

# LAPORAN HASIL ANALISA TUGAS BESAR

## STATISTIKA DATA CREDIT CARD



### **Kelompok 5 :**

Quin Derbi Kusuma	1395210071
Muhammad Ali Akbar Al-Qahri	1305210077
Aqeela Fathya Najwa	1305213021

Program Studi Data Sains

Fakultas Informatika

Universitas Telkom

Bandung

2022

## A. Latar Belakang

Dalam pengertian sehari-hari data dapat berarti fakta dari suatu objek yang diamati, yang dapat berupa angka-angka maupun kata-kata. Sedangkan jika dipandang dari sisi Statistika, maka data merupakan Fakta-fakta yang akan digunakan sebagai bahan penarikan kesimpulan. (Siswandari, 2009).

Untuk kasus kali ini kami menganalisis sebuah data yang berasal dari Kaggle. Data ini berisikan data mengenai Data kartu kredit seorang nasabah dari sebuah bank yang memuat nama, umur, pendapatan, dan lainnya. Dan dengan data-data yang sudah disajikan maka akan dianalisisnya sebuah masalah yaitu untuk melihat berapa banyak nasabah yang berhenti menggunakan kartu kredit dari bank tersebut.

## B. Deskripsi Data :

Manager yang berada dalam sebuah bank ingin mendapatkan gambaran dari para pemegang kartu kredit yang mereka terbitkan. Ia menghadapi masalah pengurangan pelanggan. Mereka ingin menganalisis data untuk mengetahui alasan di balik ini dan memanfaatkan hal yang sama untuk memprediksi pelanggan yang kemungkinan besar akan berhenti.

Kolom A-U dalam excel merupakan data yang berisi fitur-fitur pemegang kartu kredit.

Kolom V dan W merupakan data peluang pemegang kartu kredit berpindah ke bank lain.

<https://www.kaggle.com/sakshigoyal7/credit-card-customers>

- **Data Kategorikal :**

- |                   |                   |
|-------------------|-------------------|
| - Attrition_Flag  | - Marital_Status  |
| - Gender          | - Income_Category |
| - Education_Level | - Card_Category   |

- **Data Numerik (Kontinu) :**

- |                            |  |
|----------------------------|--|
| - CLIENTNUM                | - Total_Ct_Chng_Q4_Q1  |
| - Customer_Age             | - Avg_Utilization_Ratio  |
| - Dependent_count          | - Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_1 |
| - Months_on_book           | - Naive_Bayes_Classifier_Attrition_Flag_Card_Category_Contacts_Count_12_mon_Dependent_count_Education_Level_Months_Inactive_12_mon_2 |
| - Total_Relationship_Count |  |
| - Months_Inactive_12_mon   |  |
| - Contacts_Count_12_mon    |  |
| - Credit_Limit             |  |
| - Total_Revolving_Bal      |  |
| - Avg_Open_To_Buy          |  |
| - Total_Amt_Chng_Q4_Q1     |  |
| - Total_Trans_Amt          |  |
| - Total_Trans_Ct           |  |

## C. Hasil Pengolahan Data

### 1. Korelasi

	Customer Age	Dependent	Months on	Total Relati	Months Ina	Contacts C	Credit Limit	Total Revol	Avg Open T	Total Amt C	Total Trans	Total Trans Ct	Total Ct Chi	Avg Utilizati	Attrition Fla
Customer Age	1,00														
Dependent count	-0,12	1,00													
Months on book	0,79	-0,10	1,00												
Total Relationship Count	-0,01	-0,04	-0,01	1,00											
Months Inactive 12 mon	0,05	-0,01	0,07	0,00	1,00										
Contacts Count 12 mon	-0,02	-0,04	-0,01	0,06	0,03	1,00									
Credit Limit	0,00	0,07	0,01	-0,07	-0,02	0,02	1,00								
Total Revolving Bal	0,01	0,00	0,01	0,01	-0,04	-0,05	0,04	1,00							
Avg Open To Buy	0,00	0,07	0,01	-0,07	-0,02	0,03	1,00	-0,05	1,00						
Total Amt Chng Q4 Q1	-0,03	-0,05	-0,03	0,07	-0,03	0,02	0,01	0,04	0,01	1,00					
Total Trans Amt	-0,05	0,03	-0,04	-0,35	-0,04	-0,11	0,17	0,06	0,17	-0,11	1,00				
Total Trans Ct	-0,07	0,05	-0,05	-0,24	-0,04	-0,15	0,08	0,06	0,07	-0,17	0,81	1,00			
Total Ct Chng Q4 Q1	0,02	0,02	0,01	0,04	-0,01	-0,03	0,00	0,04	0,00	0,19	-0,10	-0,14	1,00		
Avg Utilization Ratio	0,01	-0,04	-0,01	0,07	-0,01	-0,06	-0,48	0,62	-0,54	0,01	-0,08	0,00	0,03	1,00	
Attrition Flag	0,02	0,02	0,01	-0,15	0,15	0,20	-0,02	-0,26	0,00	-0,03	-0,17	-0,37	-0,06	-0,18	1,00

Gambar 1.1 Hasil Korelasi antara kolom

Berdasarkan dari beberapa kolom yang disajikan di dalam data tersebut, kami membuat korelasi atau hubungan antar kolom dan mencari nilai yang paling besar. Disini kami akan memfokuskan untuk melihat ke baris atau kolom attrition flag karena ini menjadi focus utama dan di barik atau kolom ini juga yang paling banyak menampung nilai korelasi dengan kolom lainnya. Untuk mempermudah menganalisanya kami mengubah menjadi table sederhana.

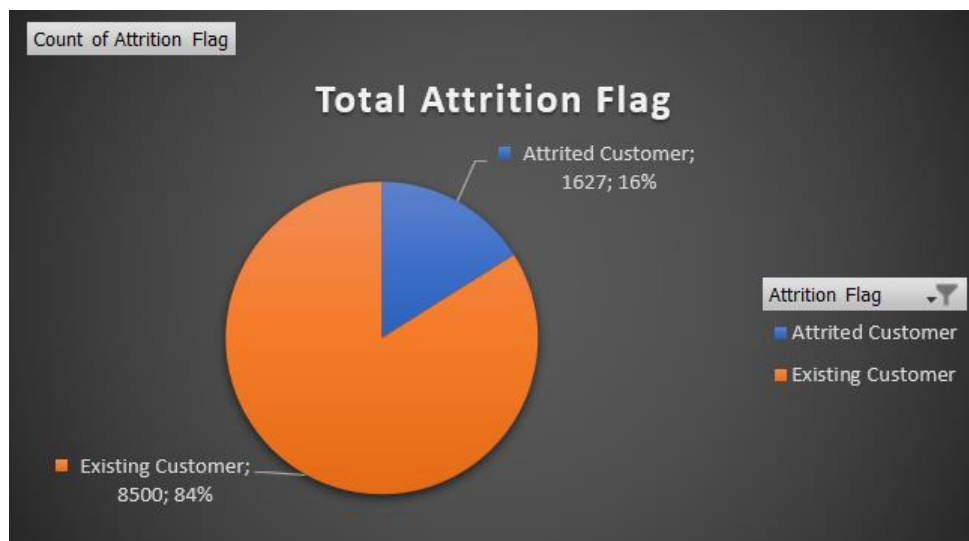
	Attrition Flag
Customer Age	0,02
Dependent count	0,02
Months on book	0,01
Total Relationship Count	-0,15
Months Inactive 12 mon	0,15
Contacts Count 12 mon	0,20
Credit Limit	-0,02
Total Revolving Bal	-0,26
Avg Open To Buy	0,00
Total Amt Chng Q4 Q1	-0,03
Total Trans Amt	-0,17
Total Trans Ct	-0,37
Total Ct Chng Q4 Q1	-0,06
Avg Utilization Ratio	-0,18

Gambar 1.2 Hasil Korelasi Attrition Flag

Dari korelasi tersebut kami akan mengambil nilai tertinggi karena jika semakin tinggi angkanya maka makin mendekati korelasi positif sempurna. Yang dimaksud korelasi positif adalah Ketika kedua variable mempunyai hubungan searah.

Disini yang menjadi nilai tertingginya adalah Contacts Count 12 mon dan Months Inactive 12 mon. dan yang paling terendah adalah Total Trans Ct. Contacts Count 12 mon yang sebesar 0,20 yang berarti termasuk kategori positif lemah dan juga sama seperti Months Inactive 12 mon.

## 2. Pengolahan data dari Attrition Flag

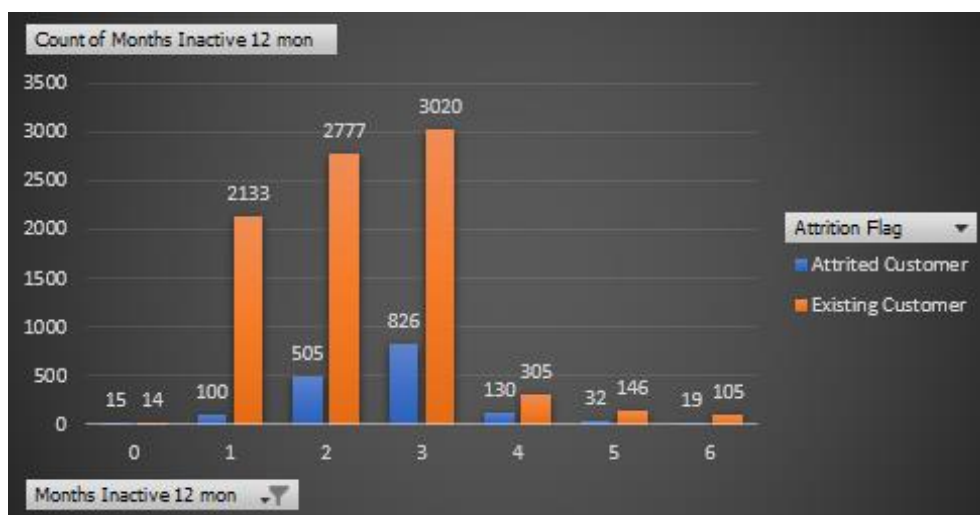


Gambar 2.1 Hasil Pengolahan Data dari Attrition Flag

Berdasarkan hasil pengolahan data yang telah dilakukan terhadap kolom attrition flag didapatkan suatu informasi yaitu dari 10127 nasabah yang menggunakan kartu kredit tersebut terdapat 1627 nasabah yang sudah meninggalkan atau berhenti menggunakan kartu kredit dari bank tersebut dan terdapat 8500 nasabah yang masih menggunakan kartu kredit ini.

Jika dilihat dari persentasenya maka nasabah yang sudah tidak menggunakan kartu kredit lagi sebesar 16% sedangkan yang masih menggunakan kartu kredit tersebut sebesar 84%.

## 3. Pengolahan data Attrited Customer dari kolom Months Inactive 12 Mon



Gambar 3.1 Hasil Pengolahan Data Attrition Flag di Months Inactive 12 month

Selanjutnya kami menganalisis kolom lainnya, dan kami mendapati data di kolom Months inactive 12 month terdapat ketimpangan yang sangat besar dari diagram batang tersebut, yaitu attrited customer di bagian ke 2 dan 3 lebih tinggi dibandingkan yang lainnya.

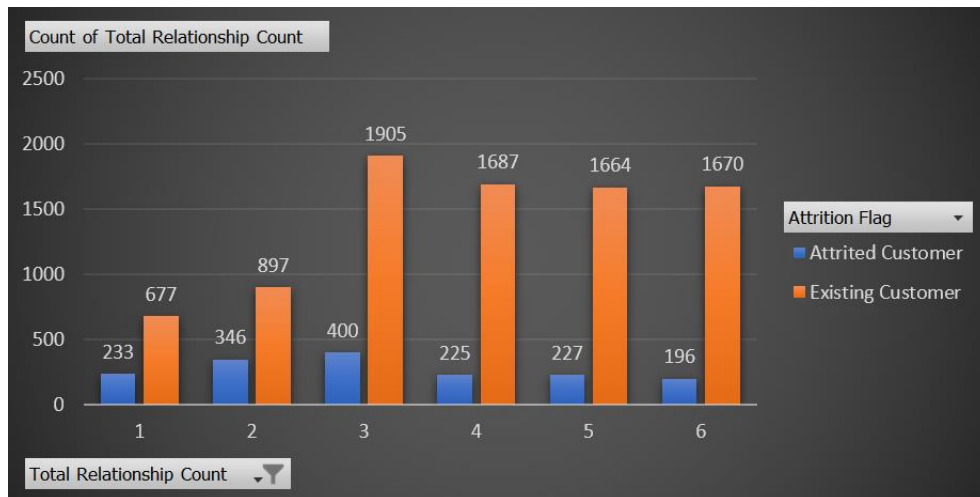
Perbedaan yang signifikan terlihat dengan bagian ke 0 dan 6 yang hanya belasan tidak seperti bagian 2 dan 3 yang terdiri dari ratusan. Oleh karena itu kami memecah data tersebut menjadi perbagiannya untuk analisis yang lebih lanjut.



Gambar 3.2 Hasil Pengolahan Data Attrition Flag di Months Inactive 12 month per kategori

Setelah kami bagi menjadi beberapa bagian untuk melihat dengan lebih jelas persentase per kategorinya, kami mendapati keadaan dimana besar presentase di bagian 0 memiliki keadaan yang lebih besar daripada bagian lainnya yaitu jumlahnya lebih dari setengah customer yaitu sebesar 52%, selanjutnya diikuti dengan bagian 4 dengan jumlah lebih dari seperempat customer yaitu sebanyak 30%. Sedangkan bagian 1 memiliki persentase terkecil yaitu hanya 4%. Dan untuk bagian 2, 3, dan 5 tidak jauh berbeda yaitu hanya sebesar 15%, 21%, dan 18%.

#### 4. Pengolahan data Attritted Customer dari kolom Total Relationship Count



Gambar 4.1 Hasil Pengolahan Data Attrition Flag di Total Relationship Count

Selanjutnya kami menganalisis kolom lainnya, dan kami mendapati data di kolom Total Relationship Count terdapat ketimpangan yang walau tidak terlalu besar dari diagram batang tersebut tetapi cukup terlihat, yaitu attrited customer di bagian ke 2 dan 3 lebih tinggi dibandingkan yang lainnya.

Oleh karena itu kami memecah data tersebut menjadi perbagiannya untuk analisis yang lebih lanjut.



Gambar 4.2 Hasil Pengolahan Data Attrition Flag di Total Relationship Count per kategori

Setelah kami bagi menjadi beberapa bagian untuk melihat dengan lebih jelas persentase per kategorinya, kami mendapati keadaan dimana besar persentase di bagian 2 memiliki keadaan yang lebih besar daripada bagian lainnya yaitu jumlahnya sekitar seperempat customer yaitu sebesar 28%, selanjutnya diikuti dengan bagian 1 yang tidak jauh berbeda yaitu dengan jumlah 26%. Sedangkan bagian 6 memiliki persentase terkecil yaitu 11%, lalu diikuti dengan bagian 4 dan 5 yang tidak jauh berbeda yaitu 12%. Dan terakhir bagian 3 yang jumlah persentasenya sebanyak 17%.

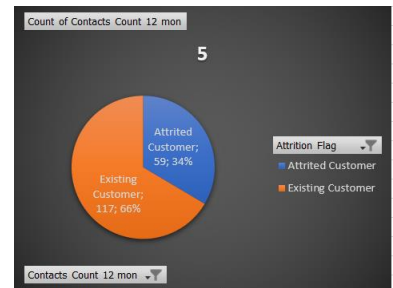
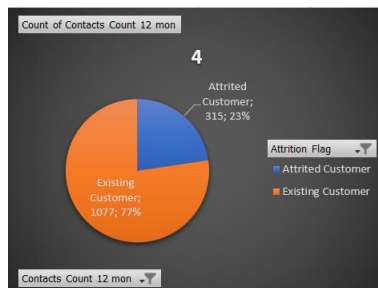
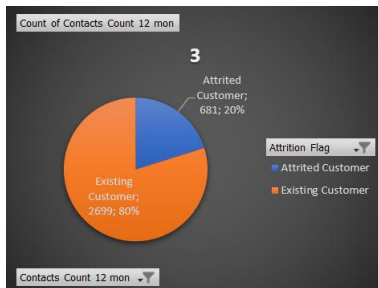
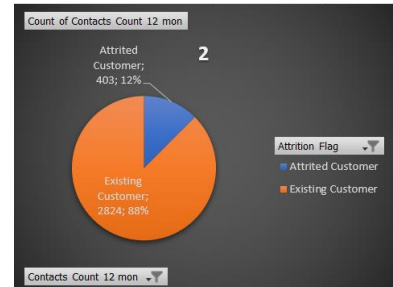
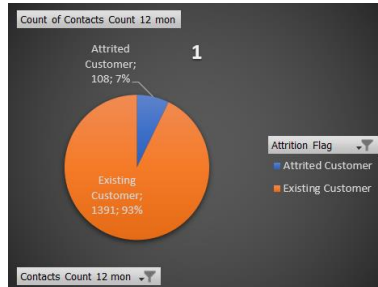
## 5. Pengolahan data Attritted Customer dari Contacts Count 12 Mon



Gambar 5.1 Hasil Pengolahan Data Attrition Flag di Contacts Count 12 Mon

Selanjutnya kami menganalisis kolom lainnya, dan kami mendapati data di kolom Contacts Count 12 Mon yaitu terdapat ketimpangan yang cukup besar dari diagram batang tersebut, yaitu attritted customer di bagian ke 2 dan 3 lebih tinggi dibandingkan yang lainnya.

Perbedaan yang signifikan terlihat dengan bagian ke 0 dan 6 yang tidak mencapai ratusan tidak seperti bagian 2 dan 3 yang terdiri dari ratusan. Oleh karena itu kami memecah data tersebut menjadi perbagiannya untuk analisis yang lebih lanjut.



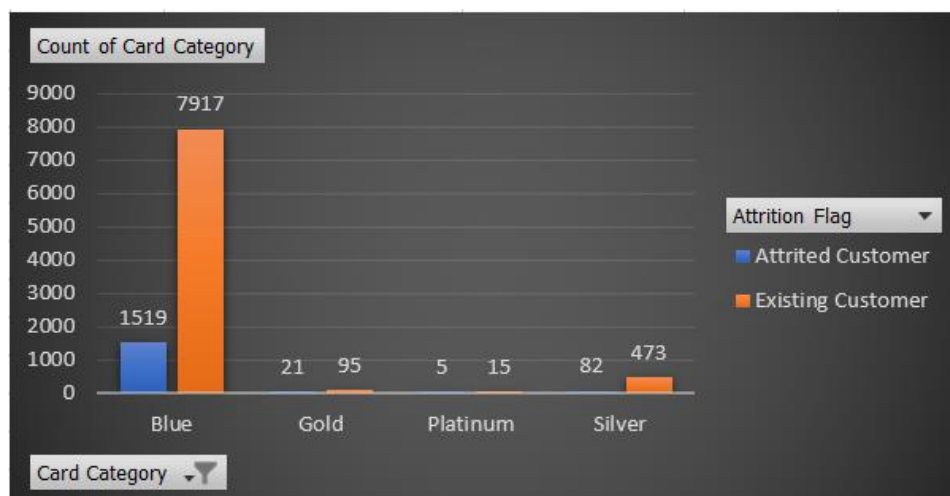


Gambar 5.2 Hasil Pengolahan Data Attrition Flag di Contacts Count 12 Mon per kategori

Setelah kami bagi menjadi beberapa bagian untuk melihat dengan lebih jelas persentase per kategorinya, kami melihat dari bagian ini bahwa semakin besar angkanya maka jumlah attrited customer nya semakin besar.

Kami mendapati keadaan dimana besar presentase di bagian 6 memiliki keadaan yang paling besar daripada bagian lainnya yaitu sebesar 100% atau artinya keseluruhan dari customer. Selanjutnya ada bagian 5 yang jumlahnya lebih dari seperempat customer yaitu sebesar 34%. Diikuti dengan bagian 3 dan 4 dengan jumlah kurang dari seperempat customer yaitu sebanyak 20% dan 23%. Sedangkan bagian 0 memiliki persentase terkecil yaitu hanya 2%. Dan untuk bagian 1 dan 2 tidak jauh berbeda yaitu hanya sebesar 7% dan 12%.

## 6. Pengolahan data Attrited Customer dari Card Category

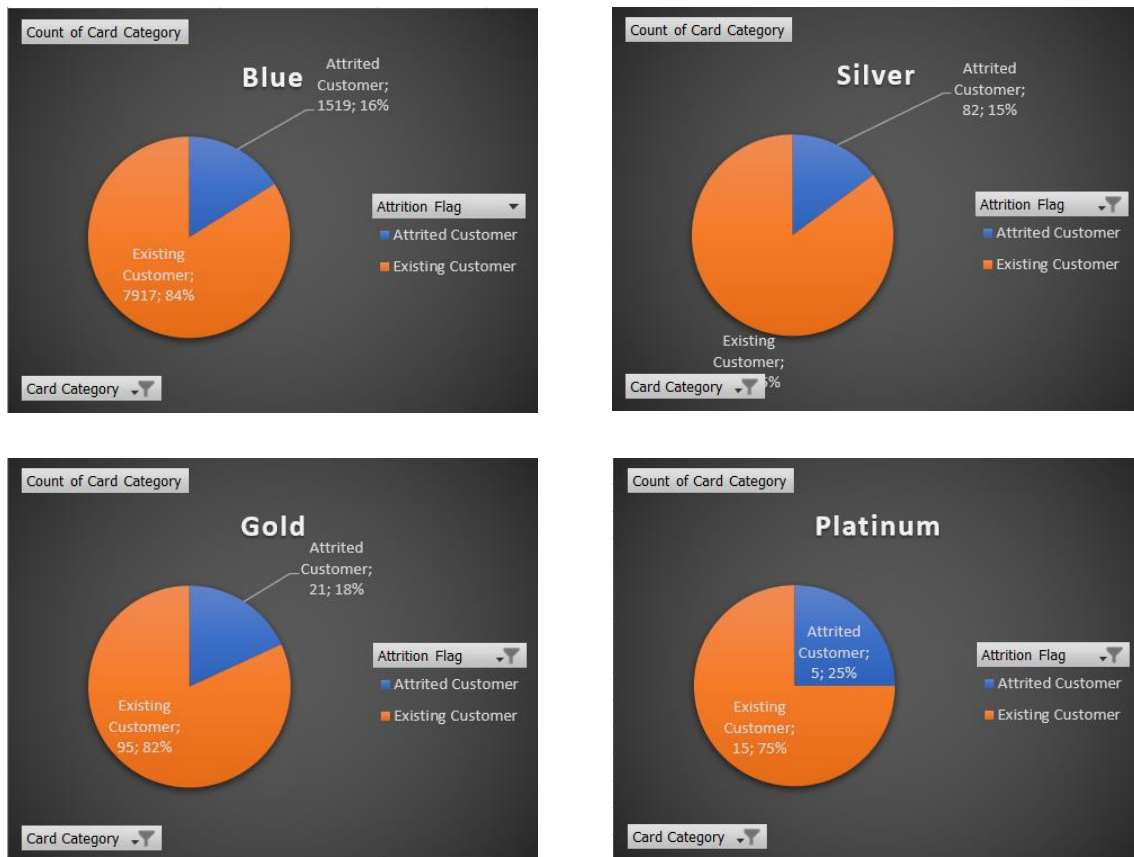


Gambar 6.1 Hasil Pengolahan Data Attrition Flag di Card Category

Selanjutnya kami menganalisis kolom lainnya, dan kami mendapati data di kolom Months inactive 12 month terdapat ketimpangan yang sangat besar dari diagram batang tersebut, yaitu attrited customer di bagian ke 2 dan 3 lebih tinggi dibandingkan yang lainnya.

Perbedaan yang signifikan terlihat dengan bagian ke 0 dan 6 yang hanya belasan tidak seperti bagian 2 dan 3 yang terdiri dari ratusan. Oleh karena itu kami memecah data tersebut menjadi perbagiannya untuk analisis yang lebih lanjut.





Gambar 6.2 Hasil Pengolahan Data Attrition Flag di Card Category per kategori

Setelah kami bagi menjadi beberapa bagian untuk melihat dengan lebih jelas persentase per kategorinya, kami mendapati keadaan dimana besar persentase di kategori kartu platinum memiliki persentase yang lebih besar daripada bagian lainnya yaitu jumlahnya sebanyak seperempat customer yaitu sebesar 25%. Sedangkan kategori silver memiliki persentase terkecil yaitu 15%. Dan diikuti kategori blue dan gold yang tidak jauh berbeda yaitu sebesar 16% dan 18%.

#### D. Kesimpulan

Berdasarkan data yang telah kita olah yaitu data yang berasal dari Bank Churners kami mendapati 10.127 data nasabah dengan jumlah kolom yaitu 23 dengan tujuan untuk mendapatkan informasi dari banyaknya customer yang berhenti berlangganan dengan kartu kredit mereka.

Pertama-tama kami mengolah data dari kolom attrition flag yang ditujukan agar kami mengetahui terlebih dahulu berapa banyak customer yang telah meninggalkan atau berhenti berlangganan kartu kredit ini dan berapa banyak yang masih menggunakannya. Maka kami mendapatkan informasi bahwa customer yang tidak lagi menggunakan kartu kredit ini sebanyak 1.627 nasabah dan jumlah tersebut merupakan 16% dari keseluruhan nasabah. Dan sebanyak 8.500 nasabah lainnya atau 84% merupakan nasabah yang masih menggunakan kartu kredit tersebut. Karena disini tujuan kita adalah mencari informasi mengenai data customer yang tidak lagi menggunakan kartu tersebut maka kami memfokuskan pengolahan data hanya dengan kategori nasabah yang sudah berhenti saja.

Selanjutnya kami mengamati dan menganalisis kolom selanjutnya yang bersifat umum seperti umur, jenis kelamin, tanggungan, pendidikan, dan lainnya namun kami tidak mendapati hasil

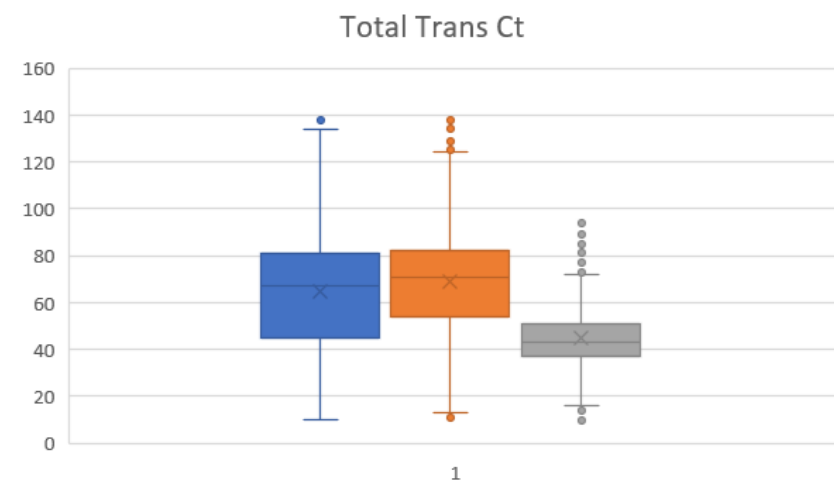
signifikan yang terlalu berbeda. Hingga akhirnya kami menganalisis kolom kategori kartu dan mendapati data yang cukup berbeda. Card category berisikan kategori blue, silver, gold, dan platinum. Kartu dengan kategori blue memiliki tingkat customer yang sangat tinggi dan memiliki jarak yang jauh dibandingkan dengan kategori lainnya. Karena kartu blue terlalu mendominasi sehingga susah untuk kami mengolah data tersebut maka kami membaginya berdasarkan kategorinya agar lebih mudah untuk membandingkannya. Setelah dilakukannya pembagian kami mendapati bahwa persentase kartu blue, silver, dan gold tidak terlalu jauh yaitu hanya sekitar 15-18%. Berbeda dengan kartu platinum yang memiliki presentase yang cukup tinggi yaitu sebesar 25%. Maka disini kami mendapati bahwa nasabah yang memiliki kartu platinum memiliki kecenderungan yang lebih tinggi untuk berhenti berlangganan kartu kredit. Dan didapatkan Platinum memiliki kecenderungan attrited customer yang lebih tinggi dibandingkan kategori kartu lainnya.

Selanjutnya kami menganalisis kolom Months Inactive 12 Mon, Total Relationship Count, Contacts Count 12 Mon. Sama seperti apa yang telah kami lakukan sebelumnya kami menganalisis secara keseluruhan lebih dahulu setelah itu baru per kategorinya.

Untuk kolom Months Inactive 12 Mon nasabah yang memiliki kecenderungan tinggi untuk berhenti menggunakan kartu kredit ini yaitu yang tidak aktif sebanyak 3 kali.

Selanjutnya untuk kolom Total Relationship Count yang memiliki kecenderungan yang tinggi ialah yang memiliki jumlah sebanyak 2. Dari data ini didapat bahwa rata-rata nasabah berhenti menggunakan kartu kredit dengan jumlah relationship count 3.8 atau jika dilakukan pembulatan sebanyak 4.

Selanjutnya untuk Contacts Count 12 Mon terdapat 1 bagian yang memiliki nasabah yang tidak lagi menggunakan kartu kredit ini sebesar 100% yaitu nasabah yang 6x kontak.



Gambar 6.1 Box Plot

Dari Box Plot diatas kami berasumsi card platinum kecenderungan lebih tinggi karena pelayanan card platinum kurang memuaskan nasabah. Dari grafik diatas dapat kita simpulkan bahawasanya nasabah yang memiliki transaksi tinggi lebih banyak exiting daripada yang memiliki transaksi rendah.

Box Plot Biru : Total Transaksi Attrited dan Existing

Box Plot Orange : Total Transaksi Attrited

Box Plot Abu : Total Transaksi Existing

all	
min	26
max	73
mode	44
mean	46,32596
median	46
range	47

Gambar 6.2 Distribusi Data Umur Seluruh Nasabah

Attrited	
min	26
max	68
mode	43
mean	46,6595
median	47
range	42

Gambar 6.3 Distribusi Data Umur Attrited Customer

Disini kami juga mendapatkan informasi mengenai umur nasabah, umur maksimal dari nasabah yang sudah meninggalkan yaitu berusia 68 yang berbeda dengan umur maksimal dari keseluruhan nasabah yaitu 73. Dan modus atau yang terbanyak usia dari nasabah yang tidak lagi menggunakan kartu kredit yaitu berusia 43 hanya berbeda 1 orang dengan modus keseluruhan nasabah yaitu 44. Dan untuk rata-rata usia tidak terdapat perbedaan untuk keseluruhan nasabah dengan attrited nasabah yaitu berusia 46.

Mungkin ini semua hasil analisa yang dapat menjadi evaluasi bagi bank terhadap pelayanan atau hubungan antara bank dengan nasabah yang bisa untuk diperbaiki kedepannya dan menghindari factor-faktor yang telah disajikan.