**LAPORAN EKSPERIMEN**

Pengantar Data Science

Eksplorasi Dataset Banking

BANK\_INFO.csv

BANK\_TELCO\_ID.csv

MOBILE\_OS.csv

AGGREGATED\_TELCO\_TRANSACTIONS.csv

Dibuat oleh

6182001004 | William Surya Wijaya

6182001005 | Kae Zarren L H

61820010XX | Ester Oka

Tanggal

Sabtu, 22 Oktober 2022

Jurusan Informatika

UNIVERSITAS KATOLIK PARAHYANGAN

1. **Hipotesa – Pertanyaan mengenai data**
   1. **Deskripsi Dataset**

* Apa saja data yang tercantum dalam tiap dataset ?
  1. **Eksplorasi Nilai Tiap Kolom Numerik**
     1. Nilai Minimum
* Adakah nilai minimum yang tidak masuk akal pada tiap dataset ?
  + 1. Nilai Maksimum
* Adakah nilai maksimum yang tidak masuk akal pada tiap dataset ?
  + 1. Nilai (null) atau NaN
* Adakah nilai (null) atau NaN pada tiap dataset ?
  + 1. Nilai Error
* Adakah nilai Error pada tiap dataset ?
  1. **Eksplorasi Nilai Tiap Kolom Non-Numerik**
     1. Nilai Un-Categorized
* Adakah nilai Non-Numerik yang tidak masuk ke dalam kategori manapun pada tiap dataset ?
  + 1. Nilai Un-Defined atau NaN
* Adakah nilai Un-Defined atau NaN pada tiap data set ?
  + 1. Nilai Error
* Adakah nilai Error pada tiap dataset ?
  1. **Hubungan Dataset dengan Penggunaan**
     1. Gender Dominan dalam DT
* Gender manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Umur Dominan dalam DT
* Rentan umur manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Domisili Dominan dalam DT
* Domisili manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Pekerjaan Dominan dalam DT
* Pekerjaan mana kah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Jabatan Dominan dalam DT
* Jabatan manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  1. **Hubungan Nilai Antar kolom**
     1. Pekerjaan dengan Rata rata Transaksi Postpaid
* Pekerjaan manakah yang memiliki rata rata Transaksi Postpaid tertinggi dan berapakah jumlahnya?
  + 1. Pekerjaan dengan Rata rata Transaksi Prepaid
* Pekerjaan manakah yang memiliki rata rata Transaksi Prepaid tertinggi dan berapakah jumlahnya?
  + 1. Pekerjaan dengan Total Transaksi Postpaid
* Pekerjaan manakah yang memiliki total Transaksi Postpaid tertinggi dan berapakah jumlahnya?
  + 1. Pekerjaan dengan Total Transaksi Prepaid
* Pekerjaan manakah yang memiliki total Transaksi Prepaid tertinggi dan berapakah jumlahnya?
  + 1. Jabatan dengan Rata rata Transaksi Postpaid
* Jabatan manakah yang memiliki Rata rata Transaksi Postpaid dan berapakah jumlahnya?
  + 1. Jabatan dengan Rata rata Transaksi Prepaid
* Jabatan manakah yang memiliki Rata rata Transaksi Prepaid dan berapakah jumlahnya?
  + 1. Jabatan dengan Total Transaksi Postpaid
* Jabatan manakah yang memiliki total Transaksi Postpaid tertinggi dan berapakah jumlahnya?
  + 1. Jabatan dengan Total Transaksi Prepaid
* Jabatan manakah yang memiliki total Transaksi Prepaid tertinggi dan berapakah jumlahnya?

1. Visualisasi data
   1. **Deskripsi Dataset**

|  |  |  |
| --- | --- | --- |
| **No** | **Dataset** | **Columns (Data tercantum)** |
| 1. | BANK\_INFO | ['BANK\_ID', 'AGE', 'MARITAL\_STATUS', 'GENDER', 'EDUCATION', 'STATE', 'OCCUPATION', 'OCC\_DESC'] |
| BANK\_ID - ID Nasabah  AGE - Usia  MARITAL\_STATUS - status  GENDER - Jenis Kelamin  EDUCATION - Pendidikan  STATE - Domisili  OCCUPATION - Pekerjaan  OCC\_DESC - Jabatan |
| 2. | BANK\_TELCO\_ID | ['BANK\_ID', 'TELCO\_ID'] |
| BANK\_ID - ID Nasabah  TELCO\_ID - ID Telco Nasabah |
| 3. | AGGREGATED\_  TELCO\_  TRANSACTION | ['TELCO\_ID', 'PREPAID\_AMT\_MTH\_1', 'PREPAID\_AMT\_MTH\_2', 'PREPAID\_AMT\_MTH\_3', 'PREPAID\_AMT\_MTH\_4', 'PREPAID\_AMT\_MTH\_5', 'PREPAID\_AMT\_MTH\_6', 'POSTPAID\_AMT\_MTH\_1', 'POSTPAID\_AMT\_MTH\_2', 'POSTPAID\_AMT\_MTH\_3', 'POSTPAID\_AMT\_MTH\_4', 'POSTPAID\_AMT\_MTH\_5', 'POSTPAID\_AMT\_MTH\_6', 'TOTAL\_POSTPAID\_AMT', 'TOTAL\_PREPAID\_AMT', 'AVG\_POSTPAID\_AMT', 'AVG\_PREPAID\_AMT'] |
| PREPAID\_AMT\_MTH\_n - Trans Pre Telko Bulan-n  PREPOST\_AMT\_MTH\_n - Trans Pre Telko Bulan-n  TOTAL\_POSTPAID\_AMT\_MTH\_n - Trans Post Telko Total  TOTAL\_PREPAID\_AMT\_MTH\_n - Trans Pre Telko Total  AVG\_POSTPAID\_AMT - AVG Trans Post Telko / Bulan  AVG\_PREPAID\_AMT - AVG Trans Pre Telko / Bulan |
| 4. | MOBILE\_OS | ['BANK\_ID', 'MOBILE', 'TABLET', 'SAMSUNG', 'OPPO', 'HUAWEI', 'SONY', 'XIAOMI', 'LENOVO', 'ASUS', 'VIVO', 'ANDROID', 'IPHONE\_IOS11'] |
| BANK\_ID - ID Nasabah  xMobile - Nama Brand Media yang Digunakan Nasabah |

* 1. **Eksplorasi Nilai Tiap Kolom Numerik**
     1. Nilai Minimum

DATASET : BANK\_INFO

|  |  |  |
| --- | --- | --- |
| No | Colum name | Minimum Value |
|  |  |  |
|  |  |  |
|  |  |  |

DATASET : AGGREGATED\_TELCO\_TRANSACTION

|  |  |  |
| --- | --- | --- |
| No | Colum name | Minimum Value |
| 1. | PREPAID\_AMT\_MTH\_1 | 0. |
|  |  |  |
|  |  |  |

* + 1. Nilai Maksimum
* Adakah nilai maksimum yang tidak masuk akal pada tiap dataset ?
* Berapa jumlah nilai maximum tersebut ?
* Apa saja nilai maximum tersebut ?
  + 1. Nilai (null) atau NaN
* Adakah nilai (null) atau NaN pada tiap dataset ?
* Berapa jumlah nilai (null) atau NaN tersebut ?
* Apa saja nilai (null) atau NaN tersebut ?
  + 1. Nilai Error
* Adakah nilai Error pada tiap dataset ?
* Berapa jumlah nilai error tersebut ?
* Apa saja nilai error tersebut ?
  1. **Eksplorasi Nilai Tiap Kolom Non-Numerik**
     1. Nilai Un-Categorized
     2. Nilai (null)-Categorized

DATASET : BANK\_INFO

|  |  |  |
| --- | --- | --- |
| No | Colum name | Un-Categorized Value |
| 1. | MARITAL\_STATUS | Not Applicable / Blank |
| 2. | EDUCATION | No Formal Education / Unknown |
|  |  |  |

* + 1. Nilai Un-Defined atau NaN
* Adakah nilai Un-Defined atau NaN pada tiap data set ?
* Berapa jumlah nilai Un-Defined atau NaN tersebut ?
* Apa saja nilai Un-Defined atau NaN tersebut ?
  + 1. Nilai Error

DATASET : BANK\_INFO

|  |  |  |
| --- | --- | --- |
| No | Colum name | Error Value |
| 1. | STATE | STATE\_INFO\_NOT\_PROVIDED |
| 2. | OCCUPATION | SKILLED OPERATOR (NOT\_HANDLING |
|  |  |  |

* 1. **Hubungan Dataset dengan Penggunaan**
     1. Gender Dominan dalam DT
* Gender manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Umur Dominan dalam DT
* Rentan umur manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Domisili Dominan dalam DT
* Domisili manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Pekerjaan Dominan dalam DT
* Pekerjaan mana kah yang lebih dominan pada dataset dan berapakah jumlahnya?
  + 1. Jabatan Dominan dalam DT
* Jabatan manakah yang lebih dominan pada dataset dan berapakah jumlahnya?
  1. **Hubungan Nilai Antar kolom**
     1. Pekerjaan dengan Rata rata Transaksi Postpaid
* Pekerjaan manakah yang memiliki rata rata Transaksi Postpaid tertinggi dan berapakah jumlahnya?
  + 1. Pekerjaan dengan Rata rata Transaksi Prepaid
* Pekerjaan manakah yang memiliki rata rata Transaksi Prepaid tertinggi dan berapakah jumlahnya?
  + 1. Pekerjaan dengan Total Transaksi Postpaid
* Pekerjaan manakah yang memiliki total Transaksi Postpaid tertinggi dan berapakah jumlahnya?
  + 1. Pekerjaan dengan Total Transaksi Prepaid
* Pekerjaan manakah yang memiliki total Transaksi Prepaid tertinggi dan berapakah jumlahnya?
  + 1. Jabatan dengan Rata rata Transaksi Postpaid
* Jabatan manakah yang memiliki Rata rata Transaksi Postpaid dan berapakah jumlahnya?
  + 1. Jabatan dengan Rata rata Transaksi Prepaid
* Jabatan manakah yang memiliki Rata rata Transaksi Prepaid dan berapakah jumlahnya?
  + 1. Jabatan dengan Total Transaksi Postpaid
* Jabatan manakah yang memiliki total Transaksi Postpaid tertinggi dan berapakah jumlahnya?
  + 1. Jabatan dengan Total Transaksi Prepaid
* Jabatan manakah yang memiliki total Transaksi Prepaid tertinggi dan berapakah jumlahnya?

1. Kesimpulan