

Data Analytics for Business 2024

Project Based Learning

Study Case

Linkit 360

CS 05 - 8 :

- | | | |
|----|-------------------------|-----------------|
| 1. | Michael Luwi Pallea | CS05-KM-CS05278 |
| 2. | Keisya Akhmala Tazkiyah | CS05-KM-CS05509 |
| 3. | Rayhan Anugrah Ramadhan | CS05-KM-CS05488 |

Kata Pengantar

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga kami dapat menyelesaikan laporan ini dengan baik. Laporan ini disusun sebagai salah satu bagian dari proses pembelajaran yang bertujuan untuk mengembangkan kemampuan kami dalam menganalisis data dan menyampaikan insight yang relevan.

Kami ingin mengucapkan terima kasih yang sebesar-besarnya kepada Kak Sarah selaku *Student Success* yang telah memberikan arahan dan dukungan sepanjang proyek ini, serta Kak Niam sebagai mentor yang dengan sabar membimbing kami dalam memahami materi dan menjawab berbagai pertanyaan yang kami ajukan. Kontribusi, waktu, dan perhatian yang diberikan sangat berharga bagi perkembangan kami selama proses pengerjaan ini.

Tidak lupa, kami juga berterima kasih kepada teman-teman dan pihak-pihak lain yang telah memberikan motivasi dan bantuan, baik secara langsung maupun tidak langsung, dalam menyelesaikan proyek ini.

Kami menyadari bahwa laporan ini masih jauh dari sempurna, oleh karena itu kami membuka diri untuk menerima kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat dan wawasan yang berguna bagi pembaca.

Demikian, kami sampaikan penghargaan dan rasa terima kasih kami.

Daftar Isi

Kata Pengantar	1
Daftar Isi	2
BAB I	4
A. Latar Belakang	4
B. Identifikasi Masalah	8
BAB II	9
A. Data dan Sumber Data	9
B. Alur Pengerjaan Study Case	10
1. Tools yang digunakan dalam proses menganalisis data	10
2. Proses Analisis Data	10
BAB III	17
A. Deskripsi Data Hasil Analisis Bisnis	17
B. Hasil Analisis	35
BAB IV	47
A. Kesimpulan	47
B. Saran	48
Lampiran	49

BAB I

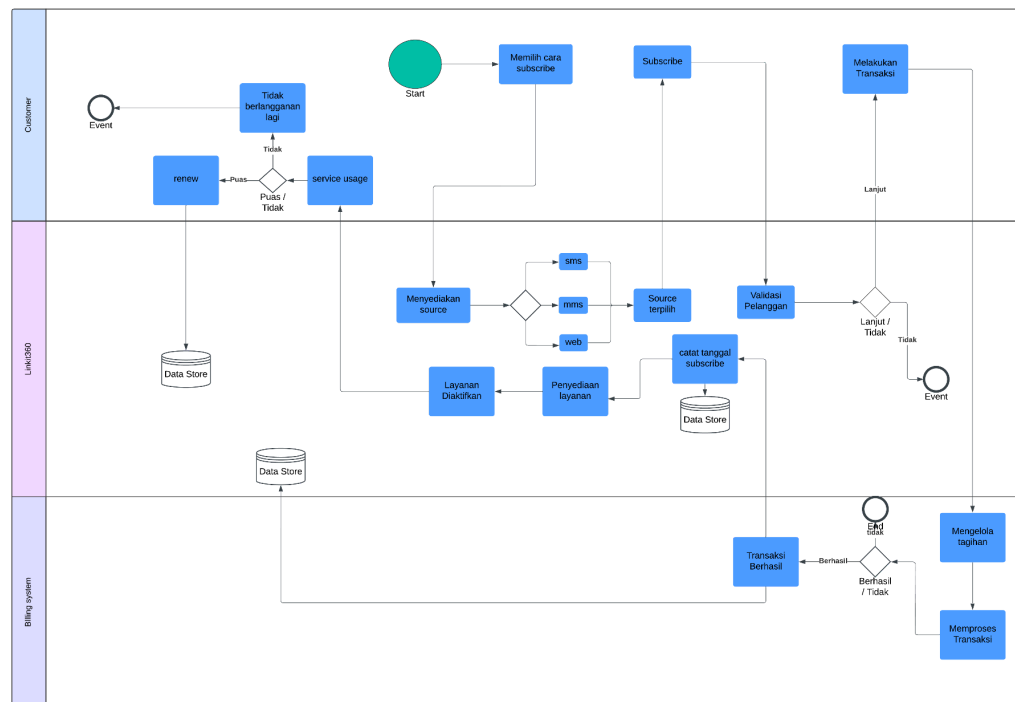
Pendahuluan

A. Latar Belakang

Perusahaan Value Added Service (VAS) Linkit360 adalah penyedia layanan tambahan berbasis mobile yang menawarkan berbagai layanan digital seperti langganan digital, konten hiburan, dan penawaran khusus untuk operator telekomunikasi di berbagai negara. Dengan meningkatnya persaingan di industri layanan digital, perusahaan berfokus pada peningkatan keterlibatan pengguna, retensi pelanggan, dan optimasi pendapatan.

Namun, perusahaan menghadapi tantangan yang kompleks, seperti tingkat churn pelanggan yang tinggi, keterlibatan pengguna yang menurun, dan optimasi revenue yang belum maksimal. Masalah ini tidak hanya berdampak pada pertumbuhan perusahaan tetapi juga pada upaya untuk mempertahankan posisi kompetitif di pasar. Oleh karena itu, analisis data yang mendalam menjadi kunci untuk memahami perilaku pelanggan, mengidentifikasi faktor-faktor yang memengaruhi churn, dan merancang strategi yang lebih efektif untuk meningkatkan revenue.

Untuk membantu memahami proses bisnis pada perusahaan, diagram BPMN (Business Process Model and Notation) berikut pada **Gambar 1** digunakan sebagai alat visualisasi yang menggambarkan alur proses utama dalam layanan perusahaan. Diagram ini mencakup langkah-langkah mulai dari pendaftaran langganan hingga pengelolaan transaksi dan keputusan pelanggan untuk memperpanjang atau menghentikan langganan.



Gambar 1 BPMN

Adapun penjelasan proses dari BPMN tersebut adalah sebagai berikut.

1. Pelanggan Memilih Langganan:

Pelanggan memulai dengan memilih metode langganan yang tersedia (misalnya SMS, MMS, atau Web). Mereka kemudian mengisi proses untuk "Subscribe" (berlangganan).

2. Validasi Pelanggan:

Setelah pelanggan mencoba berlangganan, sistem akan melakukan validasi untuk memastikan bahwa pelanggan memenuhi syarat, seperti pengecekan data atau status pembayaran.

3. Pengaktifan Layanan:

Jika validasi berhasil, layanan diaktifkan untuk pelanggan. Sistem menyediakan sumber layanan sesuai dengan metode yang dipilih pelanggan (SMS, MMS, atau Web). Proses ini juga melibatkan pencatatan data pelanggan, seperti tanggal mulai berlangganan, ke dalam basis data.

4. Penggunaan Layanan:

Setelah layanan diaktifkan, pelanggan dapat menggunakannya. Sistem memantau penggunaan ini secara otomatis.

5. Manajemen Transaksi:

Sistem billing menangani proses transaksi untuk layanan yang digunakan pelanggan. Jika transaksi berhasil diproses, layanan tetap berjalan; jika tidak, pelanggan mungkin kehilangan akses.

6. Perpanjangan atau Berhenti Berlangganan:

Di akhir siklus langganan, pelanggan dapat memilih untuk memperpanjang layanan ("Renew") atau berhenti berlangganan ("Tidak Berlangganan Lagi"). Jika pelanggan memilih untuk berhenti, layanan dihentikan dan dicatat dalam sistem.

BPMN ini tidak hanya membantu memvisualisasikan alur kerja, tetapi juga menjadi panduan penting untuk memahami titik-titik kritis dalam perjalanan pelanggan.

Dataset yang digunakan dalam analisis ini disediakan oleh perusahaan dan mencerminkan berbagai aspek perjalanan pelanggan sesuai dengan BPMN. Data ini mencakup:

1. Data langganan pelanggan: Informasi mengenai pendaftaran dan penghentian langganan.
2. Data keterlibatan pengguna: Seperti jumlah percobaan pengisian, keberhasilan transaksi
3. Data demografis pelanggan: seperti lokasi
4. Data interaksi layanan: Catatan tentang bagaimana pelanggan menggunakan layanan, termasuk metode akses (SMS, MMS, atau Web), serta layanan yang dipilih.

Adapun tahapan analisis yang dilakukan untuk menyelesaikan masalah ini meliputi:

- 1) Pemahaman Proses Bisnis: Menganalisis alur kerja yang tergambar dalam diagram BPMN untuk mengidentifikasi titik-titik kritis dalam perjalanan pelanggan.

- 2) Eksplorasi dan Pembersihan Data (EDA): Meneliti data untuk memahami pola dan tren, sekaligus membersihkan data untuk memastikan keakuratannya.
- 3) Analisis Data dengan SQL: Melakukan query untuk menggali informasi penting dari data
- 4) Visualisasi Data: Membuat dashboard interaktif menggunakan Tableau untuk menampilkan temuan dengan cara yang mudah dipahami oleh pemangku kepentingan.

Output yang diharapkan dari analisis ini adalah wawasan berbasis data mengenai pola perilaku pelanggan dan faktor-faktor yang mempengaruhi churn serta keterlibatan layanan, serta strategi untuk menaikkan revenue.

Sebagai bagian dari analisis data, *Product Canvas* pada **Gambar 2** digunakan untuk memetakan secara ringkas tujuan, solusi, dan pendekatan yang diperlukan dalam menangani tantangan dan masalah yang sebelumnya di jabarkan.

Product Canvas		
Objective <ul style="list-style-type: none"> - Meningkatkan keterlibatan layanan - Mengurangi churn - Mengoptimalkan aliran revenue 	Solution <ul style="list-style-type: none"> - Analisa data - Menemukan tren 	key metrics <ul style="list-style-type: none"> - Pendapatan (revenue) - revenue per service - revenue berdasarkan bulan - Tingkat Churn rate - churn rate by subscription cycle
Methodology <ul style="list-style-type: none"> - Pemahaman Bisnis - Analisa data dengan EDA - data extraction dengan SQL - Data Dashboarding dengan Tableau 	User Segments <ul style="list-style-type: none"> - Pelanggan berdasarkan Wilayah - Segmen berdasarkan Operator - berdasarkan service yang digunakan 	Business Benefit <ul style="list-style-type: none"> - Meningkatkan retensi melalui intervensi yang tepat sasaran - mengetahui area yang perlu dilakukan peningkatan

Gambar 2 Product Canvas

B. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan, berikut adalah penjabaran masalah yang menjadi fokus utama dalam analisis ini

1. Tingkat churn pelanggan yang tinggi

Banyak pelanggan yang berhenti menggunakan layanan dalam periode tertentu, menunjukkan tingkat retensi yang rendah. Hal ini tidak hanya berdampak langsung pada pendapatan perusahaan, tetapi juga mencerminkan adanya potensi masalah dalam pengalaman pelanggan atau relevansi layanan yang ditawarkan.

2. Rendahnya keterlibatan pengguna terhadap layanan

Meskipun layanan yang ditawarkan cukup bervariasi, tingkat keterlibatan pengguna masih rendah. Pelanggan yang jarang menggunakan layanan berpotensi menjadi pelanggan yang churn di masa mendatang.

3. Kendala dalam sistem billing

Proses billing yang tidak optimal, seperti kegagalan transaksi atau penagihan yang tidak berhasil, menimbulkan dampak ganda: pengalaman pelanggan yang buruk dan kerugian langsung pada pendapatan. Masalah ini menunjukkan perlunya perbaikan teknis dan sistematis untuk meningkatkan keberhasilan proses penagihan dan kepercayaan pelanggan terhadap layanan.

BAB II

Metodologi

A. Data dan Sumber Data

1. Sumber Data

Dataset yang digunakan diperoleh dari Learning Management System (LMS) yang telah disiapkan oleh tim Linkit 360. Data ini dirancang khusus untuk mendukung analisis dan eksplorasi dalam penelitian yang dilakukan.

2. Jenis Data

Dataset ini terdiri dari berbagai atribut yang relevan untuk analisis, mencakup data terkait pelanggan, aktivitas transaksi, dan performa bisnis. Berikut adalah daftar atribut pada dataset:

- **created_at**: Tanggal dan waktu pembuatan data.
- **country**: Negara tempat pengguna berada.
- **operator**: Operator telekomunikasi pengguna.
- **service**: Layanan yang diakses oleh pengguna.
- **source**: Sumber asal pengguna bergabung.
- **msisdn**: Nomor telepon pengguna.
- **status**: Status pelanggan (aktif, tidak aktif atau sudah tidak berlangganan).
- **cycle**: Siklus langganan pengguna.
- **adnet**: Jaringan iklan yang digunakan untuk akuisisi pelanggan.
- **revenue**: Pendapatan yang dihasilkan dari pengguna.
- **subs_date**: Tanggal langganan dimulai.
- **unsubs_from**: Sumber pemberhentian langganan.
- **unsubs_date**: Tanggal pemberhentian langganan.
- **service_price**: Harga layanan yang diberikan.
- **currency**: Mata uang yang digunakan untuk pembayaran.
- **profile_status**: Status profil pengguna.
- **trxid**: Identifikasi unik transaksi.

- **attempt_charging**: Jumlah percobaan pengisian ulang.
- **success_billing**: Jumlah keberhasilan dalam penagihan.
- **updated_at**: Waktu pembaruan data terakhir.
- **sales_per_customer**: Penjualan per pelanggan.
- **average_order_value**: Nilai rata-rata pesanan pelanggan.
- **inventory_turnover_ratio**: Rasio perputaran inventaris.

Melalui dataset ini, kelompok kami dapat melakukan berbagai analisis, termasuk eksplorasi churn risk, subscription tenure, dan metrik lain yang relevan untuk memahami faktor-faktor yang mempengaruhi performa bisnis.

B. Alur Pengerjaan Study Case

1. Tools yang digunakan dalam proses menganalisis data

Dalam pengerjaan study case ini, kelompok kami menggunakan Python, SQL, dan Tableau untuk menganalisis data secara terstruktur. Python digunakan untuk proses *data cleaning* dan *Exploratory Data Analysis (EDA)* karena fleksibilitasnya dalam manipulasi data dan kemampuan automasinya melalui pustaka seperti *pandas* dan *matplotlib*. SQL digunakan untuk menghitung *Subscription Trends* (pendapatan total, jumlah langganan, dan tingkat churn berdasarkan wilayah dan operator), *Churn and Retention* (analisis tingkat churn berdasarkan siklus langganan dan segmentasi pengguna), serta *Campaign Performance* (mengukur keberhasilan kampanye melalui tingkat penagihan sukses). Tableau digunakan untuk memvisualisasikan hasil akhir analisis, seperti tren langganan dan performa kampanye pemasaran, dengan cara yang interaktif dan mudah dipahami. Kombinasi alat ini memastikan pengolahan data dilakukan secara efisien, menghasilkan insight yang mendalam, dan memudahkan penyampaian informasi kepada audiens.

2. Proses Analisis Data

1. Pembersihan Data (Data Cleaning)

Tim kami memulai proses analisis dengan melakukan pembersihan data. Langkah pertama yang dilakukan adalah mengidentifikasi dan mengatasi data yang hilang (missing values), atau inkonsistensi dalam dataset.

- **Identifikasi Anomali pada Kolom unsub_date:** Kami melakukan pemeriksaan pada kolom unsub_date untuk mencari data yang tidak sesuai atau tidak masuk akal. Dalam proses ini, kami menemukan data dengan tanggal yang sangat tidak realistis, yaitu 0001-01-01. Tanggal ini tentu tidak valid karena merupakan tanggal yang jauh di luar rentang waktu yang diharapkan. Setelah menemukan anomali tersebut, kami menggantinya dengan nilai yang lebih relevan, yaitu Not Unsubscribed, untuk mencerminkan status yang tepat bagi entri dengan tanggal yang tidak valid tersebut.
- **Penanganan Data Kosong (Missing Data):** Kami juga menemukan beberapa kolom yang berisi nilai kosong atau NaN yang tidak dapat digunakan dalam analisis lebih lanjut. Kolom-kolom tersebut adalah: freemium_end_date, publisher, pixel, handset, dan browser. Karena informasi dari kolom-kolom ini tidak cukup penting untuk analisis lebih lanjut dan data yang hilang di dalamnya cukup signifikan, kami memutuskan untuk menghapus kolom-kolom tersebut dari dataset untuk menjaga kebersihan dan keakuratan data yang akan digunakan dalam analisis.

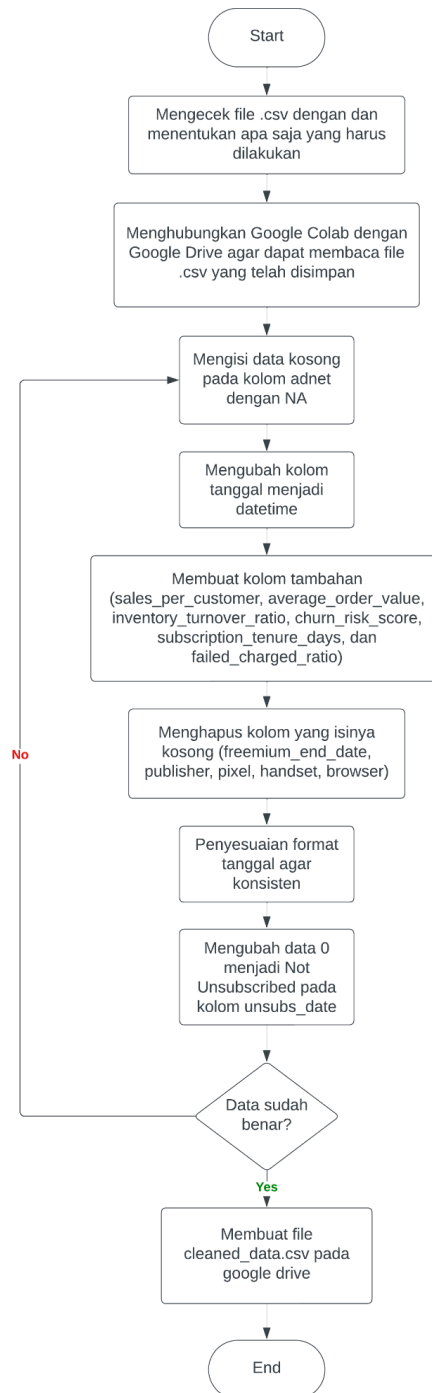
2. Transformasi Data

Setelah tahap pembersihan data, tim kami melanjutkan dengan melakukan transformasi data agar data siap untuk dianalisis lebih lanjut.

- **Penambahan Kolom Baru:** Kami diperintahkan untuk menambahkan beberapa kolom baru untuk memberikan wawasan lebih dalam terkait performa transaksi pelanggan. Kolom-kolom yang ditambahkan antara lain:

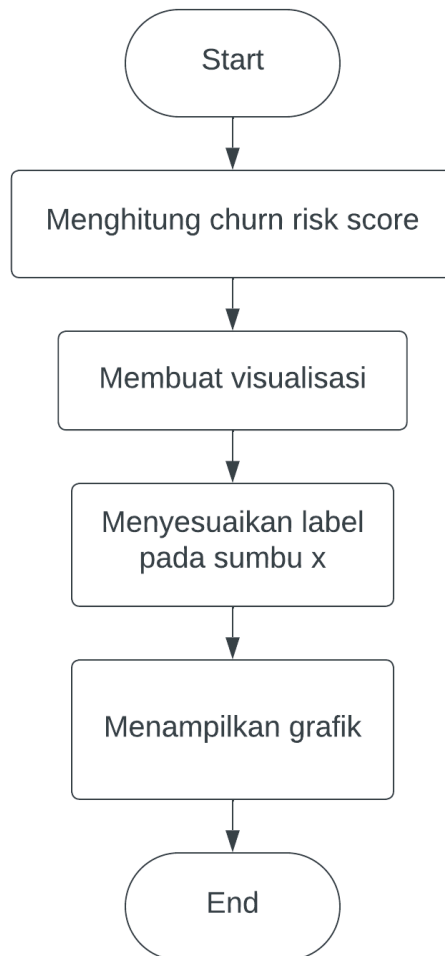
- **Sales per Customer:** Kolom ini dihitung dengan cara membagi total pendapatan (revenue) dengan jumlah pelanggan yang terdaftar dalam data. Hal ini memberikan gambaran seberapa banyak pendapatan yang dihasilkan oleh setiap pelanggan.
- **Average Order Value (AOV):** Kolom AOV dihitung dengan membagi total pendapatan dengan jumlah transaksi atau order yang terjadi. Ini memberikan informasi tentang rata-rata nilai setiap transaksi yang dilakukan oleh pelanggan.
- **Inventory Turnover Ratio:** Kolom ini dihitung untuk mengetahui seberapa efisien inventaris produk bergerak berdasarkan jumlah transaksi yang tercatat dibandingkan dengan tingkat inventaris yang ada. Dengan mengetahui rasio ini, perusahaan dapat mengevaluasi kinerja dalam pengelolaan stok dan penjualan produk.
- **Churn Risk Score:** Mengukur risiko churn (pindahnya pelanggan) berdasarkan rasio antara billing yang berhasil dan upaya penagihan. Nilai akhir dikurangi dari 1 untuk menunjukkan skor risiko (semakin tinggi, semakin berisiko pelanggan churn).
- **Subscription Tenure Days:** Menghitung jumlah hari pelanggan telah berlangganan layanan berdasarkan selisih tanggal antara renewal_date dan subs_date.
- **Failed Charged Ratio :** Mengukur rasio kegagalan penagihan berdasarkan selisih antara upaya penagihan dan billing yang berhasil, dibandingkan dengan total upaya penagihan.

Flowchart Data Cleaning



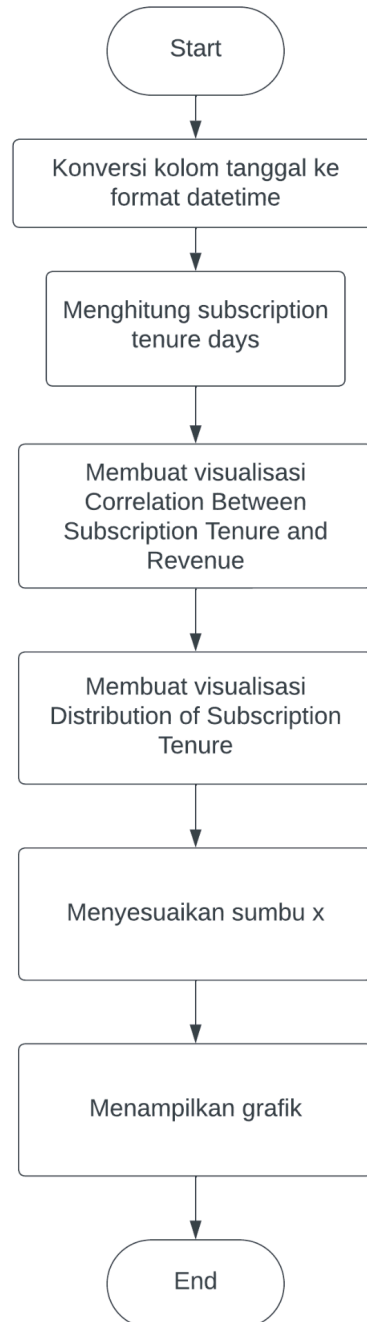
Gambar 3 Flowchart Data Cleaning

Flowchart Distribusi Churn Risk Score



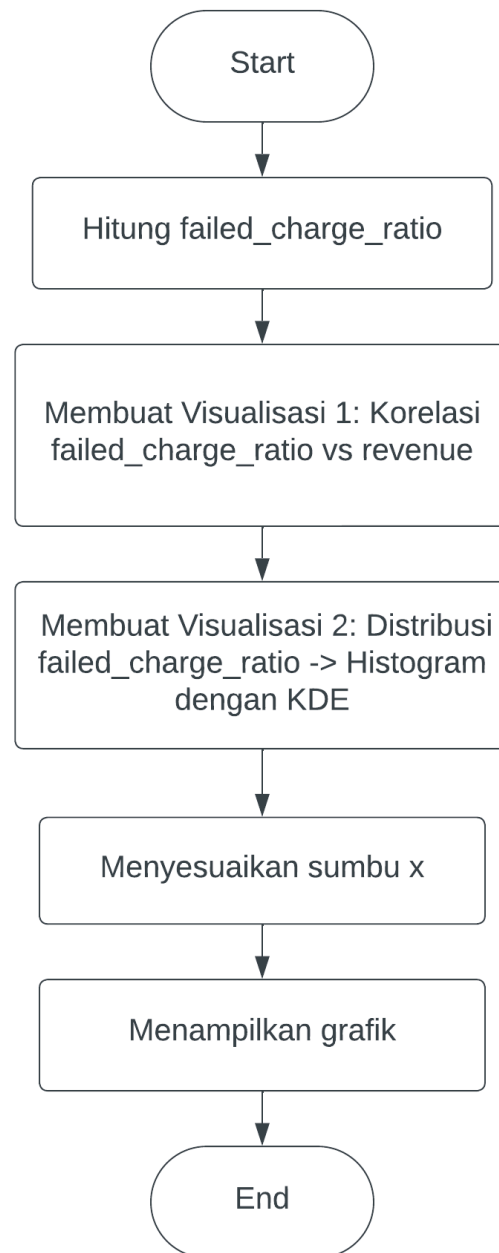
Gambar 4 Flowchart Distribusi Churn Risk Score

Flowchart Korelasi Tenure dengan Revenue dan Distribusi Subscription Tenure



Gambar 5 Flowchart Korelasi Tenure dengan Revenue dan Distribusi Subscription Tenure

Flowchart Impact of Failed Charge Ratio on Revenue dan Distribution of Failed Charge Ratio



Gambar 6 Flowchart Impact of Failed Charge Ratio on Revenue dan Distribution of Failed Charge

Ratio

BAB III

Hasil dan Pembahasan

A. Deskripsi Data Hasil Analisis Bisnis

Dataset yang dianalisis memiliki beberapa data yang tidak valid dan perlu ditangani untuk memastikan kualitas analisis. Ditemukan empat kolom, yaitu `freemium_end_date`, `publisher`, `pixel`, `handset`, dan `browser`, yang seluruh nilainya kosong dan tidak memberikan informasi berguna. Oleh karena itu, kolom-kolom tersebut dihapus (drop) dari dataset. Selain itu, kolom `adnet` memiliki beberapa nilai kosong (missing values) yang diganti dengan nilai "NA" untuk menunjukkan bahwa data tersebut tidak tersedia.

Berdasarkan dataset, ditemukan nilai revenue yang secara statistik teridentifikasi sebagai outlier. Namun, setelah dilakukan pemeriksaan lebih lanjut, nilai tersebut dinilai wajar karena mencerminkan keberhasilan strategi success billing, di mana pelanggan tertentu menunjukkan komitmen pembayaran yang signifikan terhadap produk atau layanan yang diberikan. Revenue tinggi ini tidak hanya mencerminkan pola transaksi yang besar, tetapi juga hasil dari pendekatan penagihan yang efektif, seperti pengelolaan kontrak premium, negosiasi sukses, atau loyalitas pelanggan high-value. Oleh karena itu, nilai-nilai ini tetap dipertahankan dalam dataset untuk memastikan analisis dapat menggambarkan kontribusi strategis dari pelanggan-pelanggan utama serta keberhasilan pendekatan penagihan yang diterapkan.

Langkah-langkah ini diambil untuk membersihkan dataset sehingga lebih terorganisir dan siap untuk analisis lebih lanjut. Dengan pembersihan ini, dataset menjadi lebih optimal dan akurat untuk mendukung pengambilan keputusan bisnis.

B. Hasil Analisis

Kode Program 1. Proses *data cleaning* pada Python.

```
1. # Menghubungkan Google Drive dan membaca file dataset_pbl ke dalam DataFrame
2. from google.colab import drive
3. drive.mount('/content/drive', force_remount=True)
```

```
4.
5. import pandas as pd
6. import numpy as np
7.
8. file_path = '/content/drive/MyDrive/MSIB Batch 7/Project Based
   Learning/dataset_pbl.csv'
9. df = pd.read_csv(file_path)
10. data = pd.read_csv(file_path)
11.
12. # Mengisi kolom kosong pada kolom adnet dengan "NA"
13. data['adnet'].fillna('NA', inplace=True)
14.
15. # Mengubah kolom tanggal menjadi datetime
16. data['subs_date'] = pd.to_datetime(data['subs_date'])
17. data['renewal_date'] = pd.to_datetime(data['renewal_date'], errors='coerce')
18. data['freemium_end_date'] = pd.to_datetime(data['freemium_end_date'],
   errors='coerce')
19. data['unsubs_date'] = pd.to_datetime(data['unsubs_date'], errors='coerce')
20.
21. # Membuat kolom tambahan
22. data['sales_per_customer'] = data['revenue'] / (data['success_billing'] + 1)
23. data['average_order_value'] = data['revenue'] / (data['attempt_charging'] + 1)
24. data['inventory_turnover_ratio'] = data['revenue'] / (data['service_price'] +
   1)
25.
26. # Churn Risk Score Based on Engagement
27. data['churn_risk_score'] = (data['attempt_charging'] +
   data['success_billing']) / (data['attempt_charging'] + 1)
28. data['churn_risk_score'] = 1 - data['churn_risk_score']
29.
30. # Subscription Tenure
31. data['subscription_tenure_days'] = (data['renewal_date'] -
   data['subs_date']).dt.days
32.
33. # Failed Charge Ratio
34. data['failed_charge_ratio'] = (data['attempt_charging'] -
   data['success_billing']) / (data['attempt_charging'] + 1)
35.
36. columns_to_drop = ['freemium_end_date', 'publisher', 'pixel', 'handset',
   'browser']
37. data.drop(columns=columns_to_drop, inplace=True)
38. data['unsubs_date'] = data['unsubs_date'].replace('0', 'Not Unsubscribed')
39.
40. date_columns = ['created_at', 'subs_date', 'renewal_date', 'unsubs_date',
   'updated_at',]
```

```
41. for col in date_columns:
42.
43.     df[col] = pd.to_datetime(df[col], errors='coerce')
44.
45.     df[col] = df[col].dt.tz_localize(None)
46.
47.     df[col] = df[col].dt.tz_localize('Asia/Bangkok')
48.
49.     df[col] = df[col].dt.strftime('%Y-%m-%d %H:%M:%S') + ' ' +
        df[col].dt.strftime('%z').str[:3] + ':' + df[col].dt.strftime('%z').str[3:]
50.
51. df['unsubs_date'] = pd.to_datetime(df['unsubs_date'], errors='coerce')
52. df['unsubs_date'] = df['unsubs_date'].fillna('Not Unsubscribed')
53.
54. output_path = '/content/drive/MyDrive/MSIB Batch 7/Project Based
    Learning/cleaned_dataset_v5.csv'
55. data.to_csv(output_path, index=False)
56. print(f"Data updated and saved at {output_path}")
```

Pada baris 1-10, kode digunakan untuk menghubungkan Google Drive dengan Google Colab agar dataset dapat diakses, dengan memasang Google Drive ke direktori `/content/drive/MyDrive/MSIB Batch 7/Project Based Learning/dataset_pbl.csv` dan mengimpor library pandas dan numpy untuk memanipulasi data. Baris 13 mengisi nilai kosong pada kolom `adnet` dengan "NA" menggunakan `fillna()`. Baris 16-19 mengonversi kolom tanggal seperti `subs_date`, `renewal_date`, `freemium_end_date`, dan `unsubs_date` menjadi tipe `datetime`, dengan menangani kesalahan menggunakan parameter `errors='coerce'`. Baris 22-34 membuat kolom baru, yaitu `sales_per_customer`, `average_order_value`, `inventory_turnover_ratio`, `churn_risk_score`, `subscription_tenure_days`, dan `failed_charge_ratio`, untuk memberikan wawasan tambahan terkait performa pelanggan dan metrik risiko churn. Pada baris 36-38, beberapa kolom seperti `freemium_end_date` dan lainnya dihapus menggunakan `drop()`, serta data `unsubs_date` dengan nilai "0" diganti dengan "Not Unsubscribed". Baris 40-49 menerapkan format zona waktu Asia/Bangkok pada kolom tanggal dan memformatnya menjadi string dengan timezone yang sesuai (+07:00). Baris 51-52 memastikan nilai kosong pada `unsubs_date` diganti dengan "Not Unsubscribed". Terakhir, baris 54-56 menyimpan dataset yang telah dibersihkan ke file CSV baru di Google Drive.

3. Exploratory Data Analysis (EDA)

Dalam laporan ini, kami melakukan eksplorasi data untuk mengidentifikasi pola dan hubungan yang mempengaruhi kinerja bisnis. Fokus utama analisis adalah pada Churn Risk Score, Subscription Tenure, dan Failed Charge Ratio. Setiap metrik ini dirancang untuk memberikan wawasan mendalam terkait keterlibatan pengguna, durasi langganan, dan efisiensi proses pembayaran, yang semuanya berkontribusi pada retensi pengguna dan pendapatan.

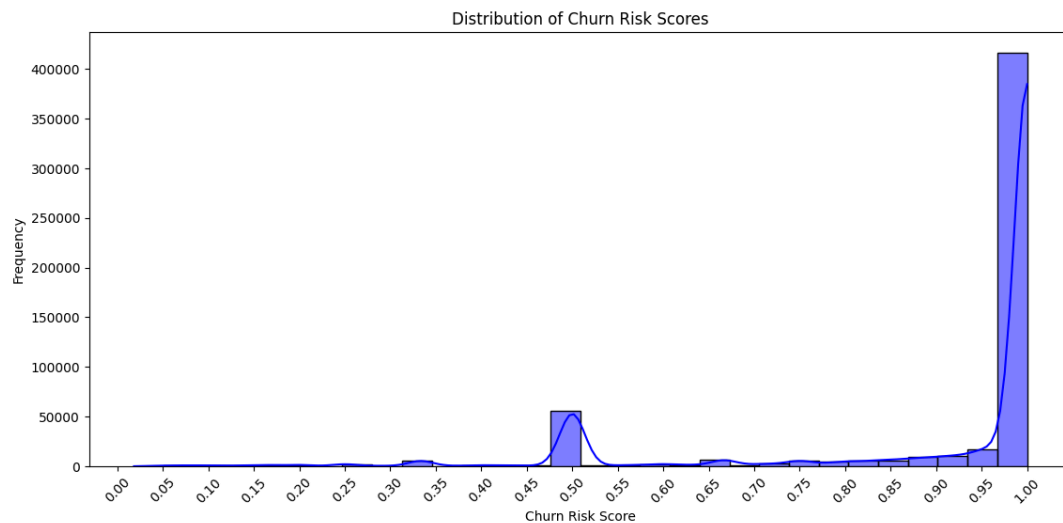
Untuk Churn Risk Score, kami menghitung skor risiko churn berdasarkan keterlibatan pengguna, menggunakan metrik seperti jumlah percobaan penagihan, keberhasilan penagihan, dan status langganan. Analisis ini bertujuan untuk mengidentifikasi segmen pengguna dengan risiko churn tertinggi.

Selanjutnya, fitur Subscription Tenure dikembangkan untuk menghitung jumlah hari pelanggan telah berlangganan, berdasarkan tanggal langganan dan perpanjangan. Kami menganalisis korelasi antara durasi langganan dengan siklus langganan dan pendapatan untuk memahami pola retensi pengguna.

Selain itu, kami membuat metrik Failed Charge Ratio, yaitu rasio antara percobaan penagihan yang gagal terhadap total percobaan penagihan. Analisis ini memberikan wawasan tentang bagaimana rasio tersebut mempengaruhi retensi pelanggan dan kemampuan bisnis untuk menghasilkan pendapatan.

Untuk memperjelas hasil analisis, kami menggunakan visualisasi data yang dibuat dengan Matplotlib dan Seaborn, yang memberikan pandangan mendalam dan intuitif terhadap tren dan hubungan dalam data.

Distribusi Churn Risk Score



Gambar 7 Grafik Distribusi Churn Risk Score

Grafik distribusi churn risk score menunjukkan bahwa sebagian besar pelanggan memiliki skor mendekati 1, menandakan risiko churn yang sangat tinggi. Distribusi ini sangat skewed ke kanan, dengan sedikit puncak pada skor risiko sedang (sekitar 0,45–0,5) dan hanya sedikit pelanggan dengan risiko rendah (skor mendekati 0). Hal ini mengindikasikan bahwa mayoritas pelanggan cenderung berisiko tinggi untuk berhenti menggunakan layanan. Untuk mengatasi hal ini, perusahaan perlu fokus pada strategi retensi, seperti memberikan insentif atau program loyalitas kepada pelanggan dengan risiko tinggi, serta melakukan intervensi preventif untuk kelompok risiko sedang agar tidak berpindah ke kategori risiko tinggi. Selain itu, menjaga kepuasan pelanggan dengan risiko rendah juga penting untuk mempertahankan basis pelanggan setia. Analisis ini menunjukkan perlunya tindakan strategis segera untuk mengurangi churn dan meningkatkan retensi pelanggan.

Kode Program 2. Distribusi Churn Risk Score.

```
1. # Churn Risk Score = 1 - (Success Billing / Attempt Charging)
2. data['churn_risk_score'] = 1 - (data['success_billing'] /
```

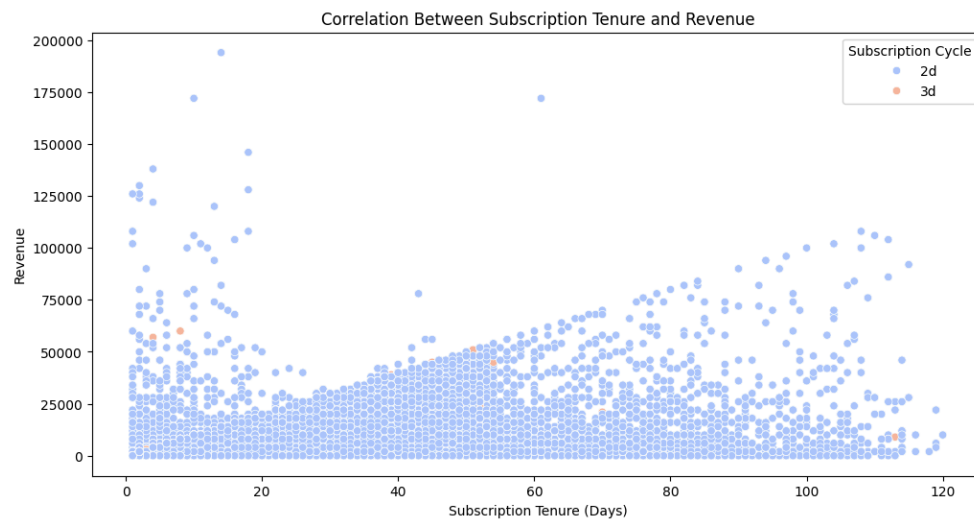
```

        (data['attempt_charging'] + 1)) # Avoid division by zero
3.
4. # Distribution of Churn Risk Scores
5. plt.figure(figsize=(13.5, 6))
6. sns.histplot(data['churn_risk_score'], kde=True, bins=30, color='blue')
7. plt.title('Distribution of Churn Risk Scores')
8. plt.xlabel('Churn Risk Score')
9. plt.ylabel('Frequency')
10.
11. # Customizing the x-axis ticks
12. x_ticks = [round(x, 2) for x in np.arange(0, 1.05, 0.05)] # Define x-axis
    ticks from 0 to 1 with 0.05 intervals
13. plt.xticks(x_ticks, rotation=45) # Add ticks and rotate them for better
    readability
14.
15. plt.show()

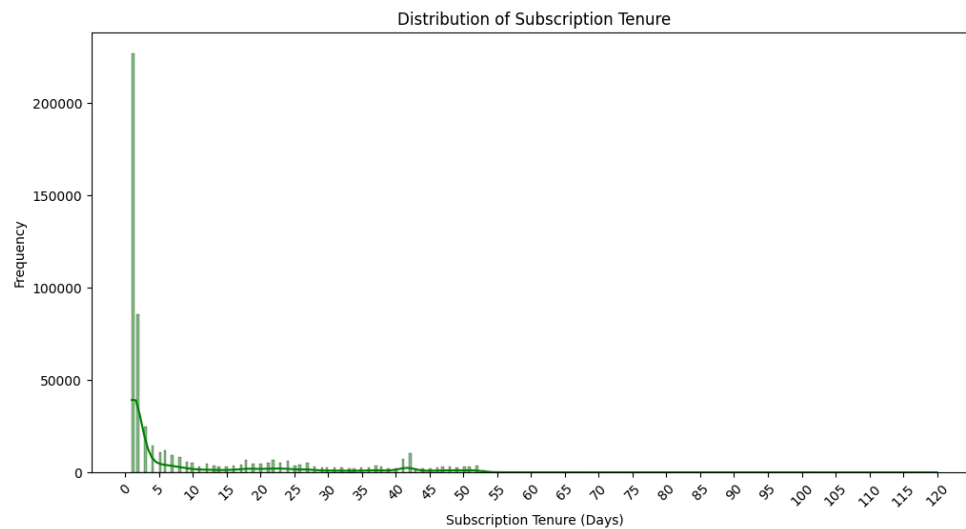
```

Kode di atas dimulai dengan menghitung Churn Risk Score menggunakan rumus: $\text{Churn Risk Score} = 1 - (\text{Success Billing} / \text{Attempt Charging})$. Baris 2 menunjukkan perhitungan ini, dengan memastikan tidak ada pembagian dengan nol dengan menambahkan 1 pada penyebut. Selanjutnya, baris 6-9 bertujuan untuk menampilkan distribusi skor risiko churn dengan membuat histogram menggunakan seaborn ('sns.histplot'). Histogram ini menggunakan 30 bin dan diberi warna biru, serta memiliki judul dan label pada sumbu x dan y. Baris 12-13 menambahkan kustomisasi pada tanda sumbu x, mendefinisikan tanda dari 0 hingga 1 dengan interval 0,05 untuk meningkatkan keterbacaan. Terakhir, baris 15 menggunakan 'plt.show()' untuk menampilkan grafik histogram tersebut.

Distribusi Subscription Tenure



Gambar 8 Grafik Correlation Between Subscription Tenure and Revenue



Gambar 9 Grafik Distribusi of Subscription Tenure

Grafik pertama, "Correlation Between Subscription Tenure And Revenue", menunjukkan hubungan positif antara masa langganan (dalam hari) di sumbu x dan pendapatan di sumbu y. Titik-titik data dalam grafik diberi kode warna berdasarkan siklus langganannya, di mana biru mewakili siklus 2 hari dan oranye mewakili siklus 3 hari.

Grafik ini mengungkapkan bahwa semakin lama masa langganan pelanggan, semakin tinggi pula pendapatan yang dihasilkan.

Grafik kedua, "Distribution of Subscription Tenure", menunjukkan distribusi frekuensi masa langganan. Grafik ini menunjukkan bahwa sebagian besar masa langganan terpusat pada rentang yang lebih rendah, dengan puncak tajam sekitar 0 hari. Hal ini mengindikasikan bahwa meskipun masa langganan yang lebih lama terkait dengan pendapatan yang lebih tinggi, sebagian besar pelanggan hanya berlangganan dalam waktu yang relatif singkat.

Kode Program 3. Subscription Tenure.

```
1. data['subs_date'] = pd.to_datetime(data['subs_date'])
2. data['renewal_date'] = pd.to_datetime(data['renewal_date'], errors='coerce')
3. data['subscription_tenure_days'] = (data['renewal_date'] -
    data['subs_date']).dt.days
4. # Correlation Analysis: Subscription Tenure and Revenue
5. plt.figure(figsize=(12, 6))
6. sns.scatterplot(x='subscription_tenure_days', y='revenue', data=data,
    hue='cycle', palette='coolwarm')
7. plt.title('Correlation Between Subscription Tenure and Revenue')
8. plt.xlabel('Subscription Tenure (Days)')
9. plt.ylabel('Revenue')
10. plt.legend(title='Subscription Cycle')
11. plt.show()
12.
13. # Distribution of Subscription Tenure
14. plt.figure(figsize=(12, 6))
15. sns.histplot(data['subscription_tenure_days'], kde=True, color='green')
16. plt.title('Distribution of Subscription Tenure')
17. plt.xlabel('Subscription Tenure (Days)')
18. plt.ylabel('Frequency')
19.
20. x_ticks = range(0, int(data['subscription_tenure_days'].max()) + 1, 5)
21. plt.xticks(x_ticks, rotation=45)
22.
23. plt.show()
```

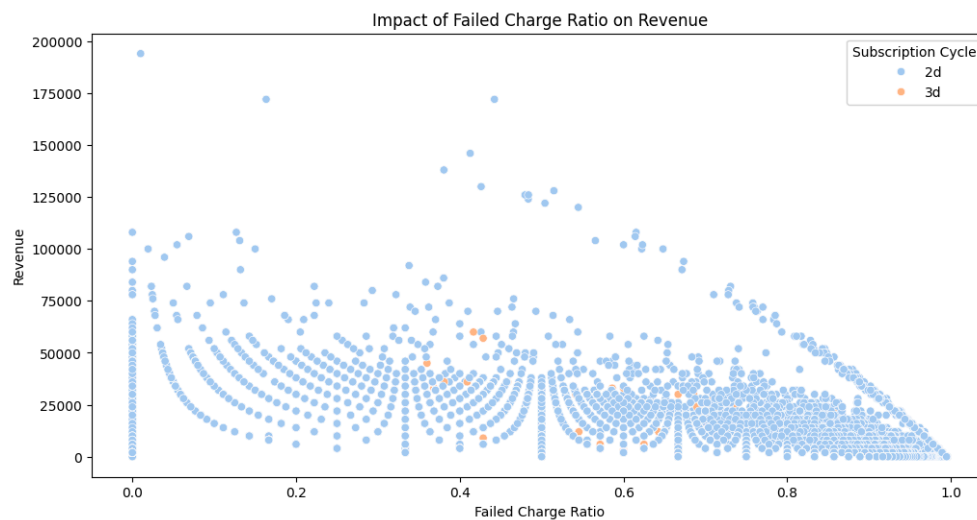
Kode di atas dimulai dengan memastikan bahwa kolom tanggal memiliki format datetime menggunakan `pd.to_datetime()` untuk kolom `subs_date` dan `renewal_date`, dengan penanganan kesalahan menggunakan `errors='coerce'` untuk menghindari kesalahan format. Selanjutnya, pada baris 3, kode menghitung lama langganan (subscription tenure) dalam hari

dengan mengurangi `subs_date` dari `renewal_date`, lalu menyimpannya dalam kolom `subscription_tenure_days`.

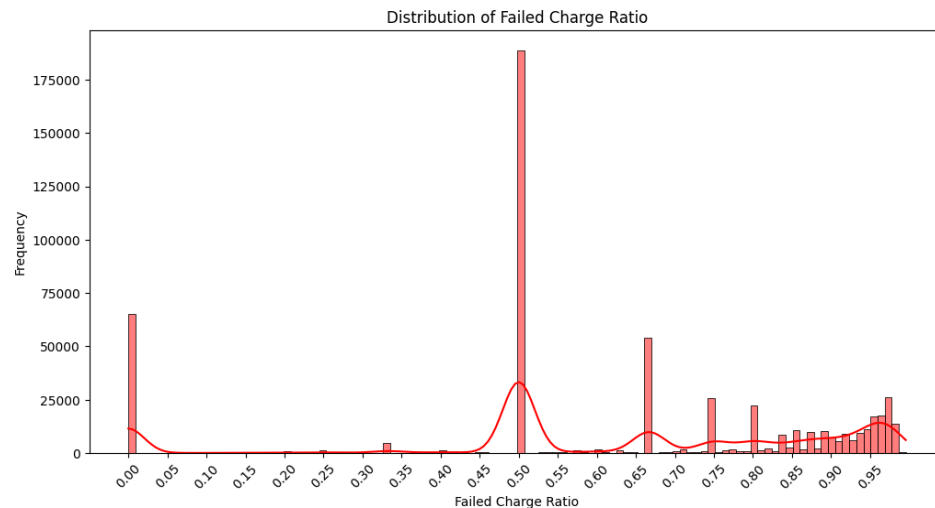
Pada baris 5-11, kode melakukan analisis korelasi antara lama langganan dan pendapatan menggunakan scatterplot dari seaborn (`sns.scatterplot`). Grafik ini menampilkan hubungan antara `subscription_tenure_days` dan `revenue` dengan penanda warna berdasarkan `cycle`.

Baris 14-18 fokus pada distribusi lama langganan dengan membuat histogram menggunakan seaborn (`sns.histplot`). Histogram ini diwarnai hijau dan menampilkan distribusi frekuensi `subscription_tenure_days`, dengan judul dan label sumbu yang sesuai. Pada baris 20-23, tanda pada sumbu x didefinisikan dari 0 hingga nilai maksimum `subscription_tenure_days` dengan interval 5, dan diputar untuk meningkatkan keterbacaan. Terakhir, grafik histogram ditampilkan menggunakan `plt.show()`.

Impact of Failed Charged Ratio on Revenue



Gambar 10 Grafik Impact of Failed Charged Ratio on Revenue



Gambar 11 Grafik Distribution of Failed Charge

Grafik Impact for Failed Charged Ratio on Revenue menunjukkan hubungan antara rasio gagal bayar (Failed Charge Ratio) dan pendapatan (Revenue) untuk siklus langganan (Subscription Cycle) yang berbeda, yaitu 2 hari (2d) dan 3 hari (3d). Sumbu horizontal (x) mewakili rasio gagal bayar yang berkisar dari 0 hingga 1, sedangkan sumbu vertikal (y) menunjukkan pendapatan yang berkisar dari 0 hingga 200.000. Titik-titik pada grafik menunjukkan data individu, dengan warna biru muda mewakili siklus langganan 2 hari dan warna oranye mewakili siklus langganan 3 hari. Grafik ini mengindikasikan bahwa semakin tinggi rasio gagal bayar, semakin rendah pendapatan yang dihasilkan, dengan sebagian besar data terkonsentrasi pada rasio gagal bayar yang rendah. Terdapat beberapa outlier yang menunjukkan pendapatan tinggi meskipun rasio gagal bayar tinggi, yang mungkin mencerminkan kasus-kasus khusus atau anomali.

Grafik Distribution of Failed Charge menunjukkan distribusi rasio kegagalan pembayaran. Sumbu X mewakili rasio kegagalan pembayaran, yang berkisar dari 0,00 hingga 0,95, sedangkan sumbu vertikal menunjukkan frekuensi kejadian untuk setiap rasio tersebut. Grafik ini memiliki beberapa puncak yang menonjol, terutama pada rasio sekitar 0,00, 0,50, dan 0,70. Puncak tertinggi terlihat pada rasio 0,50, dengan frekuensi lebih dari 175.000, menunjukkan bahwa sebagian besar data memiliki rasio kegagalan pembayaran di sekitar nilai ini. Puncak lainnya yang signifikan berada di sekitar rasio 0,00 dan 0,70, masing-masing dengan frekuensi

sekitar 75.000 dan 25.000. Grafik ini juga menunjukkan distribusi yang lebih merata pada rasio yang lebih tinggi, dengan beberapa puncak kecil yang tersebar di sepanjang sumbu horizontal.

Dari analisis distribusi rasio kegagalan pembayaran (Failed Charge Ratio), dapat diambil beberapa insight penting yang menggambarkan perilaku pengguna dalam melakukan pembayaran. Puncak terbesar pada rasio 0,50 menunjukkan bahwa sebagian besar pengguna memiliki tingkat kegagalan pembayaran yang relatif moderat, yang bisa mengindikasikan ketidakstabilan dalam proses pembayaran mereka, meskipun tidak sepenuhnya gagal. Hal ini mungkin disebabkan oleh berbagai faktor, seperti masalah teknis pada sistem pembayaran atau ketidaksesuaian data pengguna. Puncak yang lebih rendah pada rasio 0,00 menandakan bahwa transaksi berhasil dilakukan tanpa masalah, yang menunjukkan bahwa banyak pengguna berhasil melakukan pembayaran dengan lancar. Di sisi lain, puncak pada rasio 0,70 menunjukkan kelompok kecil pengguna dengan tingkat kegagalan yang lebih tinggi, yang bisa menjadi indikasi masalah lebih serius, seperti kekurangan dana atau masalah pembayaran berulang. Dengan adanya distribusi yang lebih merata pada rasio yang lebih tinggi, hal ini mengindikasikan bahwa meskipun sebagian kecil transaksi berhasil tanpa masalah, terdapat banyak transaksi yang gagal dalam berbagai tingkat kegagalan.

Kode Program 4. Impact of Failed Charged Ratio on Revenue.

```
1. # Failed Charge Ratio = (Attempt Charging - Success Billing) / Attempt
   Charging
2. data['failed_charge_ratio'] = (data['attempt_charging'] -
   data['success_billing']) / (data['attempt_charging'] + 1) # Avoid division
   by zero
3.
4. # Failed Charge Ratio Analysis by Revenue
5. plt.figure(figsize=(12, 6))
6. sns.scatterplot(x='failed_charge_ratio', y='revenue', data=data, hue='cycle',
   palette='pastel')
7. plt.title('Impact of Failed Charge Ratio on Revenue')
8. plt.xlabel('Failed Charge Ratio')
9. plt.ylabel('Revenue')
10. plt.legend(title='Subscription Cycle')
11. plt.show()
12.
13. # Distribution of Failed Charge Ratio
14. plt.figure(figsize=(12, 6))
15. sns.histplot(data['failed_charge_ratio'], kde=True, color='red')
```

```
16. plt.title('Distribution of Failed Charge Ratio')
17. plt.xlabel('Failed Charge Ratio')
18. plt.ylabel('Frequency')
19. plt.show()
```

4. Data Extraction

Dalam laporan ini, kami melakukan analisis mendalam untuk menjawab pertanyaan terkait *Subscription Trends*, *Churn and Retention*, serta *Campaign Performance*. Pendekatan ini dirancang untuk memberikan wawasan strategis yang mendukung peningkatan kinerja bisnis secara menyeluruh.

1. Subscription Trends

Analisis tren langganan melibatkan perhitungan total pendapatan, jumlah pelanggan, dan tingkat churn berdasarkan wilayah dan operator. Dengan memahami distribusi pendapatan dan pola churn di setiap wilayah atau operator, bisnis dapat mengidentifikasi area dengan performa terbaik serta wilayah yang memerlukan perhatian lebih. Wawasan ini membantu dalam pengambilan keputusan strategis untuk pengembangan wilayah bisnis.

2. Churn and Retention

Dalam analisis churn dan retensi, kami menghitung tingkat churn untuk setiap siklus langganan dan membagi pengguna berdasarkan wilayah serta operator. Proses ini bertujuan untuk memahami pola churn di berbagai segmen pelanggan. Selain itu, kami mengevaluasi keterkaitan antara siklus langganan, retensi pengguna, dan pendapatan untuk mengidentifikasi faktor-faktor yang memengaruhi keberhasilan mempertahankan pelanggan.

3. Campaign Performance

Analisis kinerja kampanye pemasaran dilakukan dengan mengukur tingkat keberhasilan penagihan untuk setiap kampanye. Kami menghitung rasio antara

keberhasilan dan total percobaan penagihan, yang membantu menilai efektivitas strategi pemasaran. Informasi ini memberikan rekomendasi praktis untuk meningkatkan hasil kampanye di masa mendatang, termasuk identifikasi kampanye yang memberikan dampak terbesar pada pendapatan.

Untuk menjawab kebutuhan analisis dan menghasilkan wawasan yang bermakna, kami terlebih dahulu menggunakan SQL untuk mengolah dan menyiapkan data. SQL memungkinkan kami untuk melakukan berbagai perhitungan, seperti mengidentifikasi tren langganan, menganalisis tingkat churn dan retensi, serta mengevaluasi performa kampanye pemasaran. Berikut adalah query yang kami gunakan yaitu:

1. Subscription Trends

Kode Program 5. Subscription Trends

```
1. SELECT
2.     country,
3.     operator,
4.     COUNT(*) AS subscription_count,
5.     SUM(revenue) AS total_revenue,
6.     SUM(CASE
7.         WHEN profile_status = 'inactive' AND status = -1 THEN 1
8.         ELSE 0
9.     END) AS subscription_churn,
10.    ROUND(
11.        (SUM(CASE
12.            WHEN profile_status = 'inactive' AND status = -1 THEN 1
13.            ELSE 0
14.        END) * 1.0 / COUNT(*)) * 100,
15.        2
16.    ) AS churn_rate_percentage
17. FROM
18.     cleaned_dataset_v6_new
19. GROUP BY
20.     country, operator;
```

Pada soal mengenai subscription trend, kami diminta untuk menyusun query SQL yang berfungsi untuk menghitung total pendapatan, jumlah langganan, dan tingkat churn

berdasarkan wilayah (region) serta operator. Pada baris ke 1 hingga ke 3 menampilkan kolom country, yang mewakili negara tempat pelanggan berada dan kolom operator yang menunjukkan penyedia layanan atau operator yang digunakan oleh pelanggan. Lalu pada baris ke 4 kita menghitung jumlah langganan (baris) dalam data untuk setiap kombinasi country dan operator. Ini akan memberi informasi tentang total pelanggan (langganan) untuk setiap negara dan operator. Pada baris ke 5 digunakan menjumlahkan total pendapatan (revenue) untuk setiap negara dan operator. Ini akan memberikan nilai total pendapatan yang dihasilkan oleh pelanggan di negara dan operator tertentu. Baris ke 6-9 digunakan untuk menghitung jumlah langganan yang mengalami churn (berhenti berlangganan) yang mana `profile_status = 'inactive'` menunjukkan pelanggan yang tidak aktif, `status = -1` mengindikasikan bahwa pelanggan tersebut berhenti berlangganan dan CASE statement digunakan untuk mengembalikan nilai 1 jika kondisi terpenuhi (pelanggan churn), dan 0 jika tidak. Fungsi SUM kemudian menghitung total langganan yang churn untuk setiap kombinasi negara dan operator.

Lalu baris ke 10-16 Menghitung persentase churn untuk setiap negara dan operator menghitung jumlah churn (`SUM(CASE WHEN profile_status = 'inactive' AND status = -1 THEN 1 ELSE 0 END)`) dibagi dengan total langganan (`COUNT(*)`). Hasilnya dikalikan dengan 100 untuk mendapatkan persentase churn. Fungsi `ROUND(..., 2)` membulatkan hasil churn rate ke dua angka desimal, sehingga menghasilkan persentase churn yang lebih mudah dibaca. Baris ke 17-20 yaitu data diambil dari tabel `cleaned_dataset_v6_new` Tabel ini berisi data pelanggan dan informasi terkait langganan, termasuk status profil, status langganan, dan pendapatan. Dan bagian selanjutnya mengelompokkan hasil query berdasarkan dua kolom: country dan operator. Ini berarti perhitungan dilakukan untuk setiap kombinasi negara dan operator secara terpisah, sehingga kita dapat mengetahui jumlah langganan, total pendapatan, churn, dan churn rate untuk setiap negara dan operator

Berikut adalah output dari query yang kita jalankan:

	A-Z country	A-Z operator	123 subscription_count	123 total_revenue	123 subscription_churn	123 churn_rate_percentage
1	ID	telkomsel	564,619	858,470,000	58,527	10.37

Gambar 12 Output SQL Query Subscription Trends

Pada output tersebut menunjukkan bahwa kolom tersebut menjelaskan antara lain:

- Kolom country menunjukkan wilayah tempat operator beroperasi. Dalam hal ini, "ID" mengacu pada Indonesia
- Kolom operator menyebutkan nama operator, yaitu "Telkomsel."
- Kolom subscription_count menampilkan jumlah total pelanggan untuk operator tersebut, yaitu 564,619 pelanggan.
- Kolom total_revenue Total pendapatan yang dihasilkan oleh operator tersebut adalah 858,470,000 (dalam satuan yang tidak ditentukan, kemungkinan dalam mata uang lokal).
- Kolom subscription_churn menampilkan jumlah pelanggan yang berhenti berlangganan (churned), yaitu 58,527 pelanggan.
- Kolom churn_rate_percentage tingkat churn adalah 10.37%

Insight yang dapat diambil dari output ini adalah:

a. Tingkat Churn yang Relatif Tinggi (10.37%)

Implikasi: Dari total 564,619 pelanggan, sebanyak 58,527 pelanggan telah berhenti berlangganan. Tingkat churn di atas 10% menunjukkan bahwa operator mungkin memiliki masalah dalam mempertahankan pelanggan, misalnya karena:

- Kurangnya kepuasan pelanggan (layanan tidak memadai, harga mahal, atau kualitas jaringan buruk).
- Persaingan dengan operator lain yang menawarkan paket lebih menarik.

Tindakan: Operator perlu mengevaluasi penyebab churn, seperti melalui survei pelanggan, dan meningkatkan layanan atau penawaran paket untuk mengurangi tingkat churn

b. Pendapatan yang Tinggi

Telkomsel berhasil menghasilkan pendapatan 858,470,000 dengan jumlah pelanggan yang relatif besar (564,619 pelanggan). Ini menunjukkan:

- Basis pelanggan yang signifikan dan kontribusi pendapatan per pelanggan yang cukup tinggi.

Tindakan:

- Fokus pada peningkatan Average Revenue Per User (ARPU), misalnya dengan menambahkan layanan premium atau fitur tambahan.
- Strategi untuk mempertahankan pelanggan lama, karena biaya mendapatkan pelanggan baru (*acquisition cost*) biasanya lebih tinggi dibandingkan mempertahankan pelanggan

c. Efek Tingkat Churn pada Potensi Pendapatan

Tingkat churn sebesar 10.37% menunjukkan potensi kehilangan pendapatan signifikan dalam jangka panjang jika masalah ini tidak ditangani. Jika churn berlanjut:

- Dengan asumsi pendapatan rata-rata dari setiap pelanggan (*ARPU*) tetap, kehilangan pelanggan 58,527 berarti potensi kehilangan pendapatan tambahan yang cukup besar.

Tindakan: Prioritaskan strategi retensi pelanggan, seperti Program loyalitas (*loyalty program*) dan penawaran khusus untuk pelanggan yang hampir churn (*retention offers*)

d. Kesempatan Ekspansi dan Optimasi

Jika wilayah ini adalah pasar utama untuk Telkomsel, ada peluang untuk meningkatkan basis pelanggan dan mengoptimalkan pendapatan dengan:

- Mengurangi churn rate untuk menjaga pelanggan yang ada.
- Melakukan penetrasi ke wilayah yang belum maksimal atau menawarkan paket menarik kepada pelanggan baru.

Tindakan: Perluasan paket layanan, terutama paket data, karena data sering menjadi kebutuhan utama pelanggan di era digital.

Kesimpulan

- Tingkat churn adalah indikator utama yang harus dikurangi untuk meningkatkan profitabilitas jangka panjang.
- Pendapatan yang tinggi menunjukkan bahwa operator mampu memanfaatkan basis pelanggan yang ada, tetapi perlu strategi untuk mempertahankan mereka.
- Prioritas tindakan Fokus pada retensi pelanggan, inovasi layanan, dan analisis lebih lanjut untuk memahami penyebab churn.

2. Churn and Retention

Kode Program 6. Churn and Retention

```

1.  SELECT
2.      country,
3.      operator,
4.      CASE
5.          WHEN FLOOR(subscription_tenure_days / 30) = 0 THEN '0-30 days'
6.          WHEN FLOOR(subscription_tenure_days / 30) = 1 THEN '31-60 days'
7.          WHEN FLOOR(subscription_tenure_days / 30) = 2 THEN '61-90 days'
8.          WHEN FLOOR(subscription_tenure_days / 30) = 3 THEN '91-120 days'
9.          ELSE '120+ days'
10.     END AS subscription_cycle,
11.     COUNT(*) AS total_subscriptions,
12.     SUM(CASE
13.         WHEN profile_status = 'inactive' AND status = -1 THEN 1
14.         ELSE 0
15.     END) AS churn_count,
16.     ROUND (
17.         (SUM(CASE
18.             WHEN profile_status = 'inactive' AND status = -1 THEN 1
19.             ELSE 0
20.         END) * 1.0 / COUNT(*)) * 100,
21.         2
22.     ) AS churn_rate
23. FROM
24.     cleaned_dataset_v6_new
25. GROUP BY
26.     country, operator, FLOOR(subscription_tenure_days / 30);

```

Pada soal mengenai churn and retention, kami diminta untuk menyusun query SQL yang bertujuan untuk menganalisis jumlah langganan, tingkat churn, serta pola retensi pelanggan berdasarkan wilayah (region) dan operator. Query ini juga mengelompokkan data langganan berdasarkan durasi waktu langganan, sehingga memungkinkan identifikasi tren churn dalam berbagai siklus waktu. Hasil dari query ini akan memberikan wawasan yang penting untuk mengevaluasi performa operator dalam mempertahankan pelanggan di masing-masing wilayah. Pada baris ke 1-3 memilih kolom country untuk mengidentifikasi wilayah tempat operator beroperasi dan kolom operator, yang menunjukkan nama penyedia layanan/operator. Selanjutnya pada baris ke 4-10 pada bagian CASE ... END AS subscription_cycle kita membagi durasi langganan (dalam hari) menjadi beberapa kategori siklus waktu berdasarkan nilai dari subscription_tenure_days. Hasilnya dikelompokkan dalam 0-30 days (jika durasi langganan kurang dari atau sama

dengan 30 hari), 31-60 days (jika durasi antara 31 hingga 60 hari), 61-90 days (jika durasi antara 61 hingga 90 hari), 91-120 days (jika durasi antara 91 hingga 120 hari), 120+ days untuk durasi lebih dari 120 hari. Kategori ini dihitung dengan membagi durasi (subscription_tenure_days) dengan 30, menggunakan FLOOR untuk pembulatan ke bawah.

Lalu untuk baris 11-15 kita menghitung total jumlah langganan untuk setiap kombinasi country, operator, dan subscription_cycle dan SUM(CASE WHEN ... END) AS churn_count kita menghitung jumlah pelanggan yang berhenti berlangganan (churned) berdasarkan kondisi. Dengan profile_status = 'inactive' pelanggan memiliki status profil tidak aktif dan status = -1 menunjukkan pelanggan telah berhenti berlangganan. Jika kondisi terpenuhi, nilai 1 dihitung; jika tidak, nilai 0 dihitung. Selanjutnya baris ke 16 - 22 ROUND(... AS churn_rate) dengan menghitung persentase churn dengan rumus $churn\ rate = (churn\ count / total\ subscriptions) \times 100$, kemudian hasilnya dibulatkan hingga dua desimal menggunakan fungsi ROUND. Dan baris ke 23-26 kita mengambil data dari tabel cleaned_dataset_v6_new, yang menyimpan informasi terkait langganan, durasi langganan, status pelanggan, dan operator lalu kita GROUP BY untuk mengelompokkan hasil berdasarkan wilayah dan operator layanan. Dan kita kelompokkan data berdasarkan kategori siklus waktu langganan menggunakan FLOOR(subscription_tenure_days / 30).

Berikut adalah output dari query yang kita jalankan:

	A-Z country	A-Z operator	A-Z subscription_cycle	123 total_subscriptions	123 churn_count	123 churn_rate
1	ID	telkomsel	0-30 days	487,047	53,446	10.97
2	ID	telkomsel	31-60 days	75,648	3,251	4.3
3	ID	telkomsel	61-90 days	1,482	1,413	95.34
4	ID	telkomsel	91-120 days	441	416	94.33
5	ID	telkomsel	120+ days	1	1	100

Gambar 13 Output SQL Query Churn and Retention

Pada output tersebut menunjukkan bahwa kolom tersebut menjelaskan antara lain:

a. Country (ID)

Menunjukkan wilayah tempat operator beroperasi. Dalam hal ini, "ID" mengacu pada Indonesia.

b. Operator (Telkomsel)

Nama operator telekomunikasi yang dianalisis, yaitu Telkomsel.

c. Subscription Cycle

Menggambarkan rentang waktu langganan pelanggan sebelum berhenti (churn) atau tetap aktif. Siklus ini dibagi menjadi:

- 0–30 days: Pelanggan baru dalam bulan pertama langganan.
- 31–60 days: Langganan bulan kedua.
- 61–90 days: Langganan bulan ketiga.
- 91–120 days: Langganan bulan keempat.
- 120+ days: Pelanggan lama (lebih dari 120 hari)

d. Total Subscriptions:

Menampilkan jumlah total pelanggan yang berlangganan pada setiap siklus.

e. Churn Count

Jumlah pelanggan yang berhenti berlangganan (churn) pada setiap siklus langganan

f. Churn Rate

Menampilkan nilai persentase nilai churn rate dalam setiap siklus waktu yang telah ditentukan

Dari output tersebut, kami memperoleh insight penting untuk menganalisis performa Telkomsel terkait churn dan retensi pelanggan, serta mengidentifikasi tantangan dalam mempertahankan pelanggan pada berbagai siklus langganan:

a. Retensi Awal yang Baik

Pada siklus 0–30 days dan 31–60 days, churn rate relatif rendah (10.97% dan 4.3%). Hal ini menunjukkan bahwa Telkomsel cukup berhasil mempertahankan pelanggan baru di fase awal langganan

Tindakan:

- Welcome Program, memberikan insentif seperti kuota tambahan atau diskon untuk pelanggan baru selama dua bulan pertama
- Promosi Paket Berkelanjutan, menawarkan paket upgrade (misalnya, kuota internet yang lebih besar) untuk memperpanjang langganan setelah bulan pertama
- Customer Onboarding, mengedukasi pelanggan baru tentang manfaat layanan Telkomsel melalui SMS, email, atau aplikasi MyTelkomsel

b. Penurunan Drastis Setelah 60 Hari

Pada siklus 61–90 days, churn rate meningkat tajam menjadi 95.34%, dan kondisi ini berlanjut pada siklus 91–120 days (94.33%). Sebagian besar pelanggan berhenti setelah dua bulan pertama.

Tindakan:

- Penawaran Paket Retensi, memberikan penawaran eksklusif seperti potongan harga atau tambahan kuota untuk pelanggan pada bulan ketiga dan keempat
- Monitoring Perilaku Pengguna, mengidentifikasi pelanggan yang jarang menggunakan layanan (low activity) di bulan kedua dan menawarkan paket yang lebih relevan untuk kebutuhan mereka
- Peringatan Sebelum Churn, mengirimkan notifikasi otomatis (via SMS atau aplikasi) kepada pelanggan yang menunjukkan tanda-tanda akan berhenti (contoh: penggunaan menurun)

c. Pelanggan Jangka Panjang Rentan Churn

Pada kategori 120+ days, churn rate mencapai 100%, yang mengindikasikan bahwa hampir tidak ada pelanggan yang bertahan dalam jangka panjang.

Tindakan:

- Program Loyalitas, meluncurkan program loyalitas berbasis poin, di mana pelanggan mendapatkan poin setiap kali memperpanjang langganannya. Poin tersebut dapat ditukarkan dengan hadiah, kuota tambahan, atau diskon
- Penghargaan untuk Pelanggan Lama, memberikan benefit khusus untuk pelanggan yang bertahan lebih dari 120 hari, seperti undian hadiah, akses eksklusif ke layanan premium, atau cashback

d. Strategi Umum untuk Semua Siklus

Data menunjukkan bahwa churn meningkat secara signifikan setelah dua bulan pertama. Oleh karena itu, diperlukan pendekatan holistik untuk semua siklus langganannya.

Tindakan:

- Peningkatan Kualitas Jaringan, menginvestasikan infrastruktur untuk memastikan pelanggan mendapatkan pengalaman jaringan yang stabil dan cepat
- Personalisasi Layanan, menggunakan data analitik untuk menawarkan paket yang sesuai dengan kebiasaan pelanggan (misalnya, pelanggan yang sering menggunakan aplikasi streaming ditawarkan paket video)

Kesimpulan:

Perusahaan menghadapi tantangan signifikan dalam mempertahankan pelanggan setelah bulan kedua berlangganan, dengan churn rate yang sangat tinggi pada siklus langganan lebih dari 60 hari. Meskipun retensi pada tahap awal langganan cukup baik, terutama pada bulan pertama dan kedua, sebagian besar pelanggan cenderung berhenti berlangganan setelah dua bulan. Untuk meningkatkan retensi, Telkomsel perlu fokus pada strategi yang lebih efektif dalam mempertahankan pelanggan jangka panjang, seperti menawarkan program loyalitas, peningkatan layanan pelanggan, dan analisis lebih lanjut untuk memahami faktor-faktor yang menyebabkan churn pada pelanggan lama.

3. Campaign Performance

Kode Program 7. Campaign Performance

```
1. SELECT
2.     cycle,
3.     country,
4.     operator,
5.     COUNT(CASE WHEN success_billing != 0 THEN 1 END) AS successful_billings,
6.     COUNT(*) AS total_attempts,
7.     ROUND(
8.         (COUNT(CASE WHEN success_billing != 0 THEN 1 END) * 1.0 / COUNT(*)) *
9.         100,
10.        2
11.    ) AS success_rate
12. FROM
13.     cleaned_dataset_v6_new
14. GROUP BY
15.     cycle, country, operator
16. ORDER BY
17.     success_rate DESC;
```

Pada soal mengenai campaign performance, kami diminta untuk menyusun query SQL yang bertujuan untuk menganalisis tingkat keberhasilan pembayaran (success rate) dalam kampanye berdasarkan siklus langganan, wilayah (country), dan operator. Query ini menghitung jumlah pembayaran yang berhasil serta total percobaan pembayaran

untuk setiap kelompok, sehingga memungkinkan identifikasi pola kinerja kampanye di berbagai wilayah dan siklus langganan. Pada baris ke 1-4 menampilkan siklus langganan pelanggan, menunjukkan wilayah geografis pelanggan yang berpartisipasi dalam kampanye dan tempat operator beroperasi dan kolom operator. Selanjutnya baris ke 5-6 kita menghitung jumlah pembayaran yang berhasil dihitung menggunakan `COUNT(CASE WHEN success_billing != 0 THEN 1 END)`, di mana logika `CASE WHEN success_billing != 0 THEN 1` memastikan hanya transaksi dengan nilai pembayaran tidak nol yang dihitung sebagai berhasil, sementara `COUNT(*)` AS total _attempts menghitung total seluruh percobaan pembayaran, baik yang berhasil maupun gagal.

Lalu pada baris 7-10 kita menghitung tingkat keberhasilan pembayaran dihitung dengan rumus $\text{Success Rate} = (\text{successful billings} / \text{total attempts}) \times 100$, dan hasilnya dibulatkan ke dua desimal menggunakan fungsi `ROUND(..., 2)` untuk memastikan presisi yang lebih baik. Dan pada baris 11-16 data diambil dari tabel `cleaned_dataset_v6_new`, yang telah dibersihkan dan berisi informasi lengkap tentang siklus, wilayah, operator, dan status pembayaran. Data ini kemudian dikelompokkan menggunakan `GROUP BY cycle, country, operator` untuk menghitung metrik seperti pembayaran berhasil dan total percobaan per kombinasi unik dari ketiga atribut tersebut. Hasil akhirnya diurutkan dengan `ORDER BY success_rate DESC` untuk menampilkan kombinasi dengan tingkat keberhasilan pembayaran tertinggi di bagian atas tabel

Berikut adalah output dari query yang kita jalankan:

	123 cycle	A-Z country	A-Z operator	123 successful_billings	123 total_attempts	123 success_rate
1	2d	ID	telkomsel	159,309	563,332	28.28
2	3d	ID	telkomsel	316	1,287	24.55

Gambar 14 Output SQL Query Campaign Performance

Pada output tersebut menunjukkan bahwa kolom tersebut menjelaskan antara lain:

- cycle

Menunjukkan siklus kampanye, di mana terdapat dua siklus yaitu 2d dan 3d. Siklus ini bisa mewakili periode tertentu (misalnya hari ke-2 dan hari ke-3)

b. country

Menunjukkan negara tempat kampanye dijalankan. Dalam hal ini, negara yang dianalisis adalah ID (Indonesia)

c. operator

Menunjukkan operator telekomunikasi yang menjadi target kampanye. Di sini, operator yang dianalisis adalah Telkomsel

d. total_attempts

Menunjukkan total percobaan pembayaran, termasuk yang berhasil dan gagal.

- Pada siklus 2d, terdapat total 563,332 percobaan.
- Pada siklus 3d, terdapat total 1,287 percobaan

e. success_rate

Menunjukkan tingkat keberhasilan pembayaran. Hasil perhitungannya telah dibulatkan hingga dua desimal menggunakan fungsi ROUND.

- Pada siklus 2d, tingkat keberhasilan adalah 28.28%.
- Pada siklus 3d, tingkat keberhasilan adalah 24.55%.

Dari output tersebut, kita dapat mengidentifikasi beberapa insight yang memberikan gambaran jelas tentang performa kampanye Perusahaan. Insight ini membantu mengevaluasi efektivitas kampanye dan mengungkap kekuatan serta kelemahan dalam strategi yang diterapkan. Dengan memahami pola keberhasilan dan tantangan yang ada, kita dapat menemukan peluang untuk meningkatkan kinerja kampanye. Data ini juga menjadi dasar untuk merumuskan strategi yang

lebih relevan dan efektif guna meningkatkan efisiensi operasional, memperluas audiens, dan memaksimalkan hasil bisnis di masa depan.

a. Tingkat Keberhasilan Pembayaran Berdasarkan Siklus Langganan

Cycle 2d memiliki tingkat keberhasilan pembayaran yang lebih tinggi (28,28%) dibandingkan dengan Cycle 3d (24,55%). Ini menunjukkan bahwa pelanggan yang berada pada siklus langganan yang lebih pendek lebih cenderung melakukan pembayaran dengan sukses.

Tindakan:

- Optimalkan Proses Pembayaran pada Siklus Lebih Lama

Telkomsel harus memfokuskan upaya untuk meningkatkan tingkat keberhasilan pembayaran pada Cycle 3d. Ini dapat dilakukan dengan memeriksa dan mengatasi hambatan teknis atau administratif

b. Peningkatan dalam Percobaan Pembayaran pada Siklus Lebih Lama

Meskipun tingkat keberhasilan di Cycle 3d lebih rendah, jumlah total percobaan pembayaran di siklus ini jauh lebih tinggi, yakni 1,287 percobaan dibandingkan dengan 563,332 percobaan pada Cycle 2d. Hal ini mengindikasikan potensi masalah dalam proses pembayaran yang perlu diperbaiki, terutama untuk pelanggan yang berada dalam siklus langganan lebih lama.

Tindakan:

- Tingkatkan Pengalaman Pengguna untuk Siklus Pendek

Dengan keberhasilan yang lebih tinggi pada Cycle 2d, penting untuk mempertahankan atau meningkatkan pengalaman pengguna di awal siklus langganan

c. Kinerja Telkomsel di Pasar Indonesia

Telkomsel menunjukkan hasil yang signifikan dalam hal pembayaran yang berhasil, dengan total 159,309 pembayaran yang berhasil pada Cycle 2d, meskipun total percobaan pada siklus lebih lama tetap cukup besar.

Tindakan:

- Analisis Kualitas Data dan Penargetan

Melakukan analisis lebih lanjut terhadap faktor-faktor yang mempengaruhi keberhasilan pembayaran untuk masing-masing siklus. Menyempurnakan data pelanggan dan meningkatkan penargetan kampanye untuk meningkatkan retensi pelanggan

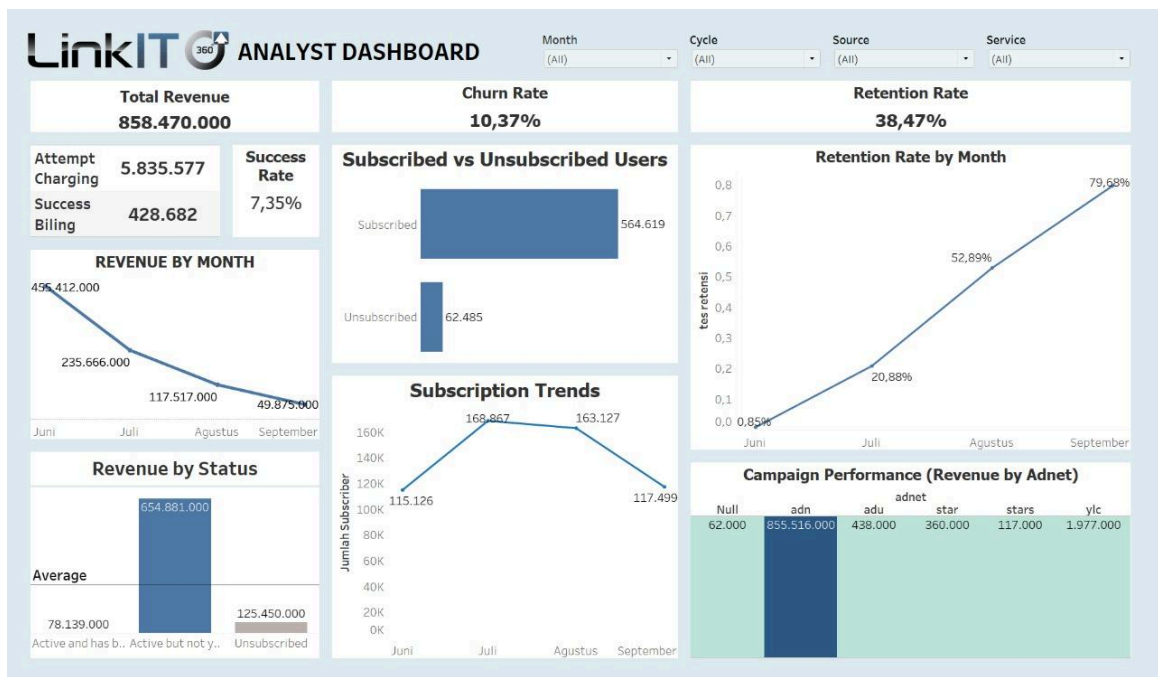
Kesimpulan:

Berdasarkan analisis performa kampanye Telkomsel, dapat disimpulkan bahwa siklus langganan yang lebih pendek (2d) memiliki tingkat keberhasilan pembayaran yang lebih tinggi, sementara siklus lebih lama (3d) menghadapi tantangan dengan tingkat keberhasilan yang lebih rendah meskipun memiliki jumlah percobaan pembayaran yang lebih banyak. Untuk mengoptimalkan hasil, Telkomsel perlu meningkatkan proses pembayaran pada siklus lebih lama dan memperkuat pengalaman pengguna pada siklus pendek, serta melakukan analisis mendalam untuk mengidentifikasi faktor-faktor yang mempengaruhi keberhasilan pembayaran guna meningkatkan retensi pelanggan

6. Pembahasan

Dashboard Creation

Berikut adalah dashboard interaktif yang telah kita buat dengan mengimpor data yang sudah dibersihkan ke dalam Tableau ini, menyajikan visualisasi yang komprehensif mengenai tren langganan, churn dan retensi pelanggan, serta kinerja kampanye. Setiap elemen dalam dashboard dirancang untuk memberikan wawasan yang lebih mendalam, memudahkan analisis data, dan mendukung pengambilan keputusan berbasis informasi yang relevan sesuai dengan kebutuhan analisis.



Gambar 15 Tableau Dashboard

Insights and Action Plans

Berdasarkan insight yang ditampilkan dalam dashboard, terdapat berbagai data yang dapat membantu untuk meningkatkan retensi pelanggan dan mengoptimalkan strategi pemasaran. Melalui analisis terhadap tren churn, tingkat retensi, serta kinerja kampanye dan billing, kita dapat menyusun rencana tindakan yang lebih terarah untuk mencapai hasil yang lebih baik. Berikut ini adalah beberapa rekomendasi yang dapat

diterapkan untuk meningkatkan retensi pelanggan, memperbaiki sistem billing, serta mengoptimalkan kinerja pemasaran berdasarkan wawasan yang didapat dari data.

a. Meningkatkan Retensi Pelanggan

Analisis dari Dashboard antara lain:

- Peningkatan retensi bulanan berdasarkan grafik Retention Ratio by Monthly, terlihat bahwa tingkat retensi pelanggan meningkat secara signifikan dari bulan ke bulan. Pada bulan Juni (0,85%) dan Juli (20,88%), tingkat retensi tergolong sangat rendah, namun Agustus mengalami lonjakan menjadi 52,89% dan puncaknya pada September di 78,68%. Walaupun demikian, tingkat retensi secara keseluruhan masih rendah dengan nilai rata-rata 38,47%
- Tingkat churn yang tinggi sebesar 10,37% menunjukkan bahwa sekitar 10% pelanggan berhenti berlangganan setiap bulannya, yang menandakan adanya masalah dalam mempertahankan pelanggan yang cukup signifikan
- Jumlah pelanggan berlangganan juga menurun sejak Juli (168.867) menjadi 117.499 pelanggan pada bulan September, yang memperlihatkan bahwa tren langganan menurun meskipun tingkat unsubscription sedikit menurun pada periode yang sama

Rekomendasi Strategi yang dapat dilakukan:

- Meningkatkan engagement pelanggan dengan menggunakan notifikasi pengingat untuk pelanggan yang mendekati masa berakhir langganan guna memastikan pembaruan langganan mereka dilakukan tepat waktu.
- Segmentasi pelanggan yang rentan churn dengan mengidentifikasi pelanggan dengan tingkat penggunaan rendah atau pelanggan baru yang

berisiko berhenti langganan, lalu buat kampanye yang dipersonalisasi untuk meningkatkan retensi pelanggan tersebut.

- Menyediakan opsi berlangganan yang lebih fleksibel (bulanan, triwulanan, atau tahunan) untuk menyesuaikan dengan preferensi dan kebutuhan finansial pelanggan.
- Kampanye retensi yang dipersonalisasi akan meningkatkan relevansi tawaran bagi pelanggan dan mengurangi tingkat churn.
- Menambahkan dukungan pelanggan berbasis digital seperti chatbot atau layanan pelanggan 24/7, yang memungkinkan pelanggan mendapatkan bantuan kapan saja, untuk meningkatkan pengalaman pelanggan dan mendorong mereka bertahan lebih lama

b. Mengoptimalkan Pemasaran dan Penagihan

Analisis dari Dashboard antara lain:

- Dengan tingkat keberhasilan penagihan hanya 7,35% meskipun ada 5,8 juta percobaan penagihan, ini menunjukkan bahwa banyak pelanggan yang tidak berhasil melakukan pembayaran meskipun sudah mencoba.
- Kampanye iklan seperti adnet (adn) memiliki kontribusi yang sangat besar terhadap pendapatan, dengan pendapatan lebih dari Rp 855.516.000, jauh lebih tinggi dibandingkan dengan kampanye lain seperti "star" dan "yle" yang hanya memberikan pendapatan jauh lebih kecil.
- Pendapatan dari Pelanggan Aktif sebesar Rp 654.881.000 berasal dari pelanggan yang aktif namun belum memperpanjang layanan mereka, menunjukkan adanya potensi besar untuk meningkatkan retensi pelanggan yang belum memperpanjang layanan mereka

Rekomendasi Strategi yang dapat dilakukan:

- Meningkatkan keberhasilan billing dengan menerapkan sistem otomatis untuk percobaan ulang pembayaran gagal dan mengirimkan notifikasi pengingat kepada pelanggan sebelum masa jatuh tempo untuk mengurangi kegagalan transaksi.
- Memperkenalkan metode pembayaran yang lebih bervariasi, seperti dompet digital, transfer bank, dan kartu kredit, untuk mempermudah pelanggan dalam menyelesaikan pembayaran
- Prioritaskan kampanye iklan yang efektif dengan fokus pada adnet (adn) yang memiliki kontribusi terbesar terhadap pendapatan, serta evaluasi kampanye dengan kinerja rendah seperti "star" dan "ylc" untuk memperbaiki materi pemasaran atau target audiens
- Meluncurkan promosi eksklusif pada Juni untuk memanfaatkan momentum kenaikan pelanggan, dan tingkatkan kampanye reaktivasi pada Agustus dan September untuk menarik pelanggan yang berhenti berlangganan

Dari analisis keseluruhan, terlihat bahwa LinkIt menghadapi tantangan besar dalam mengelola risiko churn dan kegagalan pembayaran yang memengaruhi pendapatan. Sebagian besar pelanggan memiliki risiko churn tinggi, dengan masa langganan yang cenderung singkat, meskipun pelanggan dengan masa langganan lebih lama menghasilkan pendapatan yang lebih besar. Tingkat kegagalan pembayaran yang signifikan pada rasio diatas 0,50 menunjukkan ketidakstabilan dalam proses pembayaran, mencerminkan ada masalah serius. Untuk mengatasi tantangan ini, LinkIt perlu mengimplementasikan strategi retensi pelanggan, seperti insentif atau program loyalitas, serta meningkatkan efisiensi sistem pembayaran untuk mengurangi kegagalan transaksi. Fokus pada segmen pelanggan dengan risiko churn sedang dan tinggi, disertai upaya mempertahankan pelanggan loyal, dapat membantu meningkatkan retensi dan mengoptimalkan pendapatan secara keseluruhan.

BAB IV

Kesimpulan dan Saran

A. Kesimpulan

Berdasarkan analisis yang telah dilakukan, terdapat beberapa temuan utama yang perlu diperhatikan:

1. **Tingkat Churn Pelanggan yang Tinggi**

Analisis eksplorasi data (EDA) menunjukkan bahwa risiko churn pelanggan memang tinggi. Hal ini mengindikasikan bahwa perusahaan menghadapi tantangan dalam mempertahankan pelanggan. Perlu dilakukan evaluasi terhadap pengalaman pelanggan dan relevansi layanan untuk mengurangi tingkat churn yang terjadi.

2. **Rendahnya Keterlibatan Pengguna Terhadap Layanan**

Rendahnya tingkat keterlibatan pengguna terhadap layanan kemungkinan disebabkan oleh efektivitas kampanye pemasaran yang kurang optimal. Kampanye yang lebih tepat sasaran dan menarik perlu dirancang untuk meningkatkan engagement pelanggan dan mencegah potensi churn di masa mendatang.

3. **Kendala dalam Sistem Billing**

Distribusi rasio *failed charge* sebagian besar berada pada angka 0.5 hingga 1.0, menunjukkan tingginya kegagalan proses penagihan. Hal ini tidak hanya mengurangi kepercayaan pelanggan tetapi juga berdampak pada pendapatan perusahaan. Diperlukan peningkatan pada sistem penagihan untuk memastikan transaksi berjalan lancar dan mengurangi potensi kerugian.

Dengan menyelesaikan tiga permasalahan ini, perusahaan dapat memperbaiki retensi pelanggan, meningkatkan pengalaman pengguna, dan mengoptimalkan pendapatan.

B. Saran

Untuk mengatasi tingginya tingkat churn, perusahaan harus menerapkan strategi retensi yang lebih ampuh. Misalnya, memberikan insentif kepada pelanggan yang berisiko tinggi dan melakukan intervensi preventif sebelum mereka memutuskan berhenti. Selain itu, perusahaan dapat meningkatkan keterlibatan pengguna melalui kampanye yang lebih menarik dan relevan, serta menggunakan data untuk menargetkan pelanggan berdasarkan kebutuhan dan preferensi mereka. Untuk mengatasi masalah dalam sistem billing, perlu dilakukan evaluasi dan perbaikan teknis guna meningkatkan akurasi data pembayaran dan memastikan keberhasilan transaksi. Dengan langkah-langkah ini, pelanggan akan mendapatkan pengalaman yang lebih baik dan kepercayaan yang lebih tinggi terhadap layanan yang ditawarkan.

Lampiran

A. Online Diagram

https://lucid.app/lucidchart/1dec3563-dd75-48f7-9d14-638fbccf7e19/edit?invitationId=inv_c81be53a-b236-4029-b2ff-722959c2bcf7&page=0fkvEbqIFx4y#

B. Python Code

Data Cleaning :

<https://colab.research.google.com/drive/1YevTxzOaMBVFW4TkDlpKoxaHosQLUvxt>

EDA :

<https://colab.research.google.com/drive/1JuRz5HWOtASVFM4bSajajH2eSVOUqox6>

C. Notulen

<https://docs.google.com/document/d/1pEe7IxT-PLGriqYOetc7bIK5l521wTKbGSkpN4EYro/edit?tab=t.0#heading=h.s5aekzdvuufd>

D. Tableau

https://public.tableau.com/views/PBLLINKIT360_17336430212470/Dashboard2?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

E. Recording

Link Recording:

<https://youtu.be/MoC0sun-zuM?si=psLacGtPYj69sj9B>