, 0 jika tidak'
19	Complain	1 jika customer komplain, 0 jika tidak
20	IsRetented	1 jika customer menerima tawaran lebih dari 1 kali, 0 jika tidak
21	NumPurchasesTotal	Jumlah purchases total yang dibuat

Source:

<https://www.kaggle.com/datasets/jackdaoud/marketing-data>

- **Marketing analysis** menggunakan data-data kolom jumlah purchases, kampanye (AcceptedCmp dan Response), dan komplain
- **Customer segmentation RFM** menggunakan data-data kolom jumlah produk, recency, dan jumlah purchases

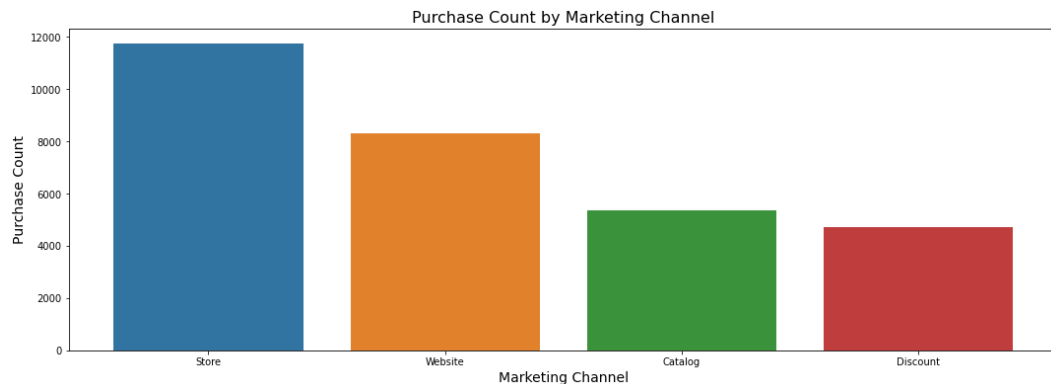




# Exploratory Data Analysis

## Purchase Count by Marketing Channel

No.	Channel	Count
0	Store	11,737
1	Website	8,317
2	Catalog	5,345
3	Discount	4,709



- Marketing channel yang menghasilkan purchase terbanyak adalah Store sebanyak 11.737 purchases
- Marketing channel yang memiliki purchases tersedikit adalah Online Discount, sebanyak 4.709 purchases
- Rata-rata purchases yang dimiliki adalah 7.527 purchases

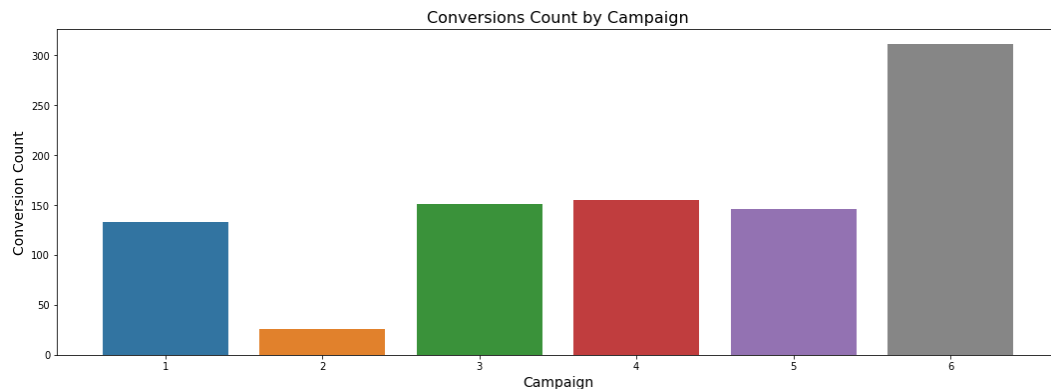


# Marketing Analysis



## Conversion Rate by Marketing Channel dan Retention Rate

No.	Campaign	Conversions	Interactions	Conversion Rate
0	1	133	2,021	6.58%
1	2	26	2,021	1.29%
2	3	151	2,021	7.47%
3	4	155	2,021	7.67%
4	5	146	2,021	7.22%
5	6	311	2,021	15.39%



No.	retented_count	converted_count	retention_rate
0	219	344	63.66%

- Kampanye yang memiliki conversion rate tertinggi adalah kampanye terakhir atau ke-6, yaitu sebesar 15,38%
- Kampanmye yang memiliki conversion rate terendah adalah kampanye ke-2, yaitu sebesar 1,29%
- Rata-rata jumlah conversions yang dihasilkan dari masing-masing kampanye adalah 153,66 conversions
- Rata-rata conversion rate yang dihasilkan dari masing-masing kampanye adalah 7,60%
- Retention rate adalah sebesar 63,66%. Artinya retention rate pada perusahaan cukup baik



# Marketing Analysis



## The Impact of Customer Complain on Customer Retention Analysis



No.	Complain	IsRetented	size	prop
0	0	0	1,785	89.16%
1		1	217	10.84%
2	1	0	17	89.47%
3		1	2	10.53%

- Jenis status Complain tidak berpengaruh signifikan terhadap IsRetented, karena proporsi dari masing-masing jenis status Complain dan IsRetented tidak berbeda jauh
- Status complain 0 cenderung dialami oleh mayoritas customer, sedangkan status complain 1 cenderung dialami yang minoritas
- Faktor-faktor yang mempengaruhi IsRetented selain Complain dapat berupa:
  1. Conversion rate, karena jika customer tidak converted, maka tidak mungkin ter-retented
  2. Marketing channel, karena mempengaruhi customer apakah akan purchase produk perusahaan atau tidak
  3. Jenis complain yang masuk, karena jenis complain dapat berupa saran atau laporan, dsb





# Customer Segmentation

## RFA Segmentation Table

10 Baris Data Pertama RFA Table

No.	recency	frequency	amount_spent	R_Quartile	F_Quartile	A_Quartile	RFAClass
0	58	25	1529	3	1	1	311
1	38	6	21	2	4	4	244
2	26	21	734	2	2	2	222
3	26	8	48	2	4	4	244
4	94	19	407	4	2	2	422
5	16	22	702	1	1	2	112
6	34	21	563	2	2	2	222
7	32	10	146	2	3	3	233
8	19	6	44	1	4	4	144
9	68	2	36	3	4	4	344

RFA Table Result:

<https://drive.google.com/file/d/1N6Lp6oUAclCMfdbKCkWcMB8mYS7WpcCL/view?usp=sharing>

- Recency menggunakan data recency (recency)
- Frequency menggunakan data jumlah total purchases yang dibuat (NumPurchasesTotal)
- Karena terdapat keterbatasan data, maka dilakukan modifikasi pada metode segmentasi RFM, dimana Monetary Value menggunakan data jumlah barang yang dibelanjakan oleh customer (MntTotal)



# Customer Segmentation

## Tipe Customer dan Jumlahnya

No.	Customer	Count	prop
0	Best Customer	48	2.38%
1	Loyal Customer	226	11.18%
2	Big Spender	267	13.21%
3	Almost Lost and Lost Good Customer	125	6.19%
4	Lainnya	1,355	67.05%

- Tipe customer yang memiliki proporsi terbanyak adalah Big Spender, yaitu sebanyak 13,21% atau 267 orang
- Tipe customer yang memiliki proporsi tersedikit adalah Best Customer, yaitu sebanyak 2,38% atau 48 orang



# Customer Segmentation



## Customer Transaction by Customer Type: Best Customer by Amount Spent

Top 10 Best Customer dengan Amount Spent Tertinggi

No.	recency	frequency	amount_spent	R_Quartile	F_Quartile	A_Quartile	RFAClass
0	16	25	2,304	1	1	1	111
1	2	23	2,092	1	1	1	111
2	20	22	2,051	1	1	1	111
3	14	26	1,719	1	1	1	111
4	20	25	1,713	1	1	1	111
5	20	27	1,673	1	1	1	111
6	2	29	1,661	1	1	1	111
7	17	29	1,612	1	1	1	111
8	14	24	1,566	1	1	1	111
9	18	22	1,552	1	1	1	111

- Amount Spent rata-rata yang dimiliki oleh top 10 Best Customer yang melakukan purchase tertinggi adalah 1.794,3 satuan amount\_spent
- Amount Spent tertinggi yang dimiliki oleh Best Customer adalah 2.304 satuan amount\_spent





# Customer Segmentation



## Customer Transaction by Customer Type: Loyal Customer by Frequency

Top 10 Loyal Customer dengan Frequency Terbanyak

No.	recency	frequency	amount_spent	R_Quartile	F_Quartile	A_Quartile	RFAClass
0	88	33	724	4	1	2	412
1	74	32	650	3	1	2	312
2	20	32	712	1	1	2	112
3	72	31	894	3	1	2	312
4	0	31	652	1	1	2	112
5	37	31	950	2	1	2	212
6	92	31	953	4	1	2	412
7	17	30	715	1	1	2	112
8	81	30	898	4	1	2	412
9	29	30	777	2	1	2	212

- Frequency rata-rata yang dimiliki oleh top 10 Loyal Customer yang melakukan transaksi paling sering adalah 31,1 kali
- Frequency tertinggi yang dimiliki oleh Loyal Customer adalah 33 kali



# Customer Segmentation



## Customer Transaction by Customer Type: Big Spender by Amount Spent

Top 10 Big Spender dengan Amount Spent Tertinggi

No.	recency	frequency	amount_spent	R_Quartile	F_Quartile	A_Quartile	RFAClass
0	29	18	2,429	2	2	1	221
1	29	18	2,429	2	2	1	221
2	56	15	2,188	3	3	1	331
3	62	18	2,130	3	2	1	321
4	73	13	2,105	3	3	1	331
5	35	21	2,068	2	2	1	221
6	25	21	2,066	2	2	1	221
7	30	21	2,036	2	2	1	221
8	32	21	1,998	2	2	1	221
9	78	21	1,988	4	2	1	421

- Amount\_spent rata-rata yang dimiliki oleh top 10 Big Spender yang melakukan purchase tertinggi adalah 2143.7 satuan amount\_spent
- Amount\_spent tertinggi yang dimiliki oleh Big Spender adalah 2.429 satuan amount\_spent



# Customer Segmentation



## Customer Transaction by Customer Type: Almost Lost and Lost Good Customer by Recency

Top 10 Almost Lost and Lost Good Customer dengan Recency Terlama

No.	recency	frequency	amount_spent	R_Quartile	F_Quartile	A_Quartile	RFAClass
0	99	26	996	4	1	1	411
1	98	27	1,446	4	1	1	411
2	98	24	1,555	4	1	1	411
3	98	22	2,068	4	1	1	411
4	98	24	1,308	4	1	1	411
5	98	28	1,180	4	1	1	411
6	98	29	1,290	4	1	1	411
7	97	25	1,577	4	1	1	411
8	97	26	992	4	1	1	411
9	97	34	1,502	4	1	1	411

- Recency rata-rata yang dimiliki oleh top 10 Almost Lost dan Lost Good Customer yang memiliki recency tertinggi adalah 97.8 hari
- Recency terlama yang dimiliki oleh Almost Lost dan Lost Good Customer adalah 99 hari



# Conclusion & Business Recommendation



## Conclusion

1. Marketing channel yang menghasilkan purchases terbanyak adalah **Store**, sedangkan kampanye marketing yang menghasilkan conversion rate tertinggi adalah **kampanye terakhir** atau **ke-6**
2. Retention rate yang dihasilkan melalui kampanye-kampanye marketing adalah sebesar **63.66%**
3. Komplain customer tidak memiliki pengaruh yang signifikan terhadap retention customer
4. Customer segmentation berhasil dilakukan untuk 2021 customers



## Business Recommendation

1. Mengurangi biaya yang diperlukan untuk marketing channel diskon dan menggunakannya untuk meningkatkan performa marketing channel Store
2. Melakukan marketing kampanye customized untuk tipe-tipe customer yang berbeda:
  - a. Memberikan lebih banyak interactions dan notifikasi belanja kepada Big Spenders
  - b. Memberikan promo diskon kepada Almost Lost and Lost Good Customer
  - c. Memberikan tawaran spesial (ongkos kirim, kecepatan pengiriman, customer service, dll.) kepada Best Customer









**Dibimbing**

**Thank You**  
**Let's Get in Touch!**



**Dartwin**

-  Email: [dartwinshu123@gmail.com](mailto:dartwinshu123@gmail.com)
-  LinkedIn: Dartwin –
-  Instagram: [dartwinshu](https://www.instagram.com/dartwinshu)
-  Github: [github.com/dartwinshu](https://github.com/dartwinshu)



**My Portfolio**