

Banking Marketing Targets

Dokumen
Laporan Final
Project (Stage 1)

By: Group 3 DS Batch 21 aka **Jump-start**



Data Scientist Team

- Irma Widiyani
- Maghfira Amalia
- Mahendra Prabaswara
- Ridwan Donovan
- Windy Nurfikri

Problem Statement

Term deposits are one of the major income sources of a bank. These days, selling deposits through telemarketing is still the main of various plans to reach out to the clients. However, the bank still needs to pay attention to the effectiveness of the direct campaign considering the high cost and the time it takes. Based on previous campaign data of 45,211 clients, **only 3.3% of them actually subscribed to a term deposit**. Despite the fact that **36,959 clients were never contacted**. Therefore, determining the right potential depositors needs to be done to save resources. Meanwhile, we also need to review the efficiency and effectiveness of telemarketing as a direct campaign method.

Goal

Increase the efficiency and effectiveness of direct campaign

Objectives

- Predict the potential depositor by classifying them (based on their background and financial history)
- Find out the success rate of telemarketing as direct campaign method

Business Metric

Conversion Rate

Descriptive Statistic

A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai? Pada kolom 'job' ada value 'admin.' yang mana seharusnya penulisan tidak perlu menggunakan tanda titik (.) seperti pekerjaan lainnya. Selebihnya, semua tipe data sudah sesuai.

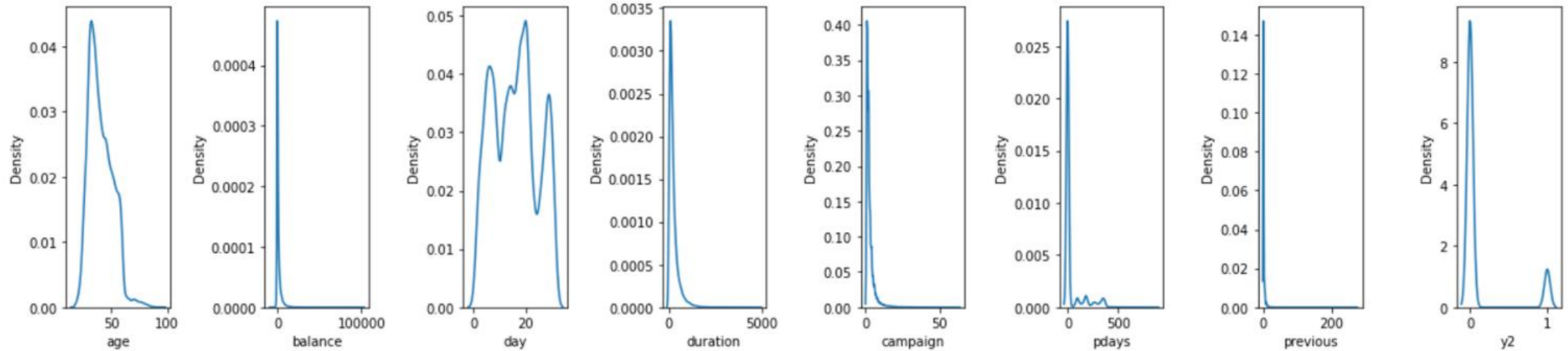
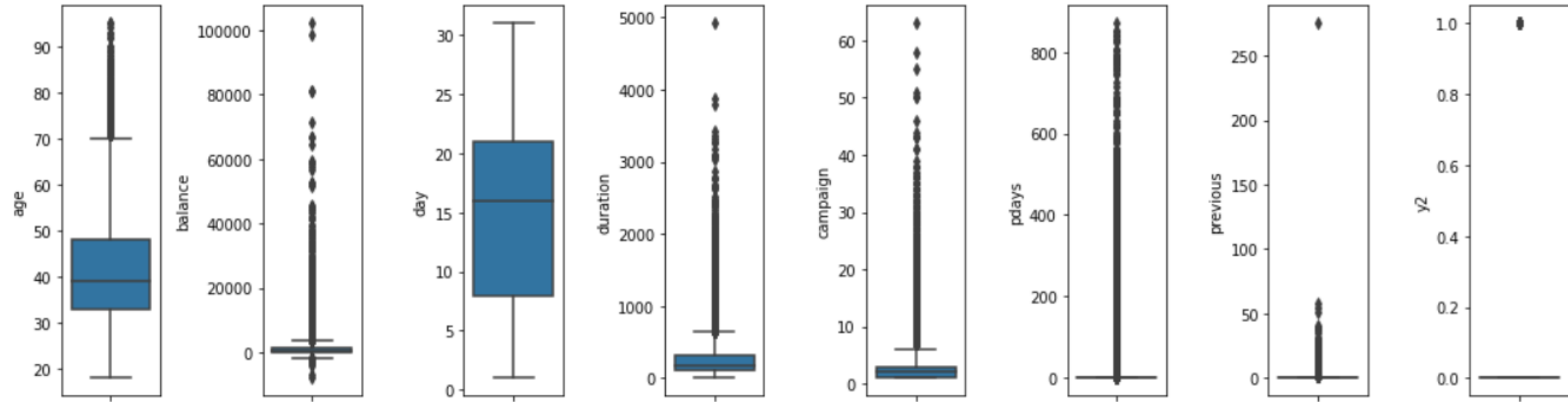
```
data = pd.read_csv('train.csv', sep = ',')
df = data.rename(columns={'y': 'subscribed'})
df.sample(5)
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	postcome	subscribed
33026	26	admin.	single	secondary	no	1375	yes	no	cellular	17	apr	238	2	-1	0	unknown	no
19393	49	blue-collar	married	secondary	no	141	no	no	cellular	6	aug	275	1	-1	0	unknown	no
16347	36	management	divorced	tertiary	no	14930	no	no	cellular	23	jul	190	1	-1	0	unknown	no
2820	42	admin.	married	secondary	no	2656	yes	no	unknown	14	may	238	1	-1	0	unknown	no
41162	89	retired	married	tertiary	no	553	no	no	telephone	19	aug	2027	5	-1	0	unknown	no

A. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?
Tidak ada kolom dengan nilai kosong.

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

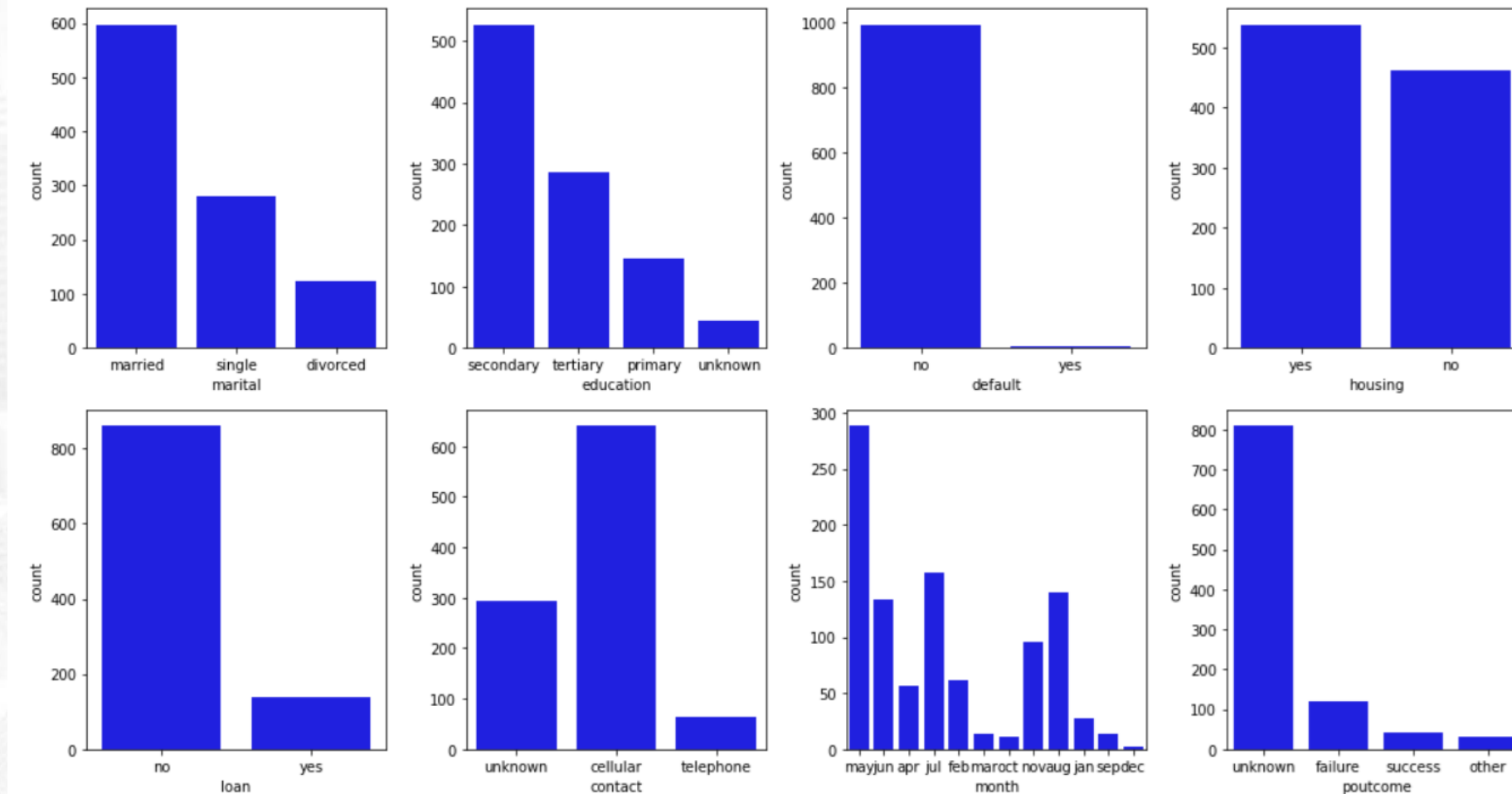
Distribusi pada setiap fitur



Persebaran data numerik berdasarkan boxplot dan kde plot pada slide sebelumnya menunjukkan:

- Adanya outlier pada kolom *previous* dan *duration*.
- Terdapat distribusi yang cukup skewed pada *balance*, *duration*, dan *campaign*.

Sedangkan untuk data kategorikal, dapat terlihat hampir seluruh data memiliki ketimpangan kecuali *housing*, yang mana secara ideal seharusnya tidak timpang



Korelasi Data

Dari correlation heatmap di atas dapat disimpulkan bahwa:

- *y2* memiliki potensi korelasi yang tinggi dengan *duration* (strong potential correlation)
- *y2* juga memiliki korelasi yang lemah dengan *previous* dan *pdays* (decent potential feature)
- *campaign* memiliki korelasi positif dengan *day*
- Sedangkan korelasi *campaign* dengan *age* sangat lemah, kemungkinan bukan fitur yang potensial (decent potential feature)
- ada korelasi antar *previous* dengan *pdays*, namun tidak cukup kuat untuk dikatakan redundant

