

TokoBli Twin Date Campaign Analysis

Satryo Sunu Prabowo





Business Understanding

Business Understanding

TokoBli, the largest e-commerce company in Indonesia, needs to optimize its campaign strategy for its upcoming "twin date" flash sale campaigns in order to maximize revenue and transactions while minimizing campaign budget.

Objective:

- To analyze the data from the past 3 campaigns and identify the campaign scenario that resulted in the highest revenue and transactions with the lowest campaign budget.
- To compare the campaign and use the insights gained from the analysis to optimize the campaign strategy for future "twin date" campaigns.

How Data Can Help Business

With the provided data, we can leverage various metrics and analyses to answer the business problem effectively. Such as:

- **Identify the most effective campaign scenario:** By comparing the key metrics across the three scenarios, TokoBli can determine which scenario generated the highest revenue, transactions, and ROI. This will allow them to focus their efforts on the most successful tactics and strategies in the future.
- **Analyze the comparison between each campaign:** By looking at the individual metrics for each campaign, such as total products sold in each category. In addition, TokoBli can gain insights into what factors contributed to each campaign's performance. For example, if Campaign A had the highest revenue but also the highest campaign budget, it may be worth investigating whether the additional spending was justified by the increased sales.

Metrics Definition

Based on the point of the business problem, the following statistics and metrics would be crucial for a comprehensive analysis:

- **Revenue:**
 - Total Revenue of each campaign
 - Average revenue of each campaign
 - ROI for each campaign
- **Transaction:**
 - Total transaction for each campaign
- **Campaign Budget:**
 - Total budget spend on each campaign
- **Categories:**
 - Total order value on each category
 - Average order value on each category
 - Average discount on each category



Data Cleaning

Data Cleaning Process

Remove Duplicate



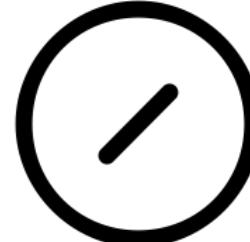
Remove/Impute Missing Data



Change Data Type



Remove Irrelevant Data



Change Data Type

	A	B	C	D	E	Automatic	F	G	H	I
Insaction	Product ID	Customer ID	Gender	St	SKU	Price				
1	72441	21 F	o	Number	1,000,12	2,10E+06				
2	72474	168 M	o	Percent	10,12%	3,05GOMBSA09H				
3	72475	168 F	o	Scientific	1.01e+04	3,97E+06				
4	72477	168 M	✓		0,05045E-01	2,00E+06				
5	72478	168 F	o		0,050543E-02	2,00E+06				
6	72478	6715 M	o	Accounting	\$ (1,000,12)	6,99E+06				
7	72442	6618 F	o	Financial	(1,000,12)	W10CFHGS4A02				
8	72443	8647 F	o		0,05045E-01	8,99E+06				
9	72462	8765 F	o	Currency	\$1,000,12	ADQAVOS94N				
11	72496	15960 M	o	Currency rounded	\$1,000	3,05GOMBSA09Z				
12	72497	15960 F	o		0,05045E-01	9,00E+06				
13	72498	15960 M	o		0,05045E-01	9,00E+06				
14	72499	15960 F	o	Date	9/26/2008	3,05GOMASAP27				
15	72499	15960 M	o	Time	3:59:00 PM	3,05GOMASAP27				
16	72501	15960 F	o	Date time	9/26/2008 15:59:00	3,05GOSND03A7				
17	72502	15960 M	o	Duration	24:01:00	1,70E+06				
18	72483	15960 F	o		0,05045E-01	9,00E+06				
19	72494	15960 M	o	Indonesian Rupiah	Rp1,000,000	3,05GOMBSA09Z				
20	72495	15960 F	o		0,05045E-01	9,00E+06				
21	724738	16033 M	o		0,05045E-01	3,05GOMASAP27				
22	72486	17634 M	o		0,05045E-01	3,05GOMASAP27				
23	72488	17634 F	o	Custom currency	0,05045E-01	4,45E+06				
24	72465	18967 M	o	Custom date and time	0,05045E-01	4,45E+06				
25	72466	18967 M	o	Custom number format	0,05045E-01	3,99E+06				
26	72467	18967 M	o	Custom text	0,05045E-01	3,99E+06				
27	72468	19100 M	o		0,05045E-01	8,00E+06				
28	72470	19100 M	o		0,05045E-01	8,00E+06				
29	72473	19100 M	o		0,05045E-01	8,00E+06				
30	72473	19100 M	o		0,05045E-01	8,00E+06				
31	72473	19100 M	o		0,05045E-01	8,00E+06				
32	72473	19100 M	o		0,05045E-01	8,00E+06				
33	72473	19100 M	o		0,05045E-01	8,00E+06				
34	72473	19100 M	o		0,05045E-01	8,00E+06				
35	72473	19100 M	o		0,05045E-01	8,00E+06				
36	72473	19100 M	o		0,05045E-01	8,00E+06				
37	72473	19100 M	o		0,05045E-01	8,00E+06				
38	72473	19100 M	o		0,05045E-01	8,00E+06				
39	72473	19100 M	o		0,05045E-01	8,00E+06				
40	72473	19100 M	o		0,05045E-01	8,00E+06				
41	72473	19100 M	o		0,05045E-01	8,00E+06				
42	72473	19100 M	o		0,05045E-01	8,00E+06				
43	72473	19100 M	o		0,05045E-01	8,00E+06				
44	72473	19100 M	o		0,05045E-01	8,00E+06				
45	72473	19100 M	o		0,05045E-01	8,00E+06				
46	72473	19100 M	o		0,05045E-01	8,00E+06				
47	72473	19100 M	o		0,05045E-01	8,00E+06				
48	72473	19100 M	o		0,05045E-01	8,00E+06				
49	72473	19100 M	o		0,05045E-01	8,00E+06				
50	72473	19100 M	o		0,05045E-01	8,00E+06				
51	72473	19100 M	o		0,05045E-01	8,00E+06				
52	72473	19100 M	o		0,05045E-01	8,00E+06				
53	72473	19100 M	o		0,05045E-01	8,00E+06				
54	72473	19100 M	o		0,05045E-01	8,00E+06				
55	72473	19100 M	o		0,05045E-01	8,00E+06				
56	72473	19100 M	o		0,05045E-01	8,00E+06				
57	72473	19100 M	o		0,05045E-01	8,00E+06				
58	72473	19100 M	o		0,05045E-01	8,00E+06				
59	72473	19100 M	o		0,05045E-01	8,00E+06				
60	72473	19100 M	o		0,05045E-01	8,00E+06				
61	72473	19100 M	o		0,05045E-01	8,00E+06				
62	72473	19100 M	o		0,05045E-01	8,00E+06				
63	72473	19100 M	o		0,05045E-01	8,00E+06				
64	72473	19100 M	o		0,05045E-01	8,00E+06				
65	72473	19100 M	o		0,05045E-01	8,00E+06				
66	72473	19100 M	o		0,05045E-01	8,00E+06				
67	72473	19100 M	o		0,05045E-01	8,00E+06				
68	72473	19100 M	o		0,05045E-01	8,00E+06				
69	72473	19100 M	o		0,05045E-01	8,00E+06				
70	72473	19100 M	o		0,05045E-01	8,00E+06				
71	72473	19100 M	o		0,05045E-01	8,00E+06				
72	72473	19100 M	o		0,05045E-01	8,00E+06				
73	72473	19100 M	o		0,05045E-01	8,00E+06				
74	72473	19100 M	o		0,05045E-01	8,00E+06				
75	72473	19100 M	o		0,05045E-01	8,00E+06				
76	72473	19100 M	o		0,05045E-01	8,00E+06				
77	72473	19100 M	o		0,05045E-01	8,00E+06				
78	72473	19100 M	o		0,05045E-01	8,00E+06				
79	72473	19100 M	o		0,05045E-01	8,00E+06				
80	72473	19100 M	o		0,05045E-01	8,00E+06				
81	72473	19100 M	o		0,05045E-01	8,00E+06				
82	72473	19100 M	o		0,05045E-01	8,00E+06				
83	72473	19100 M	o		0,05045E-01	8,00E+06				
84	72473	19100 M	o		0,05045E-01	8,00E+06				
85	72473	19100 M	o		0,05045E-01	8,00E+06				
86	72473	19100 M	o		0,05045E-01	8,00E+06				
87	72473	19100 M	o		0,05045E-01	8,00E+06				
88	72473	19100 M	o		0,05045E-01	8,00E+06				
89	72473	19100 M	o		0,05045E-01	8,00E+06				
90	72473	19100 M	o		0,05045E-01	8,00E+06				
91	72473	19100 M	o		0,05045E-01	8,00E+06				
92	72473	19100 M	o		0,05045E-01	8,00E+06				
93	72473	19100 M	o		0,05045E-01	8,00E+06				
94	72473	19100 M	o		0,05045E-01	8,00E+06				
95	72473	19100 M	o		0,05045E-01	8,00E+06				
96	72473	19100 M	o		0,05045E-01	8,00E+06				
97	72473	19100 M	o		0,05045E-01	8,00E+06				
98	72473	19100 M	o		0,05045E-01	8,00E+06				
99	72473	19100 M	o		0,05045E-01	8,00E+06				
100	72473	19100 M	o		0,05045E-01	8,00E+06				
101	72473	19100 M	o		0,05045E-01	8,00E+06				
102	72473	19100 M	o		0,05045E-01	8,00E+06				
103	72473	19100 M	o		0,05045E-01	8,00E+06				
104	72473	19100 M	o		0,05045E-01	8,00E+06				
105	72473	19100 M	o		0,05045E-01	8,00E+06				
106	72473	19100 M	o		0,05045E-01	8,00E+06				
107	72473	19100 M	o		0,05045E-01	8,00E+06				
108	72473	19100 M	o		0,05045E-01	8,00E+06				
109	72473	19100 M	o		0,05045E-01	8,00E+06				
110	72473	19100 M	o		0,05045E-01	8,00E+06				
111	72473	19100 M	o		0,05045E-01	8,00E+06				
112	72473	19100 M	o		0,05045E-01	8,00E+06				
113	72473	19100 M	o		0,05045E-01	8,00E+06				
114	72473	19100 M	o		0,05045E-01	8,00E+06				
115	72473	19100 M	o		0,05045E-01	8,00E+06				
116	72473	19100 M	o		0,05045E-01	8,00E+06				
117	72473	19100 M	o		0,05045E-01	8,00E+06				
118	72473	19100 M	o		0,05045E-01	8,00E+06				
119	72473	19100 M	o		0,05045E-01	8,00E+06				
120	72473	19100 M	o		0,05045E-01	8,00E+06				
121	72473	19100 M	o		0,05045E-01	8,00E+06				
122	72473	19100 M	o		0,05045E-01	8,00E+06				
123	72473	19100 M	o		0,05045E-01	8,00E+06				
124	72473	19100 M	o		0,05045E-01	8,00E+06				
125	72473	19100 M	o		0,05045E-01	8,00E+06				
126	72473	19100 M	o		0,05045E-01	8,00E+06				
127	72473	19100 M	o		0,05045E-01	8,00E+06				
128	72473	19100 M	o		0,05045E-01	8,00E+06				
129	72473	19100 M	o		0,05045E-01	8,00E+06				
130	72473	19100 M	o		0,05045E-01	8,00E+06				
131	72473	19100 M	o		0,05045E-01	8,00E+06				
132	72473	19100 M	o		0,05045E-01	8,00E+06				
133	72473	19100 M	o		0,05045E-01	8,00E+06				
134	72473	19100 M	o		0,05045E-01	8,00E+06				
135	72473	19100 M	o		0,05045E-01	8,00E+06				
136	72473	19100 M	o		0,05045E-01	8,00E+06				
137	72473	19100 M	o		0,05045E-01	8,00E+06				
138	72473	19100 M	o		0,05045E-01	8,00E+06				
139	72473	19100 M	o		0,05045E-01	8,00E+06				
140	72473	19100 M	o		0,05045E-01	8,00E+06				
141	72473	19100 M	o		0,05045E-01	8,00E+06				
142	72473	19100 M	o		0,05045E-01	8,00E+06				
143	72473	19100 M	o		0,05045E-01	8,00E+06				
144	72473	19100 M	o		0,05045E-01	8,00E+06				
145	72473	19100 M	o		0,05045E-01	8,00E+06				
146	72473	19100 M	o		0,05045E-01	8,00E+06				
147	72473	19100 M	o		0,05045E-01	8,00E+06				
148	72473	19100 M	o		0,05045E-01	8,00E+06				
149	72473	19100 M	o		0,05045E-01	8,00E+06				
150	72473	19100 M	o		0,05045E-01	8,00E+06				
151	72473	19100 M	o		0,05045E-01	8,00E+06				
152	72473	19100 M	o		0,05045E-01	8,00E+06				
153	72473	19100 M	o		0,05045E-01	8,00E+06				
154	72473	19100 M	o		0,05045E-01	8,00E+06				
155	72473	19100 M	o		0,05045E-01	8,00E+06				
156	72473	19100 M	o		0,05045E-01	8,00E+06				
157	72473	19100 M	o		0,05045E-01	8,00E+06				
158	72473	19100 M	o		0,05045E-01	8,00E+06				
159	72473	19100 M	o		0,05045E-01	8,00E+06				
160	72473	19100 M	o		0,05045E-01	8,00E+06				
161	72473	19100 M	o		0,05045E-01	8,00E+06				
162	72473	19100 M	o		0,05045E-01	8,00E+06				
163	72473	19100 M	o		0,05045E-01	8,00E+06				
164	72473	19100 M	o		0,05045E-01	8,00E+06				
165	72473	19100 M	o		0,05045E-01	8,00E+06				
166	72473	19100 M	o		0,05045E-01	8,00E+06				
167	72473	19100 M	o		0,05045E-01	8,00E+06				
168	72473	19100 M	o		0,05045E-01	8,00E+				

Price	I	J	K	L	M	N	O
	QTY	Discount	Shipping	Co.	Total Revenue	Category Name	Payment Method
F	2.10E+05	5	0.00E+00	0	10,500.00	Beauty & Groom COD	
E	4.00E+05	1	0.00E+00	0	4,000.00	Beauty & Groom COD	
B	2.00E+05	1	0.00E+00	0	1,200.00	Sephora	
R	3.97E+05	1	0.00E+00	0	3,963.00	Beauty & Groom COD	
B	2.70E+05	2	0.00E+00	0	540.00	Superstore	COD
F	2.00E+05	1	0.00E+00	0	200.00	Superstore	COD
J	7.99E+05	1	0.00E+00	0	7,990.00	Men's Fashion	COD
B	8.95E+05	1	0.00E+00	0	8,990.00	Women's Fashic COD	
Z	4.49E+05	1	0.00E+00	0	4,490.00	Women's Fashic COD	
G	3.70E+05	1	0.00E+00	0	3,700.00	Beauty & Groom COD	
H	3.00E+05	2	0.00E+00	0	1,500.00	Beauty & Groom COD	
D	9.00E+05	2	0.00E+00	0	1,800.00	Sephora	COD
B	6.00E+05	2	0.00E+00	0	1,800.00	Sephora	COD
B	2.20E+05	1	0.00E+00	0	2,200.00	Sephora	COD
F	4.80E+05	1	0.00E+00	0	4,800.00	Sephora	COD
B	9.00E+05	1	0.00E+00	0	900.00	Sephora	COD
J	1.70E+05	1	0.00E+00	0	1,700.00	Sephora	COD
B	2.20E+05	1	0.00E+00	0	2,200.00	Sephora	COD
H	6.00E+05	2	0.00E+00	0	1,800.00	Sephora	COD
D	6.00E+05	1	0.00E+00	0	1,800.00	Sephora	COD
B	2.10E+05	2	0.00E+00	0	1,800.00	Beauty & Groom	COD
B	5.95E+05	1	5.95E+05	0	5,391.00	Men's Fashion	COD
A	4.45E+05	1	4.45E+05	0	4,005.00	Beauty & Groom COD	
Z	5.48E+05	1	0.00E+00	0	5,480.00	Men's Fashion	COD
C	3.95E+05	1	0.00E+00	0	3,990.00	Men's Fashion	COD
B	1.20E+05	1	0.00E+00	0	1,200.00	Men's Fashion	COD
S	5.49E+05	1	0.00E+00	0	5,490.00	Entertainment	COD
B	8.00E+05	1	8.00E+05	0	7,200.00	Beauty & Groom Transfer Bank	
S	3.00E+05	1	3.00E+05	0	2,700.00	Home & Living	Transfer Bank
B	5.45E+05	1	5.45E+05	0	6,800.00	Beauty & Groom Transfer Bank	
B	1.95E+05	1	0.00E+00	0	1,590.00	Beauty & Groom COD	
A	2.95E+05	1	0.00E+00	0	2,990.00	Women's Fashic Credit Card	
B	5.53E+05	2	0.00E+00	0	11,060.00	Beauty & Groom Virtual Account	
S	3.05E+05	1	0.00E+00	0	3,050.00	Beauty & Groom Virtual Account	

SKU	Product Name	Price	GJ	TY	Discount	Shipping Co	Total Row	Category Na	Payment Me
LAUQXAG99F4E3D5CF	LAUQXAG99F4E3D5CF	2,100.00	5	0	0	0	10,500.00	Bailey & Groom	Credit Card
SOGMSAB511E1330B4E	SOGMSAB511E1330B4E	4,200.00	1	0	0	0	4,200.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	1,200.00	1	0	0	0	1,200.00	Soghat	Credit Card
BAGARASG04929549525	BAGARASG04929549525	3,963.300	1	0	0	0	3,963.300	Bailey & Groom	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	2,700.00	2	0	0	0	5,400.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	2,700.00	1	0	0	0	2,700.00	Soghat	Credit Card
METHEFS510716A9C1-1	METHEFS510716A9C1-1	7,990.00	1	0	0	0	7,990.00	Men's Fashion	Credit Card
WODHFA54527D16A9C1-6	WODHFA54527D16A9C1-6	8,990.00	1	0	0	0	8,990.00	Women's Fashion	Credit Card
WODHFA54527D16A9C1-9	WODHFA54527D16A9C1-9	4,490.00	1	0	0	0	4,490.00	Women's Fashion	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	1,200.00	1	0	0	0	1,200.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	900.00	2	0	0	0	1,800.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	900.00	2	0	0	0	1,800.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	900.00	1	0	0	0	1,800.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	1,200.00	1	0	0	0	2,200.00	Soghat	Credit Card
SOGMSAB549B237A2B	SOGMSAB549B237A2B	1,400.00	1	0	0	0	2,800.00	Soghat	Credit Card
SOGNSND54574A8B23	SOGNSND54574A8B23	900.00	1	0	0	0	900.00	Soghat	Credit Card
SOGNSND54574A8B23	SOGNSND54574A8B23	1,700.00	1	0	0	0	1,700.00	Soghat	Credit Card
SOGNSND54574A8B23	SOGNSND54574A8B23	2,200.00	1	0	0	0	2,200.00	Soghat	Credit Card
SOGNSND54574A8B23	SOGNSND54574A8B23	900.00	2	0	0	0	1,800.00	Soghat	Credit Card
SOGNSND54574A8B23	SOGNSND54574A8B23	900.00	1	0	0	0	1,800.00	Soghat	Credit Card
BAGLUXA99F4E3D5CF	BAGLUXA99F4E3D5CF	2,100.00	1	0	0	0	1,890.00	Bailey & Groom	Credit Card
METHEFS510716A9C1-6	METHEFS510716A9C1-6	5,990.00	1	0	0	0	5,310.00	Men's Fashion	Credit Card
BAGEMAK394301018E	BAGEMAK394301018E	1,450.00	1	0	0	0	4,005.00	Bailey & Groom	Credit Card
BAGEMAK394301018E	BAGEMAK394301018E	1,450.00	1	0	0	0	3,600.00	Men's Fashion	Credit Card
MEDEEASCA5190BEE88A	MEDEEASCA5190BEE88A	3,900.00	1	0	0	0	3,900.00	Men's Fashion	Credit Card
MEDEEASCA5190BEE88A	MEDEEASCA5190BEE88A	1,200.00	1	0	0	0	1,200.00	Men's Fashion	Credit Card
ENTVEY59FA8B3A9AA	ENTVEY59FA8B3A9AA	5,490.00	1	0	0	0	5,490.00	Entertainment	Credit Card
MEADEMAB9A8162204	MEADEMAB9A8162204	8,000.00	1	0	0	0	7,200.00	Bailey & Groom	Transfer Bank
MEADEMAB9A8162204	MEADEMAB9A8162204	300.00	1	0	0	0	7,200.00	Bailey & Groom	Transfer Bank
APRNEK540C4151MF	APRNEK540C4151MF	5,740.00	1	0	0	0	6,065.00	Appliance	Transfer Bank
BAGNDSY05029205673	BAGNDSY05029205673	1,900.00	1	0	0	0	1,990.00	Bailey & Groom	Transfer Bank
WOFUNPM443AC4329E	WOFUNPM443AC4329E	2,990.00	1	0	0	0	2,990.00	Women's Fashion	Credit Card
BAGMAS9C27129A1	BAGMAS9C27129A1	3,500.00	2	0	0	0	3,000.00	Bailey & Groom	Virtual Account
BAGMAS9C27129A1	BAGMAS9C27129A1	3,500.00	1	0	0	0	3,000.00	Bailey & Groom	Virtual Account
BAGMAS9C45F616B03	BAGMAS9C45F616B03	4,080.00	2	0	0	0	8,160.00	Bailey & Groom	Virtual Account
BAGMAS9C45F616B03	BAGMAS9C45F616B03	3,060.00	1	0	0	0	3,060.00	Bailey & Groom	Virtual Account
WOFDEAS23AF6D3A73	WOFDEAS23AF6D3A73	3,990.00	1	0	0	0	3,990.00	Women's Fashion	Instant Payment
WOFDEAS23AF6D3A73	WOFDEAS23AF6D3A73	9,000.00	1	0	0	0	8,990.00	Women's Fashion	Instant Payment
WEFDEAS23AF6D3A73	WEFDEAS23AF6D3A73	4,000.00	1	0	0	0	4,000.00	Women's Fashion	Instant Payment

Changed data type for variable **price** and **discount** to matches with **revenue** data type.

Changed from **Scientific** to **Number**
It also can be changed to currency, but my case
will be number for easier analysis.

Remove Duplicate Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Transaction ID	Product ID	Customer ID	Custom	Filter views	Campaign	SKU	Price	QTY	Discount	Shipping Cost	Total Revenue	Category Name	Payment Method
2	ID-2	724241	724074	4	Sort sheet	10-10 Campaign	10/10/BAGLUXSA099F	2.10+0€	5	0.00+0€	0	10.500,00	Beauty & Groom	CCD
3	ID-2	724074	724075	4	Sort range	10-10 Campaign	10/10/SOGAMBSA311	4.20+0€	1	0.00+0€	0	4.200,00	Soghaat	CCD
4	ID-3	724075	724077	4	Create a filter	10-10 Campaign	10/10/SOGAMBSA098	1.20+0€	1	0.00+0€	0	1.200,00	Soghaat	CCD
5	ID-4	724077	724077	4	Add a slicer	10-10 Campaign	10/10/BAGARASAA9	3.97+0€	1	0.00+0€	0	3.968,00	Beauty & Groom	CCD
6	ID-5	724077	724078	4	Protect sheets and ranges	10-10 Campaign	10/10/KNO5904DAF	2.70+0€	2	0.00+0€	0	5.400,00	Superstore	CCD
7	ID-6	724078	724078	4	Named ranges	10-10 Campaign	10/10/MEFHESAA59BD	2.00+0€	1	0.00+0€	0	2.000,00	Superstore	CCD
8	ID-7	724078	724079	4	Named functions	10-10 Campaign	10/10/WCHFASAA567	7.95+0€	1	1.00+0€	0	7.950,00	Women's Fashion	CCD
9	ID-8	724079	724442	4	New	10-10 Campaign	10/10/WDFEA59CE	8.95+0€	1	1.00+0€	0	8.950,00	Women's Fashion	CCD
10	ID-9	724442	724443	4	Randomize range	10-10 Campaign	10/10/BAGDOVSA099	4.49+0€	1	0.00+0€	0	4.490,00	Women's Fashion	CCD
11	ID-10	724062	724062	4	Filter views	10-10 Campaign	10/10/SOGAMBSA26	9.00+0€	2	0.00+0€	0	18.000,00	Soghaat	CCD
12	ID-11	724936	724936	4	Column stats	10-10 Campaign	10/10/SOGAMBSA26E	9.00+0€	2	0.00+0€	0	18.000,00	Soghaat	CCD
13	ID-12	724897	724897	4	Data validation	10-10 Campaign	10/10/SOGAMBSA26E	9.00+0€	2	0.00+0€	0	18.000,00	Soghaat	CCD
14	ID-13	724898	724898	4	Data cleanup	10-10 Campaign	10/10/SOGAMBSA26E	9.00+0€	2	0.00+0€	0	18.000,00	Soghaat	CCD
15	ID-14	724899	724901	4	New	Cleanup suggestions		2.20+0€	1	0.00+0€	0	2.200,00	Soghaat	CCD
16	ID-15	724901	724901	4	Data cleanup	Remove duplicates		4.80+0€	1	0.00+0€	0	4.800,00	Soghaat	CCD
17	ID-16	724902	724902	4	Data extraction	Trim whitespace		9.00+0€	1	0.00+0€	0	9.000,00	Soghaat	CCD
18	ID-17	724902	724903	4	Data extraction	Remove duplicates		1.70+0€	1	0.00+0€	0	1.700,00	Soghaat	CCD
19	ID-18	724903	724903	4	Data connectors	Trim whitespace		2.20+0€	1	0.00+0€	0	2.200,00	Soghaat	CCD
20	ID-19	724903	724904	4	New	Remove duplicates		9.00+0€	2	0.00+0€	0	18.000,00	Soghaat	CCD
21	ID-20	724904	724738	4	Data connectors	Remove duplicates		9.00+0€	1	0.00+0€	0	9.000,00	Soghaat	CCD
22	ID-21	724738	724665	4	New	Remove duplicates		9.00+0€	1	0.00+0€	0	9.000,00	Soghaat	CCD
23	ID-22	724665	17884 M	complete	10-10 Campaign	10/10/BAGLUXSA099	2.10+0€	1	2.10+0€	0	1.890,00	Beauty & Groom	CCD	
24	ID-23	724665	17884 F	complete	2023-10-10 Campaign	10/10/MEFHESAA59FD	5.98+0€	1	5.98+0€	0	5.391,00	Men's Fashion	CCD	
25	ID-24	724665	18967 M	complete	2023-10-10 Campaign	10/10/BAGKEMASAA30A	4.45+0€	1	4.45+0€	0	4.005,00	Beauty & Groom	CCD	
26	ID-25	724665	18967 M	complete	2023-10-10 Campaign	10/10/MEFDDEASAA51E	5.48+0€	1	0.00+0€	0	5.480,00	Men's Fashion	CCD	
27	ID-26	724665	18967 M	complete	2023-10-10 Campaign	10/10/MEFDDEASAA44C	3.95+0€	1	0.00+0€	0	3.950,00	Men's Fashion	CCD	
28	ID-27	724703	19100 M	complete	2023-10-10 Campaign	10/10/MEFDDEASAA51C	2.00+0€	1	0.00+0€	0	1.200,00	Men's Fashion	CCD	
29	ID-28	724513	19168 F	complete	2023-10-10 Campaign	10/10/BAGEMASAA59FB	5.49+0€	1	0.00+0€	0	5.490,00	Entertainment	CCD	
30	ID-29	724513	19168 F	complete	2023-10-10 Campaign	10/10/BAGEMASAA59FB	8.00+0€	1	8.00+0€	0	7.200,00	Beauty & Groom Transfer Bank	CCD	

Using built-in feature of Google Sheets, we can easily remove duplicates from dataset.

5 Duplicate rows found and removed.

Remove Irrelevant Data

A	B	C	D	F	G	I	J	K	L	M	N
Transaction ID	Product ID	Customer ID	Gender	Created At	Campaign	Price	QTY	Discount	Total Revenue	Category Name	Payment Meth
ID-1	724241	21 F		2023-10-10	Campaign 10/10	2,100,000	5	0	10,500,000	Beauty & Grooming	COD
ID-2	724074	168 M		2023-10-10	Campaign 10/10	4,200,000	1	0	4,200,000	Soghaat	COD
ID-3	724075	168 F		2023-10-10	Campaign 10/10	1,200,000	1	0	1,200,000	Soghaat	COD
ID-4	724076	168 F		2023-10-10	Campaign 10/10	3,968,300	1	0	3,968,300	Beauty & Grooming	COD
ID-5	724077	168 M		2023-10-10	Campaign 10/10	270,000	2	0	540,000	Superstore	COD
ID-6	724078	168 F		2023-10-10	Campaign 10/10	200,000	1	0	200,000	Superstore	COD
ID-7	724578	6715 M		2023-10-10	Campaign 10/10	7,990,000	1	0	7,990,000	Men's Fashion	COD
ID-8	724442	8616 F		2023-10-10	Campaign 10/10	8,990,000	1	0	8,990,000	Women's Fashion	COD
ID-9	724443	8616 F		2023-10-10	Campaign 10/10	4,490,000	1	0	4,490,000	Women's Fashion	COD
ID-10	724062	8765 F		2023-10-10	Campaign 10/10	3,700,000	1	0	3,700,000	Beauty & Grooming	COD
ID-11	724896	15960 M		2023-10-10	Campaign 10/10	900,000	2	0	1,800,000	Soghaat	COD
ID-12	724897	15960 F		2023-10-10	Campaign 10/10	900,000	2	0	1,800,000	Soghaat	COD
ID-13	724898	15960 M		2023-10-10	Campaign 10/10	900,000	2	0	1,800,000	Soghaat	COD
ID-14	724899	15960 F		2023-10-10	Campaign 10/10	2,200,000	1	0	2,200,000	Soghaat	COD
ID-15	724900	15960 M		2023-10-10	Campaign 10/10	4,800,000	1	0	4,800,000	Soghaat	COD
ID-16	724901	15960 F		2023-10-10	Campaign 10/10	900,000	1	0	900,000	Soghaat	COD
ID-17	724902	15960 M		2023-10-10	Campaign 10/10	1,700,000	1	0	1,700,000	Soghaat	COD
ID-18	724903	15960 F		2023-10-10	Campaign 10/10	2,200,000	1	0	2,200,000	Soghaat	COD
ID-19	724904	15960 M		2023-10-10	Campaign 10/10	900,000	2	0	1,800,000	Soghaat	COD
ID-20	724905	15960 F		2023-10-10	Campaign 10/10	900,000	1	0	900,000	Soghaat	COD
ID-21	724738	16593 F		2023-10-10	Campaign 10/10	2,100,000	1	210,000	1,890,000	Beauty & Grooming	COD
ID-22	724266	17684 M		2023-10-10	Campaign 10/10	5,990,000	1	599,000	5,391,000	Men's Fashion	COD
ID-23	724268	17684 F		2023-10-10	Campaign 10/10	4,450,000	1	445,000	4,005,000	Beauty & Grooming	COD
ID-24	724665	18967 M		2023-10-10	Campaign 10/10	5,480,000	1	0	5,480,000	Men's Fashion	COD
ID-25	724667	18967 M		2023-10-10	Campaign 10/10	3,990,000	1	0	3,990,000	Men's Fashion	COD
ID-26	724669	18967 M		2023-10-10	Campaign 10/10	1,200,000	1	0	1,200,000	Men's Fashion	COD
ID-27	724073	19100 M		2023-10-10	Campaign 10/10	5,490,000	1	0	5,490,000	Entertainment	COD
ID-28	724513	19168 F		2023-10-10	Campaign 10/10	8,000,000	1	800,000	7,200,000	Beauty & Grooming	Transfer Bank
ID-29	724516	19168 F		2023-10-10	Campaign 10/10	3,000,000	1	300,000	2,700,000	Home & Living	Transfer Bank
ID-30	724517	19168 M		2023-10-10	Campaign 10/10	6,740,000	1	674,000	6,066,000	Appliances	Transfer Bank
ID-31	724271	19977 F		2023-10-10	Campaign 10/10	1,990,000	1	0	1,990,000	Beauty & Grooming	COD
ID-32	724739	33830 F		2023-10-10	Campaign 10/10	2,990,000	1	0	2,990,000	Women's Fashion	Credit Card
ID-33	724189	35535 F		2023-10-10	Campaign 10/10	5,530,000	2	0	11,060,000	Beauty & Grooming	Virtual Account
ID-34	724190	35535 F		2023-10-10	Campaign 10/10	3,060,000	1	0	3,060,000	Beauty & Grooming	Virtual Account
ID-35	724191	35535 F		2023-10-10	Campaign 10/10	4,080,000	2	0	8,160,000	Beauty & Grooming	Virtual Account
ID-36	724192	35535 F		2023-10-10	Campaign 10/10	3,060,000	1	0	3,060,000	Beauty & Grooming	Virtual Account
ID-37	724287	39593 F		2023-10-10	Campaign 10/10	3,990,000	1	0	3,990,000	Women's Fashion	Instant Payment
ID-38	724291	39593 F		2023-10-10	Campaign 10/10	8,990,000	1	0	8,990,000	Women's Fashion	Instant Payment
ID-39	724292	39593 M		2023-10-10	Campaign 10/10	4,490,000	1	0	4,490,000	Men's Fashion	Instant Payment
ID-40	724946	41173 M		2023-10-10	Campaign 10/10	6,990,000	1	0	6,990,000	Men's Fashion	COD
ID-41	724042	41651 F		2023-10-10	Campaign 10/10	4,100,000	1	0	4,100,000	Beauty & Grooming	COD
ID-42	724043	41651 F		2023-10-10	Campaign 10/10	5,490,000	1	0	5,490,000	Beauty & Grooming	COD
ID-43	724044	41651 F		2023-10-10	Campaign 10/10	1,160,000	1	0	1,160,000	Home & Living	COD
ID-44	724561	42046 F		2023-10-10	Campaign 10/10	4,000,000	1	0	4,000,000	Beauty & Grooming	COD
ID-45	724562	42046 F		2023-10-10	Campaign 10/10	8,000,000	1	0	8,000,000	Beauty & Grooming	COD
ID-46	724835	42753 F		2023-10-10	Campaign 10/10	1,620,000	2	0	3,240,000	Beauty & Grooming	Instant Payment

Deleted **status** and **SKU** column as it is not relevant for analysis. Note that **status** column has **>50%** values missing so it will not included in the analysis.

Shipping cost also deleted because all of the shipping cost entry is **0**, meaning all items in campaigns are free-shipping.

Remove/Impute Missing Data

[W2W3]_Satryo_Sunu_Prabowo_Intermediate

	A	B	C	D	E	F	G	H	I	J
1	Transaction	Product ID	Customer ID	Gender	Status	Created At	Campaign	SKU	Price	QTY
2	ID-1	724074	21 F	complete	2023-10-10 Campaign					
3	ID-2	724074	168 M	complete	2023-10-10 Campaign	Sort A to Z				
4	ID-3	724075	168 F	complete	2023-10-10 Campaign	Sort Z to A				
5	ID-4	724076	168 F	complete	2023-10-10 Campaign	Sort by color				
6	ID-5	724077	168 M	complete	2023-10-10 Campaign					
7	ID-6	724078	168 F	complete	2023-10-10 Campaign					
8	ID-7	724578	6715 M	complete	2023-10-10 Campaign					
9	ID-8	724442	8616 F	complete	2023-10-10 Campaign	Filter by color				
10	ID-9	724443	8616 F	complete	2023-10-10 Campaign	Filter by condition				
11	ID-10	724062	8765 F	complete	2023-10-10 Campaign	Filter by values				
12	ID-11	724896	15960 M	complete	2023-10-10 Campaign	Select all/Clear				
13	ID-12	724897	15960 F	complete	2023-10-10 Campaign					
14	ID-13	724898	15960 M	complete	2023-10-10 Campaign					
15	ID-14	724898	15960 F	complete	2023-10-10 Campaign					
16	ID-15	724900	15960 M	complete	2023-10-10 Campaign					
17	ID-16	724901	15960 F	complete	2023-10-10 Campaign	✓ (Blanks)				
18	ID-17	724902	15960 M	complete	2023-10-10 Campaign	✓ 10,000				
19	ID-18	724903	15960 F	complete	2023-10-10 Campaign	✓ 50,000				
20	ID-19	724904	15960 M	complete	2023-10-10 Campaign	✓ 100,000				
21	ID-20	724905	15960 F	complete	2023-10-10 Campaign					
22	ID-21	724738	16593 F	complete	2023-10-10 Campaign					
23	ID-22	724266	17684 M	complete	2023-10-10 Campaign					
24	ID-23	724268	17684 F	complete	2023-10-10 Campaign					
25	ID-24	724665	18967 M	complete	2023-10-10 Campaign					
26	ID-25	724667	18967 M	complete	2023-10-10 Campaign					

Apply filter to search blank entries for each variable.

There are some blank datas such as **price**, **quantity** and **discounts**. Each has 2 blank entries.

Since there are variables related to price, quantity and discount, we can search for and input empty data so that it matches the data.

G	H	I	J	K	L	M	N	O
Campaign	SKU	Price	QTY	Discount	Shipping Co	Total Revenue	Category Name	Payment Method
Campaign 10/10 HALBUK5A013F87DB79F		1,760,000		0	0	1,760,000	Home & Living	Transfer Bank
Campaign 12/12 MEFMOV5B566C133E662		2,900,000		0	0	8,700,000	Men's Fashion	COD

G	H	I	J	K	L	M	N	O
Campaign	SKU	Price	QTY	Discount	Shipping Co	Total Revenue	Category Name	Payment Method
Campaign 10/10 HALBUK5A013F87DB79F		1,760,000	1	0	0	1,760,000	Home & Living	Transfer Bank
Campaign 12/12 MEFMOV5B566C133E662		2,900,000	3	0	0	8,700,000	Men's Fashion	COD

Impute data for quantity

G	H	I	J	K	L	M	N	O
Campaign	SKU	Price	QTY	Discount	Shipping Co	Total Revenue	Category Name	Payment Method
Campaign 10/10 WOFDEA59CE2178B2F7E		1,990,000	1			0	1,791,000	Women's Fashic COD
Campaign 12/12 SNO5A685DB296774		5,700,000	1			0	5,130,000	Superstore Transfer Bank

F	G	H	I	J	K	L	M	N	O
Created At	Campaign	SKU	Price	QTY	Discount	Shipping Co	Total Revenue	Category Name	Payment Method
2023-10-10	Campaign 10/10 WOFDEA59CE2178B2F7E		1,990,000	1	199,000		0	1,791,000	Women's Fashic COD
2023-12-12	Campaign 12/12 SNO5A685DB296774		5,700,000	1	570,000		0	5,130,000	Superstore Transfer Bank

Impute data for discount

G	H	I	J	K	L	M	N	O
Campaign	SKU	Price	QTY	Discount	Shipping Co	Total Revenue	Category Name	Payment Method
Campaign 10/10 SADPAR59F6C59E02C06				1	0	0	2,500,000	School & Educa COD
Campaign 12/12 KABMUM5A581B8EC67A				2	0	0	1,800,000	Kids & Baby COD

G	H	I	J	K	L	M	N	O
Campaign	SKU	Price	QTY	Discount	Shipping Co	Total Revenue	Category Name	Payment Method
Campaign 10/10 SADPAR59F6C59E02C06		2,500,000	1	0	0	2,500,000	School & Educa COD	
Campaign 12/12 KABMUM5A581B8EC67A		900,000	2	0	0	1,800,000	Kids & Baby COD	

Impute data for price



Descriptive Analytics

Remove Outlier

Using IQR, we can determine the **Lower Boundary (LB)** and **Upper Boundary (UB)** for column “Total Revenue”.

Formula for IQR = Q3 - Q1

Formula for LB = Q1 - (1.5*IQR)

Formula for UB = Q3 + (1.5*IQR)

The results for LB and UB are **-4,635,000** and **13,965,000** respectively. In the context of data, it doesn't make sense for the data to be minus therefore the lower boundary here will be **0**, and upper boundary here is **13,965,000**

By using conditional filter, we are able to remove the outliers from the dataset.

Descriptive analysis		How to find outlier using IQR	
Count	13544	IQR	4,650,000
Min	10,000	Lower Boundary	-4,635,000
Max	83,556,000	Upper Boundary	13,965,000
Mean	4,647,235		
Median	4,440,000		
Mode	7,200,000		
Q1	2,340,000		
Q3	6,990,000		
Range	83,546,000		
Variance	1007456563040		
Std Deviation	3174045.625		
Skeweness	6.487090612		

Sort A to Z

Sort Z to A

Sort by color 

Filter by color 

▼ Filter by condition

Greater than or equal to 

13965000

▶ Filter by values

	A	B	C	D	E	F	G	H	I	J	K	L
1	Transaction	Product ID	Customer ID	Gender	Created At	Campaign	Price	QTY	Discount	Total Revenue	Category Name	Payment Meth
274	ID-273	726096	53228 M		2023-10-10	Campaign 10/10	Rp7,840,000	10	Rp0	Rp78,400,000	Mobiles & Tablets	COD
305	ID-304	728166	84717 M		2023-10-10	Campaign 10/10	Rp6,990,000	10	Rp0	Rp69,900,000	Entertainment	COD
329	ID-328	727308	1476 F		2023-10-10	Campaign 10/10	Rp8,000,000	8	Rp0	Rp64,000,000	Beauty & Grooming	COD
333	ID-332	727450	8463 F		2023-10-10	Campaign 10/10	Rp2,100,000	24	Rp210,000	Rp51,190,000	Beauty & Grooming	Instant Payment
336	ID-346	727044	9351 F		2023-10-10	Campaign 10/10	Rp2,100,000	23	Rp210,000	Rp48,990,000	Beauty & Grooming	Virtual Account
343	ID-353	727086	9351 F		2023-10-10	Campaign 10/10	Rp2,100,000	8	Rp210,000	Rp16,590,000	Beauty & Grooming	Virtual Account
380	ID-370	726773	26136 F		2023-10-10	Campaign 10/10	Rp7,500,000	10	Rp750,000	Rp74,250,000	Superstore	Virtual Account
381	ID-371	727001	26136 M		2023-10-10	Campaign 10/10	Rp7,500,000	10	Rp750,000	Rp74,250,000	Superstore	Virtual Account
419	ID-426	726584	67807 F		2023-10-10	Campaign 10/10	Rp2,100,000	10	Rp210,000	Rp21,000,000	Beauty & Grooming	COD
441	ID-451	726822	83655 F		2023-10-10	Campaign 10/10	Rp6,440,000	10	Rp644,000	Rp63,556,000	Superstore	Instant Payment
818	ID-838	729004	67807 F		2023-10-10	Campaign 10/10	Rp2,100,000	31	Rp0	Rp65,100,000	Beauty & Grooming	COD
11436	ID-11849	824255	104745 M		2023-12-12	Campaign 12/12	Rp6,980,000	10	Rp0	Rp69,800,000	Men's Fashion	COD

Add more rows at the bottom

Statistical Measurement

Insights from the table:

Mean: The mean value is highest in column 11.11. This means that the average revenue per transaction in Campaign 11.11 was higher than in the other campaigns.

Range & Variance: 12.12 exhibits a higher range and variance indicating more spread or dispersion in the data compared to others.

Standard Deviation: Related with range & variance; standard deviation is also highest in 12.12. This means that the values in Campaign 12.12 were more likely to deviate from the average than in the other campaigns.

	10.10	11.11	12.12
Count	4449	4694	4389
Min	Rp180,000	Rp45,000	Rp10,000
Max	Rp13,800,000	Rp13,775,000	Rp13,940,000
Mean	Rp4,637,205	Rp4,735,207	Rp4,413,087
Median	Rp4,250,000	Rp4,896,500	Rp4,041,000
Mode	Rp7,200,000	Rp7,200,000	Rp6,990,000
Q1	Rp2,490,000	Rp2,442,500	Rp2,120,000
Q3	Rp6,990,000	Rp7,155,000	Rp6,300,000
Range	Rp13,620,000	Rp13,730,000	Rp13,930,000
Variance	6856007122479	6914377370097	7249553325750
Std Deviation	2618397.816	2629520.369	2692499.457
Skewness	0.3664557725	0.1505389849	0.5577261085



Exploratory Data Analysis (EDA)

Overall Campaign Performance

	Campaign 10/10	Campaign 11/11	Campaign 12/12
Total Revenue	Rp21,276,249,940	Rp22,227,061,190	Rp19,438,838,930
Total Campaign Budget	Rp375,125,660	Rp420,053,610	Rp309,990,670
# of Transaction	4460	4694	4390
# of Customer	2042	2098	1887
# of Products sold	5683	5700	5673
ROI	56.7176608	52.91482006	62.70781934

- Total Revenue:** Campaign 11/11 generated the highest revenue of **Rp22,227,061,190** followed by Campaign 10/10 and then Campaign 12/12. This means that **Campaign 11/11** was the **most successful in terms of generating income** from the sales events.
- Total Campaign Budget:** The budget was highest for **Campaign 11/11** at **Rp420,053,610**. This means that Campaign 11/11 invested the most money in marketing and promotion for the sales events. Despite having the **lowest budget** of **Rp309,990,670**, **Campaign 12/12** managed to generate significant revenue, indicating a **high return on investment (ROI) (62.7)**.
- Number of Transactions:** Campaign 11/11 attracted the highest number of transactions (4694), indicating a higher level of customer engagement during this campaign.

Campaign 11/11 was the **most effective in generating revenue, attracting customers, and increasing transactions**, but it also had the **highest budget**. Campaign 10/10 was the second best in terms of revenue and transactions, but it had a lower budget and customer base. **Campaign 12/12** was the **least successful** in terms of **revenue, transactions, and customers**, but it had the **lowest budget and the highest ROI**.

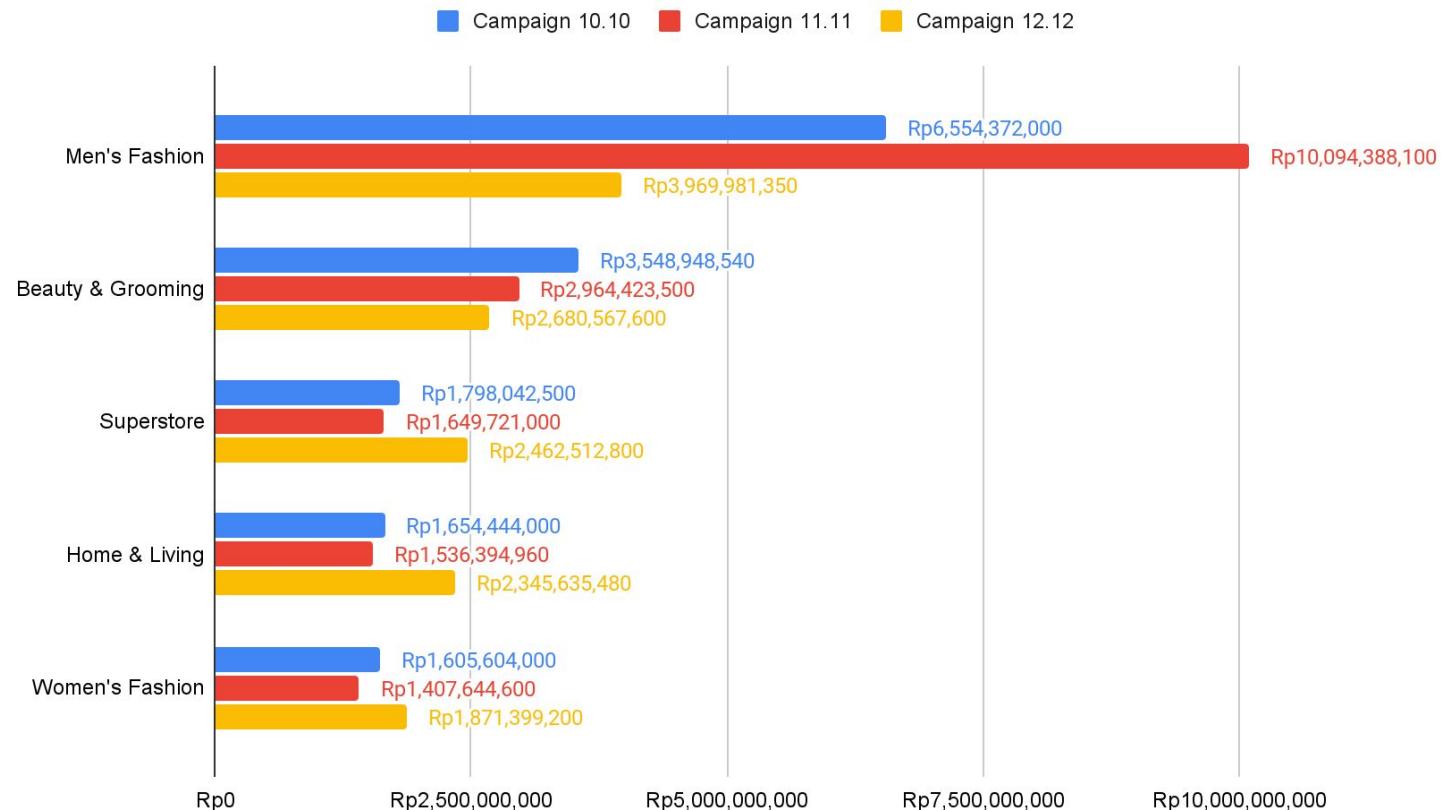
Best Seller Category

Overall Trends:

- Men's Fashion consistently drives substantial revenue** across all campaigns, demonstrating its importance in the product mix.
 - While Men's Fashion consistently generates high revenue, it also sells a large volume of items, indicating both popularity and efficiency in sales.
- Beauty & Grooming, Superstore, Health & Sports, and Women's Fashion** also maintain their positions as **significant contributors to revenue**.
 - Beauty & Grooming and Superstore demonstrate the importance of volume sales, even if their revenue contribution percentages are fluctuate.
- Campaign 12/12 has lower overall revenue** compared to previous campaigns. Men's Fashion remains the top earner but with a decrease in both revenue share and average order value. **Superstore drops further** compared with other categories like **Health & Sports, Women's Fashion and Beauty & Grooming**.
- Entertainment, School & Education, and Books** maintain the lowest revenue share and average order value, along with the lowest number of items sold per order.

Campaign	Category Name	Total Revenue	% of Revenue	Item Sold	Average Order Value
■ Campaign 10/10	Men's Fashion	Rp6,554,372,000	10.41%	1115	Rp5,878,361
	Beauty & Grooming	Rp3,548,948,540	5.64%	716	Rp4,956,632
	Superstore	Rp1,798,042,500	2.86%	426	Rp4,220,757
	Home & Living	Rp1,654,444,000	2.63%	420	Rp3,939,152
	Women's Fashion	Rp1,605,604,000	2.55%	267	Rp6,013,498
	Health & Sports	Rp1,393,861,350	2.21%	372	Rp3,746,939
	Mobiles & Tablets	Rp1,280,253,400	2.03%	315	Rp4,064,297
	Soghaat	Rp1,235,833,600	1.96%	355	Rp3,481,221
	Appliances	Rp762,733,000	1.21%	114	Rp6,690,640
	Kids & Baby	Rp724,648,750	1.15%	209	Rp3,467,219
	Computing	Rp362,712,000	0.58%	71	Rp5,108,620
	School & Education	Rp206,204,800	0.33%	64	Rp3,221,950
	Entertainment	Rp128,532,000	0.20%	10	Rp12,853,200
	Others	Rp12,500,000	0.02%	2	Rp6,250,000
	Books	Rp7,560,000	0.01%	4	Rp1,890,000
Campaign 10/10 Total		Rp21,276,249,940	33.80%	4460	Rp4,770,460
■ Campaign 11/11	Men's Fashion	Rp10,094,388,100	16.04%	1693	Rp5,962,427
	Superstore	Rp2,964,423,500	4.71%	817	Rp3,628,425
	Health & Sports	Rp1,649,721,000	2.62%	470	Rp3,510,045
	Beauty & Grooming	Rp1,536,394,960	2.44%	366	Rp4,197,800
	Women's Fashion	Rp1,407,644,600	2.24%	242	Rp5,816,713
	Home & Living	Rp1,262,908,600	2.01%	300	Rp4,209,695
	Appliances	Rp951,568,000	1.51%	146	Rp6,517,589
	Mobiles & Tablets	Rp896,112,500	1.42%	229	Rp3,913,155
	Kids & Baby	Rp535,135,930	0.85%	188	Rp2,846,468
	Soghaat	Rp495,385,000	0.79%	146	Rp3,393,048
	Computing	Rp248,940,000	0.40%	57	Rp4,367,368
	Entertainment	Rp87,590,000	0.14%	17	Rp5,152,353
	School & Education	Rp65,211,000	0.10%	17	Rp3,835,941
	Books	Rp31,638,000	0.05%	6	Rp5,273,000
Campaign 11/11 Total		Rp22,227,061,190	35.31%	4694	Rp4,735,207
■ Campaign 12/12	Men's Fashion	Rp3,969,981,350	6.31%	708	Rp5,607,318
	Health & Sports	Rp2,680,567,600	4.26%	787	Rp3,406,058
	Women's Fashion	Rp2,462,512,800	3.91%	448	Rp5,496,680
	Beauty & Grooming	Rp2,345,635,480	3.73%	542	Rp4,327,741
	Superstore	Rp1,871,399,200	2.97%	499	Rp3,750,299
	Home & Living	Rp1,647,443,400	2.62%	366	Rp4,501,211
	Mobiles & Tablets	Rp1,299,339,500	2.06%	322	Rp4,035,216
	Appliances	Rp952,159,000	1.51%	140	Rp6,801,136
	Soghaat	Rp721,804,400	1.15%	244	Rp2,958,215
	Kids & Baby	Rp554,334,700	0.88%	135	Rp4,106,183
	Computing	Rp526,549,500	0.84%	95	Rp5,542,626
	Entertainment	Rp195,327,000	0.31%	36	Rp5,425,750
	School & Education	Rp120,509,000	0.19%	51	Rp2,362,922
	Books	Rp91,276,000	0.15%	17	Rp5,369,176
Campaign 12/12 Total		Rp19,438,838,930	30.88%	4390	Rp4,427,982
Grand Total		Rp62,942,150,060	100.00%	13544	Rp4,647,235

Top 5 Product Categories for each Campaign



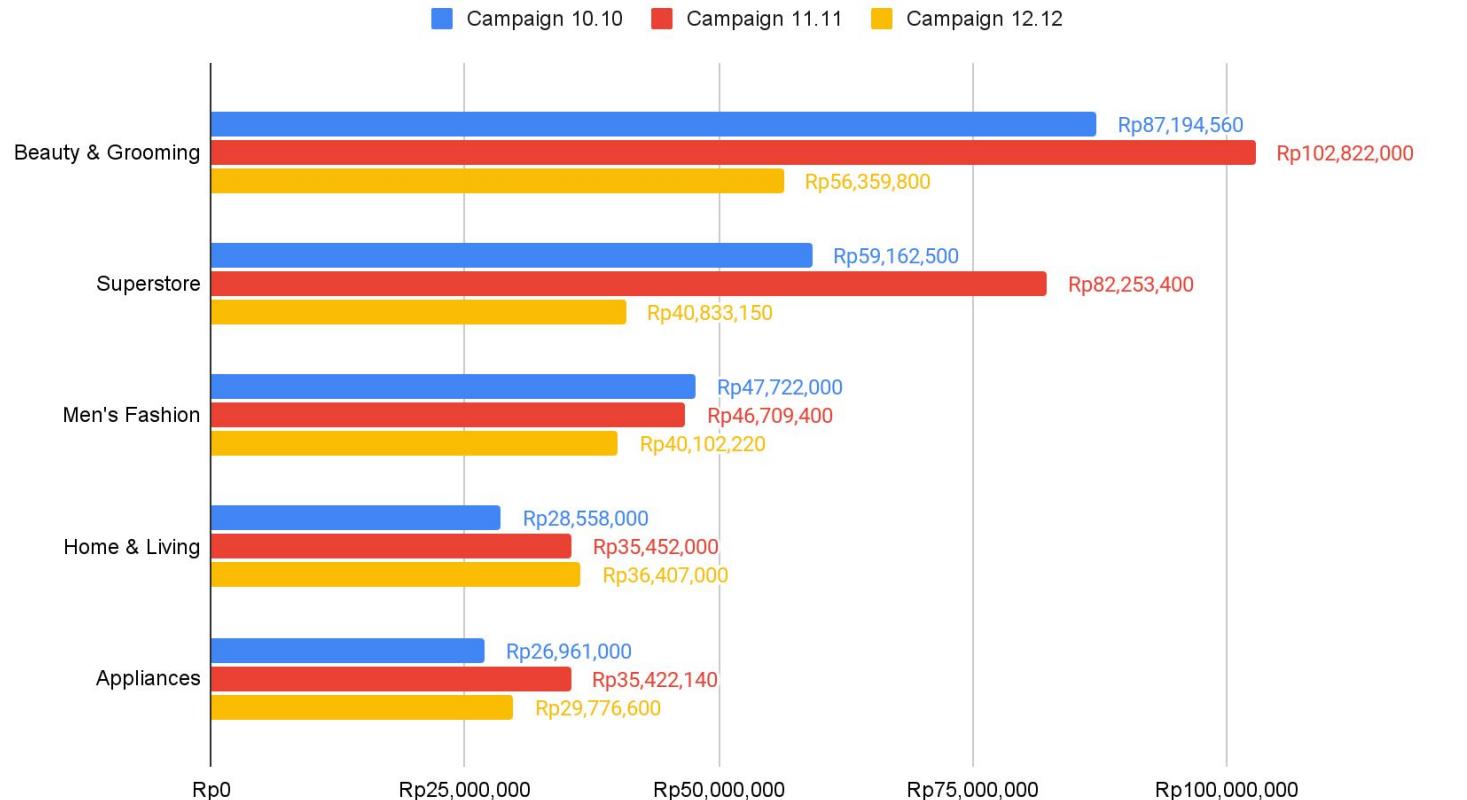
Highest discount

Overall Trends:

- **Superstore** consistently receives the **highest total discount across all campaigns**, suggesting a strategy of offering larger discounts to incentivise purchases in this category.
- **Appliances** also see **significant discounts**, but with a much lower average discount per item compared to Superstore. This implies larger discounts on fewer, high-value items.
- **Beauty & Grooming and Health & Sports** maintain consistent discounting practices across campaigns. Their average discount per item is also moderate, suggesting a balanced approach to attracting customers without sacrificing profit margins.
- **Entertainment and School & Education** consistently receive the lowest total and average discounts. This may indicate a focus on other promotional strategies for these categories.
- **Average discount per item varies greatly** across categories. **Appliances** often have **high average discounts**, while categories like **Computing and Entertainment** have **lower average discounts**.

Campaign	Category Name	Total Discount	% of Discount	Average Discount Given
Campaign 10/10	Beauty & Grooming	Rp87,194,560	7.89%	Rp121,780
	Superstore	Rp59,162,500	5.35%	Rp138,879
	Men's Fashion	Rp47,722,000	4.32%	Rp42,800
	Home & Living	Rp28,558,000	2.58%	Rp67,995
	Appliances	Rp26,961,000	2.44%	Rp236,500
	Health & Sports	Rp26,703,650	2.42%	Rp71,784
	Soghaat	Rp26,060,400	2.36%	Rp73,410
	Women's Fashion	Rp23,656,000	2.14%	Rp88,599
	Mobiles & Tablets	Rp19,951,600	1.81%	Rp63,338
	Kids & Baby	Rp15,941,750	1.44%	Rp76,276
	Computing	Rp7,788,000	0.70%	Rp109,690
	School & Education	Rp3,878,200	0.35%	Rp60,597
	Entertainment	Rp1,548,000	0.14%	Rp154,800
	Others	Rp0	0.00%	Rp0
	Books	Rp0	0.00%	Rp0
Campaign 10/10 Total		Rp375,125,660	33.94%	Rp84,109
Campaign 11/11	Superstore	Rp102,822,000	9.30%	Rp125,853
	Men's Fashion	Rp82,253,400	7.44%	Rp48,584
	Women's Fashion	Rp46,709,400	4.23%	Rp193,014
	Appliances	Rp35,452,000	3.21%	Rp242,822
	Beauty & Grooming	Rp35,422,140	3.21%	Rp96,782
	Home & Living	Rp30,343,400	2.75%	Rp101,145
	Health & Sports	Rp25,569,000	2.31%	Rp54,402
	Kids & Baby	Rp21,663,270	1.96%	Rp115,230
	Mobiles & Tablets	Rp16,873,000	1.53%	Rp73,681
	Soghaat	Rp11,145,000	1.01%	Rp76,336
	Computing	Rp7,500,000	0.68%	Rp131,579
	Entertainment	Rp2,060,000	0.19%	Rp121,176
	School & Education	Rp1,817,000	0.16%	Rp106,882
	Books	Rp424,000	0.04%	Rp70,667
Campaign 11/11 Total		Rp420,053,610	38.01%	Rp89,487
Campaign 12/12	Superstore	Rp56,359,800	5.10%	Rp112,945
	Men's Fashion	Rp40,833,150	3.69%	Rp57,674
	Beauty & Grooming	Rp40,102,220	3.63%	Rp73,989
	Women's Fashion	Rp36,407,000	3.29%	Rp81,266
	Home & Living	Rp29,776,600	2.69%	Rp81,357
	Health & Sports	Rp25,260,400	2.29%	Rp32,097
	Appliances	Rp21,311,000	1.93%	Rp152,221
	Mobiles & Tablets	Rp17,680,000	1.60%	Rp54,907
	Soghaat	Rp17,255,600	1.56%	Rp70,720
	Kids & Baby	Rp9,985,900	0.90%	Rp73,970
	Computing	Rp8,070,500	0.73%	Rp84,953
	School & Education	Rp4,196,500	0.38%	Rp82,284
	Entertainment	Rp2,133,000	0.19%	Rp59,250
	Books	Rp619,000	0.06%	Rp36,412
Campaign 12/12 Total		Rp309,990,670	28.05%	Rp70,613
Grand Total		Rp1,105,169,940	100.00%	Rp81,598

Top 5 Product Discount for each Campaign



Recommendations

- Focus marketing efforts on **Men's Fashion** to capitalize on its **consistent performance** and **high average order value**.
- Further analysis of **consumer behavior, competitor strategies**, and **external factors** impacting sales can provide **valuable insights** for **future campaign planning and decision-making**.
 - Analyze product **subcategories** within each main category to identify **specific high-performing products** or areas for improvement
- Analyzing the **number of items sold per order** can provide insights into **customer purchase behavior** and **potential upselling/cross-selling opportunities**.
 - For example, categories like **Mobiles & Tablets** and **Appliances** have **higher numbers of items sold per order**, suggesting potential for bundling accessories or related products.
- While **discounts** can attract **customers and drive sales**, it's essential to ensure that discounting strategies are **sustainable and contribute positively** to overall profitability.
 - Discounts can be used to **clear out old inventory, attract new customers, or compete with rival offerings**. Understanding the specific goals behind discounts in each category can provide valuable insights.



Hypothesis Testing

We also want to know how significant is the difference between these 3 campaigns in terms of total revenue. We can determine this by testing the hypothesis see if we have a total revenue difference in between campaign. Let's start by comparing **Campaign 10.10** with **Campaign 11.11**!

Hypothesis:

H_0 : Average Total Revenue 10.10 = Average Total Revenue 11.11

H_1 : Average Total Revenue 10.10 \neq Average Total Revenue 11.11

Alpha(α): 0.05

Statistical Test: T-test, population variance unknown

Conclusion:

No statistical difference. Based on the table , T-test showed no statistical difference between campaign 10.10 and 11.11 as **p-value = 0.07** is higher compared alpha (0.05).

We can infer that we can **fail to reject null hypothesis.**

There is insufficient evidence to conclude that one campaign is significantly better than other.

While mean Campaign 11.11 is slightly higher than Campaign 10.10 (**4,735,206.9** vs **4,637,204.75**), the difference is slim and not statistically difference. It would require further testing, investigation and/or analysis to check if this two campaign has a statistical difference.

	Campaign 10.10	Campaign 11.11
Mean	4,637,204.75	4,735,206.90
Variance	6,856,007,122,478.96	6,914,377,370,096.91
Observations	4449	4694
Hypothesized Mean Difference	0	
df	9118	
t Stat	-1.785092042	
P(T<=t) one-tail	0.037139813	
t Critical one-tail	1.6450207	
P(T<=t) two-tail	0.074279627	
t Critical two-tail	1.960224138	

Now we try to compare the significance of **Campaign 10.10** with **Campaign 12.12**!

Hypothesis:

H_0 : Average Total Revenue 10.10 = Average Total Revenue 12.12

H_1 : Average Total Revenue 10.10 \neq Average Total Revenue 12.12

Alpha(α): 0.05

Statistical Test: T-test, population variance unknown

Conclusion:

There is statistical difference. Based on the table , T-test showed no statistical difference between campaign 10.10 and 12.12 as **p-value = 0** is higher compared alpha (0.05).

We can conclude that we can **reject null hypothesis and accept alternative hypothesis**. There is sufficient evidence to conclude that one campaign is significantly better than other.

We see that the mean of Campaign 10.10 is significantly higher compared Campaign 12.12 (**4,637,204.75** vs **4,413,087.02**). That means that **Campaign 10.10** makes a **significant difference in terms of gaining revenue** compared to **Campaign 12.12**.

	Campaign 10.10	Campaign 12.12
Mean	4,637,204.75	4,413,087.02
Variance	6,856,007,122,478.96	7,249,553,325,749.67
Observations	4449	4389
Hypothesized Mean Difference	0	
df	8820	
t Stat	3.966358011	
P(T<=t) one-tail	0	
t Critical one-tail	1.645026348	
P(T<=t) two-tail	0	
t Critical two-tail	1.960232931	

Lastly, we now compare the significance of **Campaign 11.11** with **Campaign 12.12!**

Hypothesis:

H_0 : Average Total Revenue 11.11 = Average Total Revenue 12.12

H_1 : Average Total Revenue 11.11 \neq Average Total Revenue 12.12

Alpha(α): 0.05

Statistical Test: T-test, population variance unknown

Conclusion:

There is statistical difference. Based on the table , T-test showed no statistical difference between campaign 11.11 and 12.12 as **p-value = 0** is higher compared alpha (0.05).

We can conclude that we can **reject null hypothesis and accept alternative hypothesis.** There is sufficient evidence to conclude that one campaign is significantly better than other.

We see that the mean of Campaign 11.11 is significantly higher compared Campaign 12.12 (**4,735,206.9** vs **4,413,087.02**). That means that **Campaign 11.11** makes a **significant difference in terms of gaining revenue** compared to **Campaign 12.12**.

	Campaign 11.11	Campaign 12.12
Mean	4,735,206.90	4,413,087.02
Variance	6,914,377,370,096.91	7,249,553,325,749.67
Observations	4694	4389
Hypothesized Mean Difference	0	
df	9006	
t Stat	5.762458696	
P(T<=t) one-tail	0	
t Critical one-tail	1.645022779	
P(T<=t) two-tail	0	
t Critical two-tail	1.960227374	

Hypothesis Testing Insight

Based on the result of hypothesis, here are the insights that we got:

- Between **Campaign 10.10** and **Campaign 11.11**, **there is no statistical difference** between these two campaigns. Both campaigns are similarly performing in terms of average total revenue earned, however if we compare the mean and variance between campaigns the Campaign 10.10 is higher in terms of mean and lower in variance.
- **There is statistical difference** between **Campaign 10.10 to Campaign 12.12**, and **Campaign 11.11 to Campaign 12.12**. Both campaign 10.10 and campaign 11.11 outperforms campaign 12.12 in terms of average total revenue earned. Consider the magnitude of the effect and its practical implications for your business before making decisions.



Revenue Forecasting

Chapter 6

Regression Models Overdispersed Count Data

3
11
2
11

Identify all potential sources of the research problem.

To optimize the amount of spending from each customer in TokoBli E-commerce, the product manager team wants to find out what factors have a significant effect on the amount of spending from customers.

As a Data Analyst, you are asked to find these factors which will later be used to develop campaign and business strategies to increase the amount spent by customers.

Suppose the response y is a count variable, but unlike in the Poisson model, we consider four models that are extensions of the Poisson model: negative binomial, zero-inflated Poisson, zero-inflated negative binomial, and multinomial models.

According to the data set, which is shown [here](#). There are some variables that we need to analyze. The variables consist of **Customerid, Gender, Age, Income, Tenure, Avg. Session Time, Total Promo, Bounce Rate, and Yearly Spending**. Here is the sample of the data from the data set.

Customer Id	Gender	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate	Yearly Spending
100	M	30	12,000,000	10	520	12	56	4,000,000
101	F	30	8,600,000	7	744	9.46	57	4,300,000
102	M	30	7,200,000	3	114	22	40	4,800,000
103	F	30	30,000,000	9	169	18	37	5,500,000

To understand the data, among these variables we need to figure out which variables have a strong relationship with each other.

Correlation

	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate
Age						
Income	0.1407744338					
Tenure	0.1603475875	0.3789467125				
Avg. Session Time	0.1934452704	0.4836920601	0.6853319563			
Total Promo	0.207803283	0.4456727121	0.6213744703	0.7768861588		
Bounce Rate	-0.1289753817	-0.4483756005	-0.5702766089	-0.7873056093	-0.7941434615	

This is the table of correlation between each variables. Based on the table of correlation coefficients you provided, here are some insights:

- There is a **strong positive correlation** between **Average Session Time** and **Total Promo** (0.77688641588), indicating that as the session time increases, the total promo also tends to increase.
- **Tenure** and **Average Session Time** have a **moderate positive correlation** of 0.6853319563, suggesting that longer customer tenure means deep engagement, highlighting higher session time on the app.
- **Bounce Rate** has a **strong negative correlation** with **Total Promo** (-0.7873056093) and **Avg. Session Time** (-0.7873056093), suggesting that higher bounce rates are associated with lower total promos and shorter session times.

By using tools such as XLminer, we are able to utilize linear regression to model output of the analysis. Here is the the details of the analysis explained:

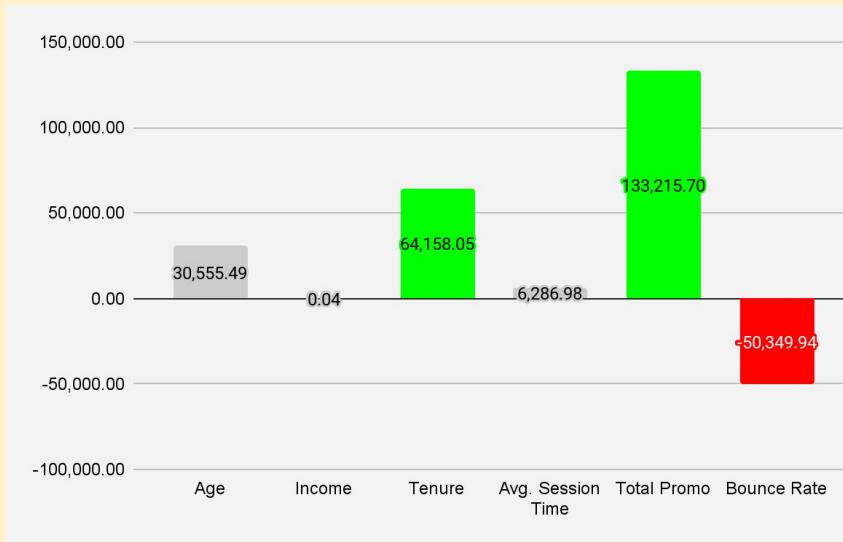
Regression Statistics	
Multiple R	0.9639907184
R Square	0.9292781052
Adjusted R Square	0.9278249156
Standard Error	1384406.116
Observations	299

Of 299 observation (samples), the table explains that **R Square (R^2): 0.92 means 92% of sales yearly spending data** can be explained by 6 variables provided (see later slides). Multiple R here explains that all of independent variables here has strong correlation to dependent variable.

ANOVA					
	df	SS	MS	F	Significance F
Regression	6	7,353,628,520,547,650.00	1,225,604,753.42	4,610.00	639.4747754
Residual	292	559641446007541	1916580294546		
Total	298	7,913,269,966,555,190.00			0

In addition, ANOVA table here shows that value for simultaneous test (**Significance F**) is **0**. When we compare this to alpha value (0.05), this implies that overall variables has a **significant effect**, meaning that **6 variables together have a significant coefficient** in explaining yearly spending.

While also using tools XLminer, we are able to determine the coefficient that affects the dependent variable (Yearly Spending) from other independent variables.



	Coefficients	P-value
Intercept	1,947,879.71	0
Age	30,555.49	0.06
Income	0.04	0
Tenure	64,158.05	0
Avg. Session Time	6,286.98	0
Total Promo	133,215.70	0
Bounce Rate	-50,349.94	0

Based on this table and graph, we can conclude that **Tenure** and **Total Promo** are variables that has most impact on Yearly Spending, especially **Total Promo**, which has coefficient of **133,215.7**. That means if total promo increases by 1, it will contribute 133,215.7 to yearly spending.

Note that P-value on variable **Age** is 0.06, which is higher than alpha value 0.05. That means that the variable age will not make a significance if we include it in the model, so we will discard the variable age on our multiple regression model.

Age Variable deleted

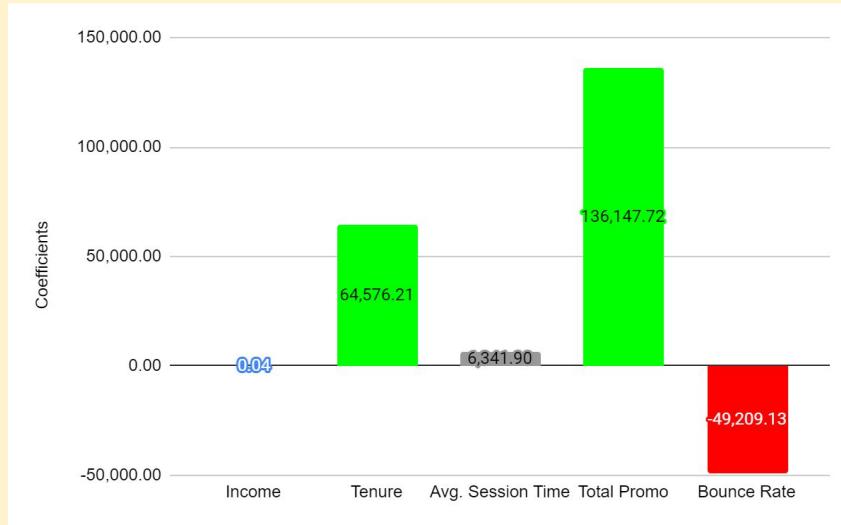
After we **delete the age variable**, we will create another linear regression that excludes variable age. The regression statistics here shows that the R square value is still similar compared to first regression statistics.

Regression Statistics	
Multiple R	0.9635715414
R Square	0.9284701154
Adjusted R Square	0.9272494689
Standard Error	1389914.044
Observations	299

ANOVA					
	df	SS	MS	F	Significance F
Regression	5	7,347,234,679,034,450.00	1,469,446,935,80	6,890.00	760.6380056
Residual	293	566,035,287,520,739.00	1931861049559		0
Total	298	7,913,269,966,555,190.00			

Anova table Significance F is still 0, which means the all of the variables here has a significance towards our regression model.

Following table shows the coefficient table and also P-value after deletion of age variable.



	Coefficients	P-value
Intercept	2,590,241.49	0
Income	0.04	0
Tenure	64,576.21	0
Avg. Session Time	6,341.90	0
Total Promo	136,147.72	0
Bounce Rate	-49,209.13	0

There are some differences in terms of coefficient values. Noticeable difference is the **Intercept** value is increased to **2,590,241.49** from **1,947,879.71**. Every variables here are slightly increased in values with exception of Income variable which stays at 0.04.

All of the P-values here are showing 0, so we can proceed to the multiple linear regression model.

Utilizing multiple linear regression model, we are able to derive the following predictive formula for yearly spending (dependent variable) based on other variables (independent variable):

$$Y(\text{yearly spending}) = 2,590,241.49 + 0.04*x1(\text{income}) + 64,576.21*x2(\text{tenure}) + 6,341.90*x3(\text{avg. session time}) + 136,147.72*x4(\text{total promo}) - 49,209.13*x5(\text{bounce rate})$$

We also want to predict potential yearly spending based on the following data:

Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate
20	20,000,000	1	1000	20	30

Therefore, we can use the formula above to predict potential yearly spending from this data:

	Intercept	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate
Value		20	20,000,000	1	1000	20	30
Coefficient	2,590,241.49	0.00	0.04	64,576.21	6,341.90	136,147.72	-49,209.13
Coefficient x Value	1,947,879.71	0.00	716,492.81	64,576.21	6,341,901.71	2,722,954.47	-1,476,273.79
Results	10,317,531.12						

Potential Yearly Spending for this new customer is **Rp10,317,531**

Regression Assumption Check

1. Non-multicollinearity

	Age	Income	Tenure	Avg. Session Time	Total Promo	Bounce Rate
Age						
Income	0.1407744338					
Tenure	0.1603475875	0.3789467125				
Avg. Session Time	0.1934452704	0.4836920601	0.6853319563			
Total Promo	0.207803283	0.4456727121	0.6213744703	0.7768861588		
Bounce Rate	-0.1289753817	-0.4483756005	-0.5702766089	-0.7873056093	-0.79414346	15

As we see on this table, **Age** and **Income** have **low correlation** between each other (<0.5). Kindly note there are variables that >0.5 such as **Tenure**, **Avg. Session Time**, and **Bounce Rate** as high correlation is subject to multicollinearity. **Critical value** for multicollinearity is **>0.8** and since the highest value here is 0.79, we keep it on the regression model.

For values are above >0.5 need to be further analyzed using VIF method after regression to see if there is potential multicollinearity between each variables.

Regression Statistics	
Multiple R	0.9635715414
R Square	0.9284701154
Adjusted R Square	0.9272494689
Standard Error	1389914.044
Observations	299

To determine VIF, we can calculate using this formula:

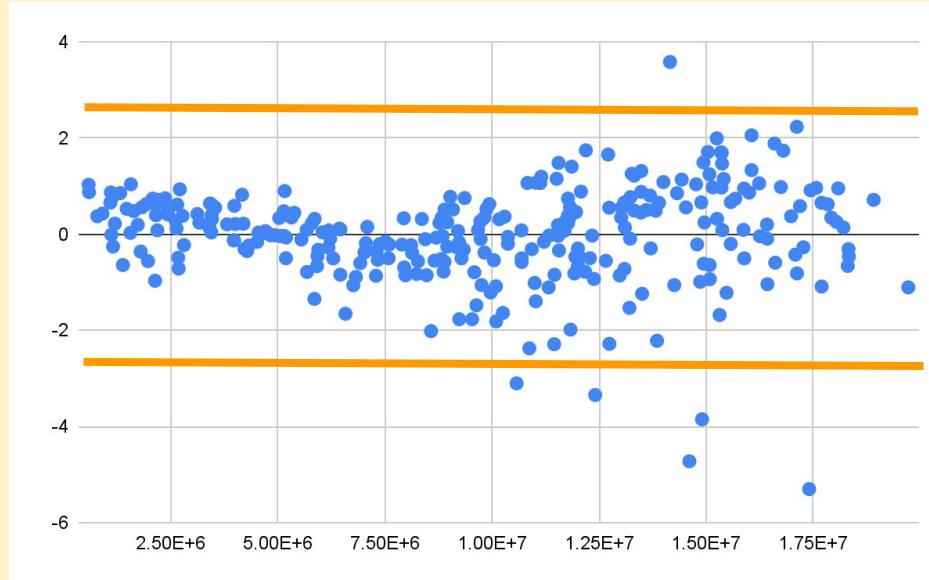
$$VIF_i = \frac{1}{1 - R_i^2}$$

Where R square can be obtained from the table above. If we calculate using the formula, the VIF value is **13.98**, which there is high possibility for multicollinearity as the **tolerance of VIF is <5**.

A value greater than 5 indicates potentially severe correlation between a given explanatory variable and other explanatory variables in the model. In this case, the coefficient estimates and p-values in the regression output are likely unreliable. The solution would be only take the variables that have correlation value **<0.5** rather than **<0.8**.

Regression Assumption Check

2. Homoscedasticity



The residual plots between standard residual and yearly spending are **not really following** a constant pattern for homoscedasticity. Therefore, scatter plot for determining for homoscedasticity is inconclusive.

Further statistical test such as **Bartlett's Test**, **Box M Test**, and/or **Levene Test** are needed to determine if the regression model is following homoscedasticity criteria.

Regression Assumption Check

3. Non-autocorrelation

As this is cross-section data, hence the data doesn't have non-correlation properties.

4. Normality

The Q-Q Plot shows that residual distribution relatively follows the z-score distribution. This can be detected via trendline as shown in the graph below. The closer pattern to trendline, most likely it will follow the normality assumption



Recommendation

Regardless of both analysis, we would like to give some recommendation for increasing yearly spending to increase our revenue:

Total Promo

- **Optimize promo targeting:** Analyze which customer segments benefit most from promotions and tailor offers accordingly. Use income, purchase history, and engagement data for segmentation.
- **Focus on personalized discounts:** Offer relevant discounts on products customers are likely to buy, not generic store-wide promotions. Utilize product recommendations and browsing history for personalization. This could include personalized discounts, exclusive offers, or bundled products.
- **Value-added promotions:** Offer promotions that go beyond discounts, such as free shipping, extended warranties, or bundled services. This can increase perceived value and encourage higher spending.

Customer Loyalty

- **Implement a loyalty program:** Offer exclusive benefits and rewards to loyal customers to incentivize continued engagement and foster loyalty.
- **Exceptional customer service:** Provide prompt, helpful, and personalized support to build positive relationships with customers.

Bounce Rate

- **Address User Experience:** Improve the overall user experience by enhancing website or app usability, speed, and functionality to reduce bounce rates.
- **Optimize Product Pages:** Implement strategies to increase user engagement and interactivity, such as interactive features, multimedia content, feature high-quality product images, detailed descriptions and clear calls-to-action.
- **A/B test different elements:** Test different layouts, headlines, and CTAs to see what resonates best with your audience and reduces bounce rate.

Dataset & Alternate Analysis

As the regression model above isn't quite reliable, attached link leads to a sheet for analysis which using correlation value <0.5 instead <0.8 .

The regression model is using only 2 variables that has correlation value of <0.5 , which are **Age and Income**.

We determine if we using only those two variables, VIF values is **1.4** which fulfills assumption of non-multicollinearity.

We also predicted the Yearly Spending for new member using other regression model, the result is **Rp9,454,235**.

The details for analysis can be checked by visiting following link directly to the sheet:

[Regression Analysis Using Correlation Value \$<0.5\$](#)

[Spreadsheet 1 - EDA](#)

[Spreadsheet 2 - Regression](#)



Thank You!

Let's have a friendly discussion.



Satryo Sunu
Prabowo