

# ABSOLUTE

Kelompok 7:

1. Fikrie Lazuardi
2. Retno Dwi
3. Yunita Rachmawati
4. Rian Dwi Haryono
5. Wahyuni Salman
6. Retno Harindhi
7. Hidayat Yatul



# STAGE 1 (EDA)

# Descriptive Statistics

## Exploratory Data Analysis

- Semua tipe data sudah sesuai
- Tidak ditemukan data duplicate
- Tidak ditemukan missing value

```
[ ] df_train.duplicated().any()

False
```

```
<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
```

```
[ ] df_train.isna().sum()

age         0
job         0
marital     0
education   0
default     0
balance     0
housing     0
loan        0
contact     0
day         0
month       0
duration    0
campaign    0
pdays      0
previous    0
poutcome    0
y           0
dtype: int64
```

# Descriptive Statistics

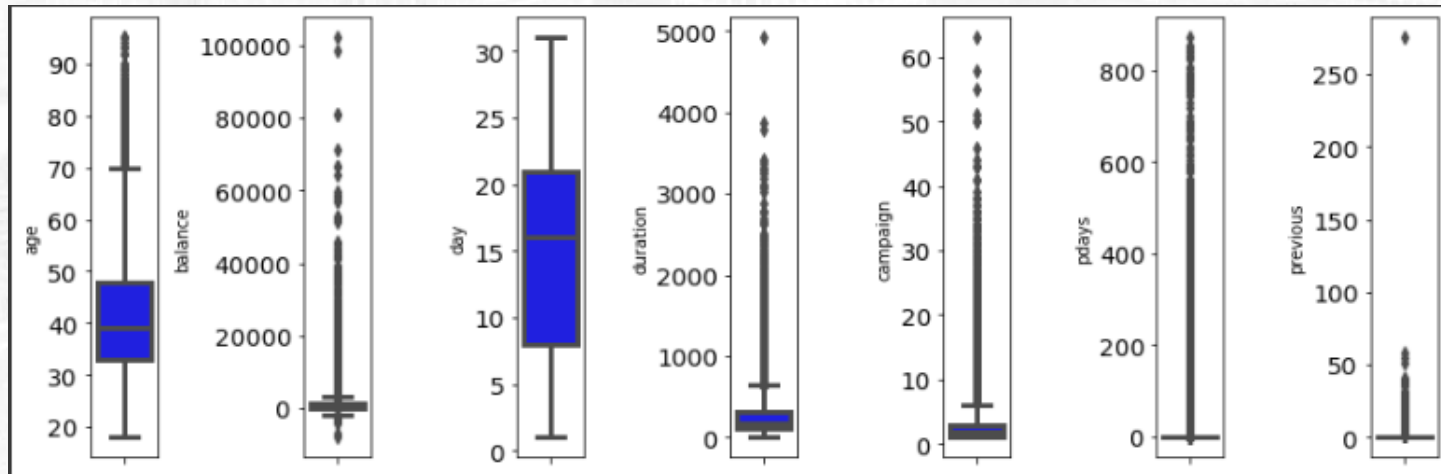
## Exploratory Data Analysis

- Karena kolom balance memiliki standar deviasi yg sangat tinggi dimana hasil describe() menyatakan terlalu jauh dari mean/ median sehingga muncul keanehan, diduga karena adanya nilai negative.
- Kolom age dan day memiliki sedikit perbedaan pada persebaran datanya, berbeda dengan kolom balance dan pdays yang memiliki persebaran data yang sangat besar, mengindikasikan outlier yang juga signifikan.

	count	mean	std	min	25%	50%	75%	max
age	45211.0	40.936210	10.618762	18.0	33.0	39.0	48.0	95.0
balance	45211.0	1362.272058	3044.765829	-8019.0	72.0	448.0	1428.0	102127.0
day	45211.0	15.806419	8.322476	1.0	8.0	16.0	21.0	31.0
duration	45211.0	258.163080	257.527812	0.0	103.0	180.0	319.0	4918.0
campaign	45211.0	2.763841	3.098021	1.0	1.0	2.0	3.0	63.0
pdays	45211.0	40.197828	100.128746	-1.0	-1.0	-1.0	-1.0	871.0
previous	45211.0	0.580323	2.303441	0.0	0.0	0.0	0.0	275.0

# Univariate Analysis

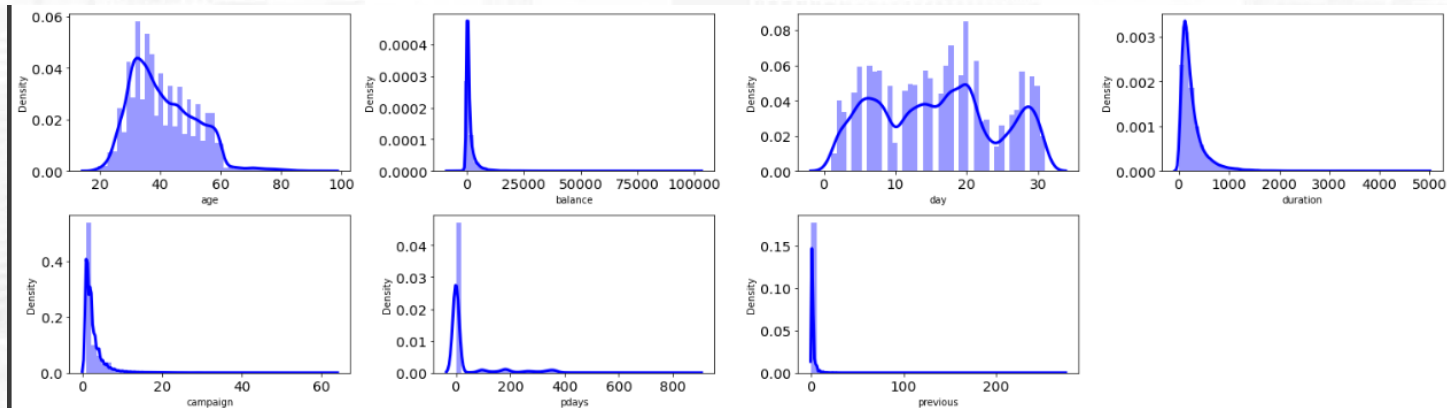
- Pada kolom "age" tidak terlalu bermasalah karena hanya sebagian kecil saja yang menjadi outlier
- Nilai -1 pada kolom "pdays" sangat mendominasi, yaitu berjumlah 36954 dari 45211 baris atau sekitar 81% data sehingga kotak pada gambar diatas terpusat pada nilai -1, sehingga nilai-nilai diatas -1 menjadi outlier yang cukup banyak yaitu sekitar 19%.
- Nilai 0 yang dimiliki kolom "previous" sangat mendominasi kolom ini dan memiliki korelasi dengan nilai -1 yang berada pada kolom "pdays". Tetapi, ada satu nilai yang sangat jauh yaitu 275 dan outlier ini akan dihapus karena hanya terdapat satu.





# Univariate Analysis

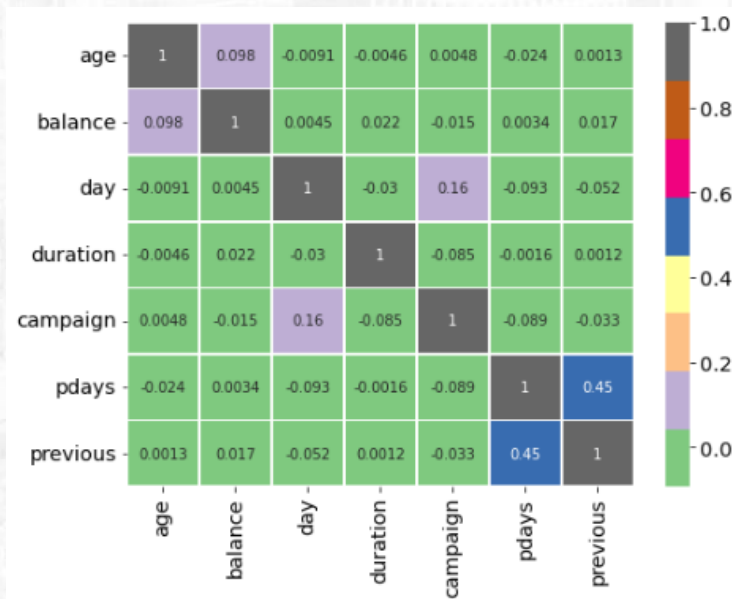
Berdasarkan nilai skewness dan visualisasi di atas dapat disimpulkan bahwa tipe data numerikal memiliki pola persebaran data yang ditunjukkan dalam bentuk skew positif yang mana nilai median lebih kecil daripada mean.



```
Skewness age : 0.6848179257252598
Skewness balance : 8.360308326166326
Skewness day : 0.09307901402122411
Skewness duration : 3.144318099423456
Skewness campaign : 4.898650166179674
Skewness pdays : 2.6157154736563477
Skewness previous : 41.84645447266292
```

# Multivariate Analysis

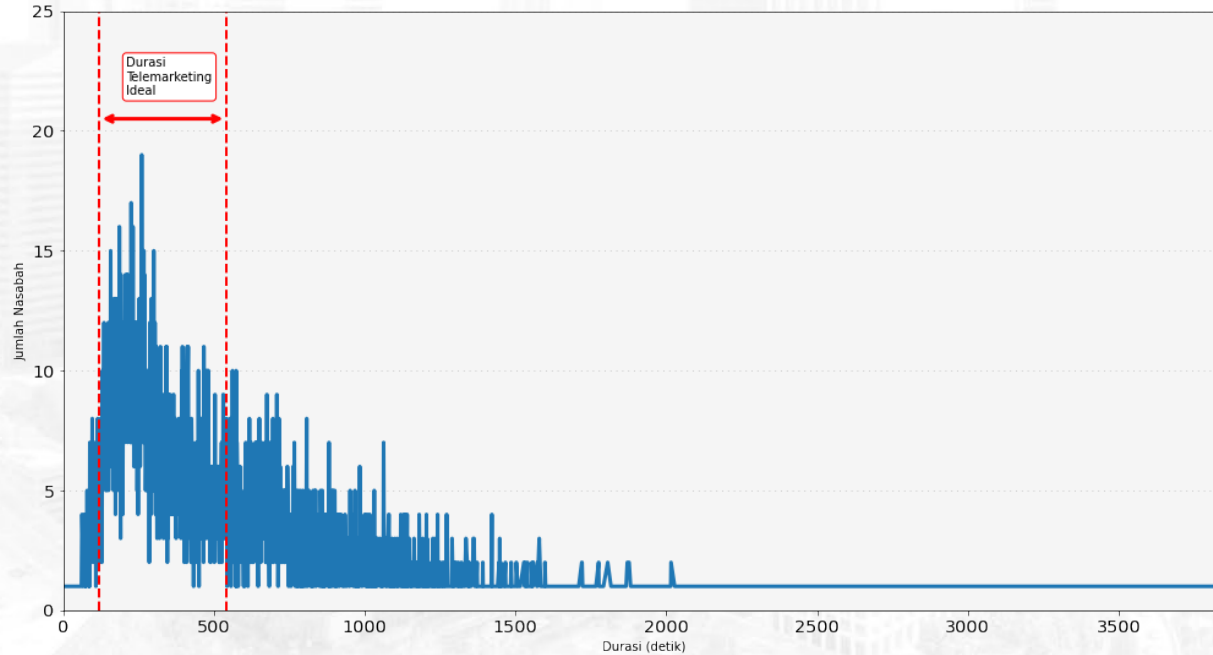
- Berdasarkan correlation heatmap dan category plots semua feature masih relevan untuk digunakan.
- Korelasi previous dan pdays cukup kuat karena bernilai 0.45. Walau begitu, semua feature akan dipertahankan untuk keperluan analisis



# Business Insight

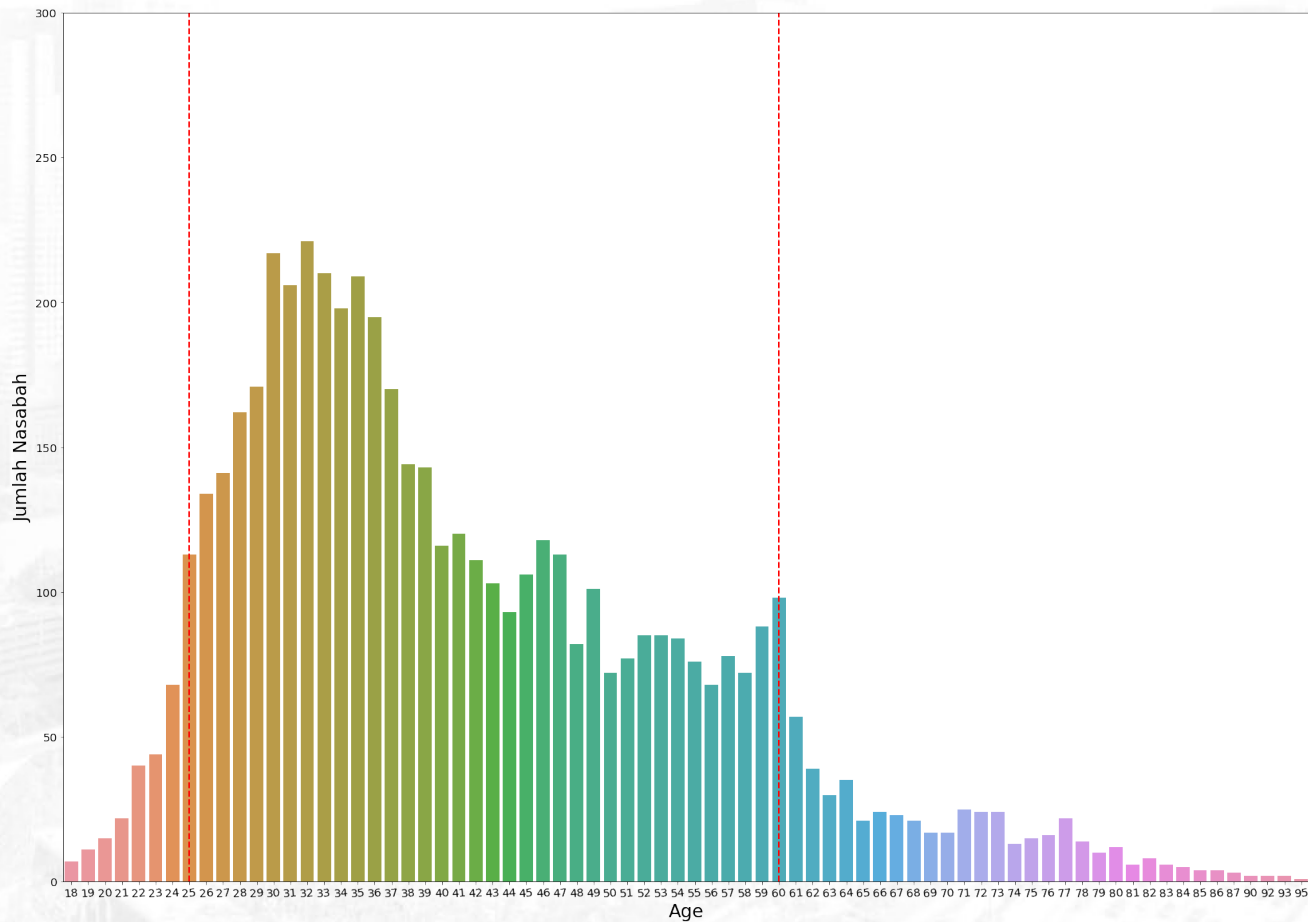
**Durasi yang cocok untuk menghubungi nasabah adalah 120 - 540 detik atau 2 - 9 menit**

56.7% dari seluruh nasabah yang membuka deposito sempat melakukan komunikasi melalui telemarketing selama 2 -9 menit





# Business Insight



# Business Insight

Berdasarkan pengolahan , terdapat business insight yang kami dapatkan:

- Durasi yang cocok untuk menghubungi nasabah adalah sebanyak 2-9 menit
- Usia 25-60 sangat cocok untuk direkomendasikan produk deposito