**Banking Dataset Marketing Target**

Pattern Pioneers

1. **Background**

1.1. Problem Statement

Bank Portugis telah menghadapi tantangan dalam memasarkan produk term deposit mereka kepada customer. Meskipun mereka telah melakukan berbagai upaya pemasaran, tingkat keberhasilan untuk mendapatkan customer baru masih belum optimal. Oleh karena itu, kami, tim data Bank Portugis diminta untuk mengembangkan sebuah model machine learning yang dapat memprediksi customer mana yang berpotensi untuk berlangganan term deposit. Peningkatan efisiensi dan profitabilitasnya campaign ingin ditingkatkan lebih dari 50 %.

1.2. Roles

Sebagai tim data analis dan data scientist untuk perusahaan bernama Bank Portugis, kami diminta untuk melakukan implementasi machine learning untuk pemasaran Bank Portugis.

1.3. Goals

* Melakukan identifikasi terhadap customer yang memiliki potensi tinggi untuk berlangganan.
* Membuat model klasifikasi customer.
* Memberikan rekomendasi bisnis kepada tim Marketing

1.4. Business Metrics

* Menghitung tingkat akurasi dari klasifikasi customer.
* Tingkat konversi customer yang ingin berlangganan deposit program.
* Jumlah laba bersih

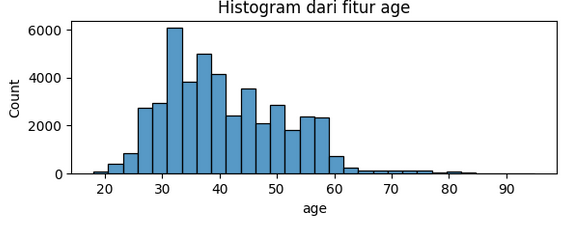
1.5. Objectives

* Meningkatkan persentase customer berlangganan term deposit.
* Mengurangi biaya operasional telemarketing.

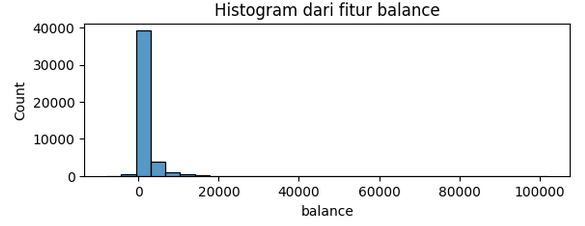
1. **EDA & Business Insights**

Pada data yang ada, memiliki 17 fitur yang dapat digunakan untuk membuat model. Berdasarkan hasil profilling, data dapat dilihat bahwa :

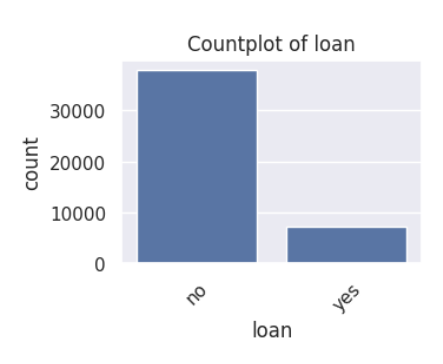
* Umur nasabah memiliki rentang dari 18 - 95 tahun



* Uang yang disimpan pada tabungan nasabah dari rentang 8019 - 102107 euro

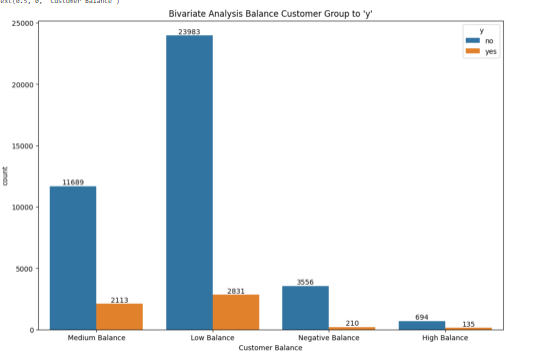


* Mayoritas nasabah bekerja sebagai *blue collar* (21,5%)
* Mayoritas nasabah tidak memiliki pinjaman pribadi (83,9%)

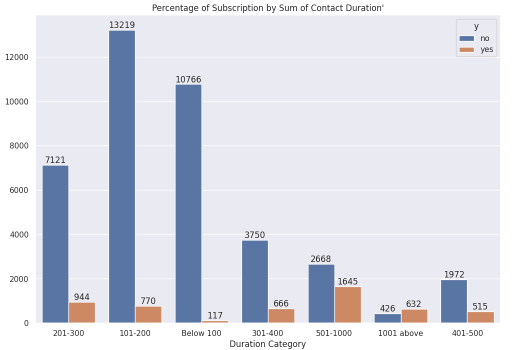


* Mayoritas nasabah sudah menikah (60,3%)
* Mayoritas nasabah memiliki pendidikan tingkat menengah (51,3%)

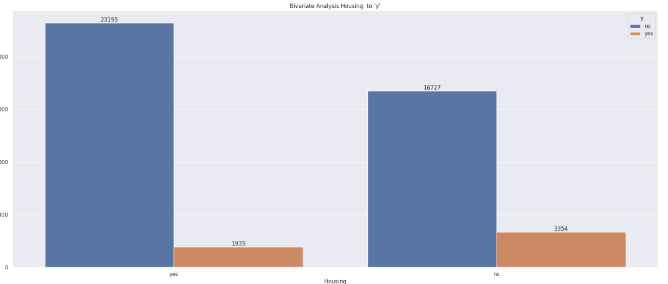
Berdasarkan dari hasil profilling nasabah Bank Portugis, bisnis insights untuk memaksimalkan nasabah mengambil program deposit berdasarkan dengan *campaign* yang telah dijalankan adalah :



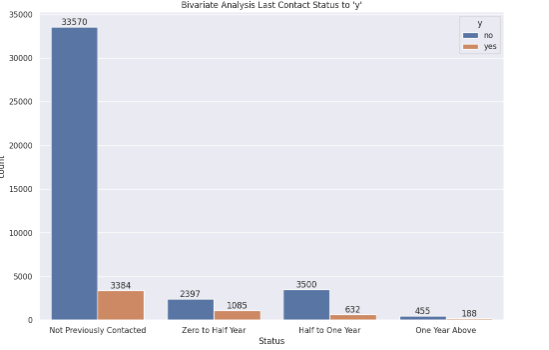
* Sebagian besar nasabah yang melakukan subs deposit berdasarkan hasil campaign kita berada pada kelompok low balance. Namun, bila dicermati presentase tertinggi pada tiap group berdasarkan hasil campaign yang kita lakukan justru berasal dari kelompok nasabah yang memiliki High Balance (16.3%) dan Medium Balance (15.4%)



* Nasabah yang kita hubungi dan durasi kontaknya diatas 500 detik memiliki presentase untuk subs deposit lebih tinggi berdasarkan kontak yang dilakukan (42,7%) dibanding dengan durasi kontak yang lebih sedikit (7,5%).



* Nasabah yang tidak memiliki kredit rumah cenderung akan melakukan subs deposit (16.7%).



* Nasabah yang pernah melakukan contact dengan tim marketing memiliki potensi lebih tinggi untuk subs deposit.

1. **Data Preprocessing**

Sebelum membuat model, kami melakukan pre-processing untuk melihat apakah ada data yang terduplikasi, hilang, dan membuat data menjadi stabil agar dapat digunakan. Berikut preprocessing yang dilakukan pada data :

* **One Hot Encoding :**

- Kolom education, default, housing, loan, dan age telah diubah menjadi beberapa kolom biner.

- Outputnya memiliki lebih banyak kolom dibandingkan input awal, dengan setiap kategori unik diwakili oleh kolom biner (0 atau 1).

* **Label Encoding:**

**-** Kolom Month, Marital, Job, Contact, dan P outcome telah diubah menjadi nilai numerik.

- Outputnya menggantikan nilai kategori dengan angka (0, 1, 2, ...).

* **Drop Feature:**

**-** Kolom Poutcome dan Contact telah dihapus.

- Outputnya memiliki lebih sedikit kolom dibandingkan input sebelum penghapusan fitur.

Standarisasi:

Kolom Duration, Pdays, Balance, dan Previous telah distandarisasi.

Outputnya memiliki nilai di kolom-kolom ini yang berada pada skala yang sama (mean 0 dan standar deviasi 1).

1. **Modelling**
2. **Business Impact & Reccomendation**