

Technical Test
Junior Data Analyst



naufal.nashif



Naufal Nashif Imanuddin

Certified Python Associate | Statistics and Data Science
Graduate | Data Enthusiast

naufal.nashif

Task 1

■ Task 1

Task 1

■ Predictive Feature Analysis For Missing Data

Description :

Identify patterns in missing values across the dataset by writing an SQL query.

Specifically:

- Find columns with more than 20% missing values.
- Calculate and return the following metrics for each such column: Table name, Column name, Total rows in the table, Number of missing values, Percentage of missing values
- Order the results by the percentage of missing values in descending order.

[Github SQL](#)

Task 1

Predictive Feature Analysis For Missing Data

```
/* TASK 1 : Predictive Feature Analysis for Missing Data */

-- Membuat tabel customer
CREATE TABLE public.customer (
  id INT PRIMARY KEY,
  name VARCHAR(255),
  email VARCHAR(255),
  phone VARCHAR(255),
  address VARCHAR(255)
);
● INSERT INTO public.customer (id, name, email, phone, address) VALUES
(1, 'Rizaldy Uto', 'uto@example.com', '1234567890', NULL),
(2, 'Caemila', NULL, NULL, NULL),
(3, NULL, 'caem@example.com', '9876543210', 'Elm Street'),
(4, NULL, NULL, NULL, NULL),
(5, 'Johan Chris', 'jo@example.com', '5555555555', 'Pine Street');

-- Membuat tabel orders
CREATE TABLE public.orders (
  id INT PRIMARY KEY,
  customer_id INT,
  order_date DATE,
  delivery_date DATE,
  tracking_number VARCHAR(128),
  FOREIGN KEY (customer_id) REFERENCES customer(id)
);
● INSERT INTO public.orders (id, customer_id, order_date, delivery_date, tracking_number) VALUES
(1, 1, '2025-01-01', '2025-01-03', '123-ABC'),
(2, 1, '2025-01-02', NULL, '456-DEF'),
(3, 2, NULL, NULL, NULL),
(4, 3, '2025-01-03', '2025-01-04', NULL),
(5, 4, NULL, NULL, NULL);
```

```
WITH cte_main AS (
  SELECT
    'customer' AS table_name,
    'name' AS column_name,
    COUNT(*) AS total_rows,
    COUNT(*) FILTER (WHERE name IS NULL) AS missing_value
  FROM public.customer
  UNION ALL
  SELECT
    'customer',
    'address',
    COUNT(*),
    COUNT(*) FILTER (WHERE address IS NULL)
  FROM public.customer
  UNION ALL
  SELECT
    'customer',
    'phone',
    COUNT(*),
    COUNT(*) FILTER (WHERE phone IS NULL)
  FROM public.customer
  UNION ALL
  SELECT
    'orders',
    'order_date',
    COUNT(*),
    COUNT(*) FILTER (WHERE order_date IS NULL)
  FROM public.orders
  UNION ALL
  SELECT
    'orders',
    'delivery_date',
    COUNT(*),
    COUNT(*) FILTER (WHERE delivery_date IS NULL)
  FROM public.orders
  UNION ALL
  SELECT
    'orders',
    'tracking_number',
    COUNT(*),
    COUNT(*) FILTER (WHERE tracking_number IS NULL)
  FROM public.orders
)
SELECT
  table_name,
  column_name,
  total_rows,
  missing_value,
  ROUND(missing_value::DECIMAL / total_rows * 100, 2) || '%' AS missing_percentage
FROM cte_main
ORDER BY missing_percentage DESC;
```

Task 2

■ Task 2

Task 2

■ Dealer Performances and Product Analysis

Scenario :

You are working as a data analyst for a company that tracks dealer performance and product sales. The management has requested a comprehensive report to gain insights into dealer performance, product profitability, and forecast accuracy. This report must include detailed segmentation and ranking based on various metrics.

[Github SQL](#)

Task 2

■ Dealer Performances and Product Analysis

```
-- 1. Identify the top-performing and least-performing dealers in terms of total sales. Include dealer ID, dealer name, total sales, and their rank.
with cte as (
  select
    dealer_id ,
    dealer_name ,
    SUM(sale_amount) as total_sales
  from public.dealers d
  left join public.sales s using (dealer_id)
  left join public.products p using (product_id)
  group by dealer_id , dealer_name
),
cte_rank as (
  select
    *,
    rank() over (order by total_sales DESC) as dealer_rank
  from cte
)
select
  dealer_id,
  dealer_name,
  total_sales,
  dealer_rank
from cte_rank
where dealer_rank = 1 or dealer_rank = (
  select
    MAX(dealer_rank)
  from cte_rank
);
```

Output :

123 dealer_id	A-Z dealer_name	123 total_sales	123 dealer_rank
25	Boyd, Rivas and Fisher	4,522	1
77	Sims-Butler	899	100

1. Identifikasi Dealer Terbaik dan Terburuk Berdasarkan Total Penjualan

Dealer terbaik dan terburuk dalam hal total penjualan telah diidentifikasi dengan menggunakan total penjualan yang dihitung dari data penjualan yang ada.

Dealer_id = 25 menunjukkan performa penjualan terbaik dengan total penjualan mencapai 4.522, sementara **Dealer_id = 77** berada di peringkat terbawah dengan hanya 899.

Task 2

■ Dealer Performances and Product Analysis

```
-- 2. Group dealers into sales deciles based on their total sales and identify the decile for each dealer.
with cte as (
  select
    dealer_id ,
    dealer_name ,
    SUM(sale_amount) as total_sales
  from public.dealers d
  left join public.sales s using (dealer_id)
  left join public.products p using (product_id)
  group by dealer_id , dealer_name
)
select
  *,
  ntile (10) over (order by total_sales desc) as decile
from cte;
```

Output :

123 dealer_id	A-Z dealer_name	123 total_sales	123 decile
25	Boyd, Rivas and Fisher	4,522	1
51	Baker-Davis	4,477	1
28	Robles Inc	4,351	1
84	Smith, Martin and Vargas	4,158	1
81	Walker Inc	3,998	1
91	Mueller Inc	3,980	1
57	Johnson, Chan and Yang	3,904	1
9	Miles Ltd	3,882	1

2. Pengelompokan Dealer ke dalam Decile Berdasarkan Total Penjualan

Dealer dikelompokkan dalam 10 kelompok (deciles) berdasarkan total penjualan mereka, yang memberikan gambaran distribusi performa penjualan

Dealer dengan penjualan tertinggi masuk ke decile 1, sementara dealer dengan penjualan terendah masuk ke decile 10.

■ Dealer Performances and Product Analysis

```
-- 3. Segment products into quartiles based on total sales and list each product's quartile.
with cte as (
  select
    product_id,
    product_name,
    SUM(sale_amount) as total_sales
  from public.dealers d
  left join public.sales s using (dealer_id)
  left join public.products p using (product_id)
  group by product_id , product_name
)
select
  product_id,
  product_name,
  total_sales,
  ntile (4) over (order by total_sales desc) as quartiles
from cte ;
```

Output :

product_id	product_name	total_sales	quartiles
40	e-enable world-class partnerships	7,010	1
37	enable B2C bandwidth	6,939	1
19	e-enable value-added experiences	6,915	1
4	aggregate seamless synergies	6,797	1
50	whiteboard one-to-one supply-chains	6,618	1
21	integrate killer users	6,572	1
26	integrate extensible markets	6,417	1

3. Segmentasi Produk ke dalam Kuartil Berdasarkan Total Penjualan

Produk dikelompokkan dalam 4 kuartil berdasarkan total penjualan. Produk dengan penjualan tertinggi berada di kuartil pertama, dan produk dengan penjualan terendah berada di kuartil keempat.

Task 2

■ Dealer Performances and Product Analysis

```
-- 4. Rank dealers based on their income percentile and list the percentile rank for each dealer.
SELECT
  dealer_id,
  dealer_name,
  dealer_income,
  percent_rank() OVER (ORDER BY dealer_income DESC) AS percentile
FROM public.dealers
```

Output :

123 dealer_id	A-Z dealer_name	123 dealer_income	123 percentile
34	Dennis Inc	499,210	0
85	Valentine-Martin	491,505	0.0101010101
6	Mcpherson, Fischer and Ramirez	490,746	0.0202020202
27	Richards-Armstrong	488,432	0.0303030303
1	Jenkins-Mcintosh	487,422	0.0404040404
45	Miranda, Brown and Cardenas	482,333	0.0505050505
68	Burns-Powers	480,775	0.0606060606

4. Peringkat Dealer Berdasarkan Persentil Penghasilan

Dealer peringkat berdasarkan penghasilan mereka menggunakan persentil. Dealer dengan penghasilan terendah berada di persentil 100.

Task 2

■ Dealer Performances and Product Analysis

```
-- 5. Categorize dealers into 5 age groups based on their age and display the group for each dealer.
select
  dealer_id,
  dealer_name,
  dealer_age,
  ntile(5) over (order by dealer_age DESC) as age_group
from public.dealers d
```

5. Kategorisasi Dealer ke dalam 5 Kelompok Berdasarkan Usia

Dealer dikelompokkan dalam 5 kelompok usia untuk memberikan gambaran mengenai usia mereka.

Output :

123 dealer_id	A-Z dealer_name	123 dealer_age	123 age_group
3	Bishop and Sons	29	1
18	Morris-Mccarthy	29	1
80	Williams, Cooper and Hoover	28	1
29	Navarro PLC	28	1
52	Crane, Baker and Bell	28	1
89	Foster-Cruz	28	1

Task 2

■ Dealer Performances and Product Analysis

```
-- 5. Categorize dealers into 5 age groups based on their age and display the group for each dealer.
select
  dealer_id,
  dealer_name,
  dealer_age,
  ntile(5) over (order by dealer_age DESC) as age_group
from public.dealers d
```

5. Kategorisasi Dealer ke dalam 5 Kelompok Berdasarkan Usia

Dealer dikelompokkan dalam 5 kelompok usia untuk memberikan gambaran mengenai usia mereka.

Output :

123 dealer_id	A-Z dealer_name	123 dealer_age	123 age_group
3	Bishop and Sons	29	1
18	Morris-Mccarthy	29	1
80	Williams, Cooper and Hoover	28	1
29	Navarro PLC	28	1
52	Crane, Baker and Bell	28	1
89	Foster-Cruz	28	1

Task 2

■ Dealer Performances and Product Analysis

```
-- 6. Rank products by sales performance within each dealer and identify the top-performing product for each dealer.
with cte as (
  select
    dealer_id,
    dealer_name,
    product_id,
    product_name,
    SUM(sale_amount) as total_sales
  from public.dealers d
  left join public.sales s using (dealer_id)
  left join public.products p using (product_id)
  group by dealer_id, dealer_name, product_id , product_name
),
cte_2 as (
  select
    *,
    rank() over (partition by dealer_name order by total_sales DESC) as rank_dealer_product
  from cte
)
select *
from cte_2
where rank_dealer_product = 1
```

6. Peringkat Produk Berdasarkan Performa Penjualan di Setiap Dealer

Setiap dealer memiliki produk yang berbeda, dan performa penjualan produk diurutkan untuk mengetahui produk mana yang terlaris dalam masing-masing dealer.

Output :

123 dealer_id	AZ dealer_name	123 product_id	AZ product_name	123 total_sales	123 rank_dealer_product
64	Allen Group	41	disintermediate end-to-end networks	463	1
82	Armstrong-Harris	17	target 24/365 solutions	638	1
51	Baker-Davis	27	orchestrate distributed platforms	852	1
65	Baker-Jones	26	integrate extensible markets	494	1
3	Bishop and Sons	8	cultivate real-time e-tailers	481	1
25	Boyd, Rivas and Fisher	43	enable next-generation solutions	814	1

Task 2

■ Dealer Performances and Product Analysis

```
-- 7. Segment monthly sales into deciles and identify the decile for each month.
WITH cte AS (
  SELECT
    EXTRACT(MONTH FROM TO_DATE(sale_date, 'YYYY-MM-DD')) AS month_date,
    SUM(sale_amount) AS total_sales
  FROM public.sales s
  GROUP BY EXTRACT(MONTH FROM TO_DATE(sale_date, 'YYYY-MM-DD'))
)
SELECT
  month_date,
  total_sales,
  NTILE(10) OVER (ORDER BY total_sales DESC) AS decile
FROM cte
ORDER BY month_date;
```

7. Segmentasi Penjualan Bulanan ke dalam Decile
Penjualan bulanan dealer dikelompokkan ke dalam decile (10 kelompok)

Output :

123 month_date	123 total_sales	123 decile
1	22,761	2
2	21,866	3
3	19,043	6
4	17,377	10
5	19,501	5
6	22,805	2
7	18,816	7

Task 2

■ Dealer Performances and Product Analysis

```
-- 8. Analyze subscription types and sales performance to categorize dealers into 5 groups.
with cte as (
  select
    dealer_id,
    dealer_name,
    subscription_service,
    SUM(sale_amount) as total_sales
  from public.dealers d
  left join public.sales s using (dealer_id)
  left join public.products p using (product_id)
  group by dealer_id, dealer_name, subscription_service
)
select
  *,
  ntile(5) over (order by total_sales DESC) as five_group
from cte
order by five_group;
```

Output :

dealer_id	dealer_name	subscription_service	total_sales	five_group
25	Boyd, Rivas and Fisher	Premium	4,522	1
51	Baker-Davis	Premium	4,477	1
28	Robles Inc	None	4,351	1
84	Smith, Martin and Vargas	None	4,158	1
81	Walker Inc	Premium	3,998	1
91	Mueller Inc	None	3,980	1
57	Johnson, Chan and Yang	Premium	3,904	1

8. Analisis Jenis Langganan dan Performa Penjualan Dealer

Dealer dikelompokkan berdasarkan jenis langganan mereka dan performa penjualan, untuk menentukan korelasi antara keduanya.

Task 2

Dealer Performances and Product Analysis

```
-- 9. Evaluate forecast accuracy for each dealer by calculating the percentage difference between forecasted and actual sales
-- forecast accuracy.

WITH cte AS (
    SELECT
        dealer_id,
        dealer_name,
        SUM(sale_amount) AS actual_sales,
        SUM(forecast_amount) AS forecasted_sales
    FROM public.dealers d
    LEFT JOIN public.sales s USING (dealer_id)
    GROUP BY dealer_id, dealer_name
),
cte_2 AS (
    SELECT
        dealer_id,
        dealer_name,
        actual_sales,
        forecasted_sales,
        ROUND(AVG(ABS(forecasted_sales - actual_sales)::decimal / NULLIF(actual_sales, 0) * 100), 2) AS forecast_error
    FROM cte
    GROUP BY dealer_id, dealer_name, actual_sales, forecasted_sales
)
select
    *,
    (100 - forecast_error) as forecast_accuracy,
    ntile(5) over (order by (100 - forecast_error) DESC) as accuracy_group
from cte_2;
```

Output :

123	A: dealer_n:	123 actu	123 forecasted_sales	123 forecast_error	123 forecast_accuracy	123 accuracy_group
28	Robles Inc	4,351	4,372	0.48	99.52	1
64	Allen Group	1,175	1,166	0.77	99.23	1
5	Deleon-Rogers	3,328	3,384	1.68	98.32	1
21	Harris-Graham	2,719	2,659	2.21	97.79	1
56	Sanders-Lewis	1,808	1,766	2.32	97.68	1
95	Ellis-Rogers	2,261	2,323	2.74	97.26	1
85	Valentine-Martir	2,371	2,304	2.83	97.17	1

9. Evaluasi Akurasi Perkiraan Penjualan untuk Setiap Dealer

Akurasi perkiraan penjualan dihitung dengan menghitung selisih antara perkiraan dan penjualan aktual.

Task 2

■ Dealer Performances and Product Analysis

```
-- 10. Rank products into deciles based on their profit margin.
select
  *,
  rank() over (order by product_margin DESC) as rank_product
from public.products p
```

10. Peringkat Produk Berdasarkan Margin Laba
Produk di peringkat berdasarkan margin laba mereka untuk mengidentifikasi produk yang paling menguntungkan.

Output :

123 product_id	A-Z product_name	123 product_margin	123 rank_product
36	enable next-generation content	49.28341	1
21	integrate killer users	47.86978	2
45	integrate user-centric content	46.82205	3
35	engage cutting-edge web services	42.82039	4
38	generate intuitive deliverables	41.076965	5
37	enable B2C bandwidth	40.766224	6
39	scale bricks-and-clicks schemas	40.669758	7

Task 3

■ Task 3

Task 3

■ Telco Company Customer Behavior Analysis

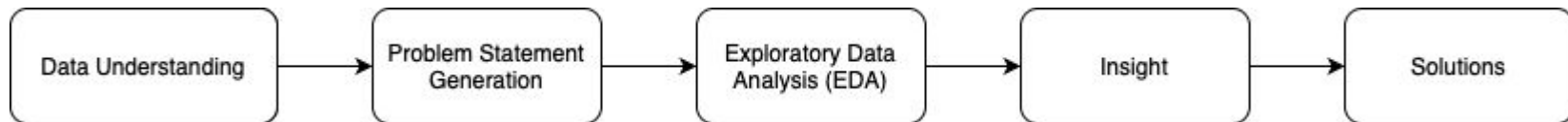
Problem Statement :

A telecommunication company wants to analyze its customer's behavior. Given the following dataset, build a comprehensive analysis that provide meaningful insights to the telecommunication company.

[Tableau Dashboard](#)

Stop

■ Step

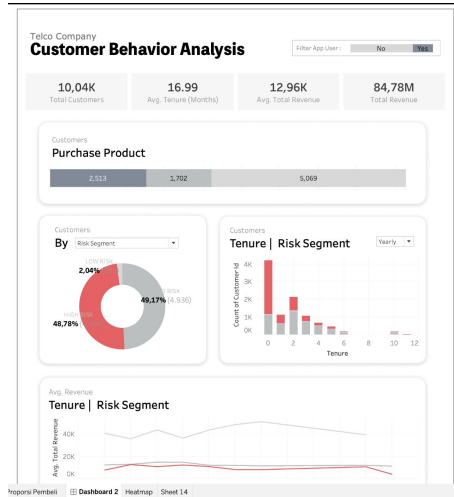


■ Problem Statement Generation

Problem Statement

How can the company leverage insights about customer segmentation, product purchasing behavior, and revenue trends to improve customer retention, maximize revenue, and tailor their offerings effectively?

EDA With Tableau Dashboard

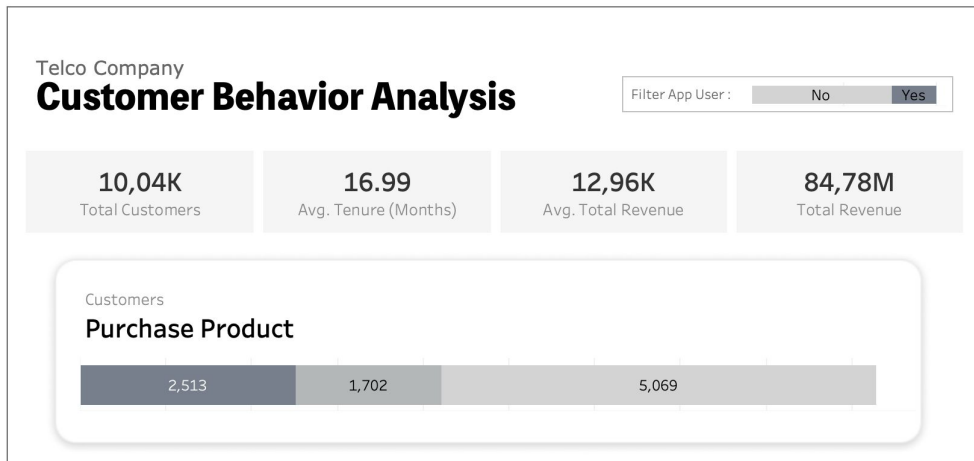


"I used Tableau for EDA, focusing on KPI indicators, pie charts, bar charts, line charts, and detailed tables to analyze customer behavior, segment distributions, and trends effectively."

[Tableau Dashboard](#)

Insight

Insight



KPI Metrics:

- Total Customers: **10K**
- Average Tenure: **16.99 months**
- Average Total Revenue: **12.96K**
- Total Revenue: **84.78M**

Product Analysis:

- **Product C** is the most purchased, followed by Product A and Product B.

Insight

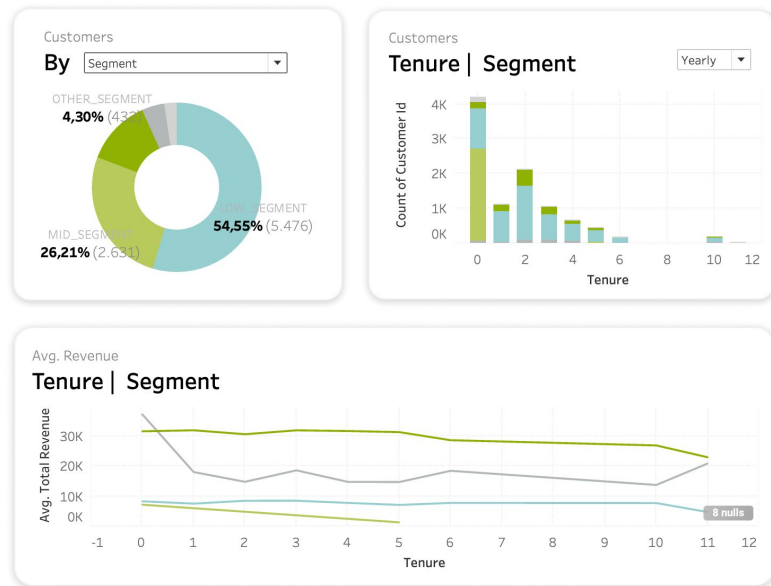
Insight



Customer by Risk Segment:

- Most customers are in the **MID RISK** category, followed by HIGH RISK and LOW RISK.
- **LOW RISK** customers generate the highest average revenue by tenure, followed by MID RISK and HIGH RISK.

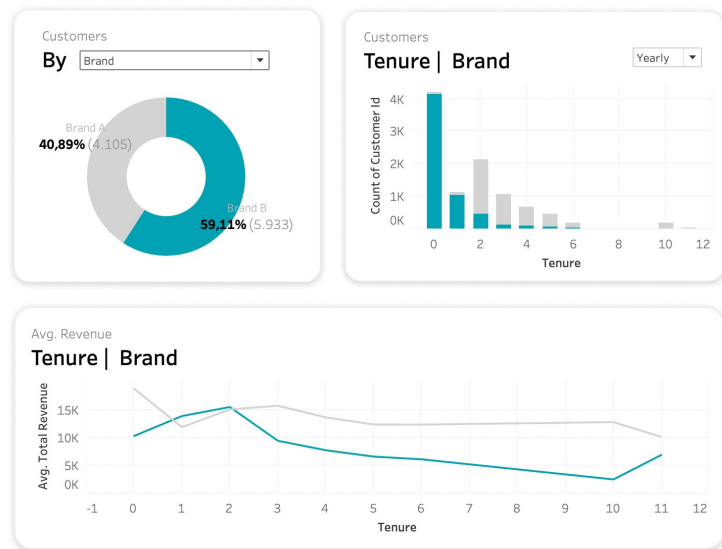
Insight



Customer by Segment:

- Most customers belong to the **LOW_SEGMENT**, followed by MID SEGMENT, HIGH SEGMENT, and OTHER SEGMENT.
- Customers in the **MID SEGMENT** have a tenure of less than one year.
- Average revenue is highest in the **HIGH SEGMENT**, followed by OTHER SEGMENT, LOW SEGMENT, and MID SEGMENT.

Insight

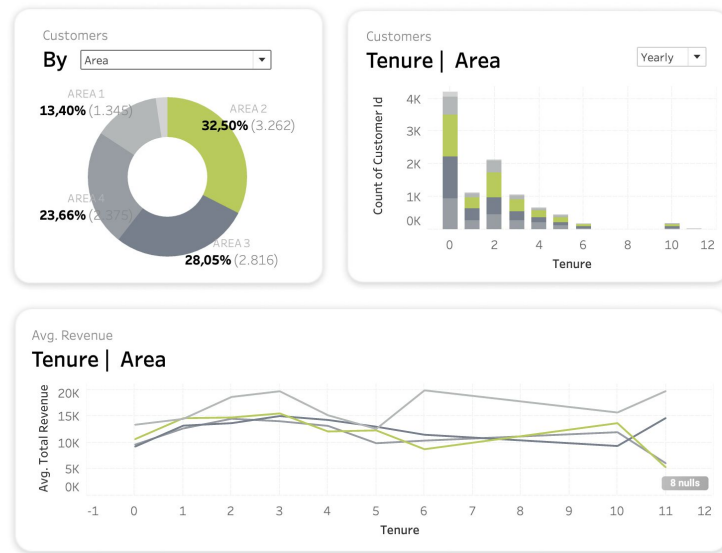


Customer by Brand:

- Most customers use **Brand B**, but **Brand A** customers show higher loyalty with longer tenure.
- **Brand A** also generates higher average revenue by tenure compared to Brand B.

Insight

Insight

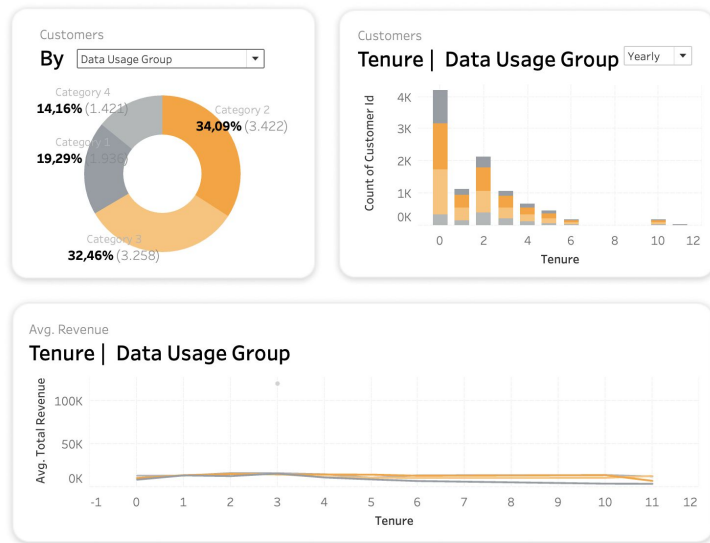


Customer by Area:

- **Area 2** has the highest total number of customers.
- **Area 1** generates the highest average revenue.

Insight

Insight



Customer by Data Usage Group:

- **Category 2** has the most customers among the data usage groups.

Solutions

■ Solutions

Solutions for the Company

1. **Product Strategy:**

- Promote **Product C** aggressively and bundle it with **Product A** or **B** to boost sales.
- Investigate and revamp marketing for **Product B** to enhance its appeal.

2. **Risk Management:**

- Retain **LOW RISK** customers with loyalty programs.
- Engage **MID RISK** and **HIGH RISK** customers with targeted retention campaigns.

3. **Segment Strategy:**

- Extend **MID SEGMENT** customer tenure with onboarding campaigns.
- Sustain **HIGH SEGMENT** revenue by offering exclusive benefits.

4. **Brand Strategy:**

- Position **Brand A** as premium to capitalize on loyalty.
- Improve **Brand B** offerings to increase retention and revenue.

5. **Regional Focus:**

- Focus on growth in **Area 2** and high-value campaigns in **Area 1**.

6. **Data Usage Optimization:**

- Tailor data plans for **Category 2** users and encourage higher usage in other categories.

Terima kasih

■ Terima kasih