**Mata Kuliah Analisis Data Eksploratif**

**Laporan Tugas Membuat Tabel Distribusi Frekuensi di Excell dan SPSS**

**Dosen Pengampu: Wahyu Sri Utami, S.Si., M.Sc.**



**Disusun oleh:**

* **Fadilah Ratu Azzahra (5232811015)**
* **Lathif Ramadhan (5231811022)**

**PROGRAM STUDI SAINS DATAFAKULTAS SAINS DAN TEKNOLOGIUNIVERSITAS TEKNOLOGI YOGYAKARTA YOGYAKARTA**

**2024**

**Pendahuluan**

Dalam era persaingan bisnis yang semakin ketat, pemahaman tentang perilaku pelanggan menjadi kunci untuk mempertahankan keberlanjutan suatu perusahaan. Salah satu aspek penting dalam analisis perilaku pelanggan adalah **churn**, yang merujuk pada fenomena di mana pelanggan memutuskan untuk menghentikan hubungan atau langganan mereka dengan perusahaan. Himpunan data churn pelanggan ini diambil dari platform Kaggle, dan berfungsi untuk memprediksi retensi pelanggan serta membantu perusahaan dalam mengembangkan strategi yang efektif untuk mempertahankan pelanggan yang ada.

Dataset ini terdiri dari **64.374 catatan pelanggan**, di mana setiap catatan mencakup berbagai fitur terkait perilaku pelanggan, termasuk kolom umur. Dengan mengolah kolom umur menjadi tabel distribusi frekuensi, kita dapat memahami sebaran usia pelanggan yang berkontribusi terhadap fenomena churn. Melalui analisis ini, perusahaan dapat mengidentifikasi pola yang mendasari keputusan pelanggan untuk bertahan atau meninggalkan layanan.

Churn pelanggan merupakan metrik yang sangat penting bagi bisnis karena dampaknya yang langsung terhadap pendapatan dan pertumbuhan. Dengan menganalisis faktor-faktor yang menyebabkan churn, perusahaan dapat mengembangkan strategi untuk meningkatkan kepuasan pelanggan dan mengurangi tingkat kehilangan pelanggan. Selain itu, teknik pemodelan prediktif dapat diterapkan untuk memproyeksikan dan mengatasi potensi churn, yang memungkinkan perusahaan untuk mengambil langkah-langkah proaktif dalam mempertahankan pelanggan yang berisiko.

Dalam laporan ini, kami akan mengeksplorasi dataset Churn Pelanggan dengan fokus pada kolom umur, serta menghitung tabel distribusi frekuensi yang memberikan wawasan berharga mengenai karakteristik demografis pelanggan. Dengan pemahaman yang lebih baik tentang umur pelanggan, perusahaan dapat merumuskan pendekatan yang lebih tepat untuk meningkatkan retensi pelanggan dan meminimalkan churn.

**Tentang Dataset**

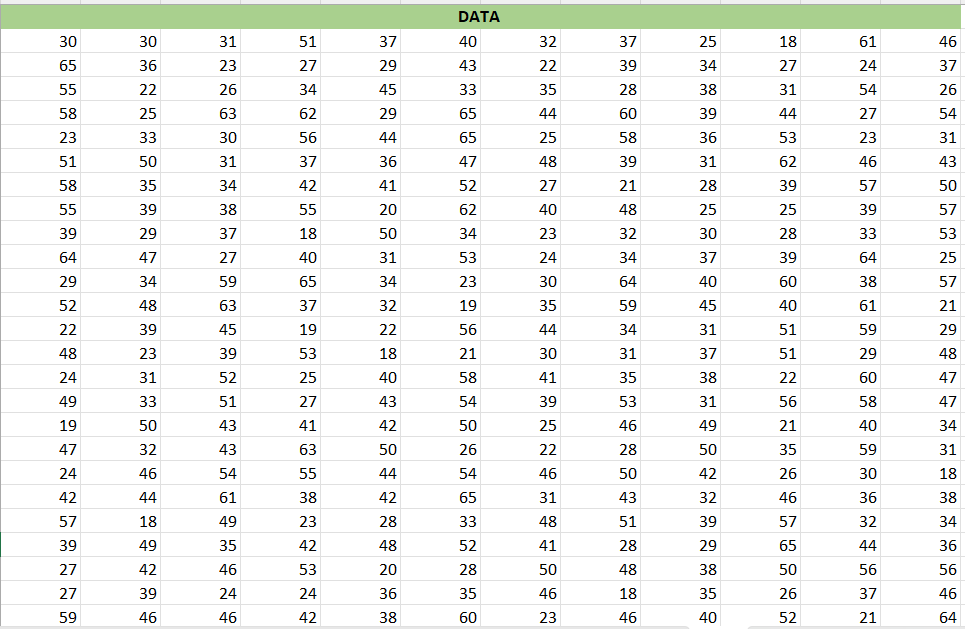
Dataset yang kami gunakan ini berasal dari situs/web Kaggle (link: <https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset>) yang terdiri dari **64.374 catatan pelanggan** dan dengan total **440.832 baris** dan **12 kolom** yang terdiri kolom *CustomerID, Age, Gender, Tenure, UsageFrequency, SupportCalls, PaymentDelay, SubcriptionType, ContractLength, TotalSpend, LastInteraction*, dan *Churn*.

Kami memilih kolom umur sebagai fokus analisis karena umur adalah salah satu faktor demografis yang paling berpengaruh terhadap perilaku pelanggan. Usia bisa mempengaruhi preferensi, kebutuhan, dan cara pelanggan berinteraksi dengan layanan yang ditawarkan. Misalnya, pelanggan yang lebih muda mungkin memiliki harapan yang berbeda dibandingkan pelanggan yang lebih tua, baik dalam hal produk maupun layanan yang mereka inginkan.

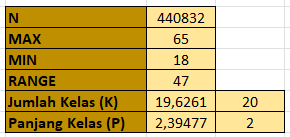
Dengan memahami distribusi umur pelanggan yang churn, kita bisa mengidentifikasi segmen-segmen tertentu yang mungkin lebih rentan terhadap kehilangan. Ini memberi kami kesempatan untuk merumuskan strategi retensi yang lebih efektif, seperti penawaran khusus atau layanan yang lebih sesuai dengan kebutuhan setiap kelompok usia. Jadi, dengan mempelajari kolom umur, kita berharap dapat menciptakan pendekatan yang lebih personal dan relevan untuk menjaga pelanggan tetap setia!

1. **Excell**

**DATA YANG DI GUNAKAN**

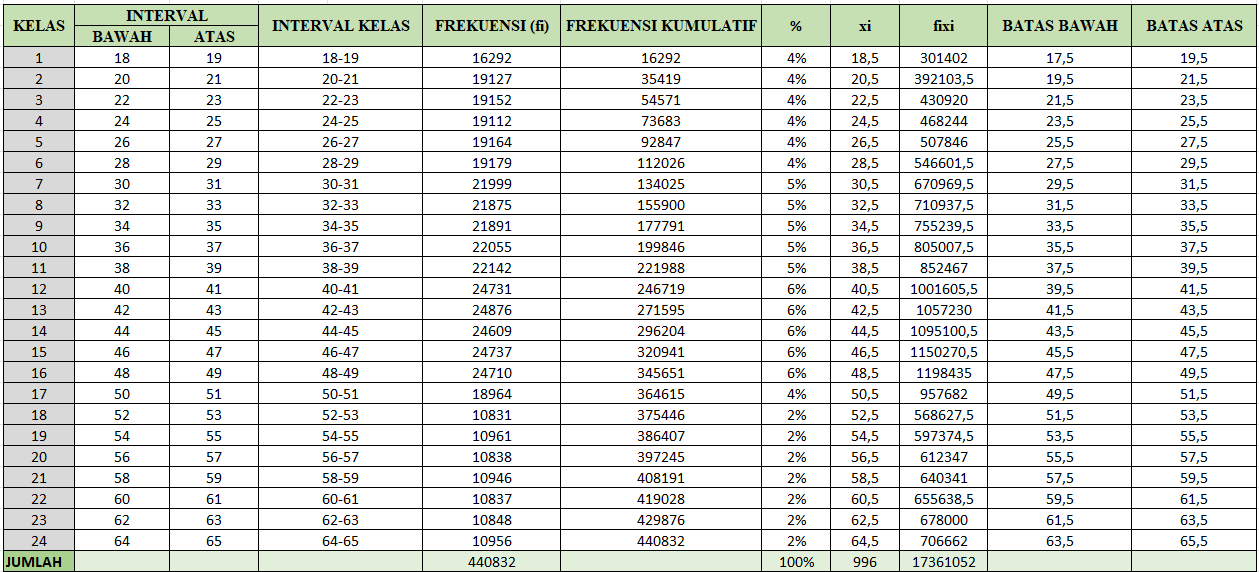
****

**Data Yang Dapat Diketahui Dari Data Di Atas**



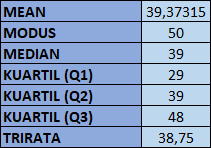
1. Jumlah Sampel (N) = 440.832
2. Nilai Terbesar (MAX) = 65
3. Nilai Terendah (MIN) = 18
4. Rentang Nilai (Range) = 47
5. Jumlah Kelas (K) = 19,6261. Tetapi karena jumlah kelas bisa ditambah atau di buat sesuai dengan yang kita inginkan maka dari itu jumlah kelas dibuat menjadi 24 kelas agar sesuai dengan nilai MAX yaitu 65.
6. Panjang Kelas (P) = 2,39477

**Tabel Distribusi Frekuensi**



Gambar ini menampilkan tabel distribusi frekuensi yang dihitung menggunakan Microsoft Excel. Tabel ini mengelompokkan umur pelanggan ke dalam 24 kelas, dengan rentang dari 18 hingga 65 tahun, memberikan gambaran umum tentang sebaran usia pelanggan.

**Hasil Yang Bisa Di Dapatkan Dari Tabel Frekuensi Yaitu :**

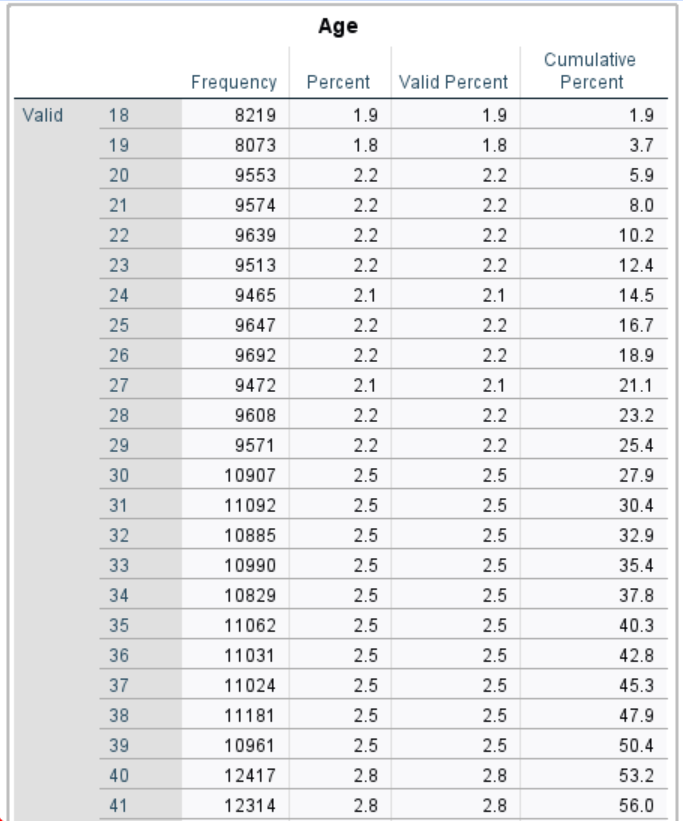


**B. SPSS**

A screenshot of a computer screen

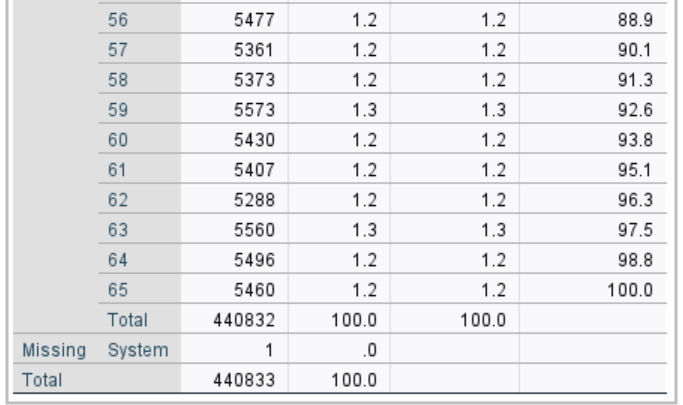
Description automatically generated

Output ini menunjukkan hasil analisis deskriptif dari dataset menggunakan SPSS, termasuk statistik penting seperti nilai minimum, maksimum, rata-rata, dan standar deviasi dari variabel umur.



A screenshot of a table

Description automatically generated



Grafik yang dihasilkan dari SPSS ini menggambarkan distribusi usia pelanggan dalam bentuk histogram. Hal ini membantu visualisasi pola sebaran usia pelanggan dan memberikan insight tentang segmen usia yang lebih berisiko mengalami churn.

**Kesimpulan**  
Dalam analisis churn pelanggan berdasarkan dataset yang diambil dari Kaggle, kolom umur menjadi fokus utama karena usia merupakan faktor demografis yang berpengaruh pada perilaku pelanggan. Dengan menggunakan tabel distribusi frekuensi, kita dapat mengidentifikasi sebaran usia pelanggan yang cenderung churn. Hasil analisis ini memberikan wawasan penting bagi perusahaan untuk mengembangkan strategi retensi yang lebih tepat dan personal, seperti penawaran khusus untuk segmen usia tertentu.