

# Laporan Final Project

Kelompok 1 - DataWhiz

Nick Wisely (Ketua Kelompok)

Faizah Dhiya Annisa

Annisa' Kunny Latifa

Mohammad Iman A

Nafla Reihan Alya S

Roberta Alfrina S

Naela Adiba

Bunga Puji Lestari

Sonia Anthera Rojak



# STAGE - 0

Preparation

- Problem
- Role
- Goal
- Objective
- Business Metric

# Problem yang mau diselesaikan dari dataset

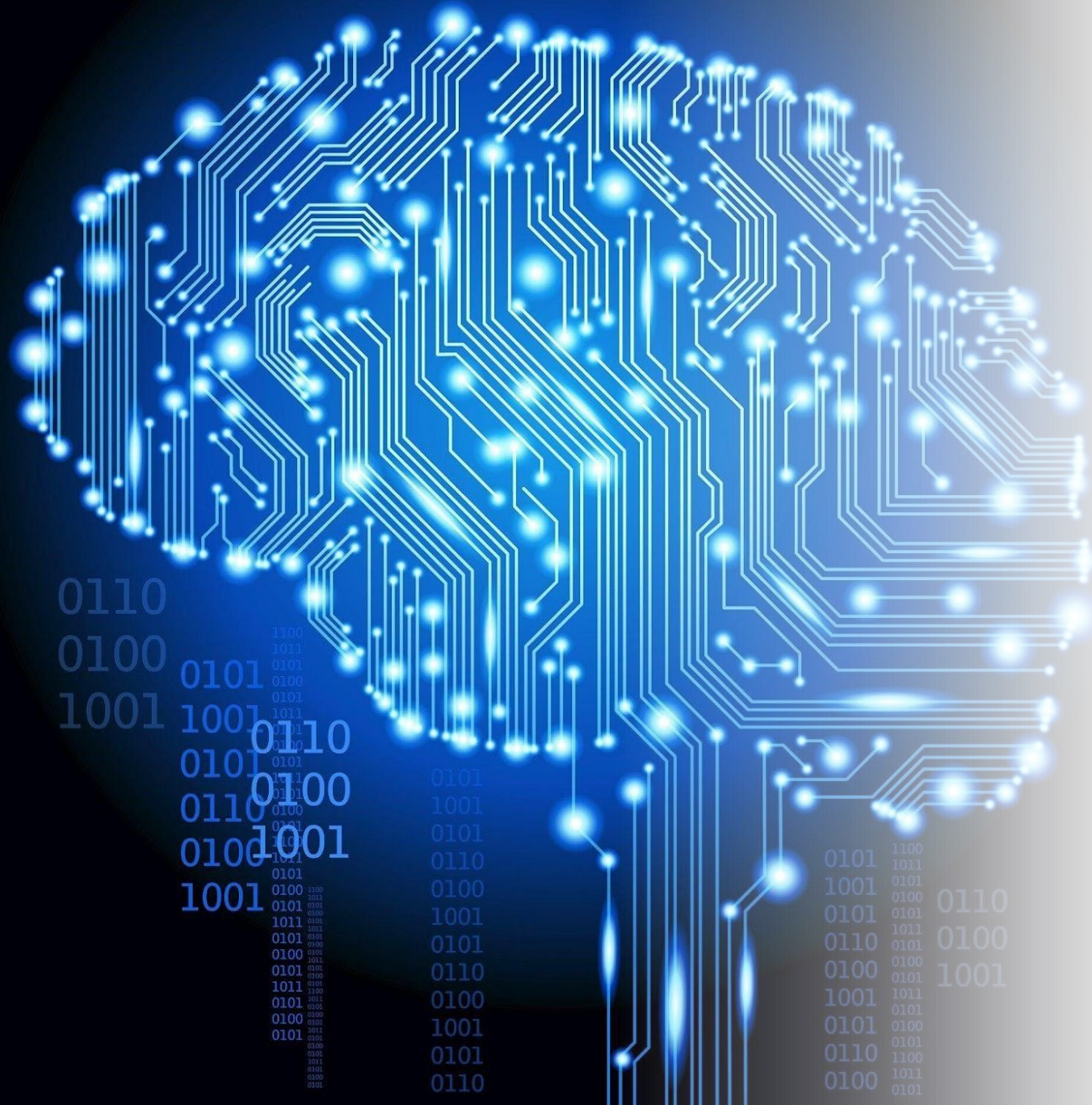


Sebuah perusahaan ingin menjual suatu produk terhadap nasabah bank yaitu deposito. **Deposito** adalah produk menguntungkan bagi nasabah untuk memperoleh bunga yang lebih besar dan bagi bank berupa tabungan dana jangka panjang.



Namun pemasaran melalui telepon memiliki keterbatasan dengan waktu dan jumlah karyawan. Tanpa strategi yang jelas, kemungkinan keberhasilan dari sebuah telepon bersifat acak dan jumlah karyawan perlu untuk ditambahkan, dan berujung pada kenaikan nilai cost.





**Strategi team data untuk  
menentukan calon nasabah  
potensial:**

### **Membuat klasifikasi Machine Learning (ML).**

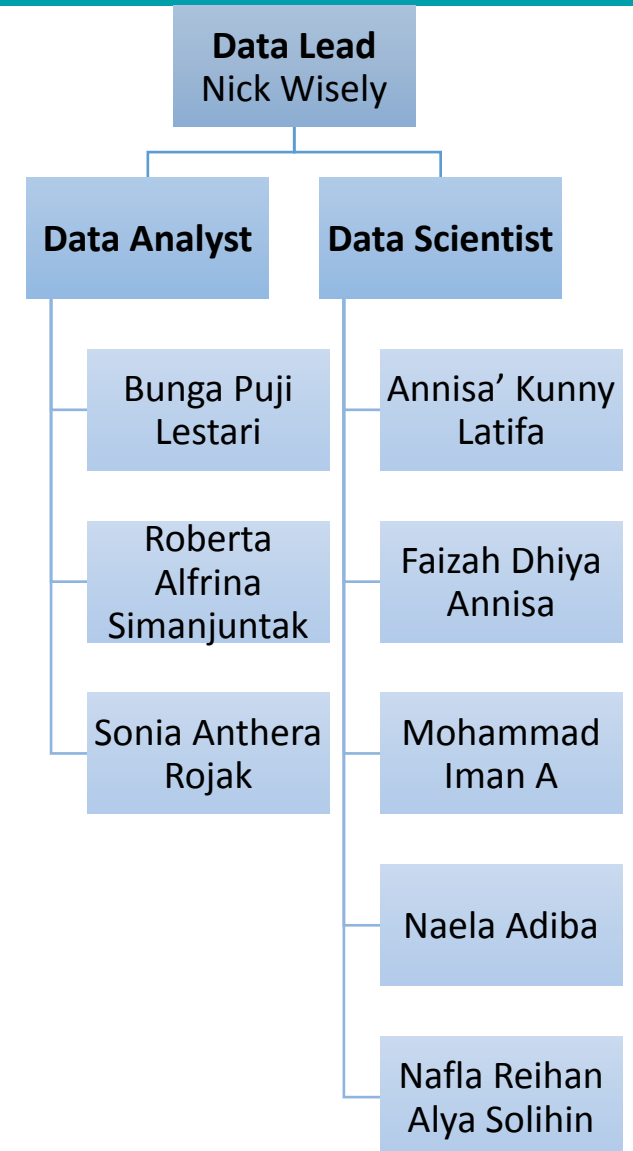
Sehingga dapat membantu tim pemasaran yang memiliki keterbatasan dengan waktu dan jumlah karyawan.

Pencapaian yang ingin diperoleh adalah tingkat produk yang disetujui oleh calon nasabah meningkat, yang juga berdampak pada kenaikan profit perusahaan.

# Sebagai siapa kalian pada dataset



**Tim DataWhiz merupakan divisi dari perbankan swasta di Jakarta, Indonesia.** Tim DataWhiz membantu operasional tim telemarketing yang bekerja dengan waktu 08.00 - 14.00. Gaji per bulan dari seorang sales telemarketing adalah Rp. 4.900.000 (sumber: [link](#)). jumlah karyawan dari tim telemarketing berjumlah 8-12 yang perhari hanya bisa menelpon sebanyak 60-70 nasabah sehingga memaksa untuk menambah personil dan menambah cost seperti penambahan space serta pelatihan yang diberikan



# Goal yang ingin dicapai

## Meningkatkan rasio keberhasilan pembukaan rekening deposito dari setiap nasabah yang di telepon

SPECIFIC

**Specific** : Jumlah karyawan yang dimiliki adalah tetap, namun nasabah yang dihubungi yang akan diatur.

MESSURABLE

**Measureable** : Conversion rate yang ada pada dataset adalah 5289 orang dari 45211 (persentase 11,69%). maka pada tahap ini ditetapkan target sebesar 30% sebagai capaian lanjutan dari penggunaan model machine learning



ACHIEVABLE

**Attainable** : Capaian ditargetkan untuk dicapai dengan membuat model Machine Learning dalam waktu 1 bulan

REALISTIC

**Relevant** : Hasil dari prediksi machine learning, secara langsung berdampak pada kenaikan profit dari perusahaan.

TIMELY

**Time based** : Pembuatan model machine learning dibatasi dalam waktu 1 bulan.

# Apa objective yang sesuai dengan goal

## Membuat Machine Learning dengan jenis Supervised Learning (tipe Classification)



**Mendapatkan wawasan faktor apa yang paling mempengaruhi seseorang untuk mau membuka rekening deposito**

Hal-hal yang bisa jadi mempengaruhi: jumlah saldo, lama pembukaan tabungan, jenis kelamin, dan hal-hal lainnya.

Wawasan ini selain berguna untuk klasifikasi nasabah oleh tim data, dapat juga berguna dalam memberikan masukan untuk input dari tim sales dan marketing untuk dilanjutkan dalam bentuk promo/kampanye yang relevan.

**Teramati adanya kenaikan pada secondary metrics**

Kenaikan nilai pada metrics utama diprediksi akan memberikan kontribusi pada nilai secondary metrics yang lainnya (cost yang menurun, CPH yang meningkat, dan lain sebagainya).

Perubahan dari indikator-indikator tersebut perlu teramati sehingga menjadi objective tambahan pada project ini.



# Business metrics yang cocok untuk mengukur ketercapaian objective



Rasio merupakan banyaknya nasabah yang membuka deposito berjangka per banyaknya nasabah yang menerima campaign.

$$\text{Conversion Rate} = \frac{\text{Amount Of Sales}}{\text{Number of Calls}}$$

Dilansir dari martech data Call Intelligence Index yang dikeluarkan oleh Invoca, conversion rate melalui telepon memiliki persentase antara 30 sampai 50 persen. artinya terdapat gap sekitar 20% dari conversion rate yang dimiliki. nilai gap ini akan menjadi tujuan akhir yang ingin dicapai. Di tahun 2014, sebanyak 40 industri melakukan bisnis melalui telepon sebanyak 32 juta. Analyst menentukan bahwa pemasaran yang dilakukan melalui telepon bertanggung jawab sekitar 54 persen dari semua bisnis. Yang dimana sebanyak 45 persen berasal dari penelusuran secara mobile. [Sumber : [Link](#)]



# STAGE - 1

Exploratory  
Data Analysis

- Descriptive Statistics
- Univariate Analysis
- Multivariate Analysis
- Business Insight
- Git

# Descriptive Statistics

Dataset ini memiliki 45211 baris dan 17 kolom (terdiri atas 7 data numerik dan 10 data kategorik).

## Data numerik

age	usia klien
balance	saldo tahunan rata-rata
day	hari kontak terakhir di bulan itu
duration	durasi kontak terakhir (sekon)
campaign	jumlah kontak yang dilakukan selama kampanye ini untuk klien ini (termasuk kontak terakhir)
pdays	jumlah hari yang berlalu setelah klien terakhir dihubungi dari kampanye sebelumnya
previous	jumlah kontak yang dilakukan sebelum kampanye ini untuk klien ini

# Descriptive Statistics

## Data kategorik

job	Jenis pekerjaan nasabah
marital	Status nasabah
education	Status pendidikan nasabah
default	Status nasabah dalam pembayaran hutang
housing	Kepemilikan rumah
loan	Status kepemilikan Hutang
contact	Tipe kontak nasabah
month	Bulan
poutcome	Hasil dari promosi sebelumnya
y	Status nasabah yang mengikuti deposito

# Descriptive Statistics

- A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?

Untuk kolom, tipe data sudah sesuai. Tetapi sebaiknya penamaan kolom seperti target deposito akan diubah, yang awalnya hanya y akan didiskusikan lebih lanjut perubahannya.

- B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?

Tidak ada. Semua baris dalam setiap kolom memiliki nilai.

- C. Apakah ada kolom yang memiliki nilai summary agak aneh?  
(min/mean/median/max/unique/top/freq)

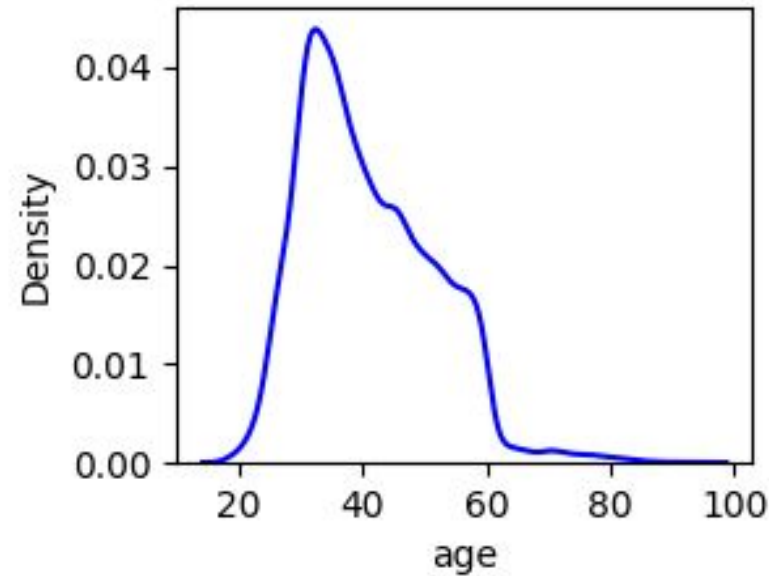
- Untuk kolom balance nilai minimal nasabah memiliki saldo sebesar -8.019 Euro, sedangkan untuk nilai maksimal nasabah memiliki saldo sebesar 102.127 Euro. Yang bisa dikatakan menjadi sebuah outlier yang dimana rata-rata nasabah memiliki saldo sebesar 449 Euro (median).
- Untuk kolom bagian campaign rata-rata nasabah dihubungi sebanyak 2 kali, tetapi ada nasabah yang dihubungi hingga 63 kali.



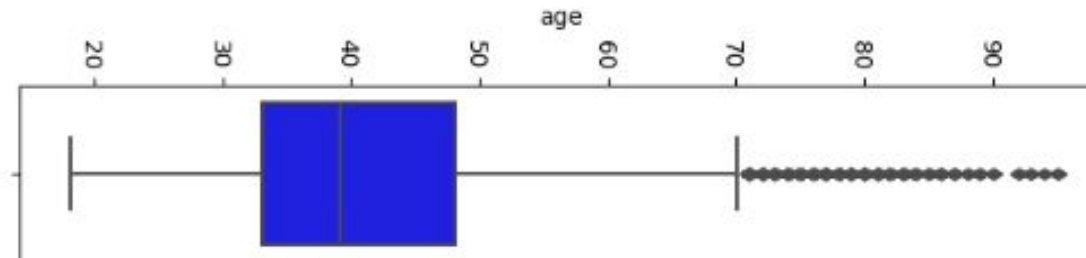
# Univariate Analysis

Feature : Age (Numerik)

	age
count	45211.000000
mean	40.936210
std	10.618762
min	18.000000
25%	33.000000
50%	39.000000
75%	48.000000
max	95.000000



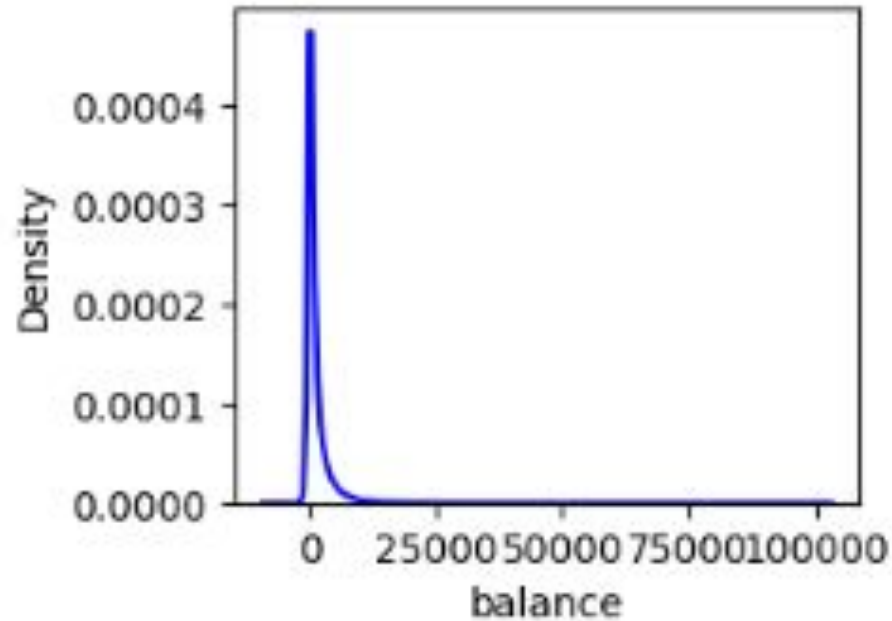
rentang umur berada di angka 18 - 95 tahun, sebaran dari umur bersifat normal.



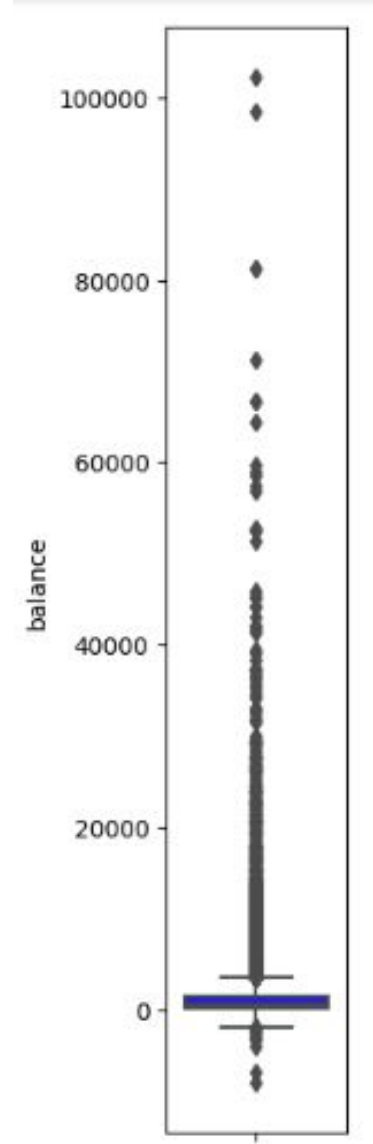
# Univariate Analysis

Feature : Balance (Numerik)

	balance
count	45211.000000
mean	1362.272058
std	3044.765829
min	-8019.000000
25%	72.000000
50%	448.000000
75%	1428.000000
max	102127.000000



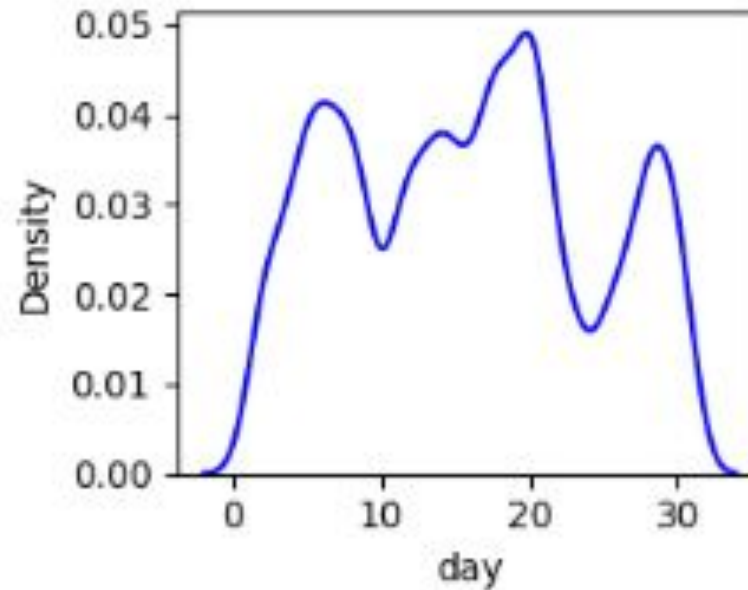
- **terdapat balance yang bernilai negatif (tidak normal).**
- harus ada treatment untuk yang nilainya negatif. Opsi yang paling relevan adalah menghapus data yang bernilai negatif.



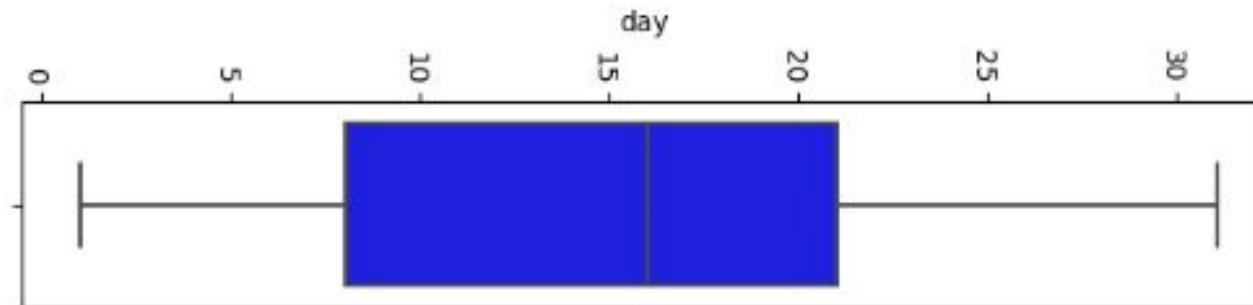
# Univariate Analysis

Feature : Day (Numerik)

	day
count	45211.000000
mean	15.806419
std	8.322476
min	1.000000
25%	8.000000
50%	16.000000
75%	21.000000
max	31.000000



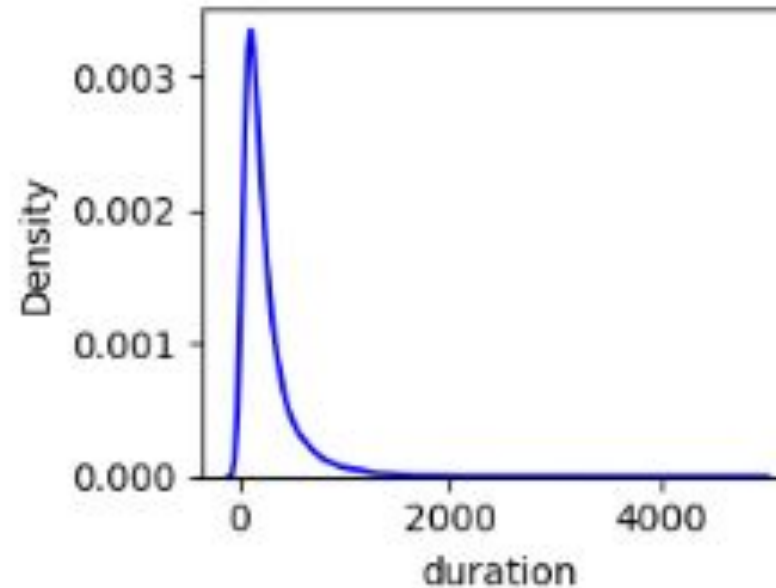
berupa tanggal kontak terakhir dalam suatu bulan, tidak ada tanggal yang bersifat anomali .



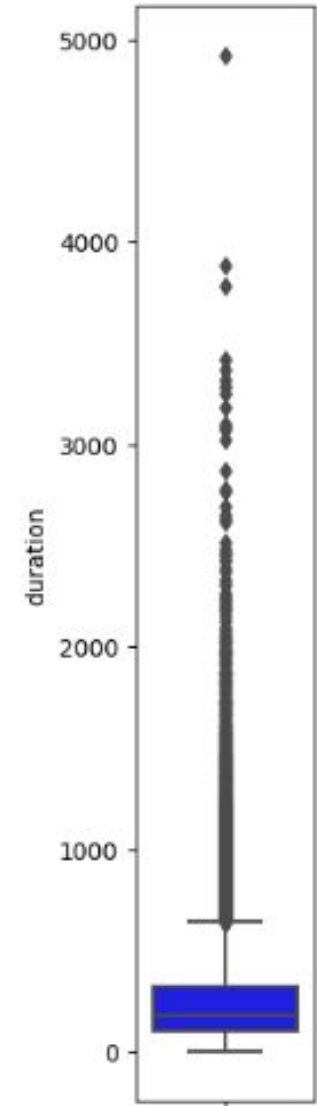
# Univariate Analysis

Feature : Duration (Numerik)

	duration
<b>count</b>	45211.000000
<b>mean</b>	258.163080
<b>std</b>	257.527812
<b>min</b>	0.000000
<b>25%</b>	103.000000
<b>50%</b>	180.000000
<b>75%</b>	319.000000
<b>max</b>	4918.000000



- lamanya durasi pemanggilan (dalam detik).
- ada yang bernilai nol, (bisa diartikan memang tidak diangkat).

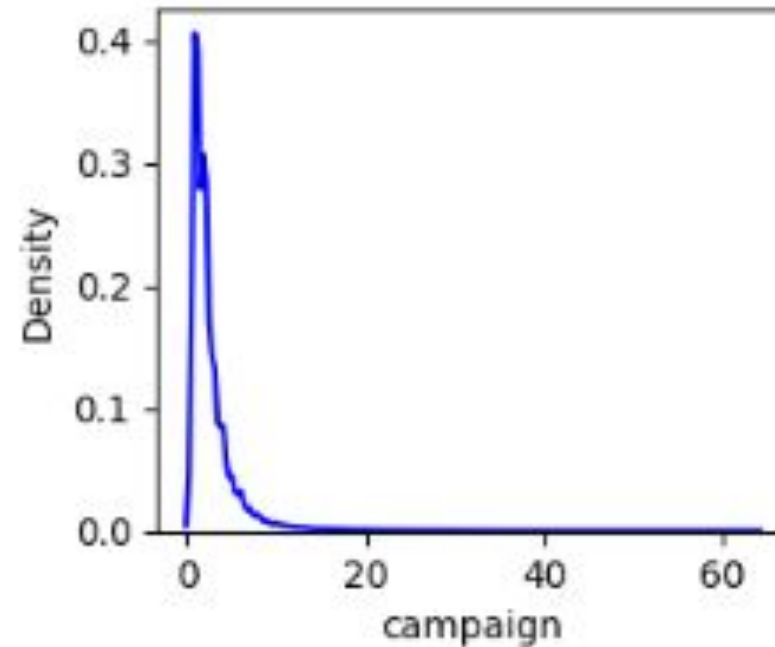




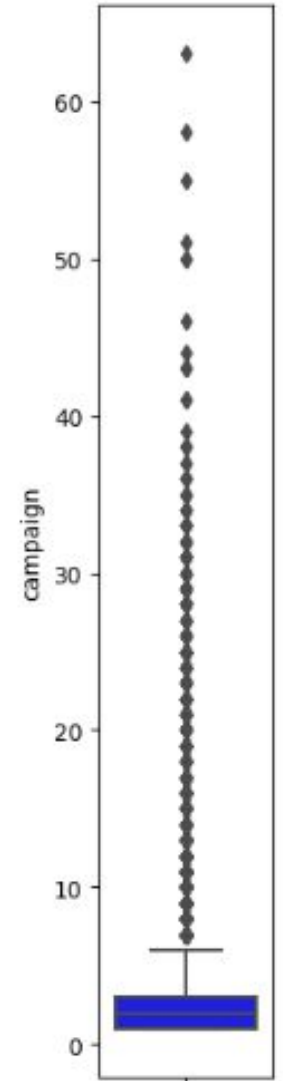
# Univariate Analysis

Feature : Campaign (Numerik)

	campaign
count	45211.000000
mean	2.763841
std	3.098021
min	1.000000
25%	1.000000
50%	2.000000
75%	3.000000
max	63.000000



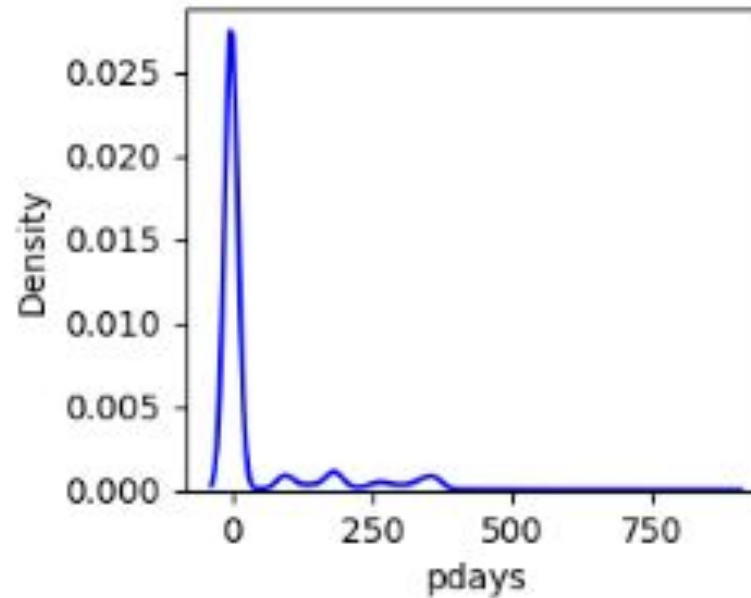
- banyaknya kontak yang dilakukan terhadap clients.
- nilai maksimal hingga 63, bernilai sangat jauh bahkan terhadap Q3. dapat diprediksi bahwa terdapat banyak outlier.



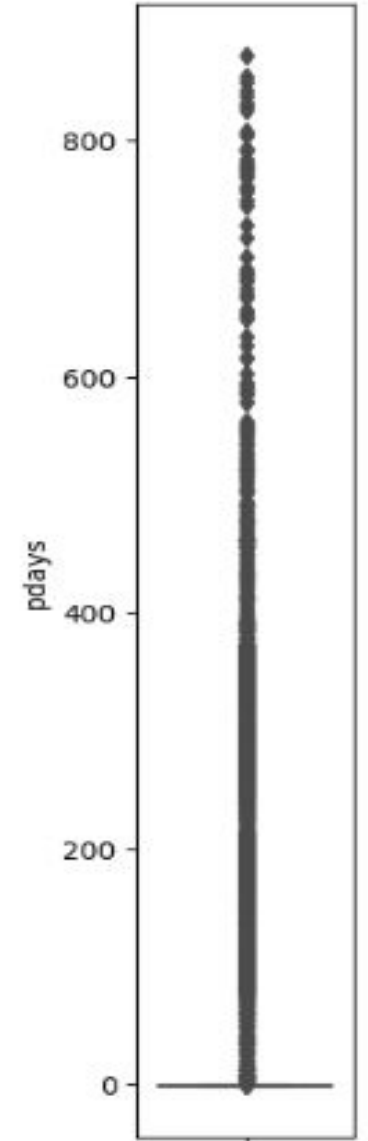
# Univariate Analysis

Feature : Pdays (Numerik)

	pdays
count	45211.000000
mean	40.197828
std	100.128746
min	-1.000000
25%	-1.000000
50%	-1.000000
75%	-1.000000
max	871.000000



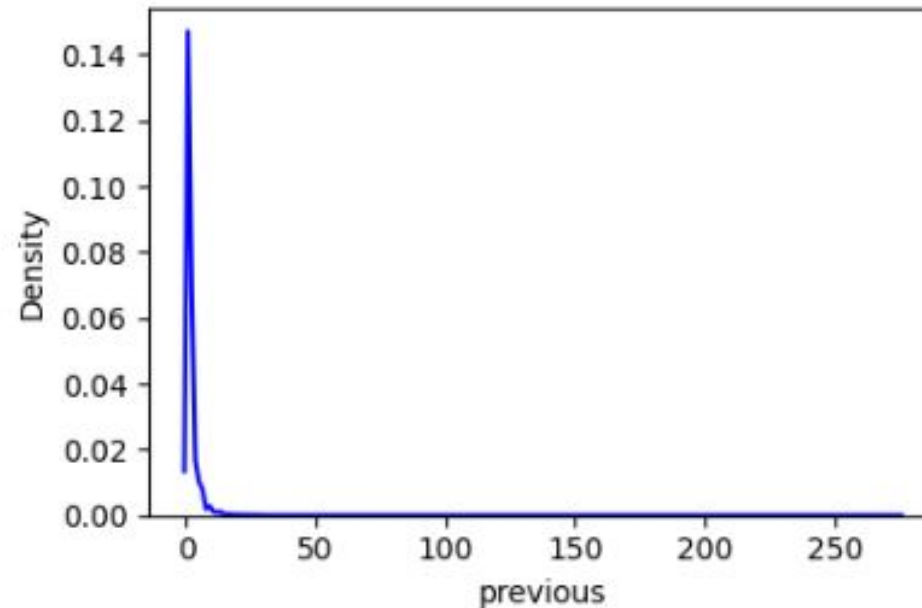
- lamanya waktu (dalam hari) terakhir kali client ini dihubungi.
- nilai -1 artinya contact belum pernah dihubungi sebelumnya.
- **kebanyakan data bernilai -1, artinya banyak orang yang baru dihubungi pertama kali (di atas 75%).**



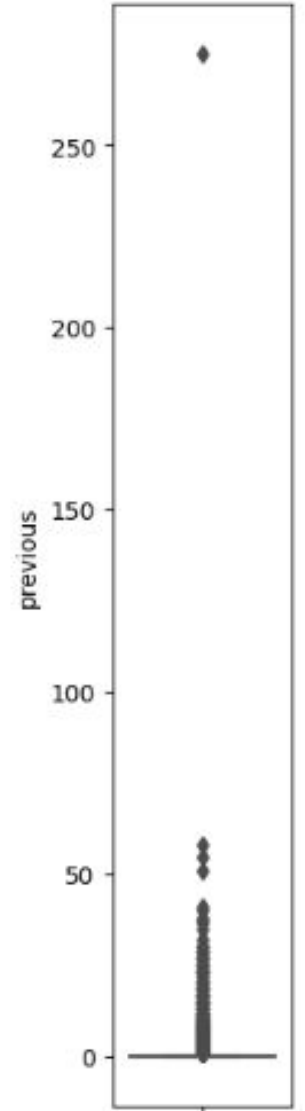
# Univariate Analysis

Feature : Previous (Numerik)

	previous
count	45211.000000
mean	0.580323
std	2.303441
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	275.000000



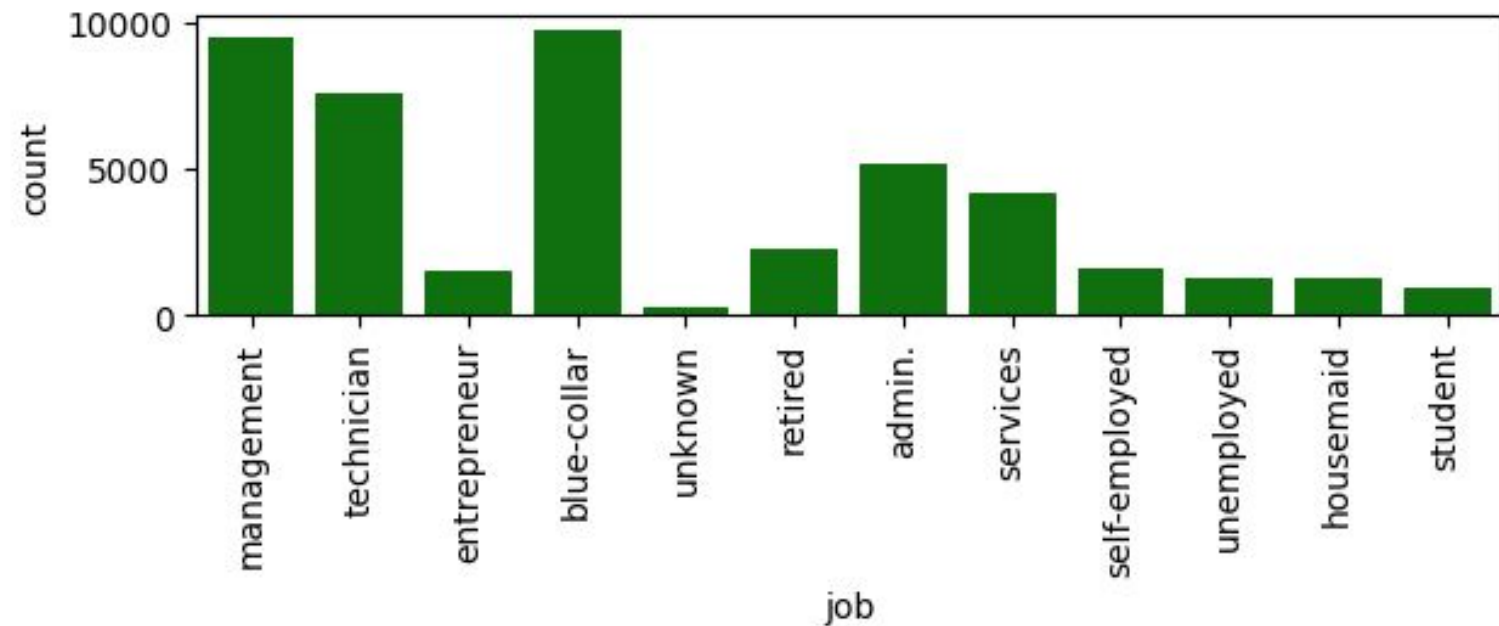
- banyaknya kontak terdahulu yang sudah dilakukan sebelum adanya pemanggilan yang saat ini.
- **diverifikasi bahwa data di atas 75% data bernilai 0, artinya di atas 75% client pada dataset baru pertama kali dihubungi.**



# Univariate Analysis

Feature : Job (Kategorik)

	job
count	45211
unique	12
top	blue-collar
freq	9732



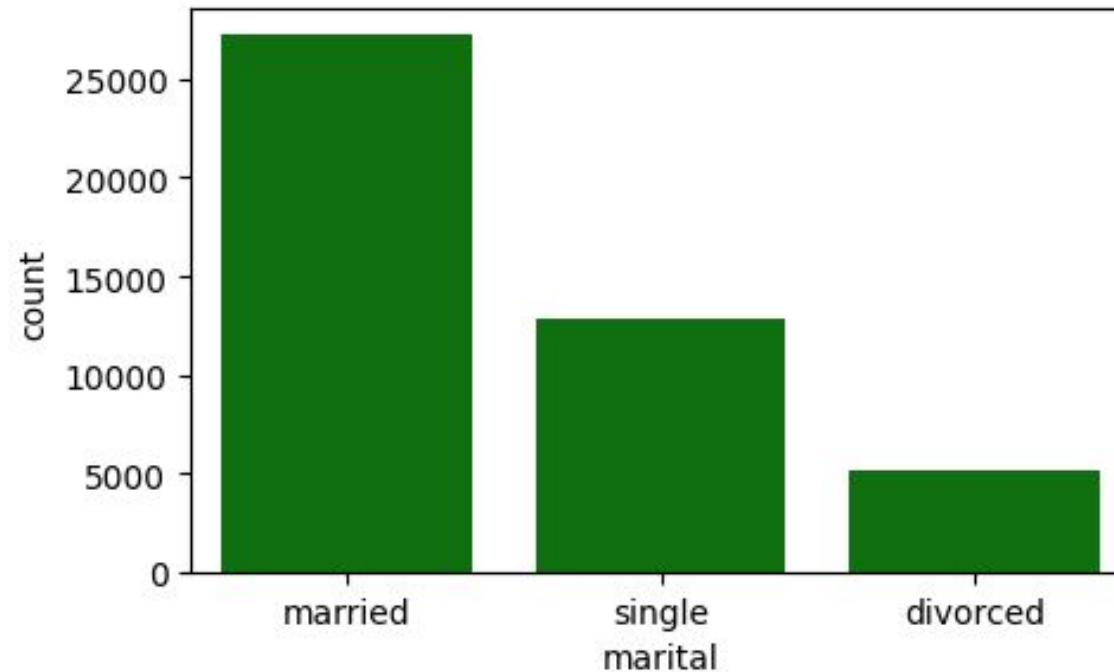
Berdasarkan pekerjaan nasabah, yang banyak dihubungi dalam promosi dengan persentase masing-masing sebesar 20% adalah pekerja kasar dengan karyawan management, sehingga mungkin ke depannya bisa ditarget khusus suatu produk untuk kedua pekerja tersebut untuk menaikkan *conversion rate* nya.



# Univariate Analysis

Feature : Marital (Kategorik)

marital	
count	45211
unique	3
top	married
freq	27214

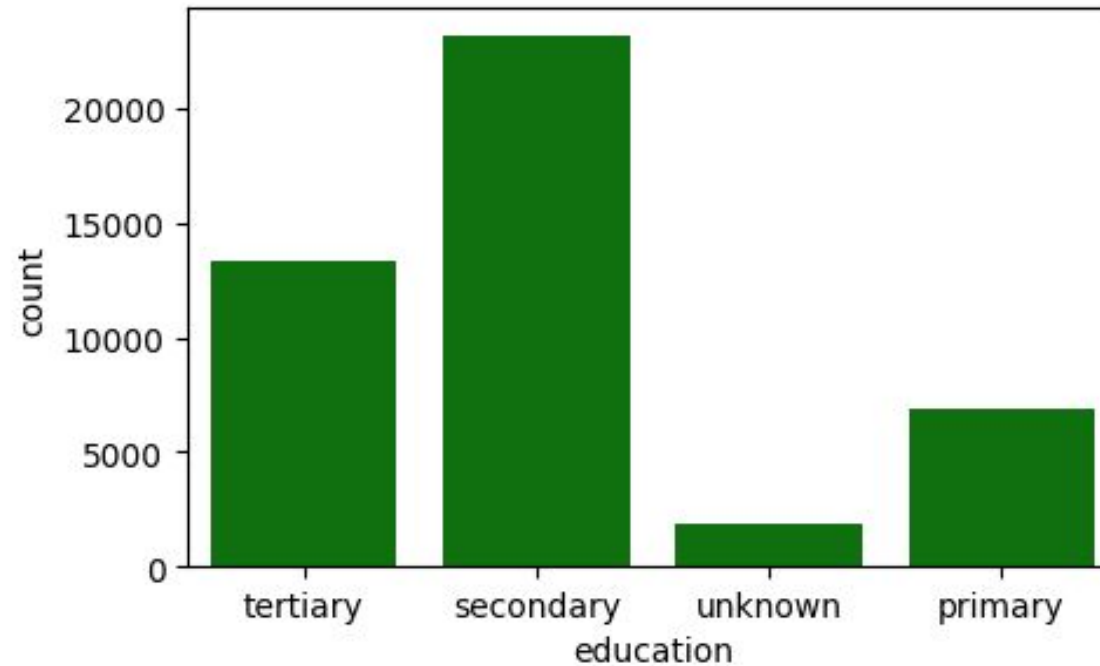


Berdasarkan status perkawinannya, jumlah nasabah yang sudah menikah lebih tinggi dibandingkan dengan nasabah yang single.

# Univariate Analysis

Feature : Education (Kategorik)

education	
count	45211
unique	4
top	secondary
freq	23202

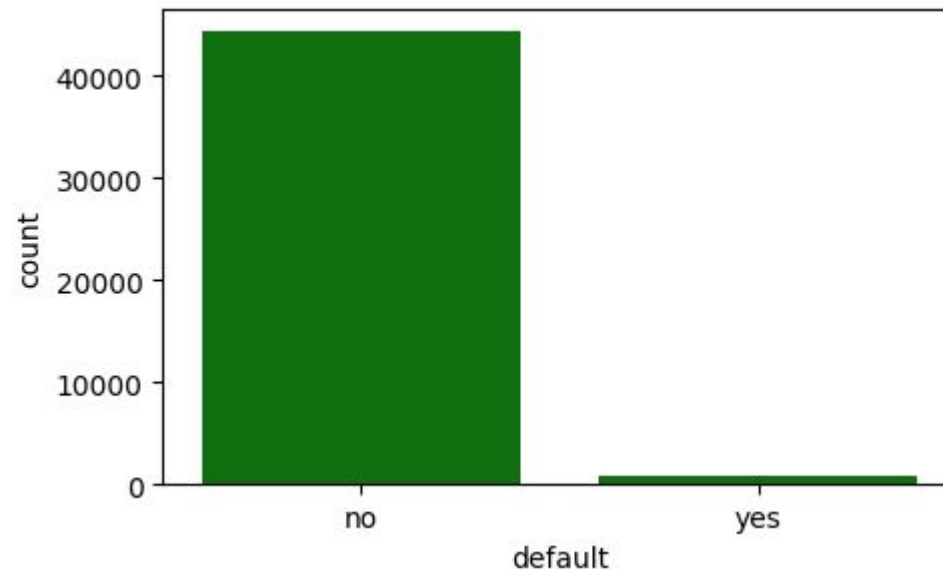


Berdasarkan tingkat pendidikannya, didapatnya jumlah tertinggi pada tingkat pendidikan secondary.

# Univariate Analysis

Feature : Default (Kategorik)

	default
count	45211
unique	2
top	no
freq	44396

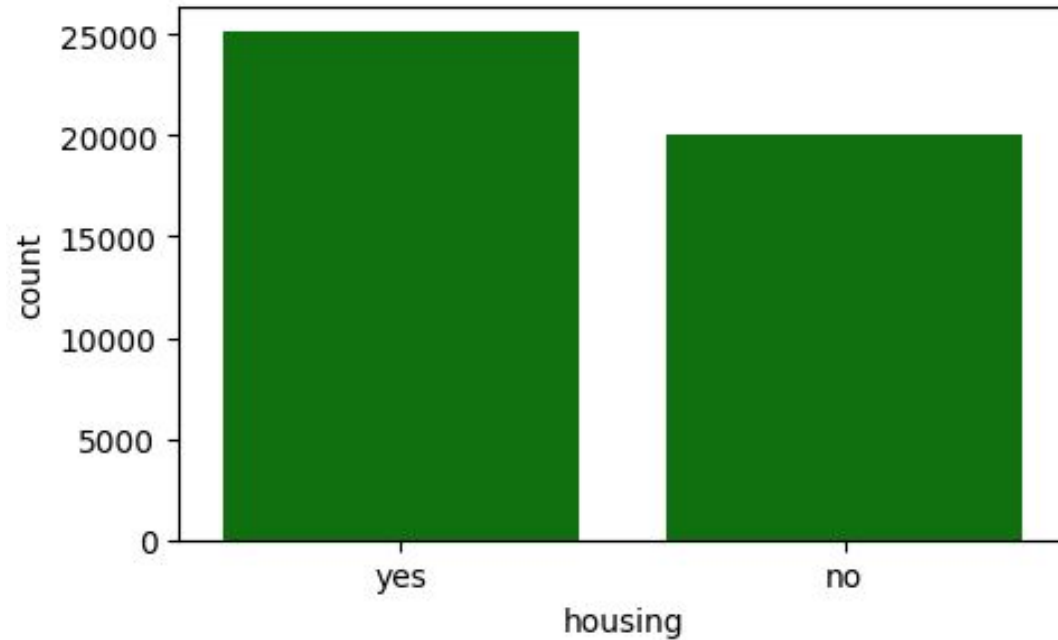


Status pembayaran pinjaman didominasi oleh yang tidak memiliki tunggakan.

# Univariate Analysis

Feature : Housing (Kategorik)

housing	
count	45211
unique	2
top	yes
freq	25130



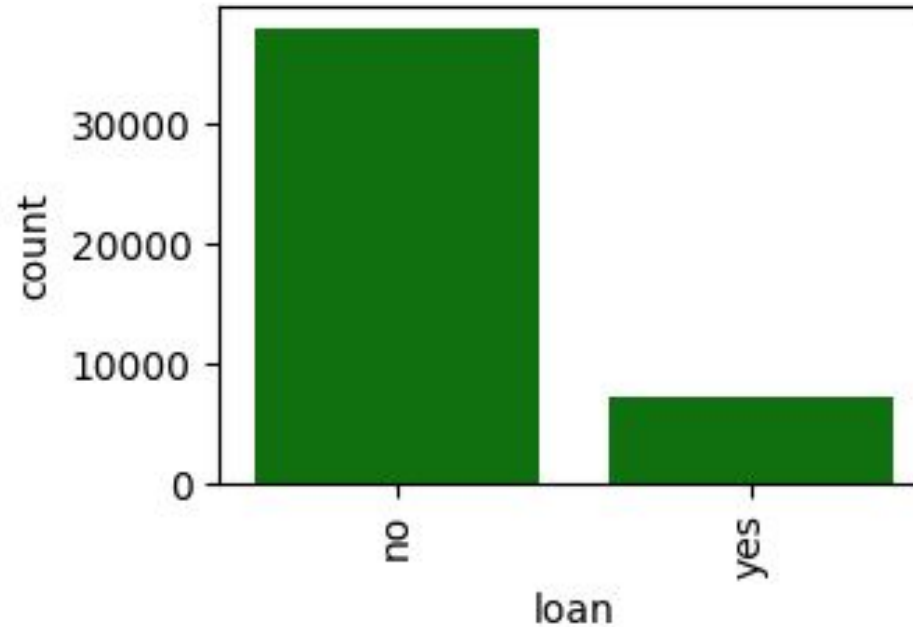
Dataset didominasi oleh nasabah yang sudah memiliki rumah.



# Univariate Analysis

Feature : Loan (Kategorik)

	loan
count	45211
unique	2
top	no
freq	37967

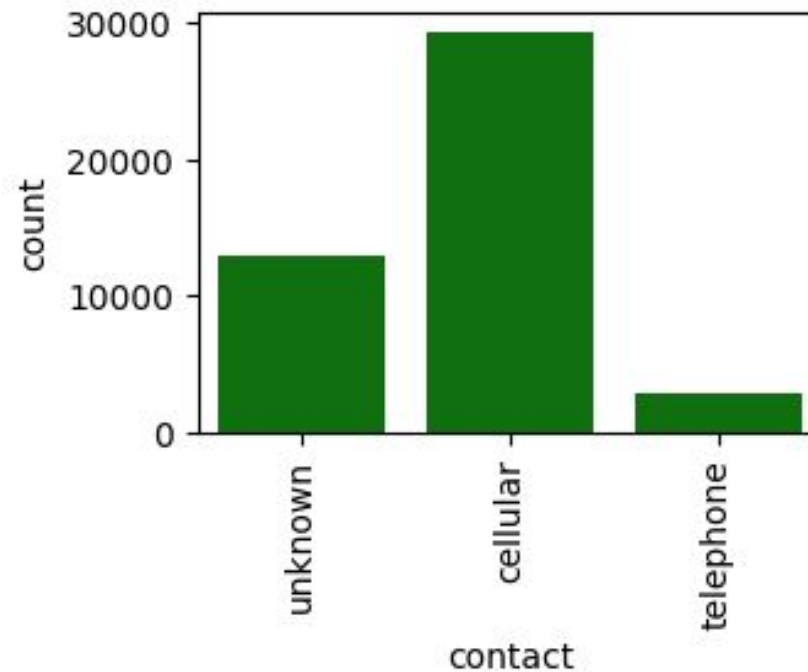


Dataset didominasi oleh nasabah yang tidak melakukan pinjaman ke bank.

# Univariate Analysis

Feature : Contact (Kategorik)

contact	
count	45211
unique	3
top	cellular
freq	29285

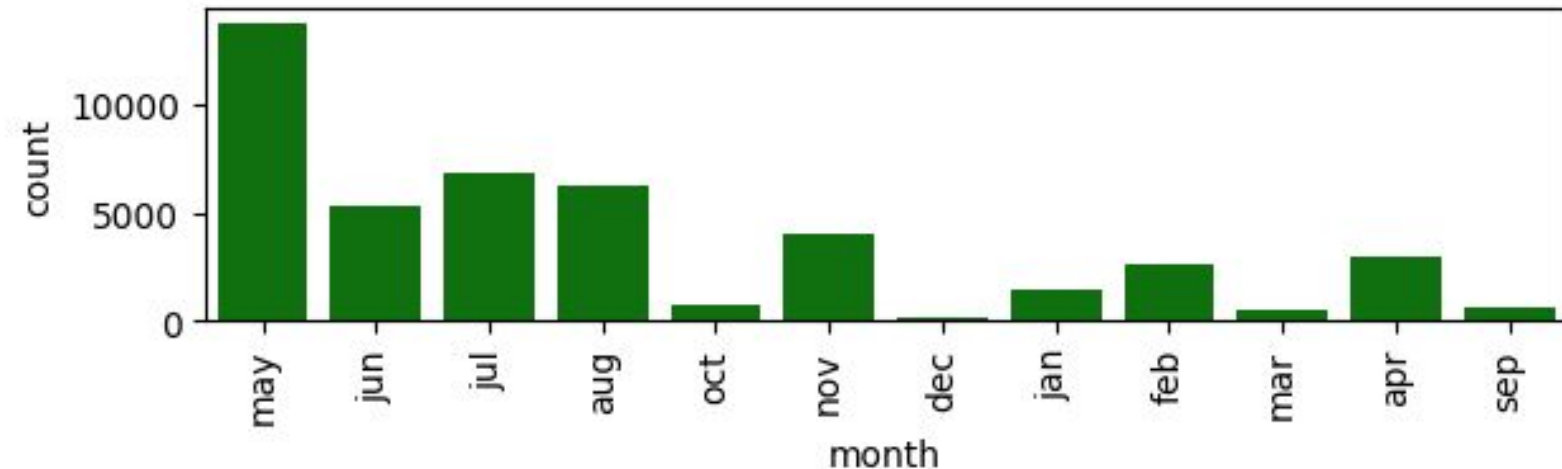


Metode yang paling umum untuk menghubungi nasabah adalah dengan menghubungi telepon selulernya

# Univariate Analysis

Feature : Month(Kategorik)

	month
count	45211
unique	12
top	may
freq	13766

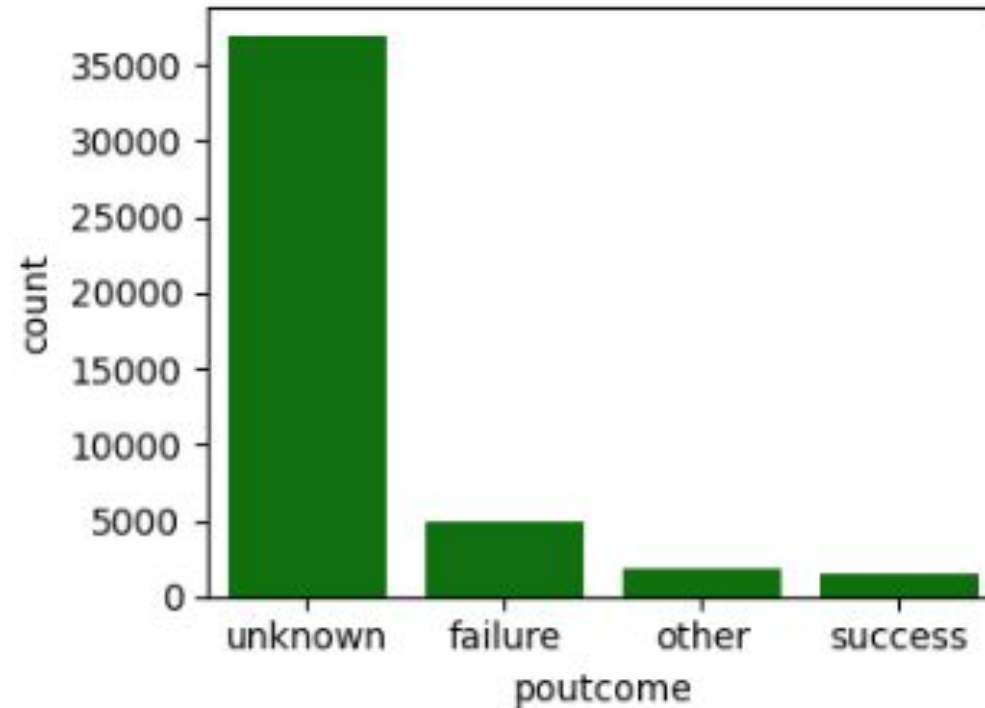


Nasabah paling banyak dihubungi di pertengahan bulan (May - Agustus).  
Tidak banyak kontak yang dilakukan pada proses pergantian tahun.

# Univariate Analysis

Feature : poutcome (Kategorik)

poutcome	
count	45211
unique	4
top	unknown
freq	36959

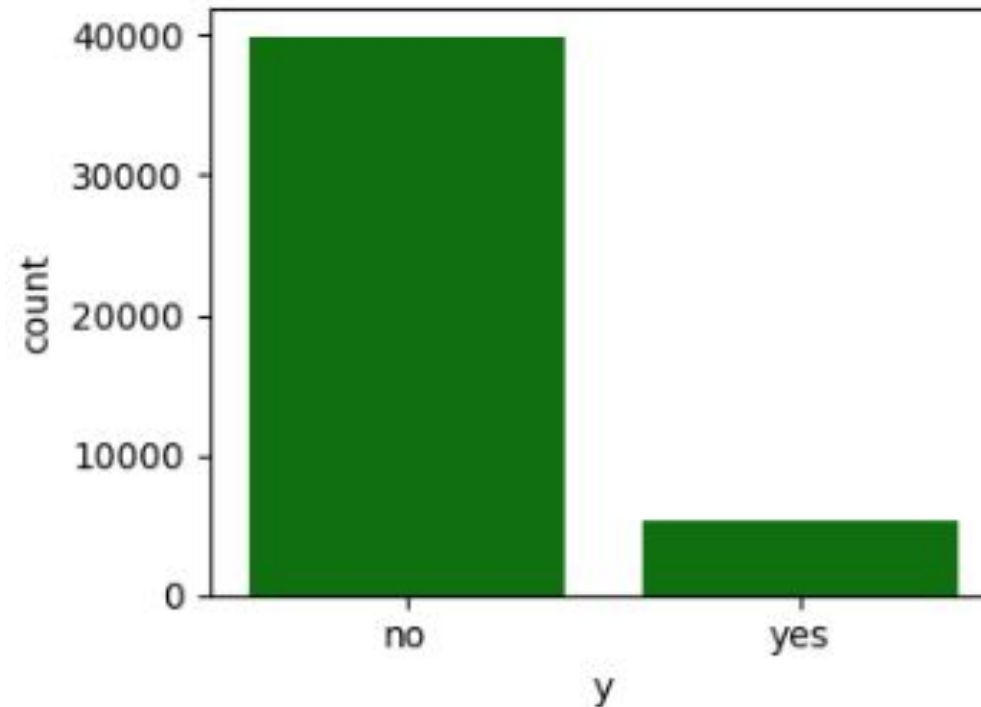


Banyak hasil kontak terakhir yang memiliki status yang tidak diketahui.

# Univariate Analysis

Feature : y (Kategorik)

	y
count	45211
unique	2
top	no
freq	39922



Sebanyak 88,3% nasabah yang mendapatkan campaign memilih untuk tidak membuka deposito.

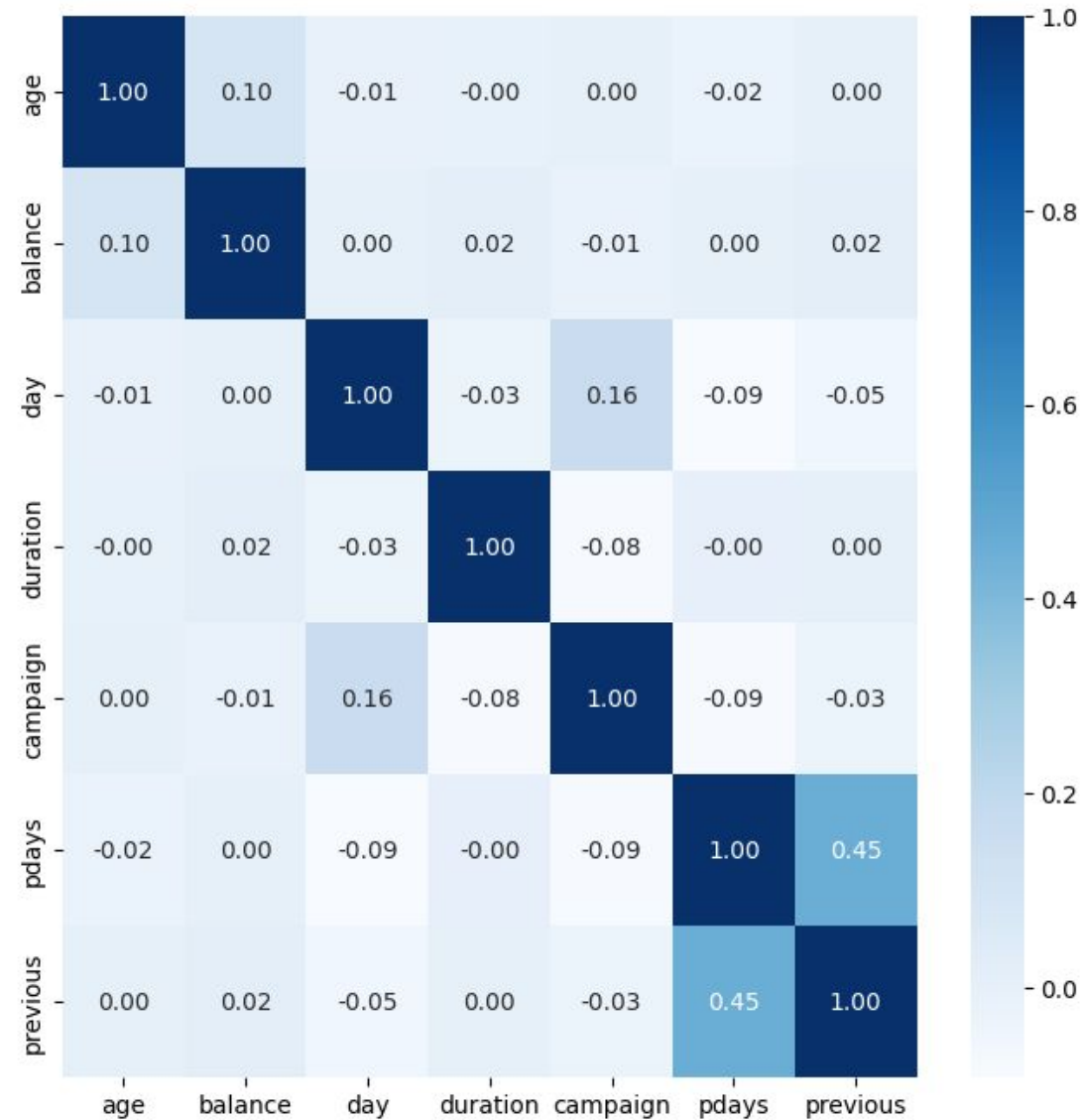
# Univariate Analysis

Dari dataset diperoleh wawasan sebagai berikut:

- **Kebanyakan nasabah yang dihubungi adalah nasabah yang belum pernah dihubungi sebelumnya** (di atas 75%).  
Penggunaan data training berdasarkan durasi, banyaknya campaign dan hasil sebelumnya perlu mewaspadai sample data yang berjumlah 25% atau kurang.
- Kebanyakan data numerik memiliki outlier di atas batas maksimal, artinya kecenderungan distribusi dari data adalah menceng kanan.



# Multivariate Analysis



# Multivariate Analysis

- A. Bagaimana korelasi antara masing-masing feature dan label. Kira-kira feature mana saja yang paling relevan dan harus dipertahankan?

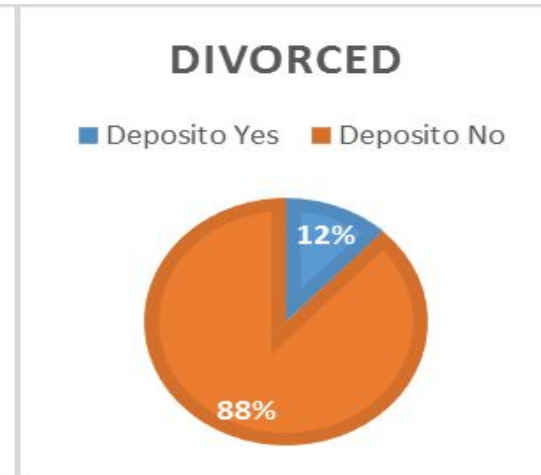
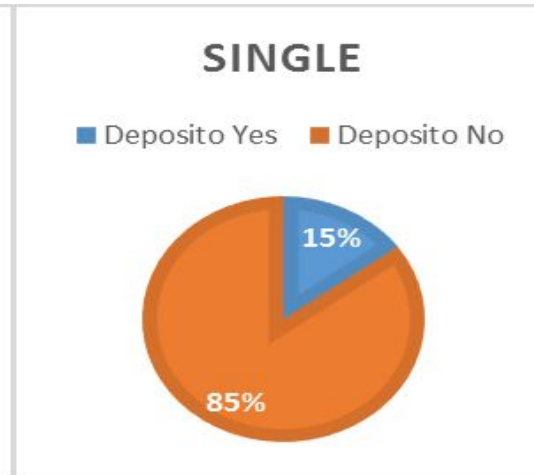
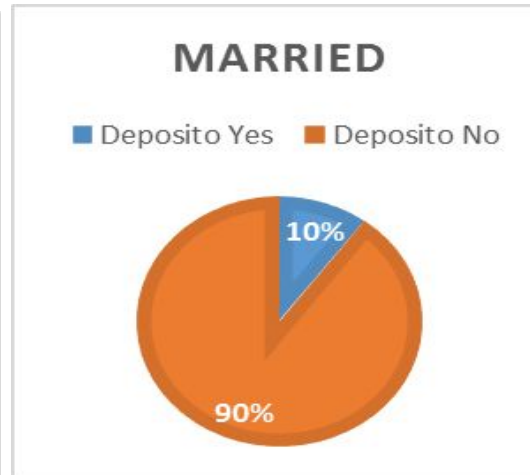
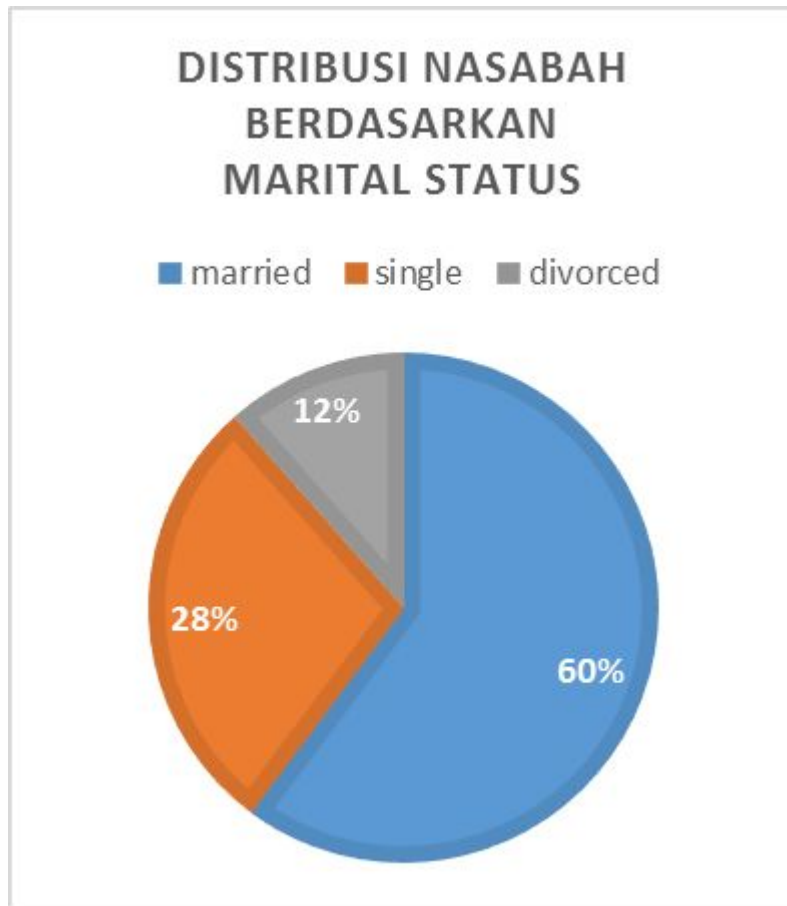
Terlihat **banyak feature pada dataset numerical yang memiliki korelasi dengan nilai 0.0** seperti days & balance, age & campaign maupun feature korelasi yang lain. Sehingga perlu dipertimbangkan kembali bila dilakukannya heatmap yang telah digabung dengan dataset kategorikal apakah masih terdapat banyak nilai dengan angka sebesar 0.0, maka perlu didiskusikan lebih lanjut apakah sebaiknya beberapa kolom yang berada dalam dataset akan didrop atau tidak.

- B. Bagaimana korelasi antar-feature, apakah ada pola yang menarik? Apa yang perlu dilakukan terhadap feature itu?

Tidak ada pola-pola yang menarik antar feature, sehingga belum ada perlakuan terhadap feature yang ada.

# Business Insight

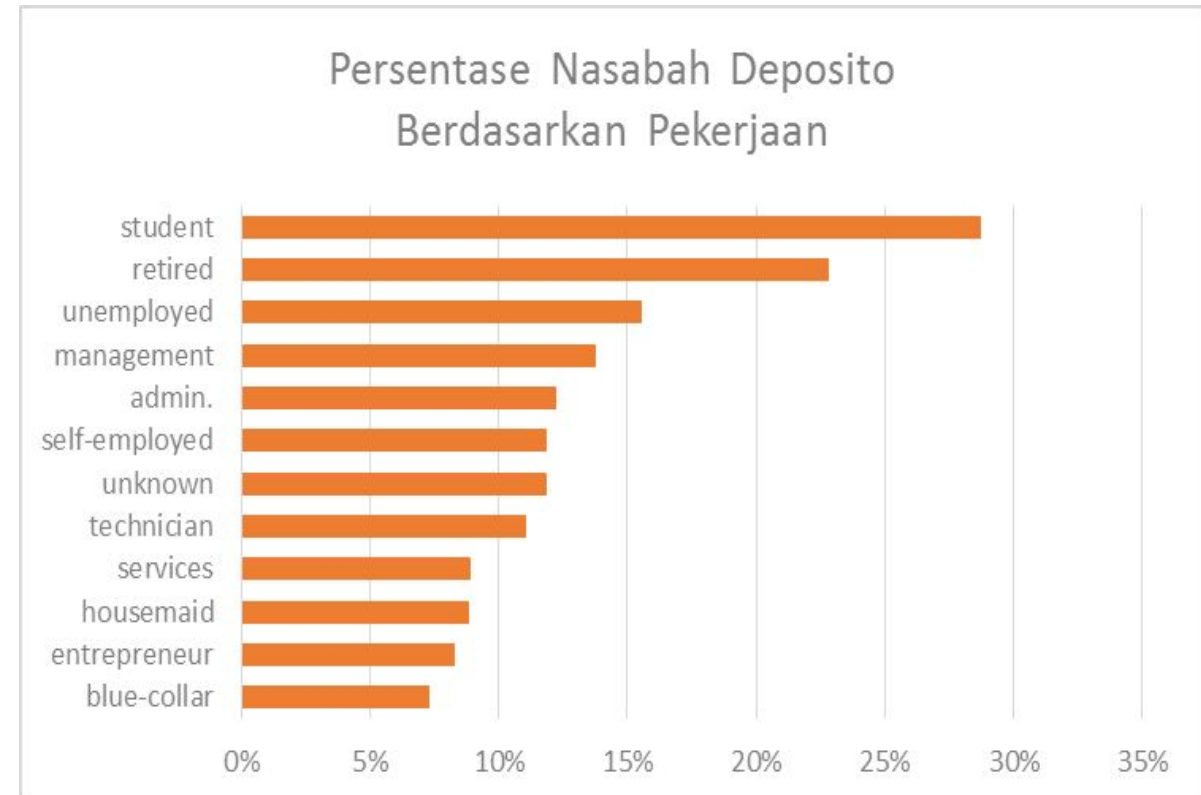
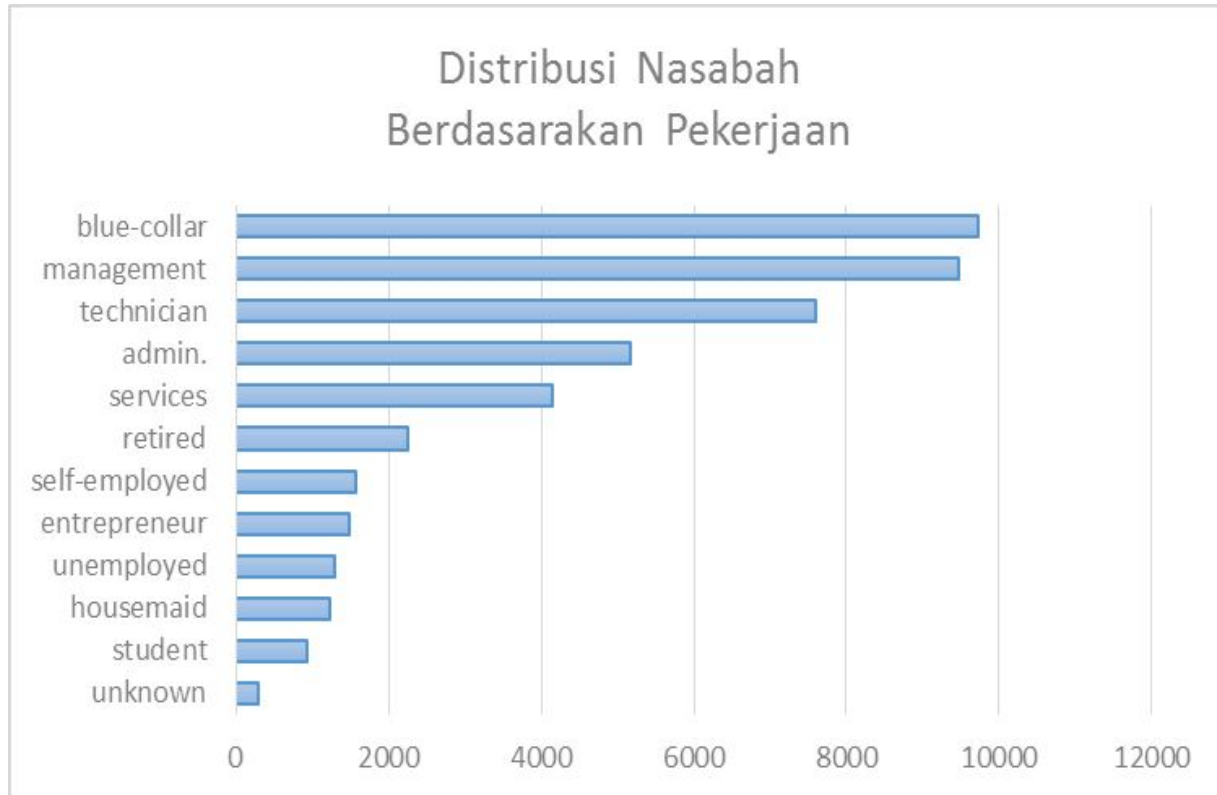
## 1. Based on “marital status”



Secara distribusi, nasabah dari bank DataWhiz didominasi oleh nasabah yang berstatus “Married”, namun persentase pembuka deposito terbanyak adalah yang berstatus “Single”.

# Business Insight

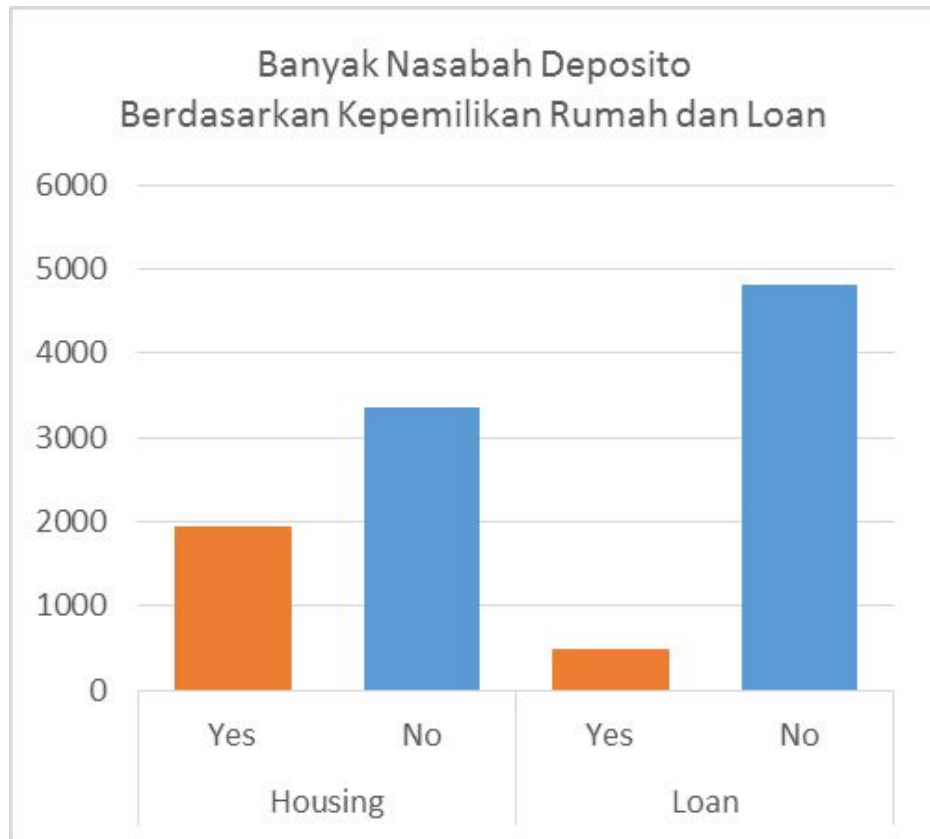
## 2. Based on “job”



Nasabah “student” memiliki persentase yang paling tinggi dalam membuka deposito dan diikuti dengan nasabah dari kategori “retired”. Sedangkan, nasabah “blue-collar” memiliki persentase yang rendah untuk membuka deposito.

# Business Insight

## 3. Based on “housing” dan “loan”



Nasabah yang memiliki tanggungan berupa rumah “housing” atau pinjaman “loan” cenderung tidak membuka akun deposito.

Pada “housing”, ada kemungkinan deposito dibuat untuk membeli rumah sebagai salah satu capaian dalam menabung.

Pada “loan”, pinjaman dapat menjadi halangan untuk nasabah bisa membuka akun deposito sehingga nilainya jauh secara signifikan.

# Business Insight

## 4 Based on “poutcome”

respons	count	
	no	yes
poutcome		
failure	4283	618
other	1533	307
success	533	978
unknown	33573	3386

Dari tabel di samping terlihat bahwa nasabah yang ‘poutcome’ nya **‘failure’**, sebanyak **87.39%** memilih untuk **tidak berlangganan** deposito berjangka, sedangkan **12.61%** sisanya memilih untuk **berlangganan**.

Di samping itu, nasabah yang ‘poutcome’ nya **‘success’**, sebanyak **64.73%** memilih untuk **berlangganan** deposito berjangka, sedangkan **35.27%** sisanya memilih untuk **tidak berlangganan**.

Artinya, kesuksesan atau kegagalan pada campaign sebelumnya cukup berpengaruh pada campaign ini.

Keterangan

poutcome : hasil dari kampanye pemasaran sebelumnya (kategori: "tidak diketahui", "lainnya", "gagal", "sukses")



# Business Insight

Berdasarkan analisa data yang telah diperoleh, diperoleh insight sebagai berikut

- **Nasabah yang single adalah nasabah dengan persentase paling banyak membuka rekening deposito.** Jumlah nasabah masih sedikit namun memiliki peluang membuka deposito lebih banyak. Hal ini mengindikasikan produk deposito konvensional cocok untuk nasabah berstatus single..
- **Nasabah yang married memiliki gap untuk membuka rekening.** secara persentase nasabah yang sudah married memiliki jumlah yang paling besar, namun persentasenya adalah yang paling kecil. Pemberian perlakuan pada deposito kepada yang berstatus 'married' akan memberikan impact yang lebih besar secara volum. sehingga akan dibahas rekomendasi business untuk nasabah berstatus married.
- **Mempertimbangkan produk deposito untuk target nasabah student, dan retired.** Student banyak melakukan investasi deposito, sehingga dapat dibuat suatu kampanye program khusus untuk nasabah student dan pensiunan.

# Business Insight

- **Deposito lebih diminati bagi orang-orang yang belum memiliki rumah.** Dapat dibuat program deposito dengan kampanye untuk program membeli rumah. Dapat disesuaikan lama waktu tabungan dengan rencana membangun rumah.
- **Hasil dari campaign sebelumnya mempengaruhi hasil saat ini.** Strategi pemasaran pada setiap pemanggilan harus mencapai titik tertentu. Usahakan untuk memperoleh kemenangan kecil pada setiap pemanggilan, karena akan mempengaruhi kesuksesan pada panggilan berikutnya. selanjutnya akan diberikan rekomendasi dari strategi pemasaran

# Business Recommendation

## Keperluan masing-masing kategori

### 1. Deposito Rumah Tangga

- Deposito khusus bagi orang yang sudah menikah.
- Keputusan dalam rumah tangga biasanya disepakati oleh suami dan istri.
- Buat program dengan goals untuk keperluan rumah tangga: Renovasi Rumah, Membeli Mobil Baru, Biaya Sekolah Anak.
- *Opsional*: dapat diintegrasikan dengan produk lain seperti KPR Rumah, Kredit Kendaraan. Hasil dari tabungan deposito dapat langsung dialihkan ke produk lain

### 2. Deposito Mahasiswa

- Deposito Khusus untuk mahasiswa
- Buat program untuk goals dari mahasiswa: Untuk membayar uang kuliah, untuk keperluan tugas mahasiswa, untuk biaya pulang kampung bagi yang di luar kota.
- bank dapat merekomendasikan jumlah dana yang dimasukan dalam deposito dan menghitung waktu yang dibutuhkan agar goals tercapai

# Business Recommendation

## Keperluan masing-masing kategori

### 3. Deposito Untuk Masa Tua

- Deposito yang dikhususkan bagi nasabah yang sudah pensiun
- Keperluan untuk menabung dengan bunga yang lebih besar dan low risk, membuat deposito menjadi produk yang cocok.
- Deposito dapat dihubungkan dengan ahli waris apabila sudah tidak ada. Hal ini untuk mengakomodasi warisan bagi para nasabah senior. Sehingga proses ahli waris dapat berjalan lebih mudah.

### 4. Deposito Untuk Membangun Rumah

- Nasabah yang membuka deposito cenderung yang belum memiliki rumah.
- Program ini dapat menghubungkan keperluan membeli rumah dengan bunga dan periode yang dibutuhkan agar dapat membeli rumah. Dapat juga dihubungkan dengan program KPR rumah, misalkan dengan bunganya dibayarkan sebagai biaya KPR.

# Business Recommendation

## Deposito terintegrasi

- Deposito dapat diintegrasikan dengan program-program lain di perbankan seperti KPR, biaya kredit kendaraan, kredit usaha, dll.
- Deposito dapat diintegrasikan dengan program sekolah, khusus untuk mahasiswa.
- Sebagian bunga dari deposito dapat dialihkan ke program lain, sehingga dapat mengalokasikan tujuan dari menabung secara langsung.
- Keuntungan bagi bank, adalah banyak produk perbankan yang digunakan.
- Mengakomodasi:
  - Nasabah married: untuk keperluan rumah tangga
  - Nasabah yang belum memiliki rumah: untuk menyicil DP KPR dengan bunga deposito.
  - Nasabah student: untuk keperluan kendaraan, membayar uang kuliah agar bisa langsung dibayar secara otomatis.

**Repository GIT:**

**[Link Github Kelompok 1 - Data Whiz](#)**