



Kelompok: Market Insider

Stage: 1

Mentor: Kevin

Pukul/ Tanggal: 20:00/ 29 November 2023

**Pembagian tugas di stage ini:**

- |  |  |
|--|--|
| 1. Achmad Hilman Shadiqin – Data Analyst | 4. Andreawan Sofian – Data Scientist   |
| 2. Riyan Maula – Data Analyst            | 5. Figo Akmal Munir – Data Scientist   |
| 3. Nabilah Astiarini – Data Analyst      | 6. Dzakwan Darussalam – Data Scientist |

**Poin pembahasan:**

Sesi mentoring ini membahas mengenai tahap *Exploratory Data Analysis, Insight & Visualization* yang telah didiskusikan oleh tim sebagai berikut.

**1. Descriptive Statistics**

Pada bagian ini, akan digali informasi mengenai dataset.

**2. Univariate Analysis**

Mengetahui distribusi numerical feature dan proporsinya untuk categorical feature

**3. Multivariate Analysis**

Melihat hubungan atau korelasi antar feature dengan target dan feature dengan feature lainnya.

**4. Business Insight**

Menggali insight yang berhubungan dengan target dan memberikan rekomendasi dalam konteks bisnis terhadap insight tersebut.



Kelompok: Market Insider

Stage: 1

Mentor: Kevin

Pukul/ Tanggal: 20:00/ 29 November 2023

## Hasil Diskusi:

### 1. Descriptive Statistics

- Tambahkan informasi mengapa kolom ID, Z\_CostContact dan Z\_Revenue tidak digunakan dalam analisis pada file presentasi.
- Jabarkan mengenai nilai *unique* pada setiap kolom, contohnya kolom Marital Status.
- Tambahkan informasi *duplicate value* dan *missing value* ada di kolom mana.
- Informasi mengenai Yes Response dari kelima Campaign dimasukkan ke tahap analisis multivariat dan dibuat grafik.
- Lakukan analisis lebih dalam mengenai data yang summarynya aneh. Keanehan dari data dapat dianalisis dari nilai minimum dan maksimum atau bisa juga dari persentilnya.

### 2. Univariate Analysis

- Setelah melakukan kdeplot untuk melihat distribusi data, tidak perlu dijabarkan lagi mengenai kategori atau pada range mana yang menyebabkan data tersebut skewed.
- Tambahkan interpretasi dari kdeplot dan boxplot.
- *Barplot* untuk fitur kategori perlu disort dari tinggi ke rendah atau sebaliknya.
- *Line Plot* untuk fitur datetime dan kategori perlu dihubungkan dengan *Response* (pada grafik Customer Registration Based on Education & Marital Status) dan dipindahkan ke bagian analisis multivariat.



Kelompok: Market Insider

Stage: 1

Mentor: Kevin

Pukul/ Tanggal: 20:00/ 29 November 2023

## Hasil Diskusi:

### 3. Multivariate Analysis

- Setiap menampilkan grafik hubungan dari dua atau lebih fitur, perlu dihubungkan dengan Response.
- Belum ada informasi atau grafik mengenai satu fitur dan target.

Untuk fitur numerik dan target, bisa menggunakan *Distribution Plot* dengan huenya Response. Sementara untuk fitur kategori dan target, bisa menggunakan *Percentage Stacked Bar* dengan menampilkan persentasenya, bukan countnya. Contoh interpretasi: pelanggan dengan umur berapa yang banyak merespon?

- Tambahkan interpretasi dari *regression plot* dan *heatmap* yang dibuat. Kolom mana yang memiliki korelasi tinggi dengan Response?
- Tampilkan Heatmap dengan menunjukkan nilai korelasi dari semua fitur dan target untuk mendapatkan dua informasi berikut:

#### a. Multikolinearitas

Apakah terjadi korelasi yang tinggi antara dua fitur? Contohnya: *Amount Spent* untuk produk *Meat* dan *Wine* atau *Catalog Purchase* dan produk *Meat*, jadi bisa disimpulkan orang yang belanja katalog itu produk daging. Dengan adanya dua fitur yang multikolinear ini, di stage selanjutnya kita dapat melakukan *Dimentional Reduction* menggunakan metode *Principle Component Analysis* (PCA) yaitu menyederhanakan 2 fitur menjadi 1 fitur.

#### b. Hubungan/korelasi dengan fitur *Response*

Nilai yang dicari yang memiliki nilai absolutenya besar.

### 4. Business Insight

- Cukup masukkan 1 atau 2 insight dari grafik yang telah dibuat pada materi presentase.
- Drop grafik yang tidak terlalu terlihat signifikan perbedaannya, misalnya pada barplot Response based on Average Year Birth by Education.
- Grafik yang dibuat bisa dijadikan business insight, contoh: grafik Average Income by Education dengan interpretasi rata-rata pelanggan yang graduation dengan income 60.000 cenderung merespon.



**Kelompok: Market Insider**

**Stage: 1**

**Mentor: Kevin**

**Pukul/ Tanggal: 20:00/ 29 November 2023**

**Tindak Lanjut:**

Setelah mendapatkan feedback dari mentor, berikut adalah perbaikan pada stage EDA, Insight & Visualization:

**1. Descriptive Statistics :**

- Menambahkan duplicated value (hanya untuk menampilkan apakah ada data yang terduplikasi)

**2. Univariate Analysis**

- Memperbaiki tampilan countplot, diurutkan dari yang terbesar ke terkecil

**3. Multivariate Analysis**

- Menampilkan lineplot untuk mengetahui perbedaan jumlah yang enrollment dari waktu ke waktu berdasarkan customer yang response dan tidak response
- Menampilkan stacked bar pada categorical feature untuk melihat percentage yang response dengan yang tidak response (normalize)
- Menampilkan heatmap dengan nilai abs pada correlation ke target (correlation < 0)
- Memperhatikan feature-feature yang saling berkorelasi kuat selain dengan target, untuk selanjutnya dipertimbangkan untuk melakukan dimension reduction (pca)

**4. Business Insight**

- Menampilkan pie chart dan bar chart untuk merepresntasikan pengaruh campaign sebelumnya, dan perbedaan antara yang accept at least 1 campaign.
- Merepresentasikan perbedaan antara customer yang response dan tidak resposen, berdasarkan demografi (income, renceny, jumlah anak), berdasarkan produk yang sering dibeli (wine, meat, fish, dll), dan berdasarkan channel pembelian (catalog, web, store) pada tingkat Pendidikan (Education) dan status perkawinan (Marital Status)