

Exploratory Data Analysis

Laporan
Homework



Dataset

Deskripsi :

Deposito berjangka adalah investasi tunai yang disimpan di bank. Kampanye pemasaran melalui telepon menjadi salah satu cara efektif untuk menjangkau nasabah, namun mereka membutuhkan investasi besar karena pusat panggilan besar disewa untuk melaksanakan kampanye. Oleh karena itu, sangat penting untuk mengidentifikasi nasabah yang kemungkinan besar akan berkonversi terlebih dahulu sehingga mereka dapat ditargetkan secara khusus melalui panggilan.

Dataset :

Memprediksi pelanggan yang berpotensi untuk men-deposito uangnya (berlangganan) atau tidak (tidak berlangganan)

Data :

Setiap satu baris data mewakili satu nasabah bank, satu kolom berisi data dari nasabah tersebut



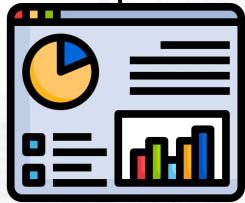
Dataset

Kolom	Deskripsi (Mewakili)
Age	Usia nasabah bank
Job	Pekerjaan nasabah bank
Marital	Status pernikahan nasabah bank
Education	Pendidikan terakhir nasabah bank
Default	Apakah nasabah mempunyai saldo di rekening ?
Balance	Saldo tahunan nasabah bank (Euro)
Housing	Apakah nasabah memiliki pinjaman rumah ?
Loan	Apakah nasabah memiliki pinjaman pribadi ?
Contact	Jenis komunikasi kontak

Dataset

Kolom	Deskripsi (Mewakili)
Day	Hari kontak terakhir dalam sebulan
Month	Bulan kontak terakhir dalam setahun
Duration	Durasi kontak terakhir (detik)
Campaign	Jumlah kontak yang dilakukan selama kampanye untuk nasabah bank
Pdays	Jumlah hari yang berlalu setelah klien terakhir dihubungi dari kampanye sebelumnya (numerik, -1 berarti klien tidak dihubungi sebelumnya)
Previous	Jumlah kontak yang dilakukan sebelum kampanye untuk nasabah bank
Poutcome	Hasil dari kampanye pemasaran sebelumnya
Y	Apakah klien sudah pernah melakukan deposito di bank ?

STAGE 1 - *EDA, INSIGHT & Visualisasi*



Eksplorasi Data



Exploratory Data Analysis



Insight Bisnis dan Visualisasi



EXPLORATORY DATA ANALYSIS

Descriptive Statistics

Sekilas tentang dataset :

Terdapat 17 kolom (16 kolom fitur dan 1 kolom target)

```
df = pd.read_csv('banking_dataset_train.csv', delimiter = ";")  
df.sample(5)
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
23547	32	management	married	tertiary	no	0	no	no	cellular	28	aug	15	13	-1	0	unknown	no
16662	40	blue-collar	married	secondary	no	3131	yes	no	cellular	24	jul	401	1	-1	0	unknown	no
11145	48	management	single	tertiary	no	0	no	no	unknown	18	jun	96	3	-1	0	unknown	no
24101	31	admin.	married	secondary	no	352	no	no	telephone	28	oct	60	1	-1	0	unknown	no
2632	52	admin.	divorced	secondary	no	26	yes	no	unknown	13	may	215	1	-1	0	unknown	no

Descriptive Statistics

```
df.info()

<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
```

Terdapat 2 jenis tipe data, yaitu **int64** dan **object**. Semua tipe data sudah sesuai, dengan kolom **fitur**.

Kolom **categorical** dengan tipe data **object**, dan kolom **numerical** dengan tipe data **int64**, karena kolom **numerical mempunyai nilai yang bulat**.

Descriptive Statistics

```
df.isna().sum()/len(df)*100  
#dalam bentuk percentage  
  
age           0.0  
job           0.0  
marital       0.0  
education     0.0  
default        0.0  
balance        0.0  
housing        0.0  
loan           0.0  
contact        0.0  
day            0.0  
month          0.0  
duration       0.0  
campaign       0.0  
pdays          0.0  
previous       0.0  
poutcome       0.0  
y               0.0  
dtype: float64
```

Berdasarkan df.isna() diketahui bahwa **tidak ada data yang kosong** pada dataset yang digunakan, sehingga nilai persentase missing value terhadap keseluruhan data adalah 0

Descriptive Statistics

```
df[nums].describe()
```

	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

- Berdasarkan hasil perhitungan statistika, beberapa variabel nilai mean **lebih besar** dari median-nya, mengindikasikan bahwa **grafik distribusi frekuensi menceng kanan atau kemencenggan positif**.
- Dalam hal nilai minimum, kolom 'balance' memiliki nilai **negatif (-8019)**, yang mungkin tidak sesuai untuk saldo rekening bank.
- Kolom 'duration' memiliki nilai **minimum 0**, yang mungkin tidak sesuai untuk durasi panggilan. Hal ini mungkin menunjukkan panggilan yang terlewat atau masalah lain.
- Kolom pdays memiliki nilai **minimum -1**, nilai tersebut merupakan representasi nasabah yang belum pernah dihubungi di campaign sebelumnya.

Descriptive Statistics

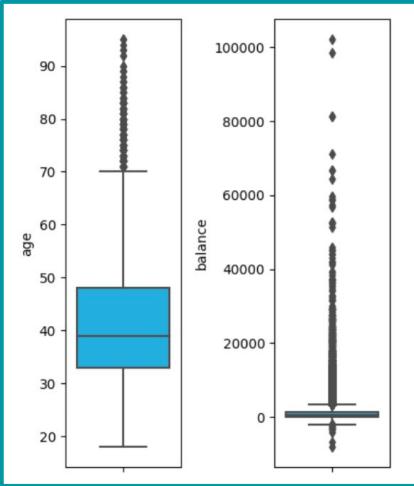
```
df[cats].describe()
```

	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

- Berdasarkan observasi kolom-kolom categorical, kebanyakan dari nasabah bank adalah orang yang **memiliki pekerjaan “blue-collar”** atau para pekerja kasar, yang **sudah menikah dengan pendidikan menengah**.
- Para nasabah tersebut sebagian besar **memiliki pinjaman rumah**, yang dapat **dihubungi lewat telepon seluler**. Namun, sebagian besar dari para nasabah tersebut **tidak mendepositkan uang mereka pada bank sebelumnya**.
- Dalam hal unique data, **tidak ada data yang terlalu beragam** (karena tidak ada data id), sehingga masih bagus untuk digunakan semua.
- Dalam hal frequency, variabel '**default**' **memiliki jumlah nilai "no" yang terlalu banyak**, hal ini juga terjadi pada variabel 'month', 'poutcome' dan 'y' yang cukup ada ketimpangan data.

Univariate Analysis

Boxplot - Numerical Columns



"day":

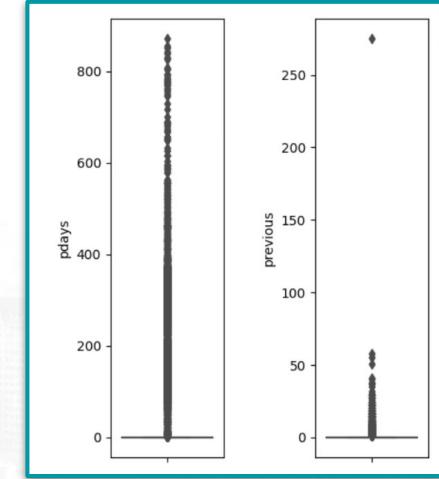
- Secara umum, distribusi data cenderung **normal** tanpa terlihat outliers.

"duration":

- **50%** durasi panggilan **sekitar 100 hingga 400 detik**
- Terdapat banyak **outlier dan nilai ekstrem** pada sisi atas distribusi

"campaign":

- **50% data** terpusat pada jumlah campaign antara **1 hingga 4**.
- Terdapat **banyak outlier dan nilai ekstrem** pada sisi atas distribusi, menunjukkan adanya beberapa responden dengan jumlah kontak yang sangat tinggi.

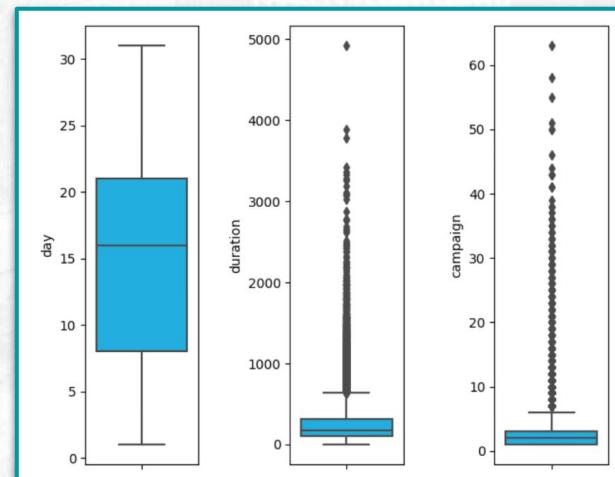


"age" :

- **50%** dari keseluruhan data terpusat pada usia antara **33 hingga 48 tahun**.
- Terdapat **beberapa outlier** pada sisi atas distribusi dengan usia **lebih dari 70 tahun**.

"balance" :

- **Sebagian besar responden** memiliki saldo di **bawah 1428**.
- Terdapat outlier ekstrem pada sisi atas distribusi, menunjukkan adanya responden dengan saldo rekening yang sangat tinggi.
- Saat pre-processing, disarankan untuk mengubah **data minus menjadi 0**, dan **menghapus** atau **mentransformasi** data untuk menangani **outliers** yang ekstrem



"pdays":

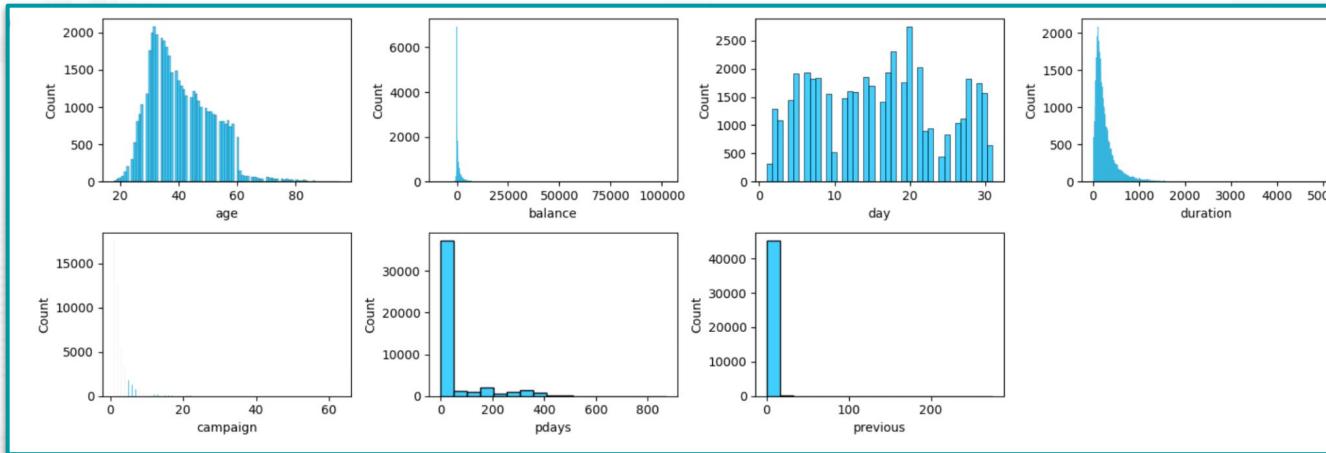
- Terdapat **banyak nilai ekstrem dan outlier** pada sisi atas distribusi
- Saat melakukan **pre-processing data**, nilai **-1** dapat diganti dengan nilai yang lebih bermakna seperti **0** untuk menandai klien yang tidak pernah dihubungi sebelumnya.

"previous":

- Terdapat **beberapa outlier dan nilai ekstrem** yang sangat tinggi pada sisi atas distribusi.
- Saat melakukan **pre-processing data**, dapat dilakukan **penghapusan outlier** yang berada di luar kisaran nilai yang masuk akal.

Univariate Analysis

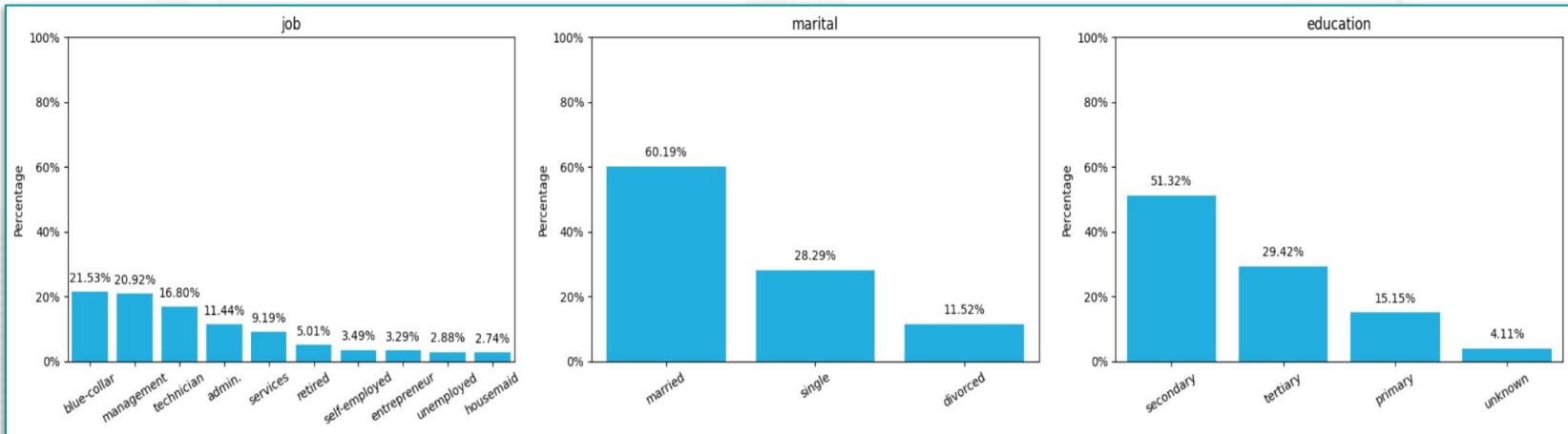
Histplot - Numerical Columns



- "age": Distribusi **cukup normal**, tidak ada indikasi skewness yang signifikan.
- "balance": Distribusi sangat **skew ke kanan (positively skewed)**. Perlu dilakukan **penghapusan outlier** dengan **transformasi data (log transformation)** atau **IQR**.
- "day": Distribusi **tidak menunjukkan karakteristik** yang mencolok.
- "duration": Distribusi sangat **skew ke kanan**. Perlu dilakukan **penanganan outlier** dengan **transformasi data (log transformation)** atau **IQR**.
- "campaign": Distribusi cenderung **positively skewed**, dengan sebagian besar nasabah menerima panggilan dalam jumlah yang sedikit. Terdapat **nilai maksimum** yang **jauh lebih tinggi** dari nilai-nilai lainnya, menunjukkan adanya beberapa nasabah yang menerima panggilan kampanye dalam jumlah yang sangat banyak.
- "pdays": Distribusi nilai pdays sangat **skew ke kanan**. Saat melakukan pra-pemrosesan data, **nilai -1** dapat diganti dengan nilai yang lebih bermakna seperti **NaN** untuk menandai klien yang tidak pernah dihubungi sebelumnya.
- "previous": Distribusi sangat **skew ke kanan**. Perlu dilakukan **penghapusan outlier** yang berada di luar kisaran nilai.

Univariate Analysis

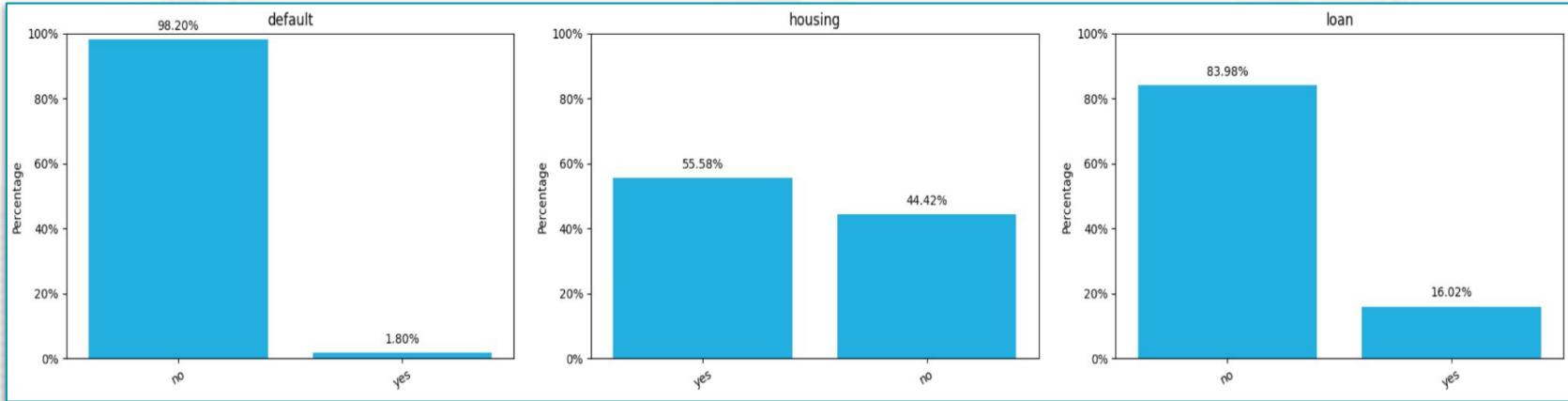
Barplot - Categorical Columns



- Feature "job":
 - **Blue-collar** dan **management** menjadi pekerjaan paling **umum** dalam dataset ini.
- Feature "marital":
 - Mayoritas responden (60%) dalam dataset ini adalah yang sudah **menikah**.
- Feature "education":
 - Sekitar **setengah** dari responden memiliki pendidikan tingkat menengah (**secondary**)

Univariate Analysis

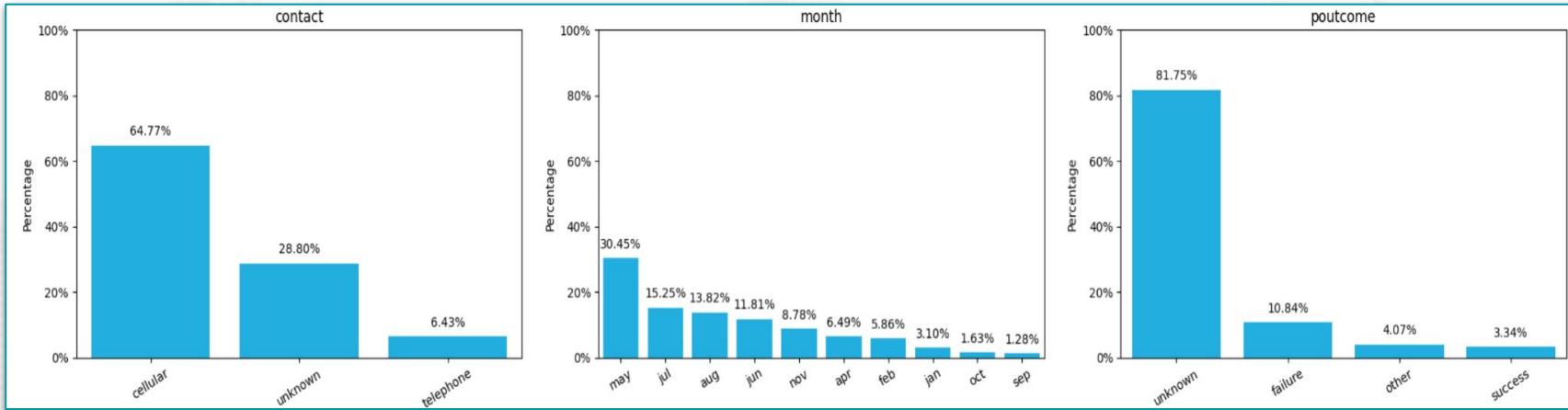
Barplot - Categorical Columns



- Feature "default":
 - Mayoritas responden (98%) **tidak memiliki masalah default** pada pinjaman atau kredit macet.
- Feature "housing":
 - Lebih dari setengah responden (55%) memiliki **kepemilikan rumah (housing)**.
- Feature "loan":
 - Mayoritas responden (83%) **tidak memiliki pinjaman**.

Univariate Analysis

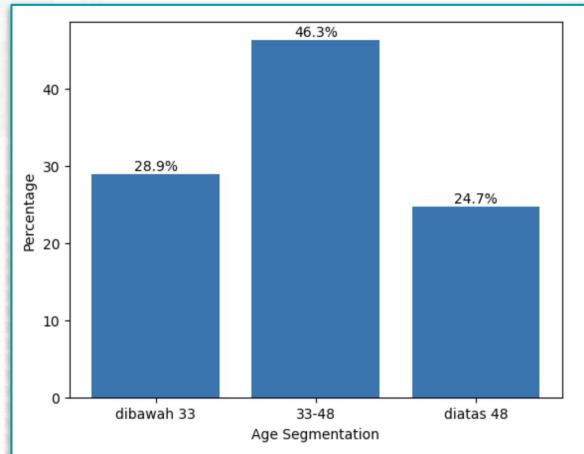
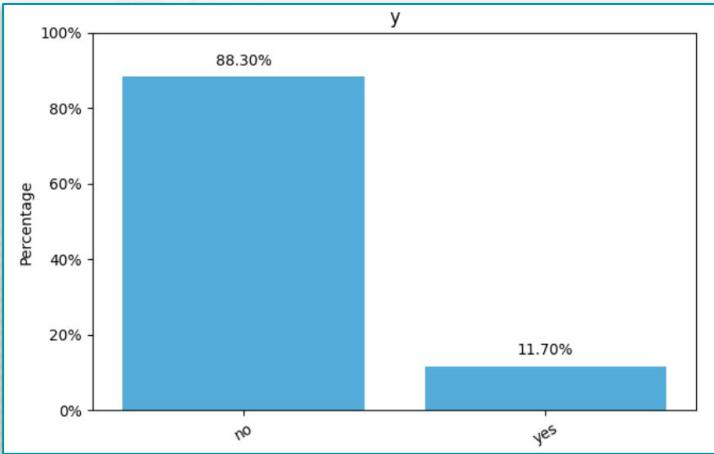
Barplot - Categorical Columns



- Feature "contact":
 - Sebagian besar komunikasi (64%) dilakukan melalui **telepon seluler (cellular)**
- Feature "month":
 - Nasabah paling banyak dihubungi pada bulan **Mei (30%)**, diikuti oleh bulan **Juli, Agustus** dan **Juni**
 - Saat **data pre-processing**, pertimbangkan untuk **menggabungkan** bulan-bulan ke dalam kategori yang lebih luas (**misalnya musim atau kuartal**) untuk analisis atau pemodelan yang lebih baik.
- Feature "poutcome":
 - Sebagian besar responden(80%) memiliki hasil pemasaran sebelumnya yang tidak diketahui (**unknown**).

Univariate Analysis

Barplot - Categorical Columns

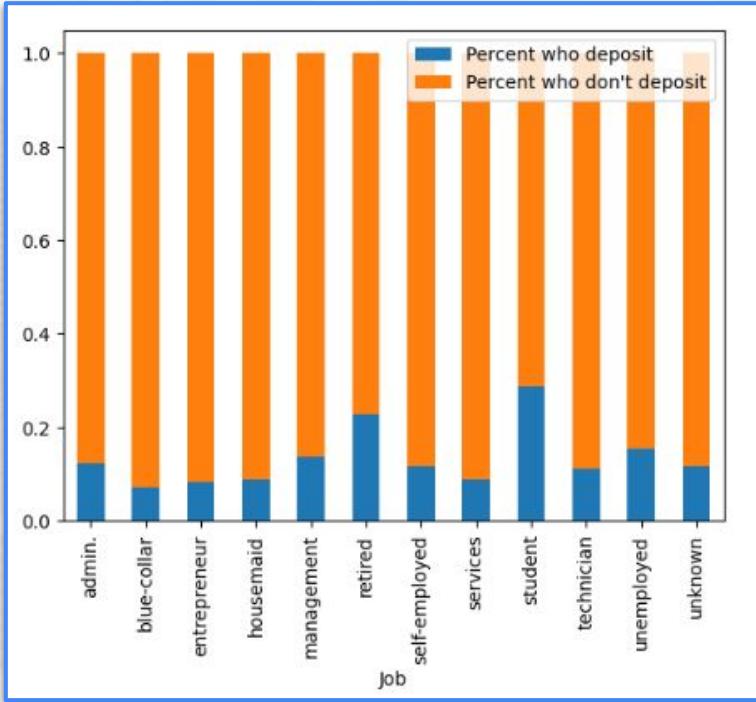


- Feature "y":
 - Mayoritas (88%) responden **tidak berlangganan deposito**.
 - Saat **data pre-processing**, memastikan pemrosesan kelas target seperti mengubah **nilai "yes"** menjadi **1** dan **"no"** menjadi **0**,
 - selanjutnya berdasarkan data yang **imbalanced**, perlu dipertimbangkan juga untuk melakukan **undersampling**.
- Feature "age segmentation":
 - Mayoritas (46%) responden berada di rentang **umur 33-48 tahun**

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

JOB



Job	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
admin.	4540	631	12.202669	87.797331
blue-collar	9024	708	7.274969	92.725031
entrepreneur	1364	123	8.271688	91.728312
housemaid	1131	109	8.790323	91.209677
management	8157	1301	13.755551	86.244449
retired	1748	516	22.791519	77.208481
self-employed	1392	187	11.842939	88.157061
services	3785	369	8.883004	91.116996
student	669	269	28.678038	71.321962
technician	6757	840	11.056996	88.943004
unemployed	1101	202	15.502686	84.497314
unknown	254	34	11.805556	88.194444

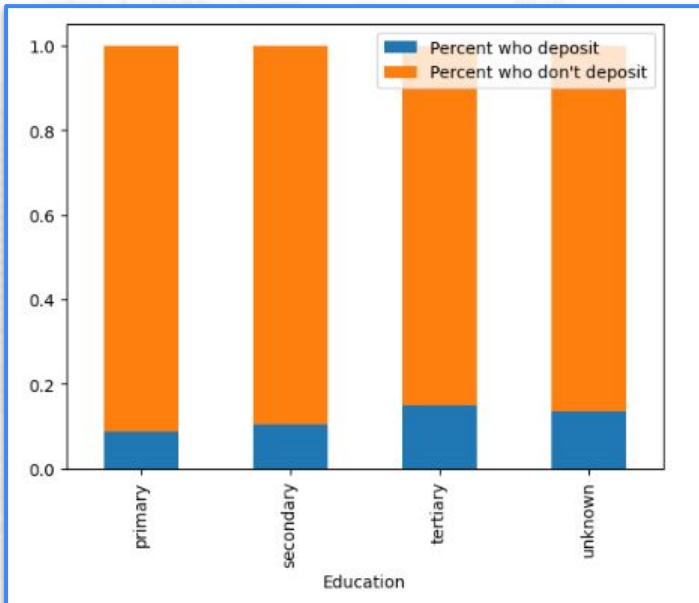
Insight:

Berdasarkan **conversion-ratanya**, persentase **tertinggi** nasabah yang berlangganan deposit memiliki pekerjaan sebagai **pelajar (28.67%)**.

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

EDUCATION



Education	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
primary	6260	591	8.626478	91.373522
secondary	20752	2450	10.559435	89.440565
tertiary	11305	1996	15.006390	84.993610
unknown	1605	252	13.570275	86.429725

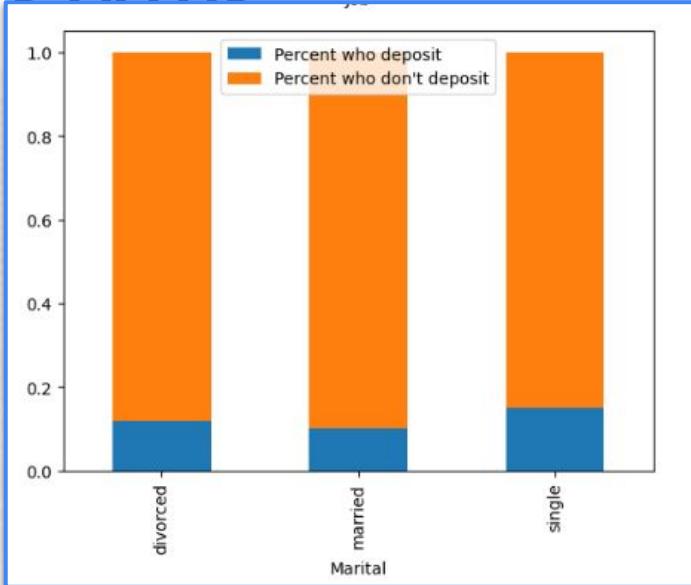
Insight:

- Berdasarkan **conversion-ratanya**, pendidikan **tertiary** memiliki **persentase tertinggi (15%)**.
- Semakin tinggi pendidikan** nasabah, peluang **convertnya** akan **semakin tinggi**.

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

MARITAL STATUS



Marital	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
divorced	4585	622	11.945458	88.054542
married	24459	2755	10.123466	89.876534
single	10878	1912	14.949179	85.050821

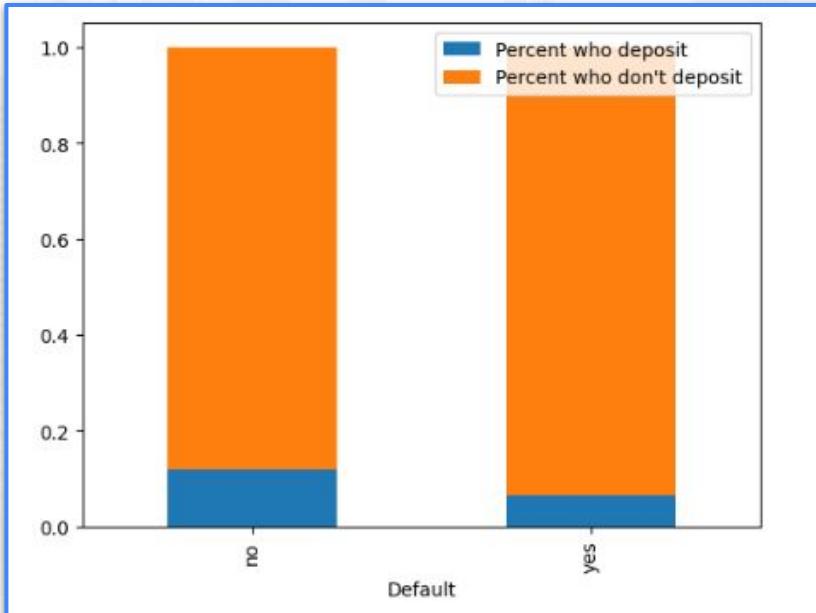
Insight:

Nasabah dengan status '**single**' memiliki persentase conversion rate **paling tinggi (14.95%)**, meskipun selisihnya tidak terlihat signifikan.

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

DEFAULT



	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
Default				
no	39159	5237	11.796108	88.203892
yes	763	52	6.380368	93.619632
Housing and Y relationship				

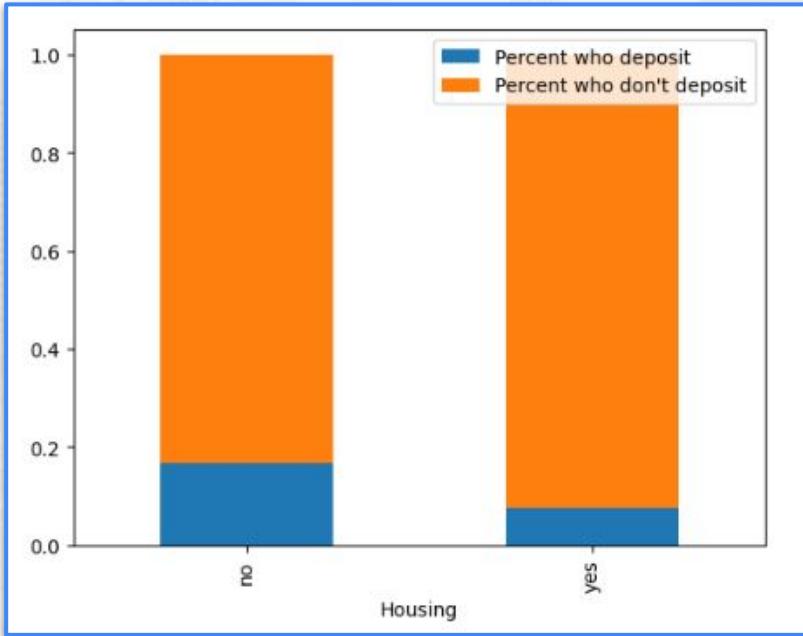
Insight:

Nasabah yang **tidak memiliki masalah dengan default** atau kredit macet lebih banyak yang **berlangganan deposit (11.79%)** daripada yang terkena masalah.

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

HOUSING



	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
Housing				
no	16727	3354	16.702355	83.297645
yes	23195	1935	7.699960	92.300040

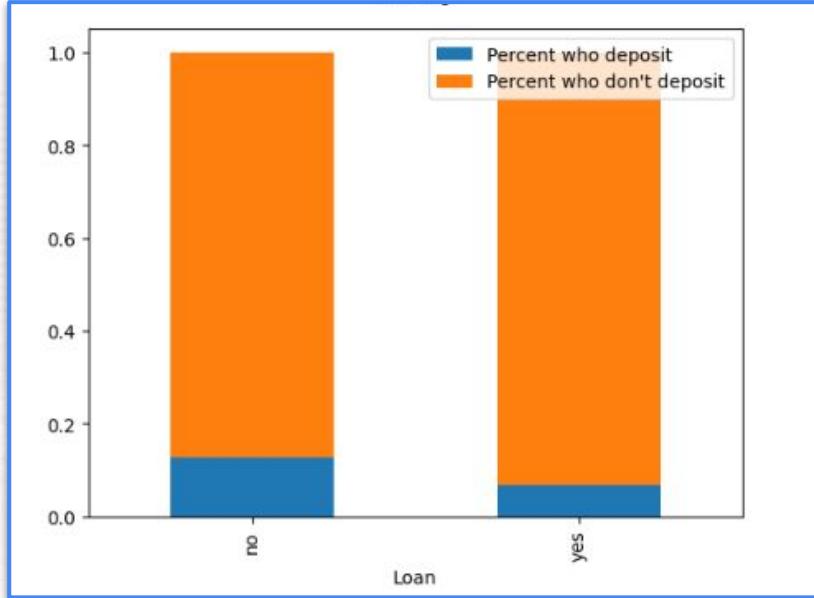
Insight:

- Nasabah yang **belum pernah mengajukan KPR** cenderung lebih banyak berlangganan produk deposit (**16.70%**).
- Conversion rate nasabah yang **tidak memiliki KPR** memiliki nilai lebih dari **2 kali lipat** dari yang memiliki KPR

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

LOAN



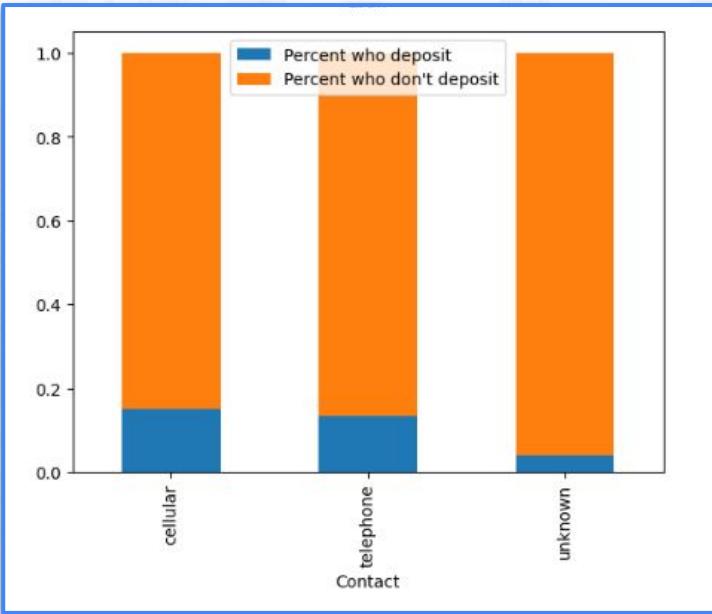
Loan	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
no	33162	4805	12.655727	87.344273
yes	6760	484	6.681391	93.318609

Insight:

- Nasabah yang belum pernah mengajukan pinjaman pribadi lebih tinggi persentase berlanggannya (**12.65%**).
- Perbedaan persentase yang tidak memiliki pinjaman hampir 2 kali lipat dari nasabah yang memiliki pinjaman.

Multivariate Analysis

CONTACT



Contact	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
cellular	24916	4369	14.918900	85.081100
telephone	2516	390	13.420509	86.579491
unknown	12490	530	4.070661	95.929339

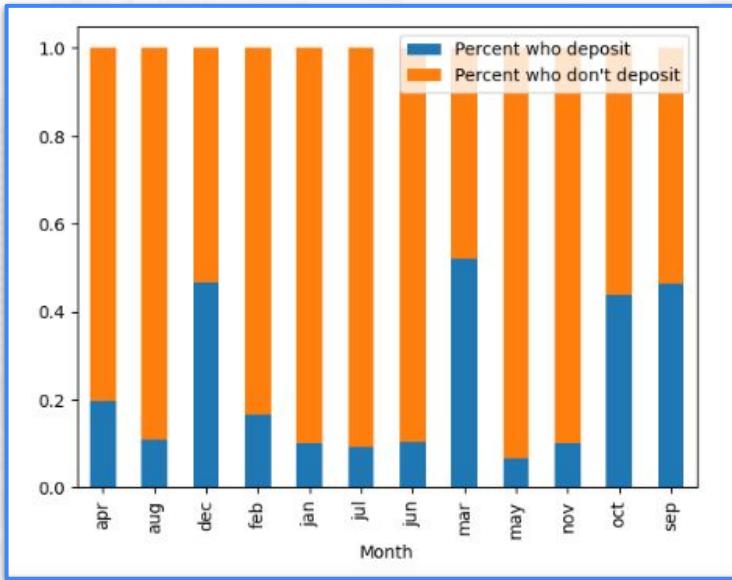
Insight:

- Jenis kontak melalui **seluler** memiliki **persentase tertinggi (14.91%)**.
- Meskipun tidak berbeda jauh dengan telepon, namun perbedaannya cukup signifikan dibandingkan dengan metode lain atau unknown.

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

MONTH



Month	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
apr	2355	577	19.679400	80.320600
aug	5559	688	11.013286	88.986714
dec	114	100	46.728972	53.271028
feb	2208	441	16.647792	83.352208
jan	1261	142	10.121169	89.878831
jul	6268	627	9.093546	90.906454
jun	4795	546	10.222805	89.777195
mar	229	248	51.991614	48.008386
may	12841	925	6.719454	93.280546
nov	3567	403	10.151134	89.848866
oct	415	323	43.766938	56.233062
sep	310	269	46.459413	53.540587

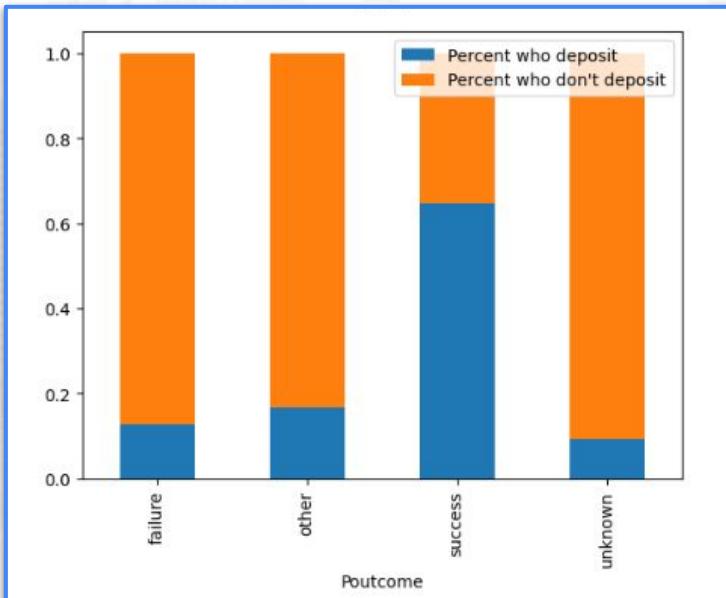
Insight:

- Bulan **Maret** menjadi bulan terakhir kontak dengan **hasil konversi berlangganan tertinggi (51.99%)**.
- Bulan September, Oktober dan Desember juga cukup tinggi; sisanya rendah.

Multivariate Analysis

Stacked Bar Chart - Categorical Columns

P-OUTCOME



Poutcome	Do not deposit	Deposit	Percent who deposit	Percent who don't deposit
failure	4283	618	12.609671	87.390329
other	1533	307	16.684783	83.315217
success	533	978	64.725347	35.274653
unknown	33573	3386	9.161503	90.838497

Insight:

Hasil **sukses** pada campaign terakhir menjadi tanda **nasabah yang berlangganan deposit (64.72%)**, memiliki persentase tertinggi dan selisih yang signifikan dari hasil lainnya.

Multivariate Analysis

Heatmap - Numerical Columns

	age	balance	day	duration	campaign	pdays	previous
age	1.000000	0.097783	-0.009120	-0.004648	0.004760	-0.023758	0.001288
balance	0.097783	1.000000	0.004503	0.021560	-0.014578	0.003435	0.016674
day	-0.009120	0.004503	1.000000	-0.030206	0.162490	-0.093044	-0.051710
duration	-0.004648	0.021560	-0.030206	1.000000	-0.084570	-0.001565	0.001203
campaign	0.004760	-0.014578	0.162490	-0.084570	1.000000	-0.088628	-0.032855
pdays	-0.023758	0.003435	-0.093044	-0.001565	-0.088628	1.000000	0.454820
previous	0.001288	0.016674	-0.051710	0.001203	-0.032855	0.454820	1.000000

OBSERVATIONS:

Feature 'pdays' dan 'previous' memiliki korelasi yang masuk dalam kategori **moderate correlation (0.45)**, feature lainnya berkorelasi lemah / sangat lemah.

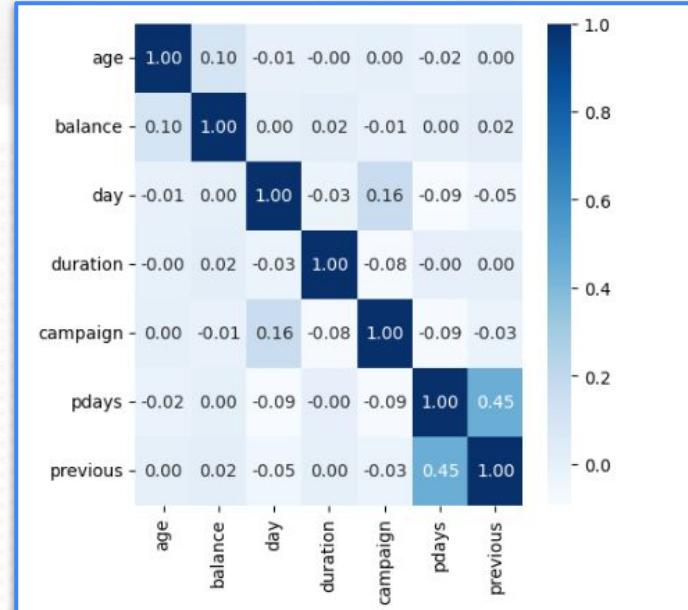
FEATURES SELECTION:

1. Feature yang perlu dipertahankan:

- Previous

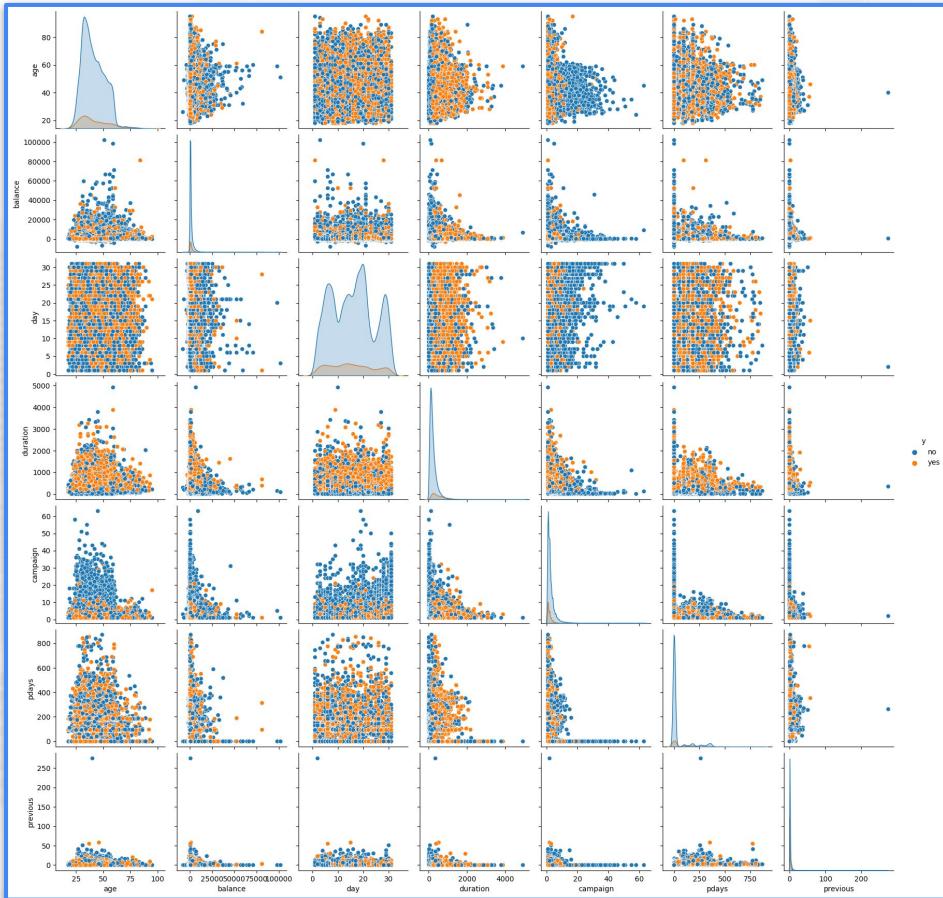
2. Feature yang akan ditakeout:

- Pdays : Memiliki banyak **outliers** yang ekstrim, memiliki nilai **majoritas -1**, memiliki **standar deviasi** yang sangat besar ketimbang 'previous'!
- Duration: Memiliki banyak nilai 0 (masih dalam pertimbangan)



Multivariate Analysis

Scatter Plot - Numerical Columns vs Y



OBSERVATIONS:

- Korelasi feature 'age' dan 'duration': Semakin lama durasi, kemungkinan deposit menjadi semakin tinggi tidak peduli berapapun usianya.
- Korelasi feature 'day' dan 'duration': Semakin lama durasi semakin tinggi kemungkinan nasabah untuk deposit tidak peduli di hari apa kontak terakhir dilakukan.
- Korelasi feature 'age' dan 'campaign': Semakin banyak campaign, semakin kecil kecenderungan nasabah untuk berlangganan deposit tidak peduli berapapun usia nasabah.
- Sisanya tidak menunjukkan hubungan dan pola yang menarik terhadap label target.

CONCLUSION:

Selain feature 'pdays' dan 'duration', sisanya cukup relevan untuk dipertahankan.

3. Business Insight & Recommendation

1. Segmentasi dan Personalisasi.

- Fitur 'job', 'default', 'loan', 'housing' dan 'contact' mengindikasikan bahwa kita bahwa nasabah yang **berprofesi sebagai pelajar atau pensiunan yang tidak memiliki kredit macet, pinjaman pribadi, dan KPR** jauh lebih tinggi kemungkinannya untuk berlangganan deposito berjangka yang ditawarkan melalui telepon maupun seluler.

Recommendation:

- Targetkan upaya marketing kepada kelompok nasabah yang tidak memiliki masalah atau beban pinjaman apapun.
- Untuk target pelajar, buatkan program khusus dan lakukan kerja sama dengan institusi pendidikan.
- Untuk pensiunan, ciptakan program tabungan pensiun dengan resiko yang lebih minim dan suku bunga yang lebih besar, serta jangka waktu investasi yang lebih singkat.
- Metode kontak bisa dilakukan melalui seluler atau telepon karena masih sangat efektif dari segi conversion rate..

3. Business Insight & Recommendation

2. Analisis Mendalam Durasi Panggilan.

- Hubungan antara fitur 'duration' dengan fitur 'age' dan 'day' menunjukkan bahwa **semakin lama durasi panggilan, lebih tinggi kemungkinan dan cenderung berlangganan deposito berjangka**, hal ini mungkin karena nasabah yang tertarik dengan deposito akan lebih banyak bertanya dan melakukan percakapan lebih lama sebelum akhirnya berlangganan.

Recommendation:

- Fokuskan pada peningkatan kualitas dan efisiensi waktu interaksi pelanggan selama panggilan.
- Buatlah guidelines dan skenario interaksi pelanggan yang berfokus pada elemen-elemen yang telah terbukti berhasil (misal menargetkan untuk pembicaraan maksimal 30 menit (1.800 detik) karena berdasarkan persebaran data, cukup banyak nasabah yang tertarik dengan durasi tidak lebih dari 2.000 detik). Dengan ini, diharapkan biaya telepon ataupun seluler juga bisa terpangkas, dan marketer bisa fokus ke nasabah lain tanpa ada pemborosan jumlah marketer yang dibutuhkan.

3. Business Insight & Recommendation

3. Analisis Musiman dan Optimalisasi Kontak.

- Fitur 'month' menunjukkan pola menarik, dengan **beberapa bulan (terakhir dilakukan kontak)** **memiliki tingkat langganan yang jauh lebih tinggi** dibanding yang lainnya.
- Hubungan antara fitur 'age' dan fitur 'campaign' menunjukkan bahwa **semakin banyak kontak** yang dilakukan saat ini berapapun usianya, nasabah cenderung **enggan berlangganan**.
- Fitur 'poutcome' menjadi indikasi bahwa hasil pada campaign sebelumnya mencerminkan keputusan berlangganan.

Recommendation:

- Ekstrapolasi analisis di sekitar bulan Maret, September, Oktober dan Desember seperti analisis peristiwa tertentu atau trend ekonomi yang terkait. Pertimbangkan untuk menyesuaikan strategi marketing dan alokasi sumber daya untuk lebih efektif melakukan kontak di bulan-bulan tersebut dan di bulan sebelumnya seperti Februari, Agustus dan November.
- Evaluasi efektivitas jumlah kontak dan hasil sebelumnya. Targetkan lebih pada nasabah yang pada campaign sebelumnya memang tertarik atau 'sukses' saat dikontak. Batasi kontak pada rentang 2 bulan, jika tidak berhasil beralih ke nasabah lain di bulan lain yang potensial (memiliki konversi tinggi). Hal ini diharapkan bisa mengurangi intensitas dan frekuensi kontak yang bisa berdampak pada berkurangnya biaya investasi campaign dan meningkatkan conversion rate.

3. Business Insight & Recommendation

4 Maksimalkan Penggunaan Machine Learning untuk Beberapa Fitur

- Untuk variabel numerik, variabel '**pdays' berkorelasi dengan 'previous' sehingga 'pdays' kemungkinan akan di takeout**, sedangkan variabel 'duration' masih dalam tahap pertimbangan karena memberikan insight yang cukup menarik terkait karakteristik dan segmentasi calon nasabah. Selebihnya, variabel numerik bisa digunakan untuk analisis ML.
- Untuk variabel kategorikal, semua variabel bisa dipakai untuk analisis ML, namun akan dipertimbangkan untuk **diencoding**, karena variabel '**month','loan', 'housing' dan 'default'** memberikan info penting bagaimana karakteristik nasabah yang berlangganan.

Recommendation:

- Mencoba menganalisis menggunakan ML dengan memasukkan semua variabel numerik yang tersisa ditambah beberapa variabel kategorikal yang akan di One Hot Encoding.
- Diharapkan gabungan dari variabel-variabel yang akan dipakai modelling ML ini bisa menghasilkan rekomendasi hasil ML yang paling efektif dan efisien untuk mengurangi biaya investasi panggilan, campaign dan karyawan serta meningkatkan conversion rate nasabah yang berlangganan.