TUGAS MISI 4 PERGUDANGAN DATA
“Perancangan Data Warehouse untuk Analisis Customer Churn pada Industri Telekomunikasi”

KELOMPOK 8 DW RA

ANGGOTA KELOMPOK :	
EKSANTY F SUGMA ISLAMIATY	122450001
RESIDEN NUSANTARA R M	122450080
AISYAH TIARA PRATIWI	121450074
UKASYAH MUNTAHA	122450028
RENDRA EKA PRAYOGA	122450112

PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN

2025

DAFTAR ISI

1. Ringkasan Proyek dan Latar Belakang..... 1
2. Tujuan dan Ruang Lingkup Sistem.....1
2.1 Tujuan utama..... 1
2.2 Ruang lingkup.....1
3. Metodologi Proyek..... 2
4. Analisis Kebutuhan.....2
5. Desain Konseptual, Logikal, dan Fisik.....3
5.1 Desain Konseptual.....3
5.2 Desain Logikal.....4
5.3 Desain Fisik..... 4
6. Proses Implementasi..... 5
7. Hasil Implementasi.....9
8. Evaluasi..... 11
8.1 Keberhasilan..... 11
8.2 Kendala..... 11
9. Rencana Pengembangan ke Depan.....11
10. Tim Proyek.....11

1. Ringkasan Proyek dan Latar Belakang

Perusahaan telekomunikasi bernama ‘Pythagoras’ mengalami tingkat churn pelanggan yang tinggi, yang berdampak langsung pada penurunan pendapatan dan loyalitas pelanggan. Churn adalah kondisi ketika pelanggan berhenti menggunakan layanan. Untuk mengatasi masalah ini, dibutuhkan sistem Data Warehouse (DW) yang mampu mengintegrasikan berbagai data pelanggan, layanan, lokasi, serta pembayaran guna mendukung analisis churn secara menyeluruh dan prediktif. Proyek ini bertujuan membangun sistem Data Warehouse berbasis SQL Server yang mampu menyajikan informasi analitik berbasis OLAP guna mendukung keputusan strategis perusahaan dalam menurunkan churn.

2. Tujuan dan Ruang Lingkup Sistem

2.1 Tujuan utama

- Membangun sistem Data Warehouse untuk integrasi dan analisis churn.
- Mengidentifikasi segmen pelanggan yang lebih cenderung untuk churn atau tetap bertahan berdasarkan faktor-faktor seperti usia, jenis kelamin, metode pembayaran, dan lokasi.
- Mengidentifikasi hubungan antara pengeluaran pelanggan dan tingkat churn untuk merancang strategi retensi yang lebih efektif.

2.2 Ruang lingkup

- Ekstraksi data dari CSV sumber.
- Desain dan implementasi skema bintang (star schema).
- Transformasi dan pemuatan data (ETL).
- Pembuatan query analitik (OLAP) untuk analisis churn.
- Dokumentasi sistem dan hasil implementasi

2.3 Identifikasi Kebutuhan Data Pengguna

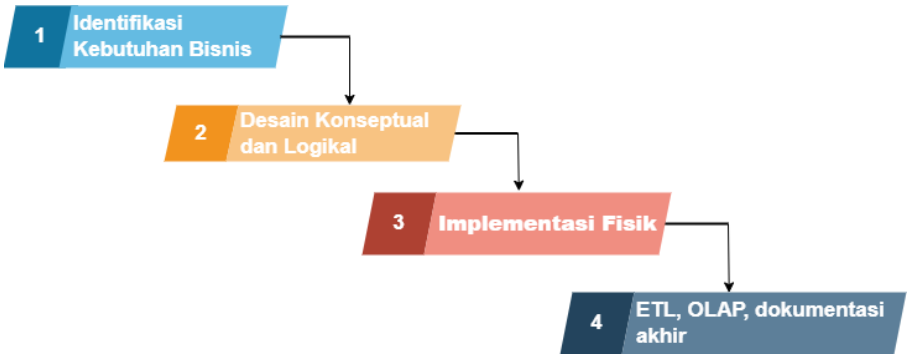
- Data Demografis Pengguna yang menyimpan informasi dasar mengenai identitas pelanggan untuk mempermudah segmentasi pasar, seperti Nama, usia, jenis kelamin, status pernikahan, jumlah tanggungan, dan lokasi berdasarkan kota. Dengan data ini, perusahaan bisa mengidentifikasi segmen pelanggan yang lebih cenderung churn.
- Data Penggunaan Layanan dan Pengeluaran yang menyimpan informasi terkait penggunaan layanan dan pengeluaran pelanggan untuk menganalisis hubungan antara pengeluaran dan keputusan churn, seperti Total biaya bulanan (*Monthly Charge*), total biaya layanan (*Total Charges*), jenis layanan yang digunakan (misalnya, internet, TV kabel), pembayaran bulanan, dan durasi langganan. Data ini digunakan untuk menganalisis apakah pelanggan dengan pengeluaran lebih tinggi atau lebih rendah cenderung churn, serta apakah jenis layanan tertentu lebih rentan menyebabkan churn. Data ini juga membantu untuk merancang penawaran khusus bagi pelanggan dengan pengeluaran lebih rendah agar tetap bertahan.

- Data Perilaku dan Interaksi Pelanggan yang menyimpan informasi tentang interaksi pelanggan dengan perusahaan, seperti alasan churn dan tingkat kepuasan, seperti Alasan churn (*Churn Reason*), jumlah referensi yang diberikan (*Number of Referrals*), dan keluhan. Hal ini untuk mengidentifikasi alasan pelanggan churn agar dapat merancang strategi yang tepat untuk memperbaiki produk dan layanan yang tidak memadai.

3. Metodologi Proyek

Metodologi yang digunakan adalah model Waterfall, dengan tahapan sebagai berikut:

- Misi 1: Identifikasi kebutuhan bisnis dan stakeholder.
- Misi 2: Desain konseptual dan logikal Data Warehouse.
- Misi 3: Implementasi fisik, optimasi indeks, dan partisi tabel.
- Misi 4: ETL, pengisian data, analitik OLAP, dan dokumentasi akhir.



Tools:

- SQL Server (DBMS)
- Python (ETL)
- Excel/CSV (Sumber data)
- GitHub (repository skrip dan dokumentasi)
- Tableau (opsional, untuk visualisasi)

4. Analisis Kebutuhan

Berdasarkan Misi 1, kebutuhan sistem Data Warehouse pada proyek ini dirancang berdasarkan peran dan tujuan dari stakeholder sebagai berikut:

1. **Sponsor Eksekutif:**
    a. Kebutuhan: Laporan ringkasan churn dan insight strategis.
    b. Tujuan: Meningkatkan retensi pelanggan, menekan biaya, dan meningkatkan laba.
2. **Manajer Proyek:**
    a. Kebutuhan: Informasi kemajuan proyek dan efektivitas implementasi.
    b. Tujuan: Menyelesaikan proyek tepat waktu dan berkualitas.
3. **Manajer Penghubung Pengguna:**
    a. Kebutuhan: Validasi kebutuhan pengguna dari hasil laporan churn.
    b. Tujuan: Menjembatani kebutuhan pengguna dan tim teknis.

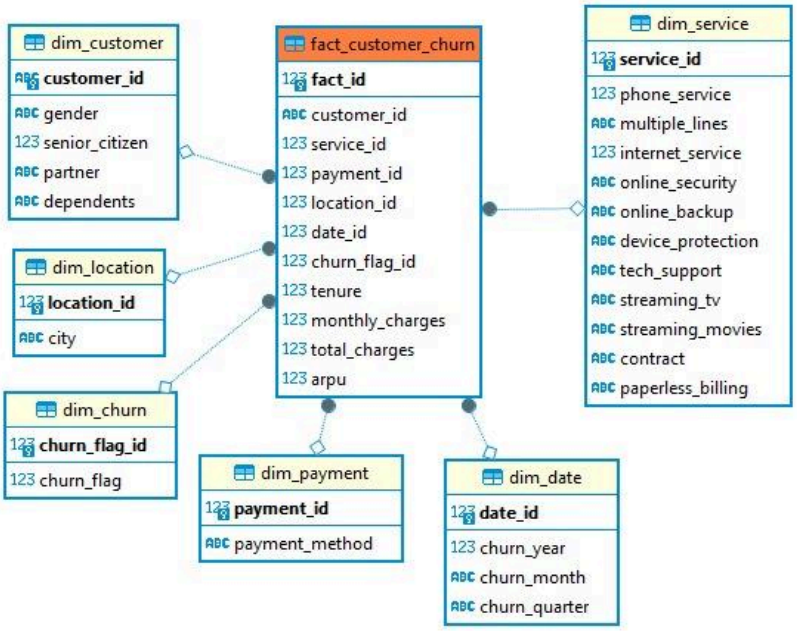
4. **Arsitek Utama:**
    a. Kebutuhan: Rancangan skema dan infrastruktur DW.
    b. Tujuan: Membangun sistem yang skalabel dan cepat.
5. **Analis Bisnis:**
    a. Kebutuhan: Akses ke data dimensional dan metrik churn.
    b. Tujuan: Menyediakan insight dari data dan dasar pengambilan keputusan.
6. **Administrator DW:**
    a. Kebutuhan: Infrastruktur yang aman dan stabil.
    b. Tujuan: Menjaga kualitas, ketersediaan, dan integritas data.
7. **Spesialis Infrastruktur:**
    a. Kebutuhan: Arsitektur penyimpanan dan partisi data.
    b. Tujuan: Memastikan sistem dapat diakses secara efisien.
8. **Analis Penjaminan Mutu:**
    a. Kebutuhan: Verifikasi kualitas dan keakuratan data.
    b. Tujuan: Menjamin validitas hasil analisis.
9. **Koordinator Pengujian:**
    a. Kebutuhan: Data uji untuk simulasi dan validasi hasil kueri.
    b. Tujuan: Menjamin sistem berjalan stabil dan bebas dari kesalahan.
10. **Pelatih Utama:**
    a. Kebutuhan: Sistem yang mudah dipahami dan digunakan.
    b. Tujuan: Meningkatkan literasi data dan penggunaan DW oleh pengguna akhir.

Dengan kebutuhan yang beragam dari berbagai pemangku kepentingan, sistem DW dirancang secara multidimensional agar mampu memenuhi analisis operasional hingga strategis yang mendalam.

5. Desain Konseptual, Logikal, dan Fisik

5.1 Desain Konseptual

- Skema bintang dengan satu tabel fakta dan enam dimensi.



5.2 Desain Logikal

- Tabel Fakta: fact_customer_churn
- Tabel Dimensi: dim_customer, dim_service, dim_payment, dim_location, dim_date, dim_churn

Facts	Measures	Dimensions and Cardinalities	Hierarchies and Levels
Customer Churn	- Monthly Charge - Total Charges - Tenure in Months - Avg Monthly GB Download - Churn Status	Customer (1:n)	**Demographics:** - Gender → Age Group → Marital Status - Number of Dependents
		Services (1:n)	**Service Type:** - Internet Type (DSL/Fiber/Cable) - Phone Service (Yes/No) - Additional Features (Security, Backup, etc.)
		Contract (1:n)	**Billing Structure:** - Contract Type - Payment Method
		Location (1:n)	**Geography:** - City → State → Region
		Time (1:n)	**Temporal:** - Tenure Months → Contract Period - Join Date → Churn Date

5.3 Desain Fisik

- Tabel dibuat di SQL Server.
- Data di partisi berdasarkan tahun churn.
- Indeks dibuat pada kolom filter dan agregasi umum.

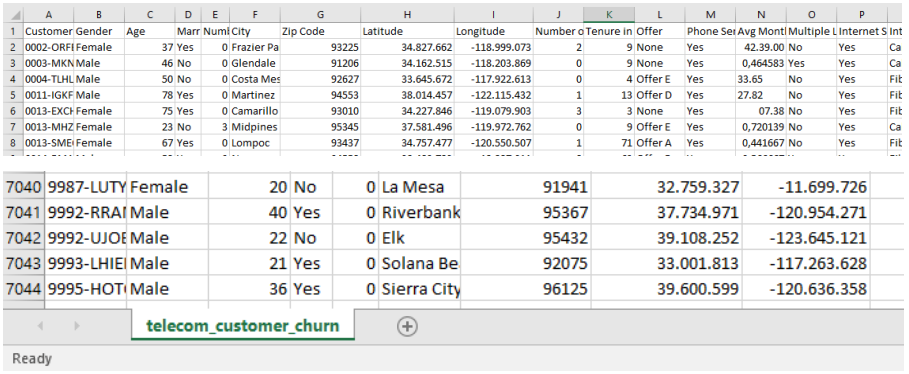
```
CREATE CLUSTERED INDEX idx_fact_customer_churn
ON fact_customer_churn (Customer_ID, Date_Key);
```

Kode di atas merupakan langkah implementasi indeksing untuk menunjang efisiensi query, dibuat clustered index pada kolom Customer_ID dan Date_Key guna mempercepat pencarian berdasarkan pelanggan dan waktu.

6. Proses Implementasi

Langkah-langkah implementasi proyek secara sistematis:

1. Ekstraksi Data:
   a. Dataset telecom_customer_churn.csv diunduh dari Kaggle, memuat 7.043 baris data pelanggan.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Customer	Gender	Age	Marr	Numl	City	Zip Code	Latitude	Longitude	Number o	Tenure in	Offer	Phone Ser	Avg Monti	Multiple L	Internet S	Int
2	0002-ORFI	Female	37	Yes	0	Frazier Pa	93225	34.827.662	-118.999.073	2	9	None	Yes	42.39.00	No	Yes	Ca
3	0003-MKN	Male	46	No	0	Glendale	91206	34.162.515	-118.203.869	0	9	None	Yes	0,464583	Yes	Yes	Ca
4	0004-TLHL	Male	50	No	0	Costa Mes	92627	33.645.672	-117.922.613	0	4	Offer E	Yes	33.65	No	Yes	Fib
5	0011-IGKF	Male	78	Yes	0	Martinez	94553	38.014.457	-122.115.432	1	13	Offer D	Yes	27.82	No	Yes	Fib
6	0013-EXCI	Female	75	Yes	0	Camarillo	93010	34.227.846	-119.079.903	3	3	None	Yes	07.38	No	Yes	Fib
7	0013-MHZ	Female	23	No	3	Midpines	95345	37.581.496	-119.972.762	0	9	Offer E	Yes	0,720139	No	Yes	Ca
8	0013-SMEI	Female	67	Yes	0	Lompoc	93437	34.757.477	-120.550.507	1	71	Offer A	Yes	0,441667	No	Yes	Fib
7040	9987-LUTY	Female	20	No	0	La Mesa	91941	32.759.327	-11.699.726								
7041	9992-RRAI	Male	40	Yes	0	Riverbank	95367	37.734.971	-120.954.271								
7042	9992-UJOI	Male	22	No	0	Elk	95432	39.108.252	-123.645.121								
7043	9993-LHIE	Male	21	Yes	0	Solana Be	92075	33.001.813	-117.263.628								
7044	9995-HOT	Male	36	Yes	0	Sierra City	96125	39.600.599	-120.636.358								

Gambar 1. Jumlah Dataset

2. Transformasi Data (ETL):
   a. Penambahan kolom baru Churn_Flag untuk klasifikasi pelanggan churn (1) atau aktif (0).
   b. Perhitungan kolom baru ARPU dengan rumus Total Revenue / Tenure in Months.
   c. Pembersihan data: nilai tenure = 0 diganti 1 untuk menghindari pembagian nol.
   d. Penambahan kolom waktu: churn_year = 2022, churn_month = 'June', churn_quarter = 'Q2' karena data berasal dari kuartal II 2022.
   e. Semua transformasi dilakukan dengan Python menggunakan Pandas.
3. Load ke SQL Server:
   a. Tabel fakta dan dimensi dibuat menggunakan create_tables.sql.
   b. Data hasil transformasi dimuat ke dim_customer, dim_service, dim_payment, dim_location, dim_date, dim_churn, dan fact_customer_churn.
   c. Data disisipkan menggunakan insert_data.sql atau wizard SSMS.
4. Implementasi Query Analitik (OLAP):
   a. Menggunakan SQL Query:
      Total Revenue
      Query:

```
SELECT
   SUM(f.total_charges) AS ‘Total Revenue’
FROM
   fact_customer_churn f
JOIN
   dim_date d ON f.date_id = d.date_id
WHERE
   d.churn_year = 2022
   AND d.churn_month = 'June'
   AND d.churn_quarter = 'Q2';
```

Hasil:

Total Revenue
1170609.2980766296

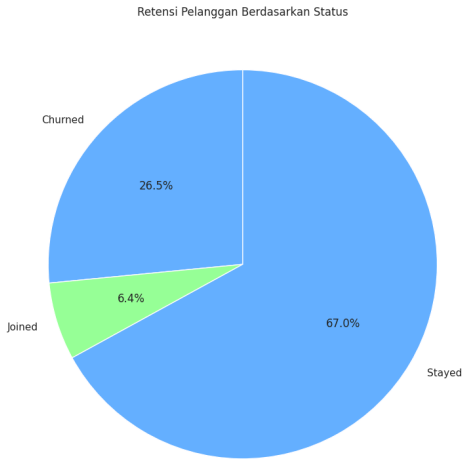
Analisis menggunakan query MySQL untuk menghitung total Revenue (pendapatan) pada bulan Juni Tahun 2022 didapatkan bahwa total pendapatannya adalah sebesar USD 1170609.2980766295.

Melihat Retensi Pelanggan

Query:

```
SELECT
   dc.churn_flag,
   COUNT(*) AS total_customers
FROM dim_churn dc
JOIN fact_customer_churn fcc ON dc.churn_flag_id = fcc.churn_flag_id
GROUP BY dc.churn_flag;
```

Hasil:



Retensi pelanggan cukup kuat (lebih dari dua pertiga tetap bertahan). Namun, churn rate 26.5% masih tergolong tinggi. Untuk mengatasi masalah ini perlu dilakukan analisis lanjut mengenai alasan kenapa pelanggan tidak melanjutkan langganannya.

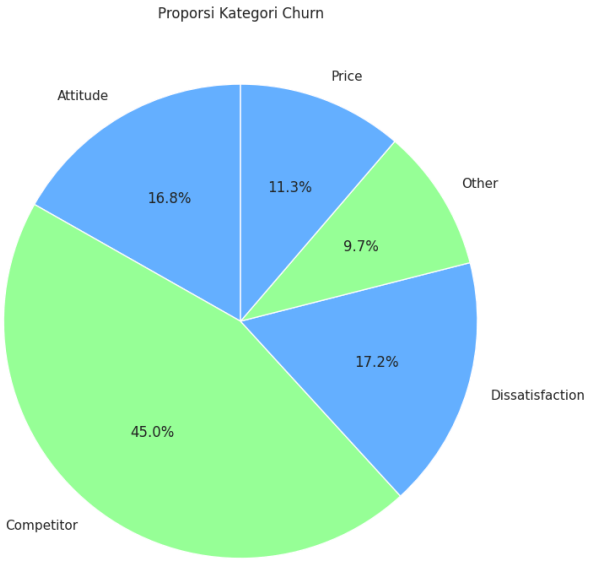
Proporsi Kategori Churn

Query:

```
SELECT
```

```
    churn_category,
    COUNT(*) AS jumlah_churn,
    ROUND(COUNT(*) * 100.0 / (
        SELECT COUNT(*)
        FROM dim_churn
        WHERE churn_flag = 1
    ), 2) AS persentase_churn
FROM dim_churn
WHERE churn_flag = 1
GROUP BY churn_category
ORDER BY persentase_churn DESC;
```

Hasil:



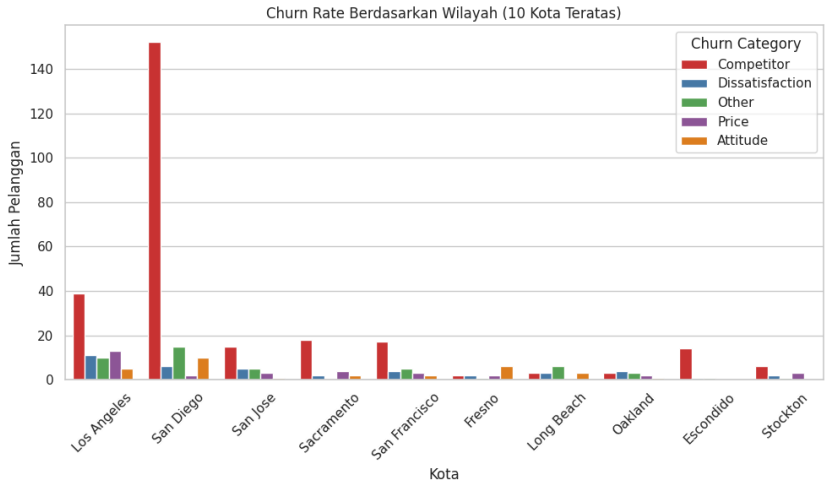
Kompetitor (45%) menjadi penyebab utama churn. Perusahaan kehilangan hampir setengah pelanggan karena pesaing. Hal ini bisa disebabkan oleh penawaran lebih menarik dari kompetitor, atau kurangnya diferensiasi layanan.

Churn Rate Berdasarkan Wilayah (10 Kota Teratas)

Query:

```
SELECT
    dl.city,
    COUNT(*) AS total_customers,
    SUM(CASE WHEN ch.churn_flag = 1 THEN 1 ELSE 0 END) AS
total_churn,
    ROUND(SUM(CASE WHEN ch.churn_flag = 1 THEN 1 ELSE 0 END)
/ COUNT(*) * 100, 2) AS churn_rate_percentage
FROM fact_customer_churn fcc
JOIN dim_churn ch ON fcc.churn_flag_id = ch.churn_flag_id
JOIN dim_location dl ON fcc.location_id = dl.location_id
GROUP BY dl.city
ORDER BY total_customers DESC
LIMIT 10;
```

Hasil:



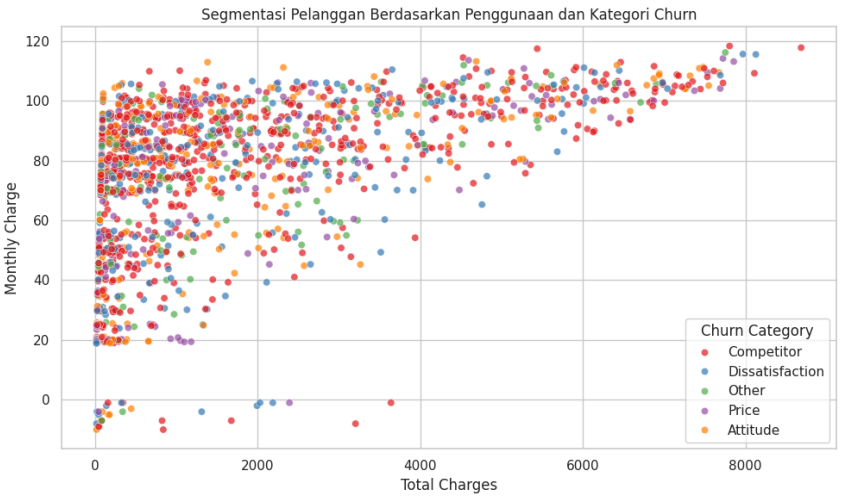
Pesaing (Competitor) adalah penyebab utama churn, terutama di San Diego. Beberapa kota menunjukkan churn yang disebabkan oleh kombinasi faktor, bukan hanya satu penyebab dominan.

Segmentasi pelanggan berdasarkan penggunaan layanan dan churn

Query SQL:

```
SELECT
    ds.internet_service,
    ds.phone_service,
    ds.tech_support,
    ch.churn_flag,
    COUNT(*) AS total_customers
FROM fact_customer_churn fcc
JOIN dim_service ds ON fcc.service_id = ds.service_id
JOIN dim_churn ch ON fcc.churn_flag_id = ch.churn_flag_id
GROUP BY ds.internet_service, ds.phone_service, ds.tech_support,
ch.churn_flag
ORDER BY total_customers DESC;
```

Hasil:

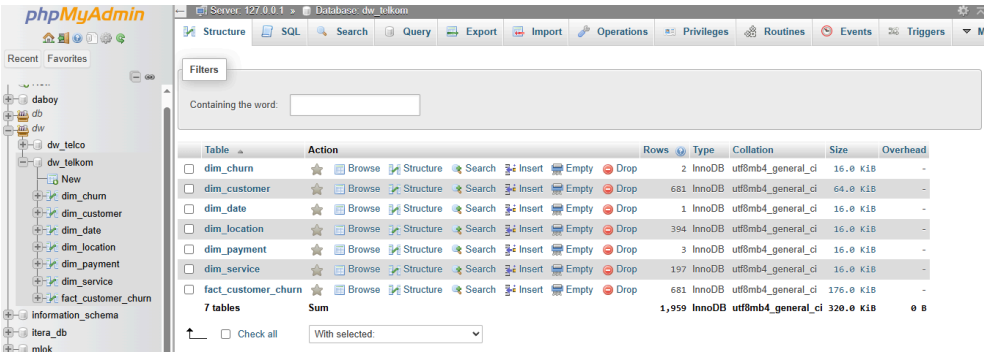


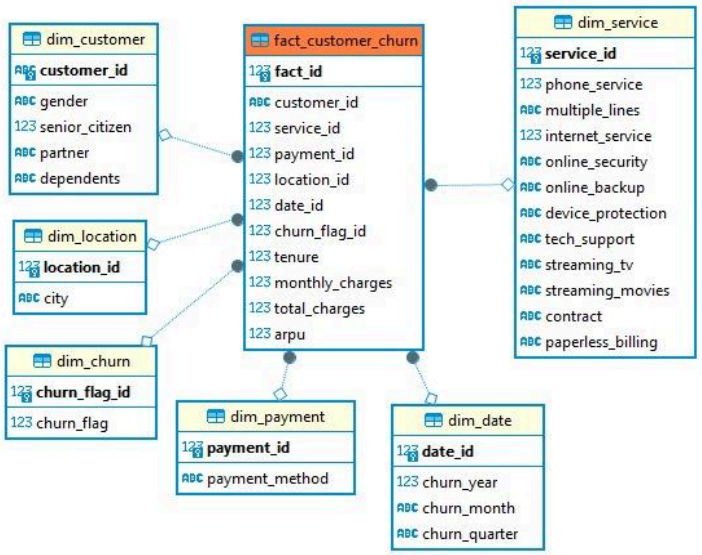
Terlihat pola hubungan positif, di mana pelanggan dengan biaya bulanan tinggi cenderung memiliki total pengeluaran yang juga tinggi. Kategori churn tersebar di seluruh grafik, namun churn karena Price lebih banyak muncul pada pelanggan dengan biaya bulanan tinggi, sedangkan Attitude dan Competitor mendominasi berbagai segmen.

Terdapat pula konsentrasi pelanggan baru dengan total charges rendah namun biaya bulanan tinggi. Beberapa outlier juga terlihat, yang mungkin disebabkan oleh anomali data atau kondisi khusus.

7. Hasil Implementasi

- Tabel dimensi dan fakta berhasil dibuat dan terisi di SQL Server.

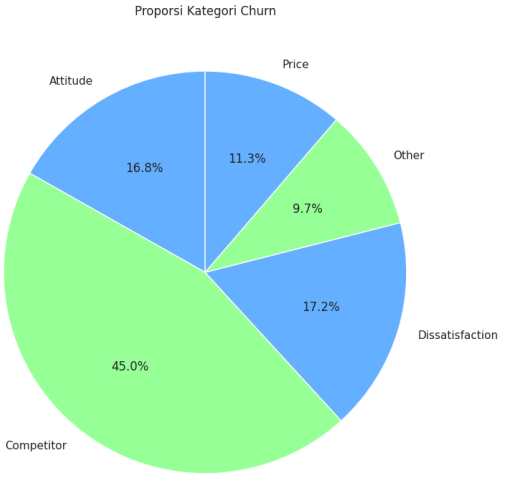




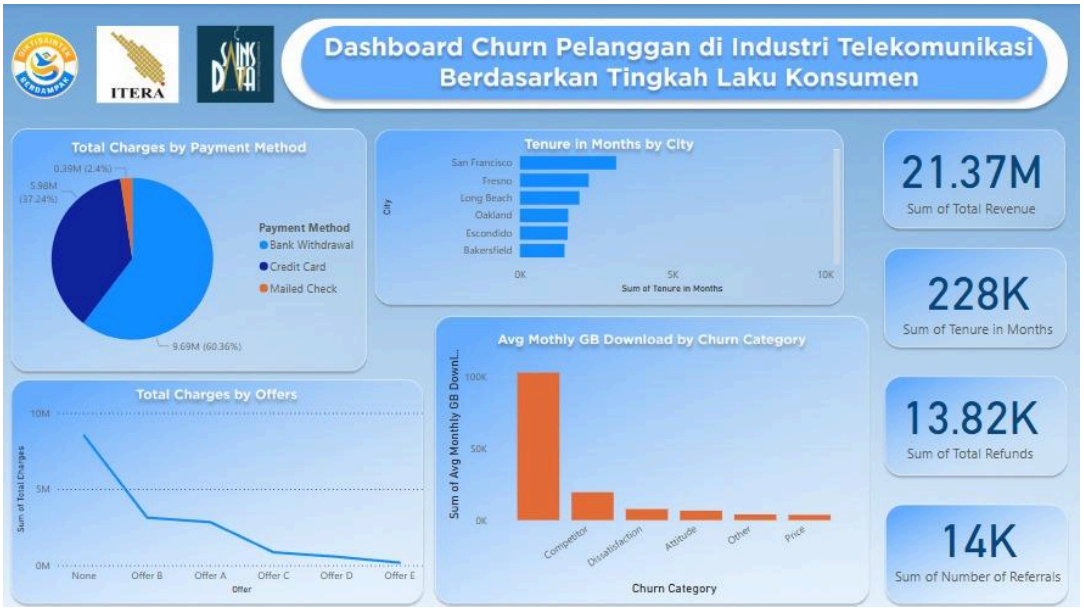
- ARPU berhasil dihitung dan dapat digunakan sebagai indikator loyalitas pelanggan.

Kota	Kontrak	Rata_Rata_ARPU ▾ 1
Castro Valley	Month-to-Month	165.99
Elk Grove	Two Year	156.12
Santa Ana	One Year	155.65
San Ramon	Month-to-Month	155.61
Martinez	Two Year	155.51
Santa Barbara	Two Year	154.94
Clarksburg	One Year	154.49
Alpaugh	One Year	152.28
Templeton	One Year	150.16
King City	Month-to-Month	150.09
Redway	Two Year	149.59
Dobbins	One Year	149.06
Rancho Cordova	One Year	147.72
Acampo	One Year	147.44
Meadow Vista	Month-to-Month	147.27
Butte City	Month-to-Month	145.66
Mojave	Month-to-Month	145.41
Corona	Month-to-Month	145.08
Redlands	Month-to-Month	144.20
La Grange	Month-to-Month	143.45
Burney	Month-to-Month	142.29
Escondido	One Year	142.14
Tollhouse	One Year	141.64

- Query analitik menunjukkan churn paling tinggi pada pelanggan berdasarkan alasannya.



- Dashboard



8. Evaluasi

8.1 Keberhasilan

- Skema star schema bekerja dengan baik untuk OLAP.
- ETL menghasilkan data bersih dan siap analisis.
- Query OLAP dapat dijalankan cepat dengan indexing.

8.2 Kendala

- Tidak ada

9. Rencana Pengembangan ke Depan

- Integrasi model machine learning untuk prediksi churn.
- Ekstraksi data real-time dari sistem operasional (streaming data).

10. Tim Proyek

ANGGOTA TIM	PERAN
Eksanty F Sugma Islamiaty	Koordinasi keseluruhan proyek, pengelolaan tim, dan dokumentasi akhir
Residen Nusantara R M	Implementasi database SQL Server dan optimasi indexing
Aisyah Tiara Pratiwi	Desain konseptual dan logikal skema star schema, serta pembuatan script SQL

Ukasyah Muntaha	Pengolahan data dan pembuatan ETL script menggunakan Python
Rendra Eka Prayoga	Pengujian query analitik, validasi hasil implementasi, dan pembuatan analisis OLAP