

Analisis Pengaruh Kampanye Marketing terhadap Pembukaan Akun Deposito

EDA, Analisis Korelasi, Chi-Square, dan Regresi Linier

Hijir Della Wirasti
Business Intelligence
Batch 13
14 Desember 2024



<https://github.com/hijirdella/Bank-Campaign-EDA>
linkedin.com/in/hijirdella/



Important Resources



Google Collab (Code) https://colab.research.google.com/drive/1ABc4G0Dv_nvuV3Fkw4ghEFvspqwbjhla?usp=sharing

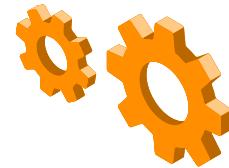
Dataset <https://www.kaggle.com/datasets/krantiswalke/bankfullcsv>

Github <https://github.com/hijirdella/Bank-Campaign-EDA>

LinkedIn linkedin.com/in/hijirdella/

Email hijirdw@gmail.com

Table of contents



01 Objectives

02 Data Profiling

03 EDA - Pivot Table & Data Visualization

04 Analisis Korelasi, Chi-Square, & Regresi Linier

01

Objectives



Objectives



Mission

Analisis ini bertujuan untuk memahami pengaruh jumlah **kampanye (campaign)** terhadap **pembukaan akun deposit (y)** menggunakan dataset bank marketing.



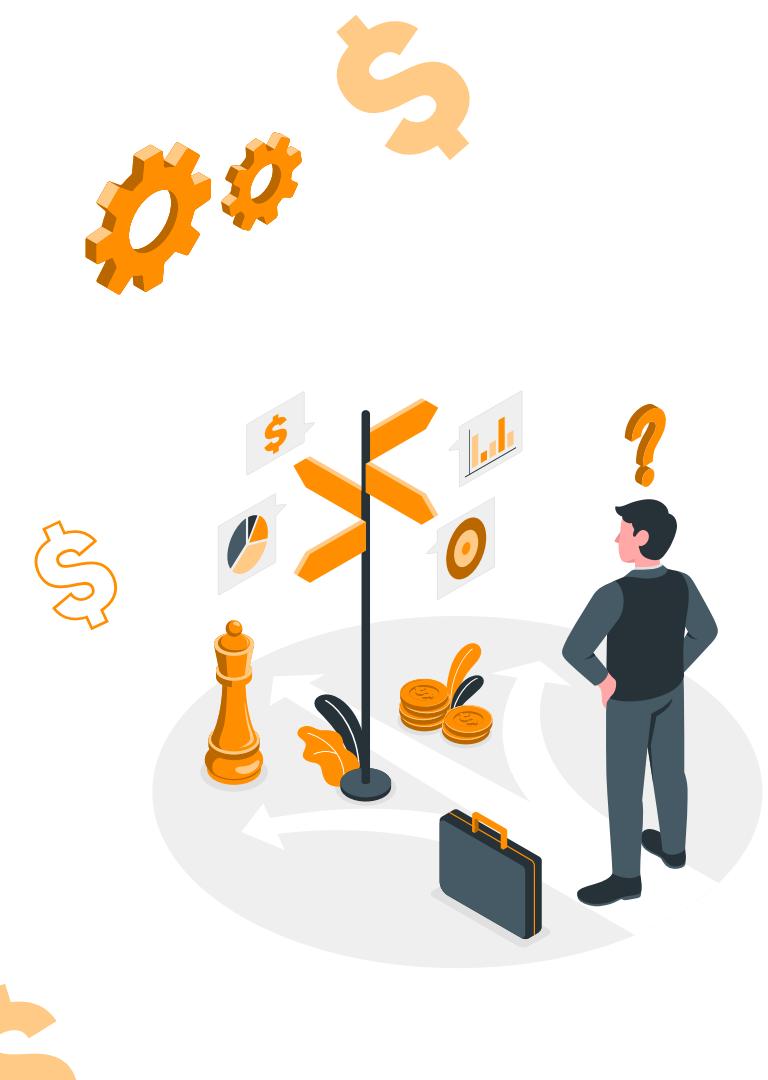
Objectives

1. Melakukan Data Profiling
2. Melakukan EDA untuk memahami pola data melalui pivot table dan visualisasi.
3. Melakukan analisis statistik (korelasi, uji chi-square, regresi linier).
4. Memberikan insight dan rekomendasi berdasarkan hasil analisis.



02

Data Profiling



Data Profiling

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45211 entries, 0 to 45210
Data columns (total 17 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   age      45211 non-null   int64  
 1   job      45211 non-null   object 
 2   marital  45211 non-null   object 
 3   education    45211 non-null   object 
 4   default   45211 non-null   object 
 5   balance   45211 non-null   int64  
 6   housing   45211 non-null   object 
 7   loan      45211 non-null   object 
 8   contact   45211 non-null   object 
 9   day       45211 non-null   int64  
 10  month     45211 non-null   object 
 11  duration  45211 non-null   int64  
 12  campaign  45211 non-null   int64  
 13  pdays     45211 non-null   int64  
 14  previous  45211 non-null   int64  
 15  poutcome  45211 non-null   object 
 16  y         45211 non-null   object 
dtypes: int64(7), object(10)
memory usage: 5.9+ MB
```

		Column	Data Type	Non-Null Count	Unique Values
0		age	int64	45211	77
1		job	object	45211	12
2		marital	object	45211	3
3		education	object	45211	4
4		default	object	45211	2
5		balance	int64	45211	7168
6		housing	object	45211	2
7		loan	object	45211	2
8		contact	object	45211	3
9		day	int64	45211	31
10		month	object	45211	12
11		duration	int64	45211	1573
12		campaign	int64	45211	48
13		pdays	int64	45211	559
14		previous	int64	45211	41
15		poutcome	object	45211	4
16		y	object	45211	2

Sample Values

0	[58, 44, 33, 47, 35]
1	[management, technician, entrepreneur, blue-co...]
2	[married, single, divorced]
3	[tertiary, secondary, unknown, primary]
4	[no, yes]
5	[2143, 29, 2, 1506, 1]
6	[yes, no]
7	[no, yes]
8	[unknown, cellular, telephone]
9	[5, 6, 7, 8, 9]
10	[may, jun, jul, aug, oct]
11	[261, 151, 76, 92, 198]
12	[1, 2, 3, 5, 4]
13	[-1, 151, 166, 91, 86]
14	[0, 3, 1, 4, 2]
15	[unknown, failure, other, success]
16	[no, yes]

Categorical Data

	job	marital	education	default	housing	loan	contact	month	poutcome	y
count	45211	45211	45211	45211	45211	45211	45211	45211	45211	45211
unique	12	3	4	2	2	2	3	12	4	2
top	blue-collar	married	secondary	no	yes	no	cellular	may	unknown	no
freq	9732	27214	23202	44396	25130	37967	29285	13766	36959	39922

Numerical Data

	age	balance	day	duration	campaign	pdays	previous
count	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000
mean	40.936210	1362.272058	15.806419	258.163080	2.763841	40.197828	0.580323
std	10.618762	3044.765829	8.322476	257.527812	3.098021	100.128746	2.303441
min	18.000000	-8019.000000	1.000000	0.000000	1.000000	-1.000000	0.000000
25%	33.000000	72.000000	8.000000	103.000000	1.000000	-1.000000	0.000000
50%	39.000000	448.000000	16.000000	180.000000	2.000000	-1.000000	0.000000
75%	48.000000	1428.000000	21.000000	319.000000	3.000000	-1.000000	0.000000
max	95.000000	102127.000000	31.000000	4918.000000	63.000000	871.000000	275.000000

Ukuran Dataset: 45,211 baris, 17 kolom, tanpa missing values, dan tanpa duplicate

Numerik: `balance` (mean: €1,362), `campaign` (mean: 2.76 kontak), `duration` (mean: 258 detik).

Kategorikal: `job` (12 kategori), `marital` (3 kategori), `y` (hasil kampanye: yes/no).

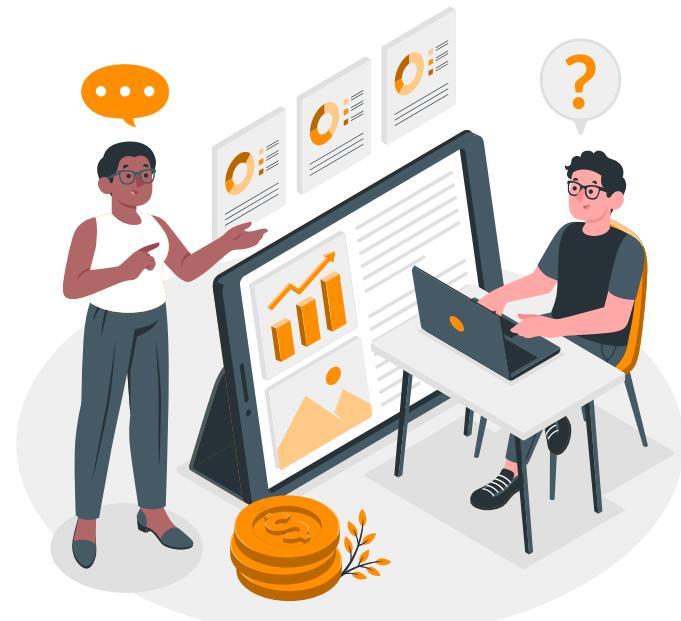
3

EDA - Pivot Table & Data Visualization

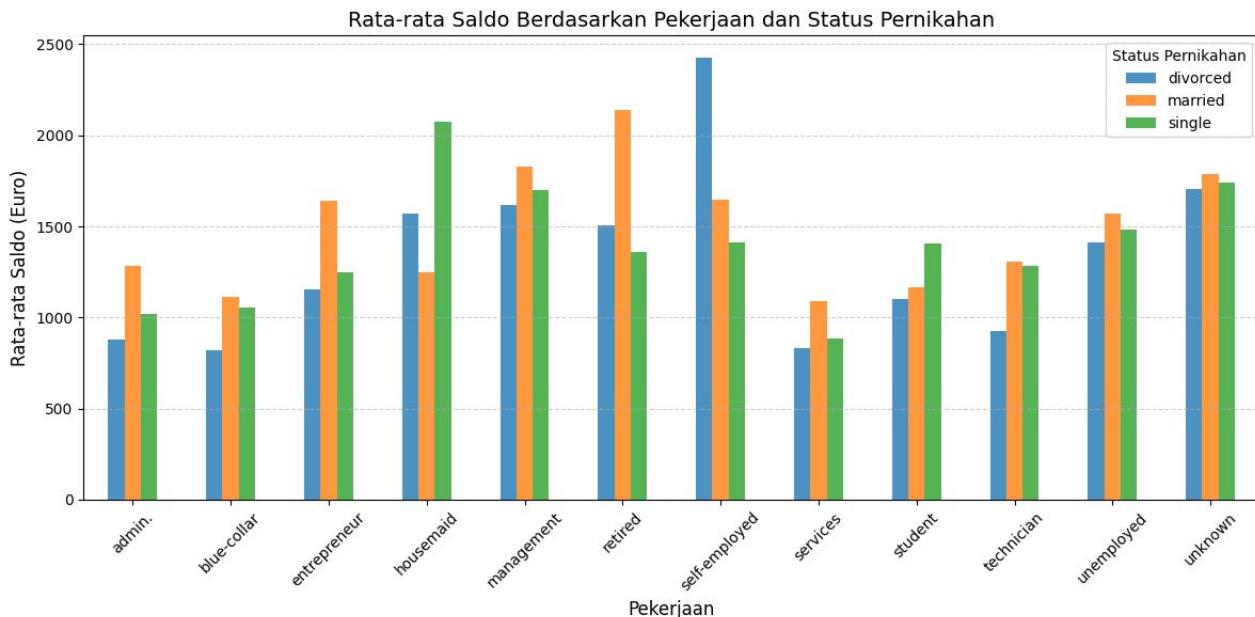
Exploratory Data Analysis (EDA)

Terdiri dari 20 Pivot Table & Visualization

Univariate Analysis



1. Rata-rata saldo berdasarkan pekerjaan dan status pernikahan



Self-employed: Saldo tertinggi di semua kategori pernikahan.

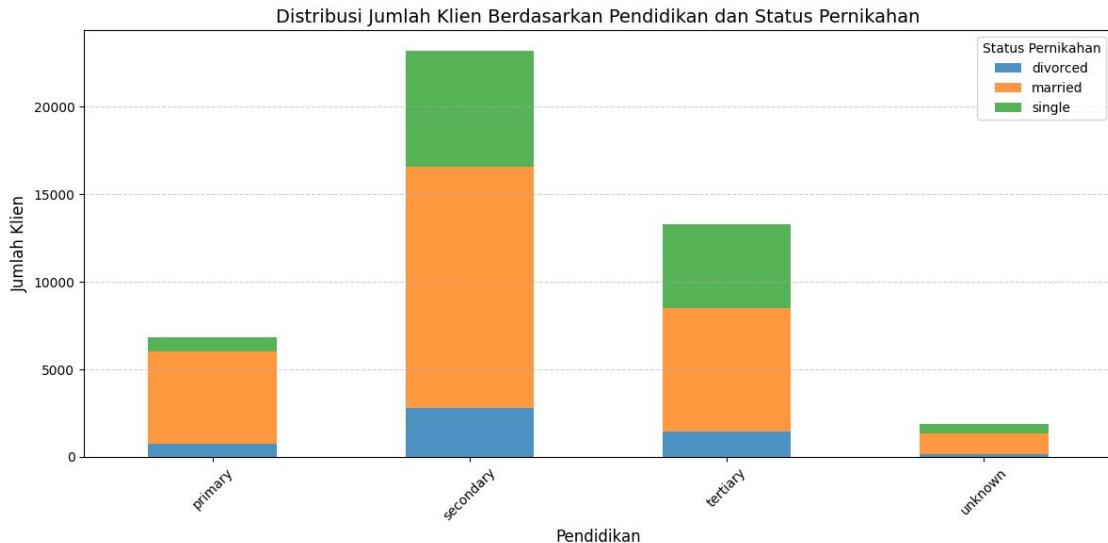
Blue-collar: Saldo terendah, khususnya pada **divorced**.

Married: Saldo rata-rata lebih tinggi dibanding **single** dan **divorced**.

Fokus kampanye pada nasabah dengan saldo tinggi (self-employed, married)

	marital	divorced	married	single
job				
admin.	878.333333	1281.409580	1020.739005	
blue-collar	820.806667	1113.165901	1056.105263	
entrepreneur	1155.983240	1643.387850	1248.235294	
housemaid	1573.222826	1248.173246	2074.743056	
management	1618.073807	1828.159444	1700.219545	
retired	1507.835294	2140.098787	1360.388889	
self-employed	2426.350000	1644.948640	1410.365471	
services	834.304189	1088.847528	887.324708	
student	1101.000000	1164.796296	1403.753986	
technician	924.048649	1307.399556	1283.938168	
unemployed	1409.643275	1568.632011	1484.079800	
unknown	1706.294118	1788.128079	1741.794118	

2. Distribusi jumlah klien berdasarkan pendidikan dan status pernikahan



	marital	divorced	married	single
education				
primary	752	5246	853	
secondary	2815	13770	6617	
tertiary	1471	7038	4792	
unknown	169	1160	528	

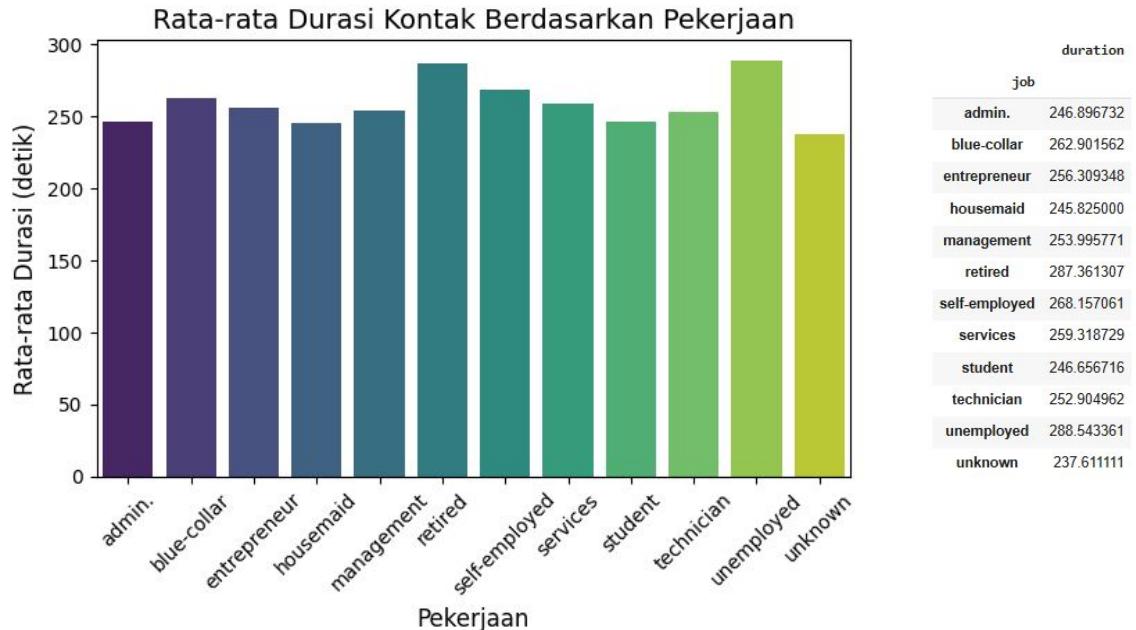
Secondary Education: Memiliki jumlah klien tertinggi di semua kategori pernikahan, terutama pada status **married**.

Primary Education: Klien didominasi oleh status **married**, dengan jumlah jauh lebih kecil dibanding kategori pendidikan lainnya.

Tertiary Education: Klien dengan status **single** lebih banyak dibandingkan pada tingkat pendidikan lainnya.

Kesimpulan: Fokus kampanye dapat diarahkan pada klien dengan pendidikan **secondary** dan status **married** karena jumlahnya dominan.

3. Rata-rata durasi kontak terakhir berdasarkan pekerjaan

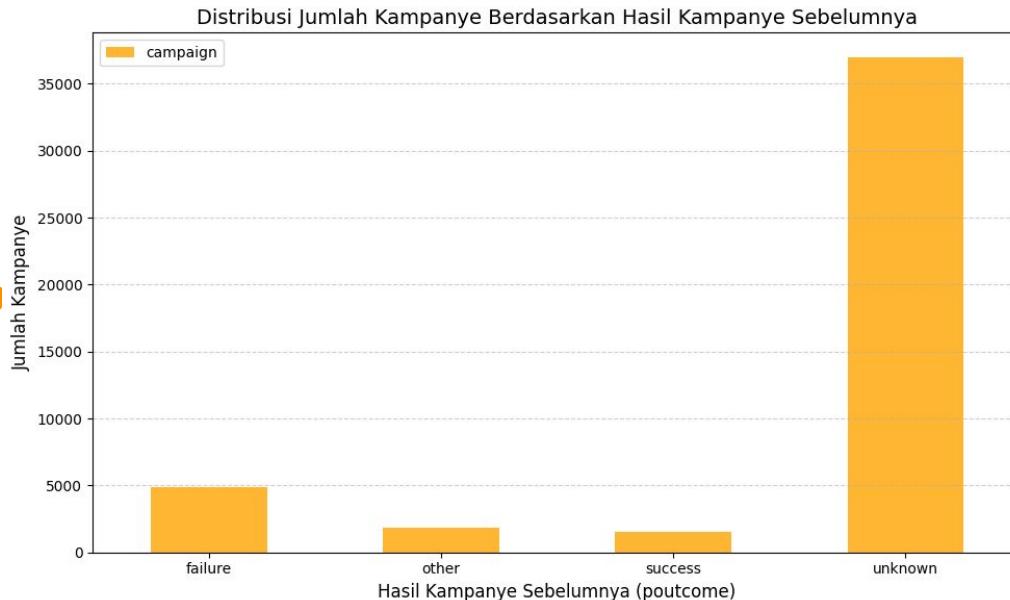


Pekerjaan dengan Durasi Tertinggi: Retired (287 detik) dan Unemployed (288 detik).

Pekerjaan dengan Durasi Terendah: Unknown (237 detik) dan Admin (246 detik).

Kampanye: Fokuskan kontak lebih lama pada kategori pekerjaan dengan durasi pendek untuk meningkatkan efektivitas.

4. Distribusi jumlah kampanye berdasarkan hasil kampanye sebelumnya (poutcome)



campaign	poutcome
failure	4901
other	1840
success	1511
unknown	36959

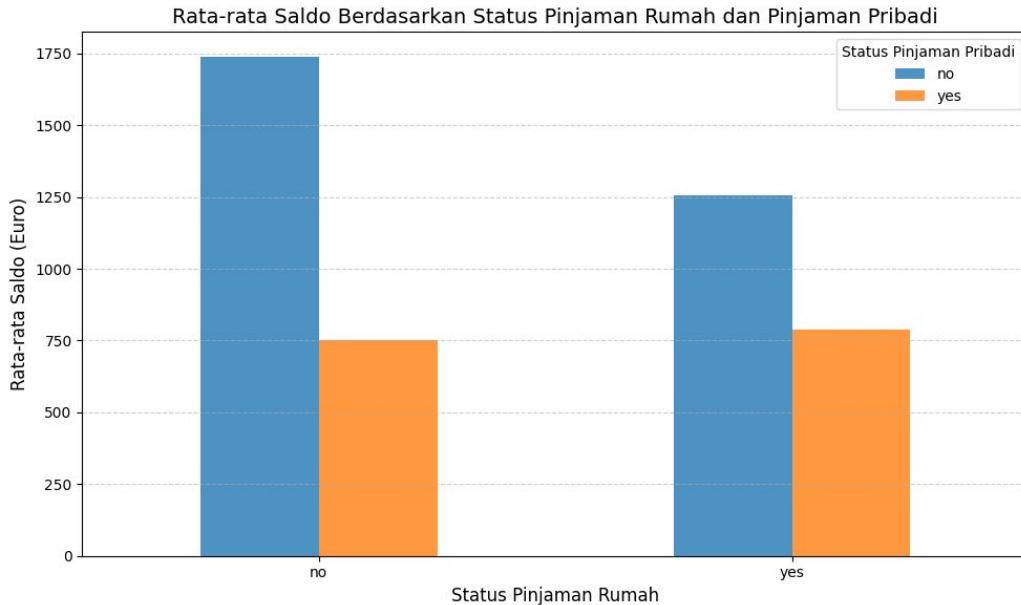
Unknown: Sebagian besar kampanye (36,959) dilakukan tanpa informasi hasil kampanye sebelumnya.

Failure: Sebanyak 4,901 kampanye sebelumnya gagal.

Success: Hanya 1,511 kampanye sebelumnya berhasil.

Rekomendasi: Fokus pada evaluasi hasil kampanye sebelumnya untuk meningkatkan efektivitas.

5. Rata-rata saldo berdasarkan status pinjaman rumah dan pinjaman pribadi



loan	no	yes
housing		
no	1737.606719	752.713243
yes	1256.407793	788.537898

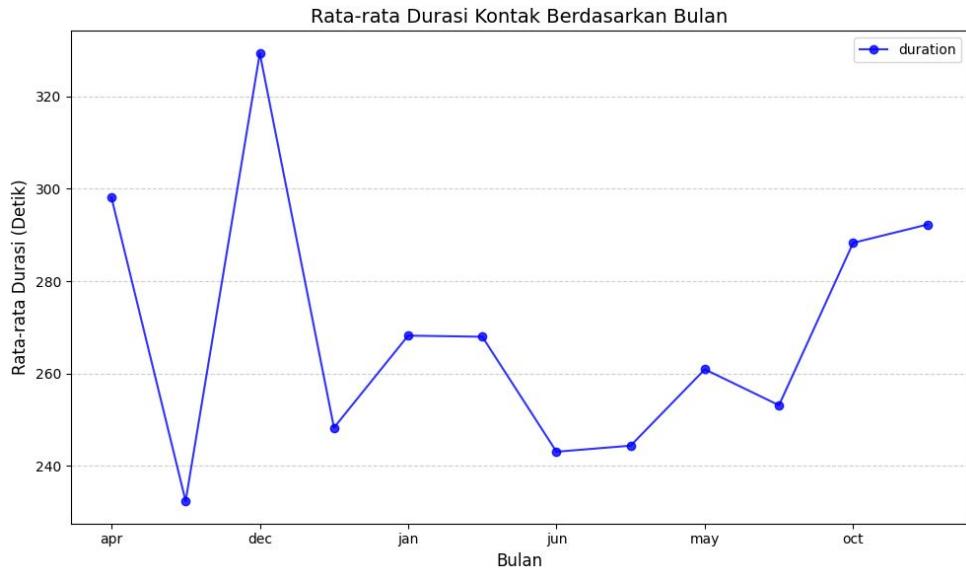
Pinjaman Rumah (No): Saldo rata-rata lebih tinggi, terutama bagi nasabah tanpa pinjaman pribadi (€1,737).

Pinjaman Rumah (Yes): Saldo rata-rata lebih rendah dibandingkan yang tidak memiliki pinjaman rumah.

Pinjaman Pribadi: Nasabah dengan pinjaman pribadi memiliki saldo rata-rata lebih rendah di kedua kategori.

Rekomendasi: Fokuskan kampanye pada nasabah tanpa pinjaman pribadi dan rumah untuk potensi hasil yang lebih baik.

6. Rata-rata durasi kontak berdasarkan bulan



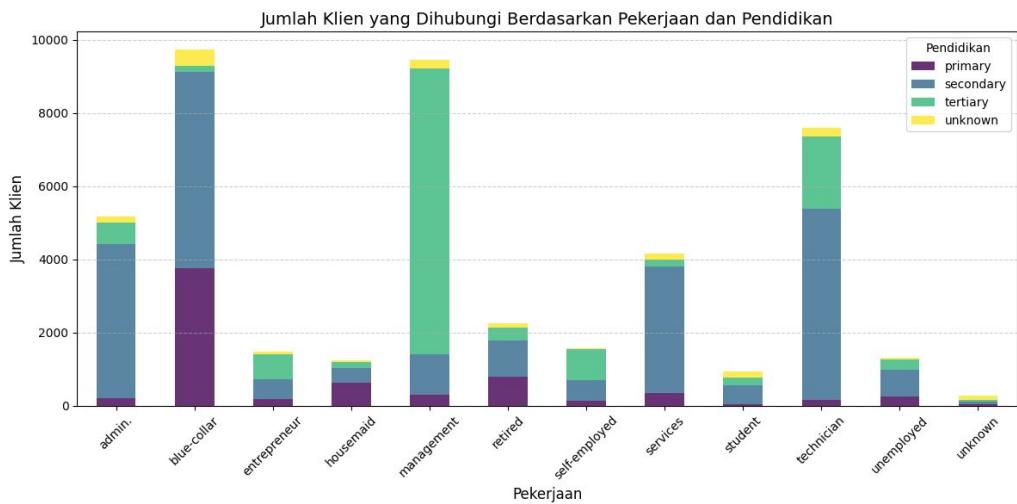
Pinjaman Rumah (No): Saldo rata-rata lebih tinggi, terutama bagi nasabah tanpa pinjaman pribadi (€1,737).

Pinjaman Rumah (Yes): Saldo rata-rata lebih rendah dibandingkan yang tidak memiliki pinjaman rumah.

Pinjaman Pribadi: Nasabah dengan pinjaman pribadi memiliki saldo rata-rata lebih rendah di kedua kategori.

Rekomendasi: Fokuskan kampanye pada nasabah tanpa pinjaman pribadi dan rumah untuk potensi hasil yang lebih baik.

7. Jumlah klien yang dihubungi berdasarkan pekerjaan dan pendidikan



	education	primary	secondary	tertiary	unknown
job					
admin.	209	4219	572	171	
blue-collar	3758	5371	149	454	
entrepreneur	183	542	686	76	
housemaid	627	395	173	45	
management	294	1121	7801	242	
retired	795	984	366	119	
self-employed	130	577	833	39	
services	345	3457	202	150	
student	44	508	223	163	
technician	158	5229	1968	242	
unemployed	257	728	289	29	
unknown	51	71	39	127	

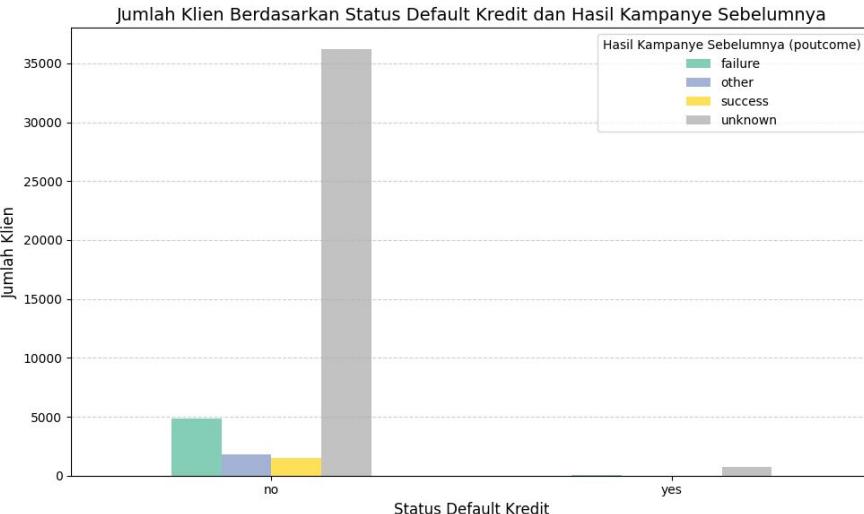
Pekerjaan yang Paling Sering Dihubungi: Management, technician, dan blue-collar.

Tingkat Pendidikan Dominan: Secondary education mendominasi hampir semua pekerjaan.

Kategori Tertinggi di Tertiary Education: Management dan self-employed.

Rekomendasi: Optimalkan kampanye pada pekerjaan dengan distribusi pendidikan yang lebih tinggi.

8. Jumlah klien berdasarkan status default kredit dan hasil kampanye sebelumnya



poutcome	failure	other	success	unknown
default				
no	4862	1824	1509	36201
yes	39	16	2	758

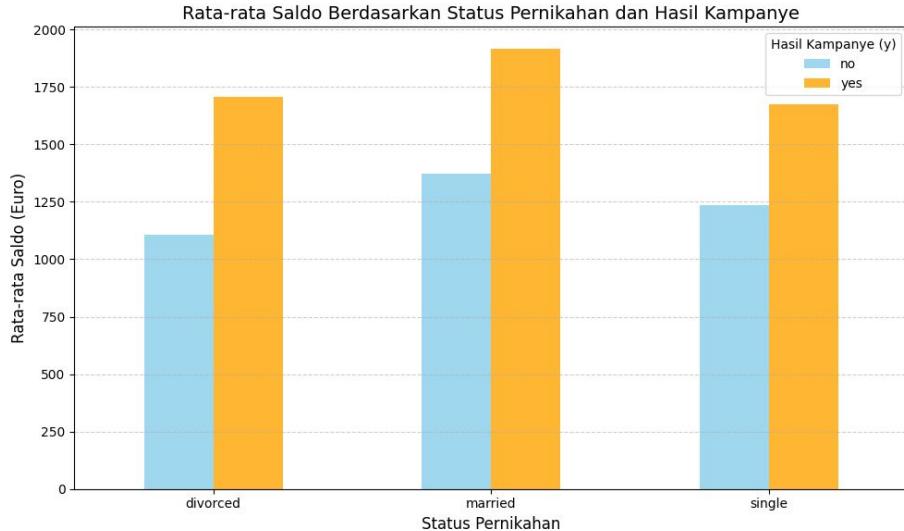
Mayoritas Klien Tanpa Default Kredit: Sebagian besar klien tidak memiliki riwayat default kredit (36,201).

Default Kredit dan Kampanye: Klien dengan default kredit memiliki partisipasi kampanye jauh lebih rendah (758).

Hasil Kampanye Tidak Diketahui: Sebagian besar kampanye berakhir tanpa informasi hasil (kategori "unknown").

Rekomendasi: Fokus pada klien tanpa default kredit untuk meningkatkan efektivitas kampanye.

9. Rata-rata saldo berdasarkan status pernikahan dan hasil kampanye



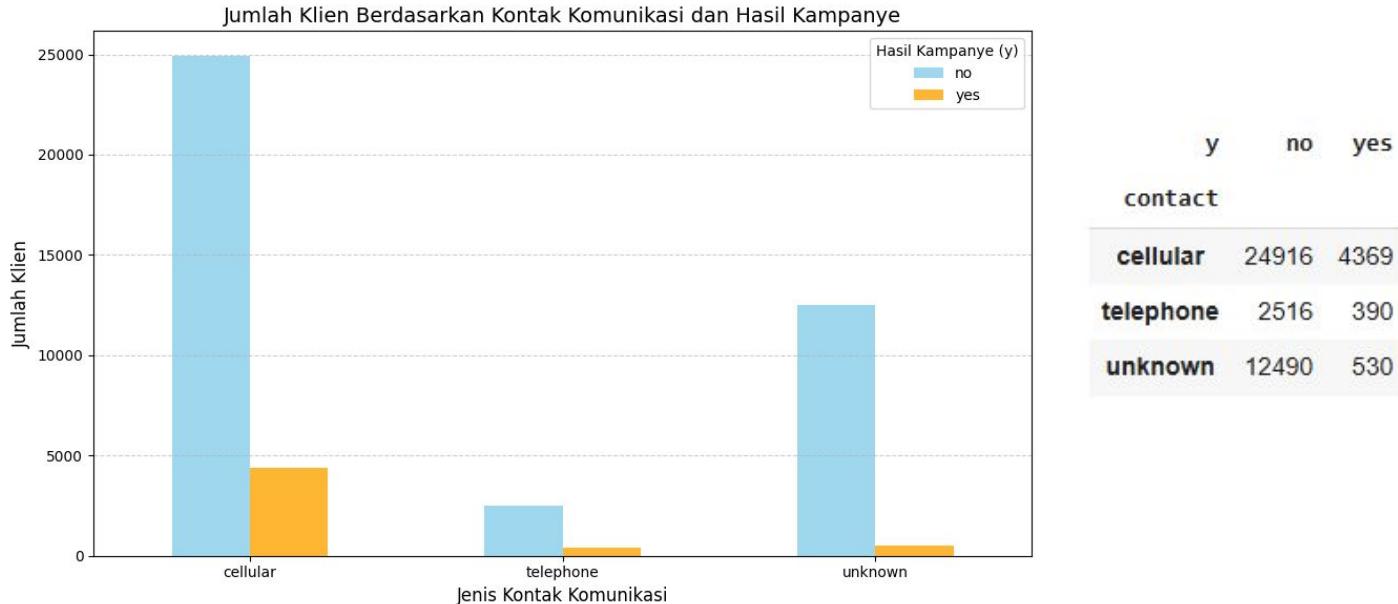
y	no	yes
marital		
divorced	1107.095747	1707.964630
married	1370.746228	1915.810163
single	1235.869921	1674.875523

Rata-rata Saldo Tertinggi: Klien yang menikah memiliki rata-rata saldo tertinggi baik yang merespons kampanye maupun tidak.

Peningkatan Saldo: Rata-rata saldo lebih tinggi untuk klien yang merespons kampanye (kategori "yes") di semua status pernikahan.

Rekomendasi: Fokuskan kampanye pada klien menikah untuk potensi keberhasilan lebih tinggi.

10. Jumlah klien berdasarkan kontak komunikasi dan hasil kampanye

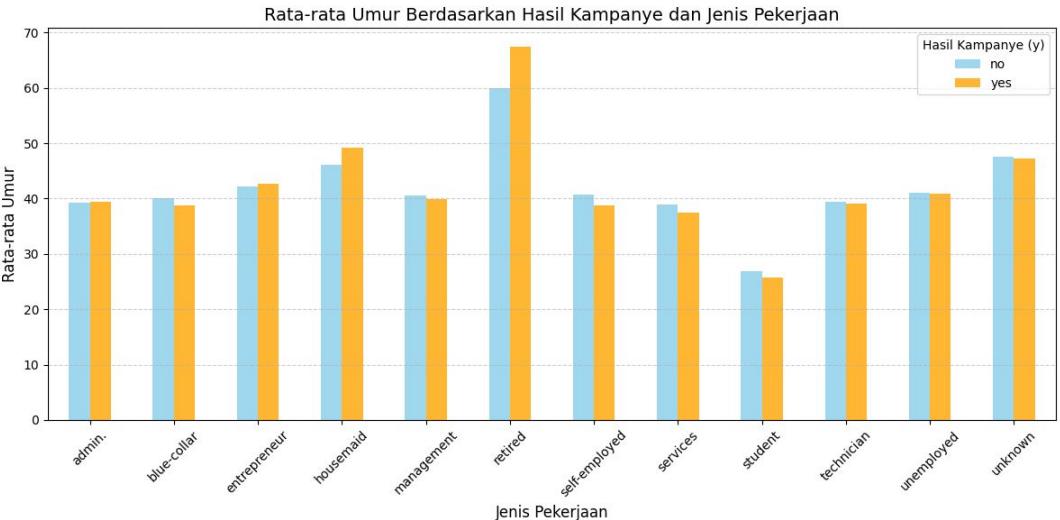


Kontak "Cellular" Dominan: Sebagian besar klien dihubungi melalui cellular dengan keberhasilan tertinggi.

Efektivitas Kontak: Cellular lebih efektif dibandingkan telephone dan kategori unknown.

Rekomendasi: Prioritaskan penggunaan kontak cellular untuk kampanye mendatang.

11. Rata-rata umur berdasarkan hasil kampanye (y) dan jenis pekerjaan (job)



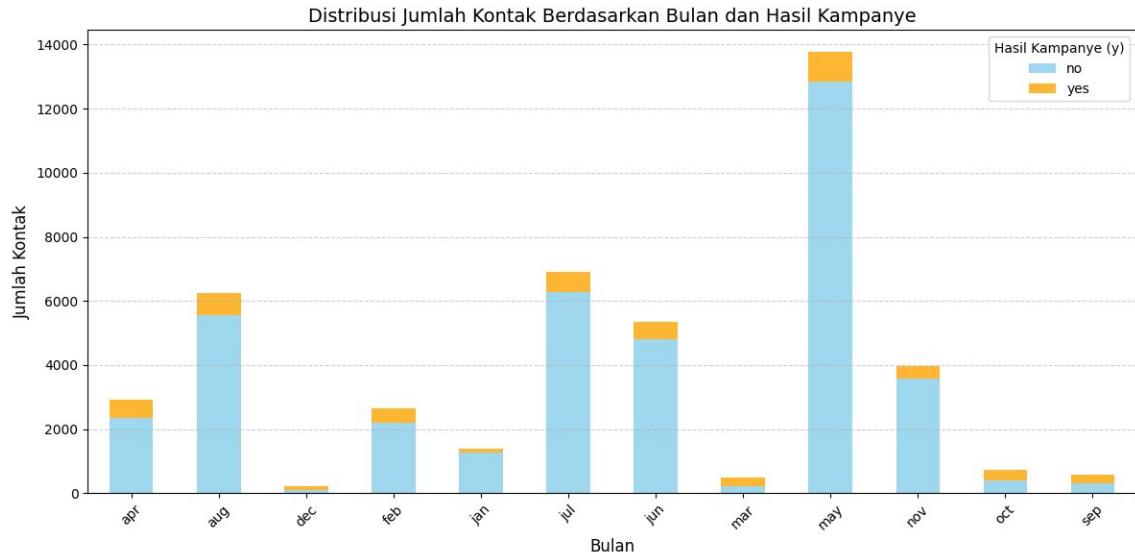
Pekerjaan "**Retired**" memiliki usia rata-rata tertinggi untuk hasil kampanye "yes" (67 tahun).

Pekerjaan "**Student**" memiliki usia rata-rata terendah untuk hasil "yes" (25 tahun).

Fokuskan kampanye pada pekerjaan dengan hasil positif tinggi seperti "**Retired**" dan "**Housemaid**".

job	y	no	yes
admin.	39.266300	39.459588	
blue-collar	40.143506	38.776836	
entrepreneur	42.145894	42.691057	
housemaid	46.140584	49.266055	
management	40.534878	39.914681	
retired	59.912471	67.434109	
self-employed	40.712644	38.786096	
services	38.863144	37.479675	
student	26.862481	25.747212	
technician	39.344532	39.073810	
unemployed	40.978202	40.871287	
unknown	47.637795	47.264706	

12. Distribusi jumlah kontak berdasarkan bulan dan hasil kampanye



Bulan Mei memiliki jumlah kontak tertinggi dengan hasil kampanye "yes" mencapai 925.

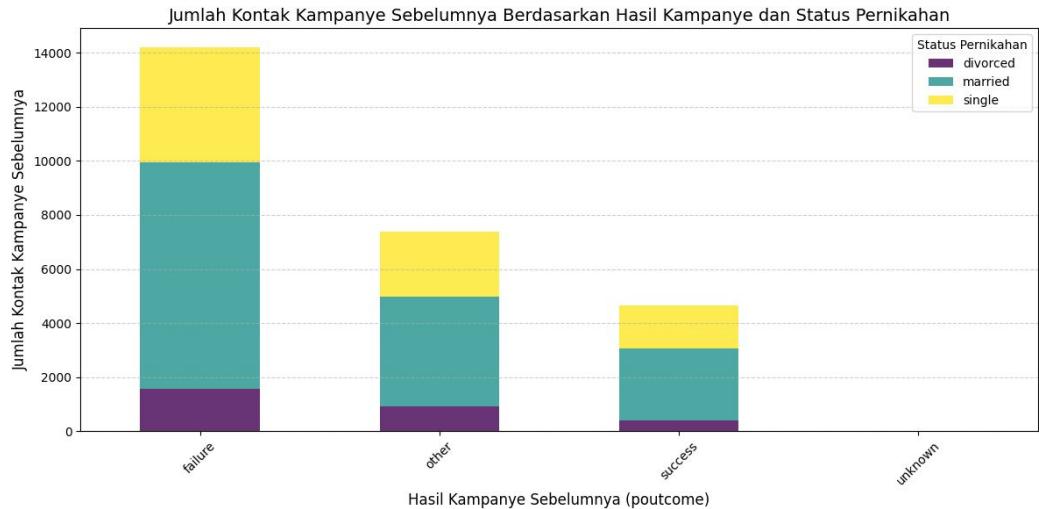
Bulan Juli menempati posisi kedua dengan jumlah kontak "yes" sebanyak 627.

Bulan Desember memiliki rasio keberhasilan tinggi meskipun jumlah kontak rendah.

Fokus pada bulan dengan tingkat keberhasilan tinggi seperti **Mei** dan **Juli** untuk optimasi kampanye.

y	no	yes
month		
apr	2355	577
aug	5559	688
dec	114	100
feb	2208	441
jan	1261	142
Jul	6268	627
jun	4795	546
mar	229	248
may	12841	925
nov	3567	403
oct	415	323
sep	310	269

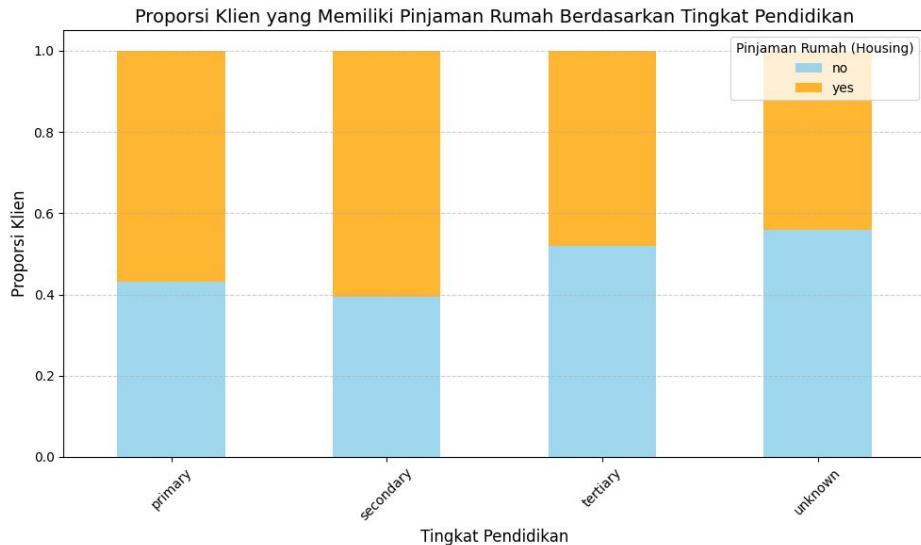
13. Jumlah kontak kampanye sebelumnya (previous) berdasarkan hasil kampanye sebelumnya (poutcome) dan status pernikahan



	marital	divorced	married	single
poutcome				
failure	marital	1574	8374	4248
other	divorced	907	4073	2397
success	married	390	2686	1571
unknown	single	0	13	4

Hasil kampanye sebelumnya "**failure**" memiliki jumlah kontak tertinggi untuk semua status pernikahan. Nasabah dengan status "**married**" lebih sering dihubungi pada semua kategori hasil kampanye sebelumnya. Kampanye yang sukses sebelumnya memiliki jumlah kontak paling sedikit, terutama untuk nasabah "**divorced**". Fokus pada nasabah dengan **status "married"** dan hasil kampanye sukses sebelumnya untuk meningkatkan efektivitas.

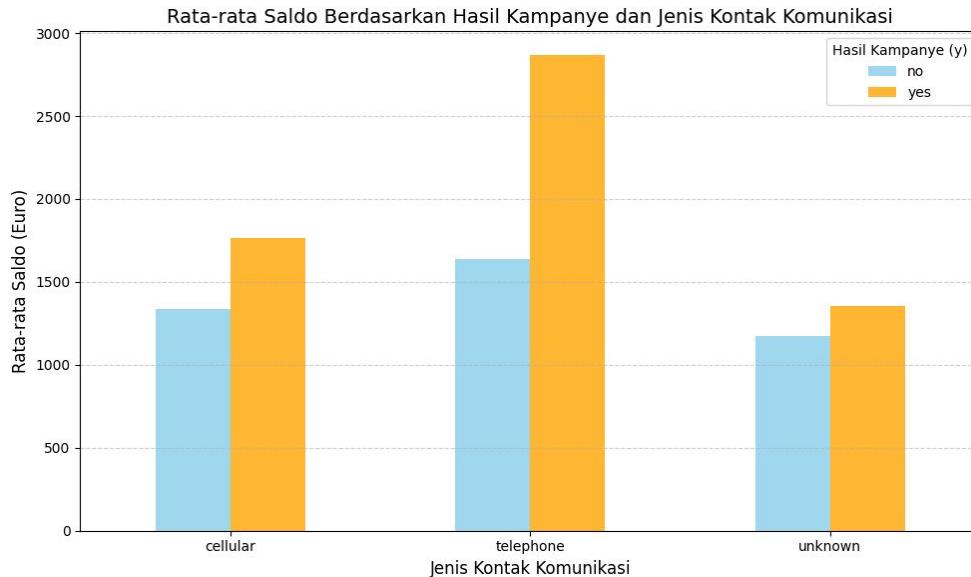
14. Proporsi klien yang memiliki pinjaman rumah berdasarkan tingkat pendidikan



housing	no	yes
education		
primary	2957	3894
secondary	9164	14038
tertiary	6923	6378
unknown	1037	820

Klien dengan pendidikan **secondary** memiliki proporsi tertinggi dalam memiliki pinjaman rumah. Pendidikan **primary** dan **tertiary** memiliki distribusi pinjaman rumah yang hampir seimbang. Kategori **unknown** menunjukkan proporsi lebih besar tanpa pinjaman rumah. Fokus pemasaran dapat diarahkan pada nasabah dengan pendidikan **secondary** untuk kampanye yang melibatkan produk terkait pinjaman rumah.

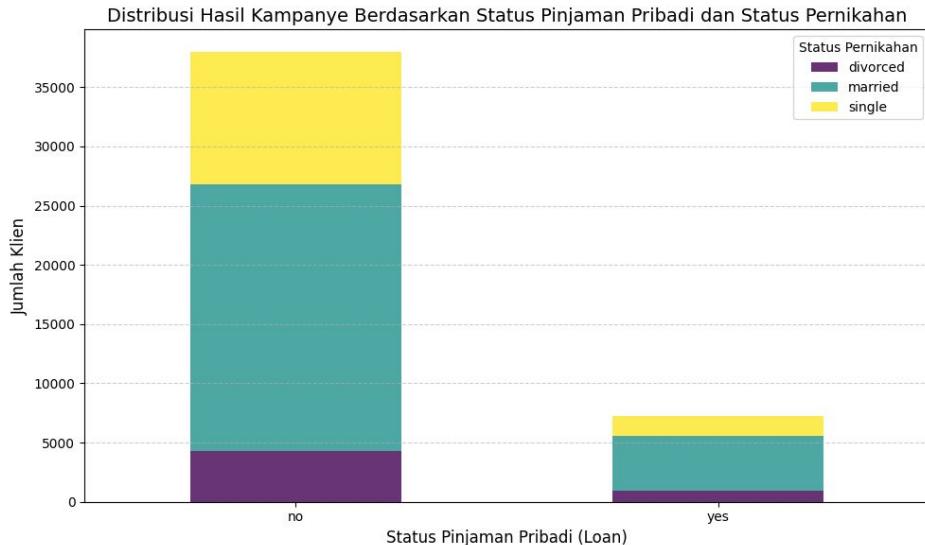
15. Rata-rata saldo berdasarkan hasil kampanye (y) dan jenis kontak komunikasi (contact)



y	no	yes
contact		
cellular	1333.772195	1763.795834
telephone	1638.855326	2867.189744
unknown	1176.243475	1355.745283

- **Kontak melalui telepon** memiliki saldo rata-rata tertinggi, terutama untuk hasil kampanye "yes".
- **Kontak seluler** memiliki perbedaan saldo antara hasil "yes" dan "no", namun lebih rendah dibandingkan telepon.
- Kategori **unknown** memiliki saldo rata-rata terendah untuk kedua hasil kampanye.
- Strategi pemasaran dapat difokuskan pada nasabah dengan preferensi kontak melalui telepon untuk hasil kampanye yang lebih efektif.

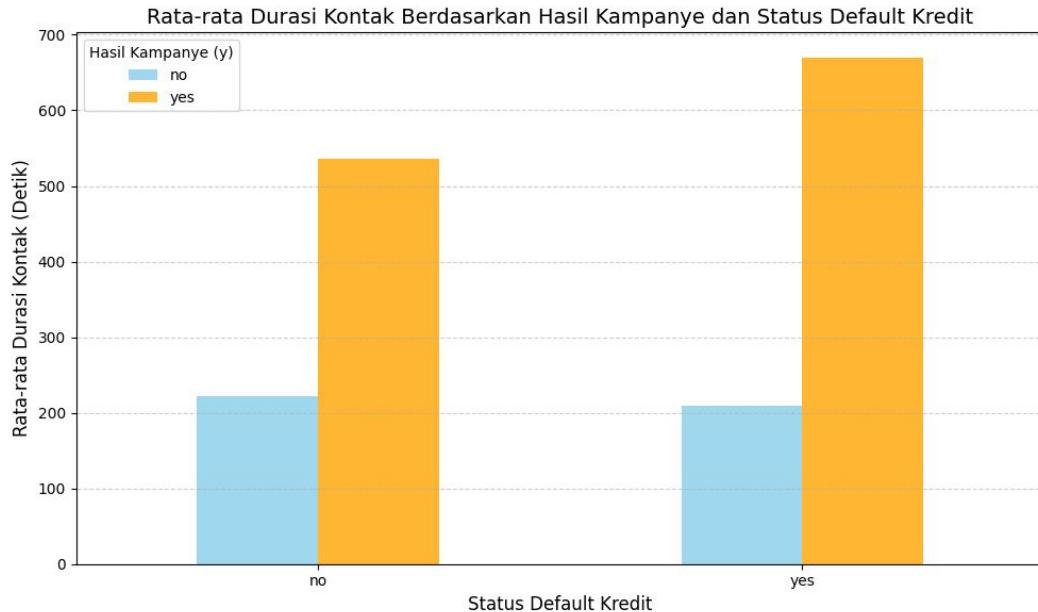
16. Distribusi hasil kampanye (y) berdasarkan status pinjaman pribadi (loan) dan status pernikahan



marital	divorced	married	single
loan			
no	4286	22554	11127
yes	921	4660	1663

- **Majoritas klien** yang tidak memiliki pinjaman pribadi lebih sering dihubungi dalam kampanye.
- Klien yang **menikah** mendominasi kelompok baik dengan maupun tanpa pinjaman pribadi.
- Klien **single** memiliki proporsi yang lebih kecil di kedua kategori.
- Fokus strategi kampanye dapat diarahkan pada klien tanpa pinjaman pribadi, khususnya yang sudah menikah, untuk efektivitas lebih tinggi.

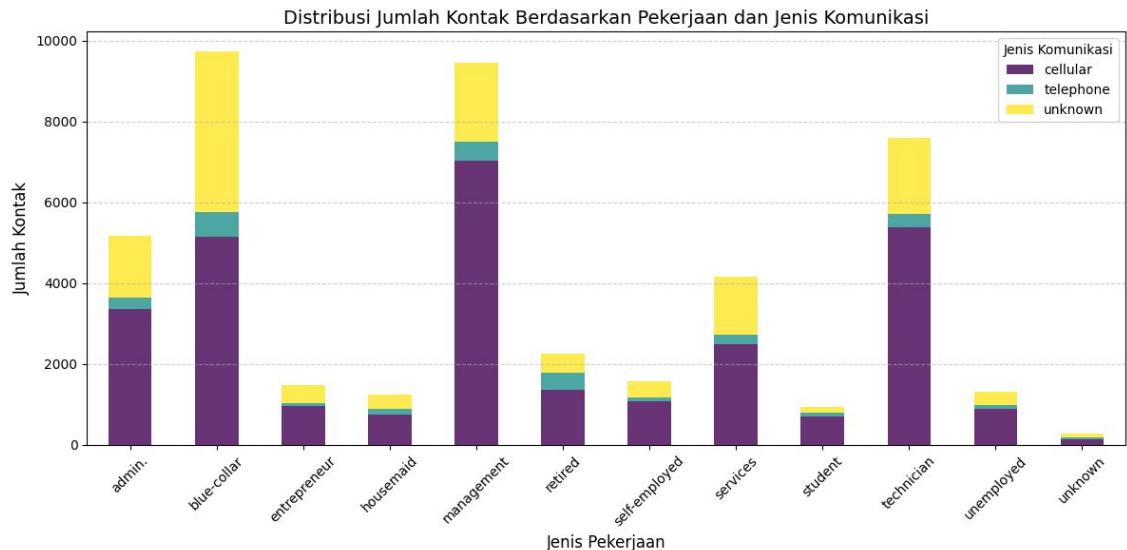
17. Rata-rata durasi kontak berdasarkan hasil kampanye (y) dan status default kredit



y	no	yes
default		
no	221.404862	535.982242
yes	209.786370	669.461538

- Klien tanpa default kredit memiliki durasi kontak rata-rata lebih pendek untuk hasil kampanye gagal.
- Durasi kontak lebih panjang untuk hasil kampanye berhasil, baik pada klien dengan maupun tanpa default kredit.
- Fokus pada klien tanpa default kredit dengan durasi kontak lebih panjang dapat meningkatkan keberhasilan kampanye.

18. Distribusi jumlah kontak berdasarkan pekerjaan dan jenis komunikasi

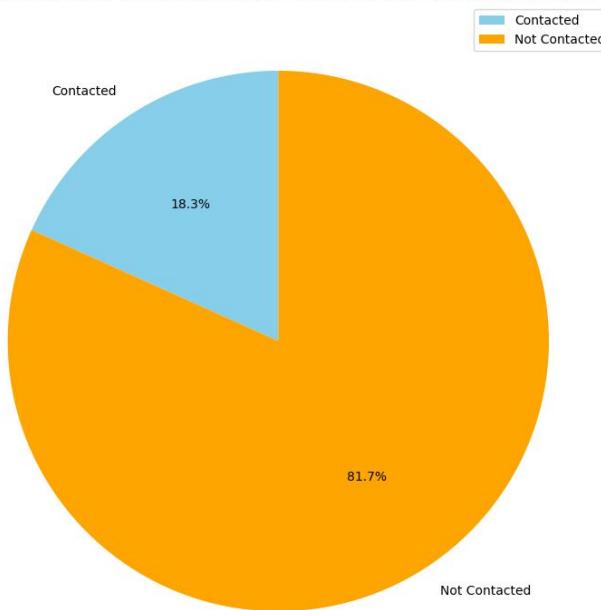


contact	cellular	telephone	unknown
job			
admin.	3354	289	1528
blue-collar	5156	611	3965
entrepreneur	954	87	446
housemaid	748	151	341
management	7021	471	1966
retired	1370	405	489
self-employed	1084	98	397
services	2498	228	1428
student	697	99	142
technician	5381	321	1895
unemployed	891	100	312
unknown	131	46	111

- Mayoritas komunikasi dilakukan melalui **cellular**, terutama pada pekerjaan **management** dan **blue-collar**.
- Jenis komunikasi **telephone** paling sedikit digunakan di semua pekerjaan.
- Jenis komunikasi **unknown** masih signifikan pada beberapa pekerjaan, terutama **technician** dan **admin**.
- Fokus peningkatan komunikasi cellular dapat memperluas jangkauan ke klien pada pekerjaan dominan.

19. Distribusi Jumlah Klien Berdasarkan Status Kontak Sebelumnya (pdays)

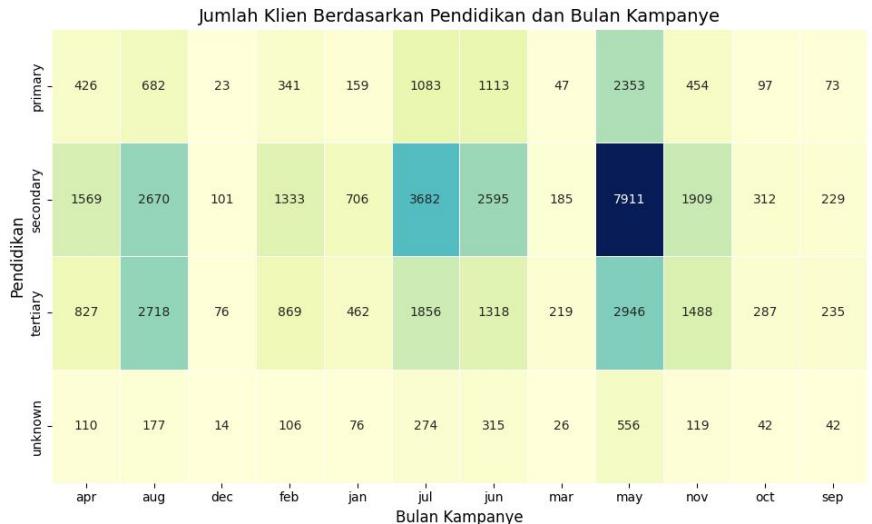
Distribusi Jumlah Klien Berdasarkan Status Kontak Sebelumnya (pdays)



pdays	
Contacted	8257
Not Contacted	36954

- Sebagian besar klien (81.7%) belum pernah dihubungi sebelumnya.
- Hanya 18.3% klien yang sudah dihubungi pada kampanye sebelumnya.
- Strategi dapat difokuskan untuk menghubungi klien baru guna memperluas jangkauan.

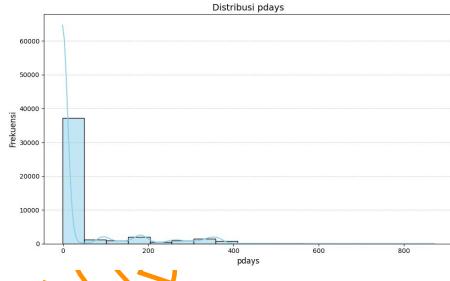
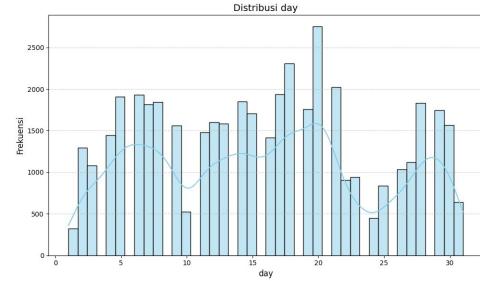
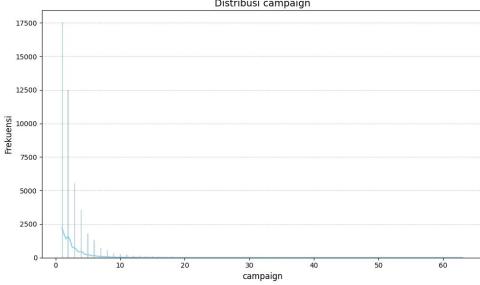
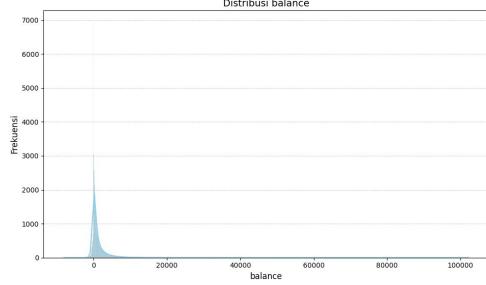
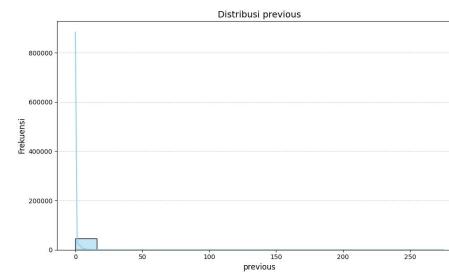
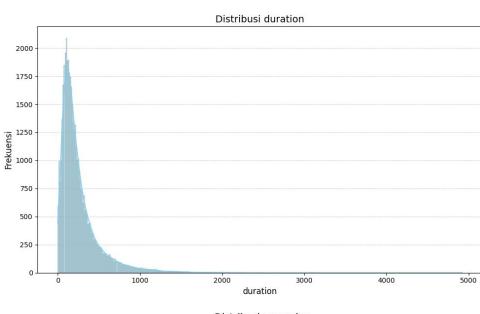
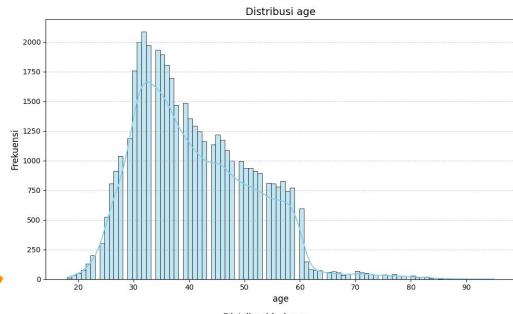
20. Jumlah klien berdasarkan pendidikan dan bulan kampanye



	month	apr	aug	dec	feb	jan	jul	jun	mar	may	nov	oct	sep
education	primary	426	682	23	341	159	1083	1113	47	2353	454	97	73
secondary	1569	2670	101	1333	706	3682	2595	185	7911	1909	312	229	
tertiary	827	2718	76	869	462	1856	1318	219	2946	1488	287	235	
unknown	110	177	14	106	76	274	315	26	556	119	42	42	

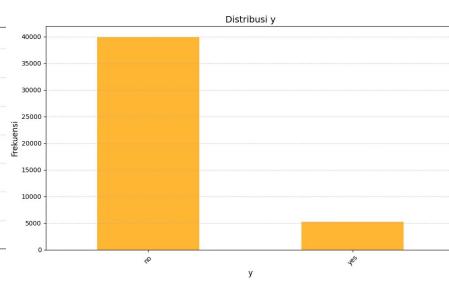
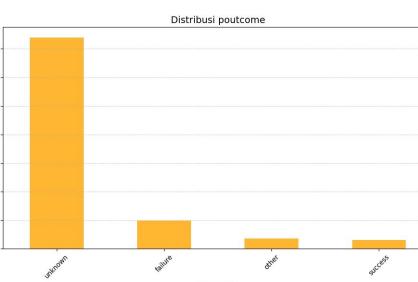
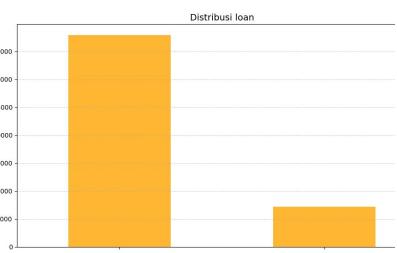
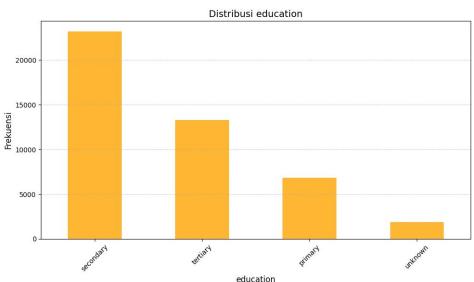
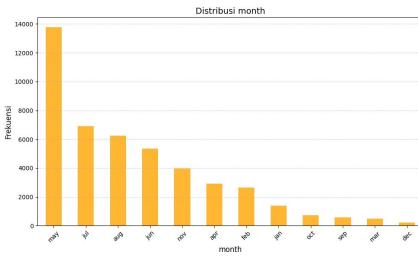
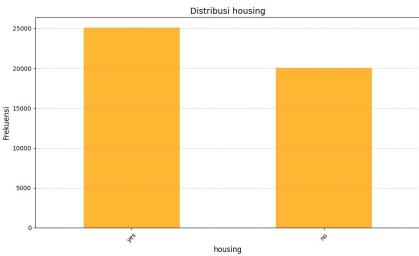
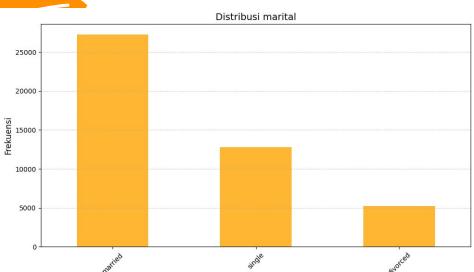
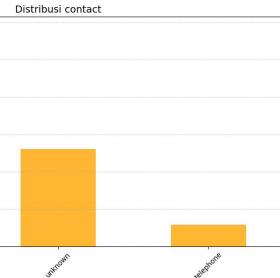
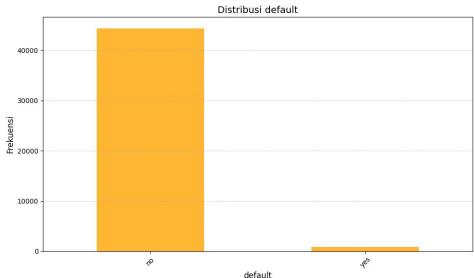
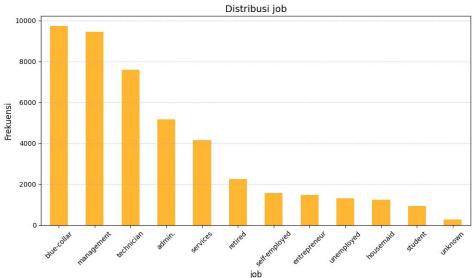
- Klien dengan tingkat pendidikan **secondary** paling banyak dihubungi pada bulan Mei (7911).
- Tingkat pendidikan **tertiary** dan **primary** memiliki tren serupa di bulan Juli dan Mei.
- Klien dengan pendidikan **unknown** menunjukkan jumlah paling sedikit di semua bulan.
- Bulan **Mei** adalah periode dengan aktivitas kampanye tertinggi untuk semua tingkat pendidikan.

Univariate Analysis (Numerical Data)



Age: Mayoritas klien berusia 30-40 tahun.
Balance: Didominasi nilai kecil, ada outlier tinggi.
Duration: Durasi kontak pendek mendominasi.
Campaign: Sebagian besar klien mengikuti 1-2 kampanye.
Pdays & Previous: Mayoritas klien belum dihubungi sebelumnya.

Univariate Analysis (Categorical Data)



Job: Mayoritas blue-collar, management, technician.

Marital: Mayoritas married.

Education: Mayoritas secondary.

Default: Hampir semua no.

Housing Loan: Sedikit lebih banyak yes.

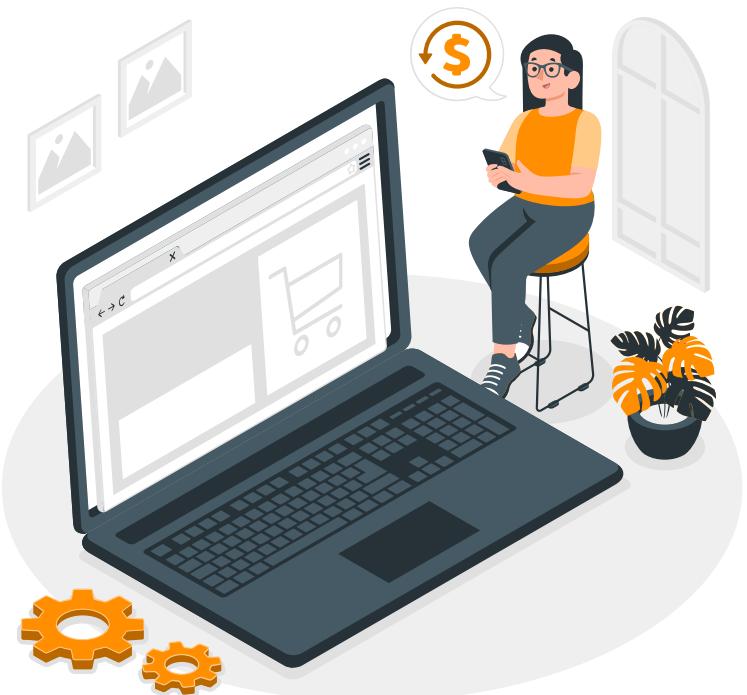
Personal Loan: Mayoritas no.

Contact: Dominan melalui cellular.

Month: Kontak tertinggi di May.

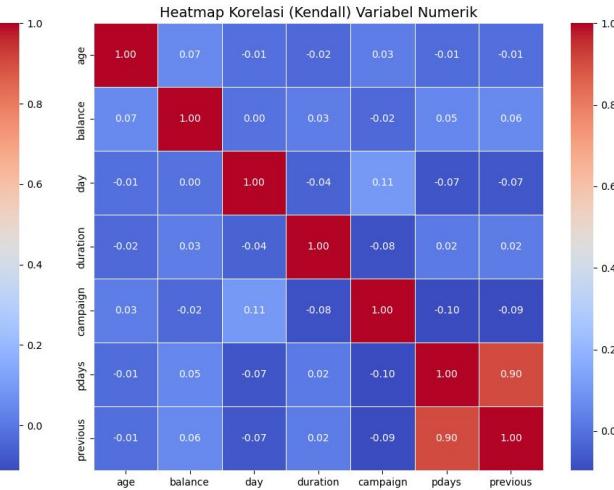
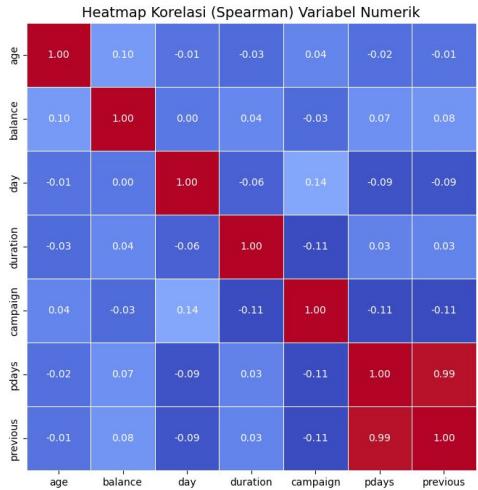
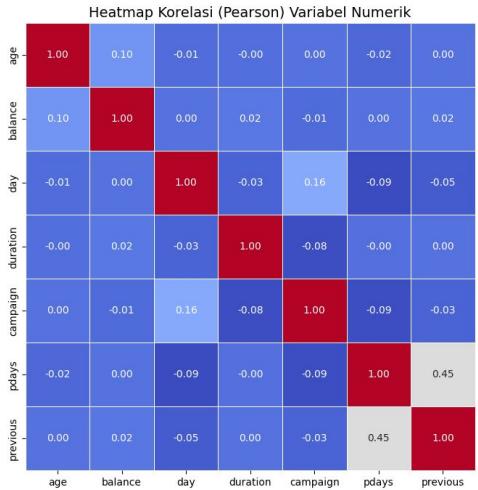
Outcome: Mayoritas unknown.

Target (y): Dominan no.



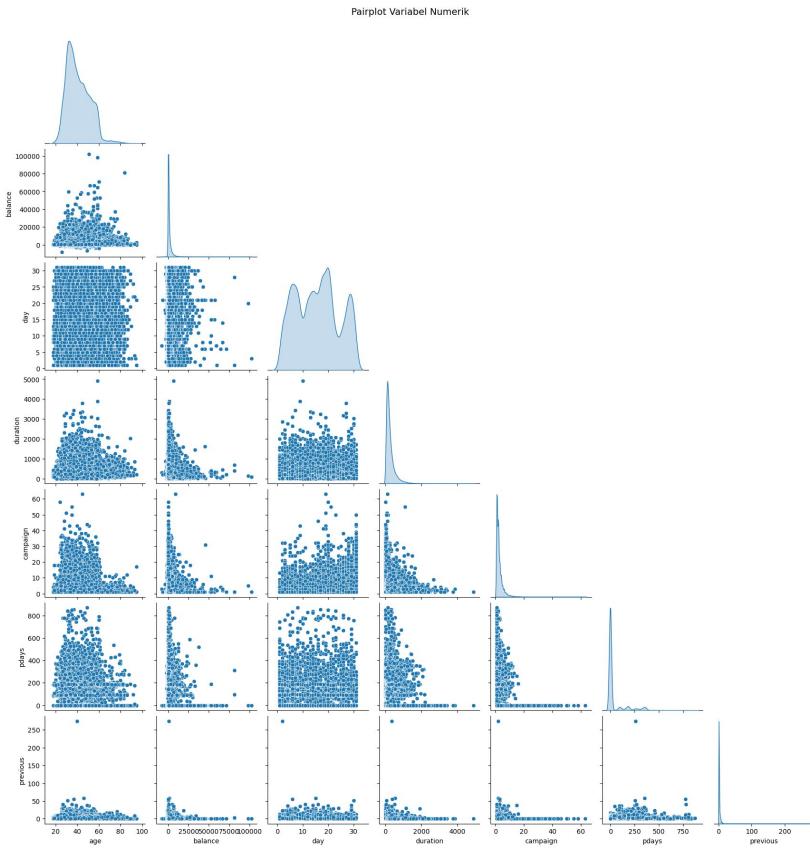
4. Analisis Korelasi, Chi-Square, & Regresi Linier

Multivariate Analysis (Heatmap)



- Hubungan antara variabel numerik dalam dataset ini umumnya **sangat lemah** atau **tidak signifikan**.
- Satu-satunya hubungan kuat ditemukan antara **pdays** dan **previous**, yang logis karena keduanya berkaitan dengan riwayat kontak sebelumnya.

Pairplot untuk Hubungan antar Variabel Numerik



- **Distribusi Variabel Numerik:**

Sebagian besar variabel memiliki distribusi yang tidak merata (skewed), terutama balance, duration, dan pdays.

- **Hubungan Antar Variabel:**

Tidak ada pola hubungan yang jelas di antara sebagian besar variabel numerik, kecuali hubungan lemah antara pdays dan previous.

- **Korelasi Lemah:**

Scatter plot menunjukkan korelasi yang sangat rendah antara variabel seperti campaign terhadap variabel lainnya, mengonfirmasi hasil heatmap.

Pengaruh Jumlah Kampanye terhadap Pembukaan Akun Deposit

Definisi Hipotesis

Hipotesis Nol (H_0)

Definisi: Jumlah kampanye yang dilakukan oleh bank tidak mempengaruhi keputusan nasabah untuk membuka akun deposit. Dalam istilah statistik, ini berarti bahwa tidak ada perbedaan signifikan dalam tingkat keberhasilan pembukaan akun deposit antara kelompok nasabah yang menerima jumlah kampanye yang berbeda.

Hipotesis Alternatif (H_1)

Definisi: Jumlah kampanye yang dilakukan oleh bank mempengaruhi keputusan nasabah untuk membuka akun deposit. Dalam istilah statistik, ini berarti bahwa terdapat perbedaan signifikan dalam tingkat keberhasilan pembukaan akun deposit antara kelompok nasabah yang menerima jumlah kampanye yang berbeda.

Participant

Peserta eksperimen adalah nasabah bank yang terdaftar dalam dataset dan menjadi target dari kampanye pemasaran.

Variabel yang Diuji

Variabel Independen: jumlah kampanye (campaign) yang diterima oleh nasabah.

Variabel Dependen: Keputusan nasabah untuk membuka akun deposit (kolom y, yang berisi "yes" atau "no").

Metrics yang Digunakan

Conversion Rate mengukur persentase peserta yang membuka akun deposit setelah berpartisipasi dalam kampanye. Ini dihitung dengan menghitung proporsi peserta yang membuka akun deposit (nilai yes dalam kolom y) dari total peserta.

Statistical Significance mengacu pada p-value yang diperoleh dari pengujian hipotesis yang digunakan untuk menentukan apakah ada hubungan yang signifikan antara jumlah kampanye dan keputusan membuka akun deposit.

Chi-Square

Jumlah total data: 45211

Tabel Kontingensi:

y	no	yes
campaign		
False	13835	1327
True	26087	3962

Hasil Uji Chi-Square:

Chi-Square Statistic: 191.28

P-value: 0.0000

Degrees of Freedom: 1

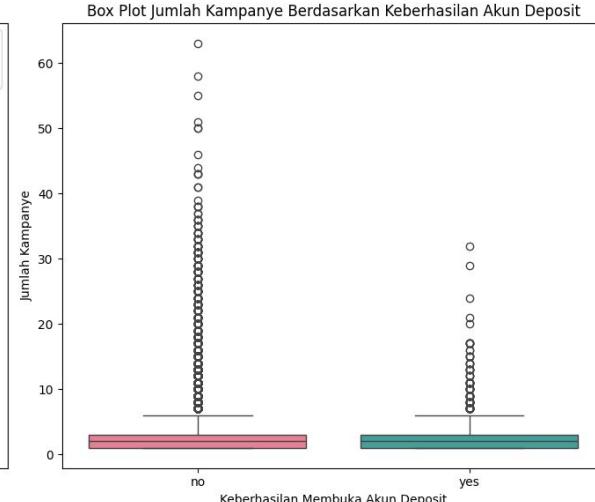
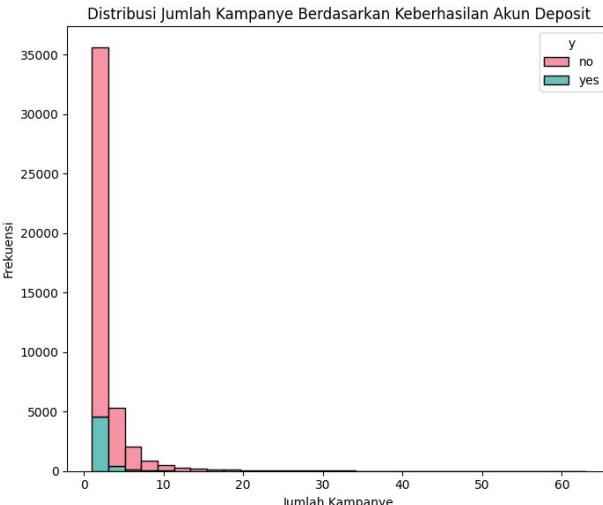
Expected Frequencies:

[[13388.27639291 1773.72360709
[26533.72360709 3515.27639291

Kesimpulan:

Hipotesis Alternatif (H_1): Jumlah kampanye memengaruhi keputusan nasabah untuk membuka akun deposit.

- Dengan p-value = 0.0000 (< 0.05), kita menolak Hipotesis Nol (H_0).
- Hipotesis Alternatif (H_1) diterima: Jumlah kampanye secara signifikan memengaruhi keputusan nasabah untuk membuka akun deposit.
- Namun, arah pengaruhnya negatif: Jumlah kampanye yang lebih tinggi justru menurunkan kemungkinan pembukaan akun deposit.



Analisis Regresi Linear

Hasil Regresi Linear Menggunakan Scikit-Learn:

Mean Squared Error (MSE): 0.1055

R-squared (R²): 0.0055

Hasil Regresi Linear Menggunakan Statsmodels:

OLS Regression Results

Dep. Variable:	y_binary	R-squared:	0.005		
Model:	OLS	Adj. R-squared:	0.005		
Method:	Least Squares	F-statistic:	243.4		
Date:	Sat, 14 Dec 2024	Prob (F-statistic):	1.01e-54		
Time:	10:12:50	Log-Likelihood:	-12713.		
No. Observations:	45211	AIC:	2.543e+04		
Df Residuals:	45209	BIC:	2.545e+04		
Df Model:	1				
Covariance Type:	nonrobust				
coef	std err	t	P> t	[0.025	0.975]
const	0.1380	0.002	68.290	0.000	0.134 - 0.142
campaign	-0.0076	0.000	-15.600	0.000	-0.009 - 0.007
Omnibus:	19794.814	Durbin-Watson:	1.532		
Prob(Omnibus):	0.000	Jarque-Bera (JB):	66791.488		
Skew:	2.363	Prob(JB):	0.00		
kurtosis:	6.623	Cond. No.	5.71		

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Kesimpulan Analisis:
Hipotesis Alternatif (H_1): Jumlah kampanye memiliki hubungan negatif terhadap pembukaan akun deposit.

- Koefisien Campaign:**

- Nilai: -0.0076

b. Interpretasi: Setiap peningkatan 1 unit kampanye menurunkan peluang membuka akun deposit sebesar 0.76%.

- Signifikansi Statistik:**

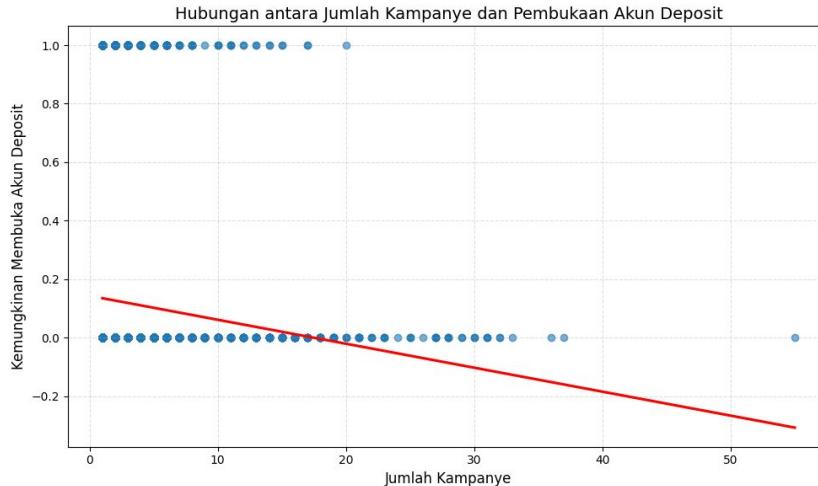
P-value: 0.000 (< 0.05), hubungan signifikan secara statistik.

- R-squared (R²):**

Nilai: 0.005 (0.55% variasi keputusan dijelaskan oleh jumlah kampanye).

- Kesimpulan:**

Hubungan negatif signifikan, tetapi jumlah kampanye bukan faktor utama.



Kesimpulan

1. **Hubungan Signifikan:**
 - o Hasil Chi-Square dan regresi linear menunjukkan bahwa jumlah kampanye memiliki hubungan signifikan dengan keputusan membuka akun deposit.
2. **Arah Hubungan (Negatif):**
 - o Berdasarkan regresi linear, hubungan jumlah kampanye terhadap pembukaan akun deposit adalah **negatif**.
 - o Semakin banyak kampanye dilakukan, semakin kecil kemungkinan nasabah membuka akun deposit.
3. **Efek Negatif Kampanye Berlebihan:**
 - o Hal ini mengindikasikan bahwa terlalu banyak kampanye dapat menjadi kontraproduktif dan menyebabkan penurunan efektivitas.

Rekomendasi Bisnis

1. **Optimalisasi Kampanye**
 - o Hindari terlalu banyak menghubungi nasabah, terutama mereka yang telah dihubungi lebih dari median jumlah kampanye.
 - o Fokus pada kualitas kontak daripada kuantitas.
2. **Optimalkan Targeting:**
 - o Fokus pada kualitas interaksi dalam kampanye, bukan kuantitas.
3. **Segmentasi Nasabah:**
 - o Identifikasi segmen nasabah yang lebih responsif terhadap kampanye awal untuk mengurangi biaya dan meningkatkan efektivitas.
4. **Analisis Faktor Lain:**
 - o Tambahkan variabel lain seperti balance, duration, atau age dalam model regresi untuk mengeksplorasi faktor-faktor yang lebih signifikan.

Thanks !

Any questions?

<https://www.linkedin.com/in/hijirdella/>

<https://github.com/hijirdella>

hijirdw@gmail.com

