



Databricks AI Challenge

Day 09

SQL Analytics & Dashboards

databricks
14 DAYS
AI CHALLENGE



A step-by-step guide to transforming raw data into actionable insights.

#DatabricksWithIDC

The Workflow: From Warehouse to Widget



Ingest & Analyze: The Gold Table

Before analysis, we validate the schema using the `display()` function. We are working with pre-cleaned 'Gold' level data containing high-value commerce metrics.

```
df = spark.read.format('delta').load('/volumes/workspace/ecommerce/ecommerce_data/gold_products')  
display(df.limit(10))
```

product_id int	views long	purchases long	revenue double	conversion_rate double
8500290	1	1	10602.30	100

Key Analysis: The Conversion Funnel

$$\text{Conversion Rate} = \frac{\text{Total Purchases}}{\text{Total Views}} \times 100$$

```
SELECT product_id,  
       SUM/views) as views,  
       SUM/purchases) as purchases,  
       ROUND(SUM/purchases)*100.0/SUM/views, 2) as conversion_rate ←  
FROM gold_products  
GROUP BY product_id  
LIMIT 10;
```

Normalizes
performance across
traffic levels

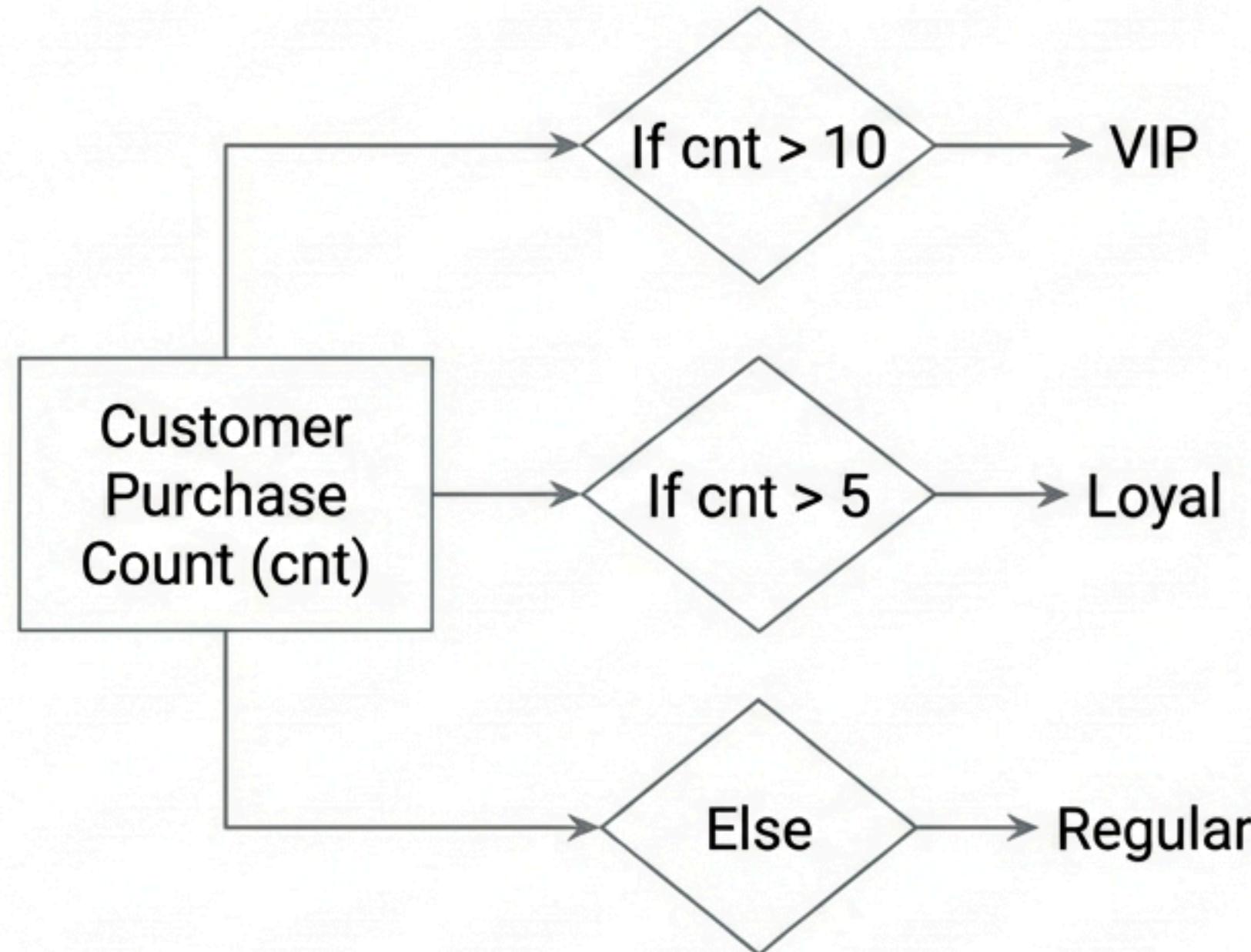
Key Analysis: Smoothing Trends

Using Window Functions to reduce daily volatility.

```
WITH daily AS (
    SELECT product_id, SUM(revenue) as rev
    FROM gold_products GROUP BY product_id
)
SELECT product_id, rev,
    AVG(rev) OVER (ORDER BY product_id ROWS BETWEEN 6 PRECEDING AND
CURRENT ROW) as ma7
FROM daily
LIMIT 10;
```

7-Day Moving Average: Looks at current row + previous 6

Key Analysis: Customer Segmentation



```
CASE WHEN cnt > 10 THEN 'VIP'  
      WHEN cnt > 5 THEN 'Loyal'  
      ELSE 'Regular' END as tier,  
COUNT(*) as customers,  
AVG(total_spent) as avg_ltv
```

Key Analysis: Identifying Top Performers

Aggregating revenue to find the 'Critical Few' products driving the business.

```
SELECT product_id, SUM(revenue) AS total_revenue  
FROM gold_products  
GROUP BY product_id  
ORDER BY total_revenue DESC  
LIMIT 10;
```

product_id	total_revenue
1005115	20625574.32
1005105	11445354.69
1005135	7086522.13
1004249	6815294.62
1004767	5574025.06
1005116	5186187.77
1004856	4118976.32

The Dashboard Canvas

1. Create

Click 'New' -> 'Dashboard'.

2. Visualize

Use the 'Canvas' tab to add charts.

3. Interact

Add filters for dynamic slicing.

4. Arrange

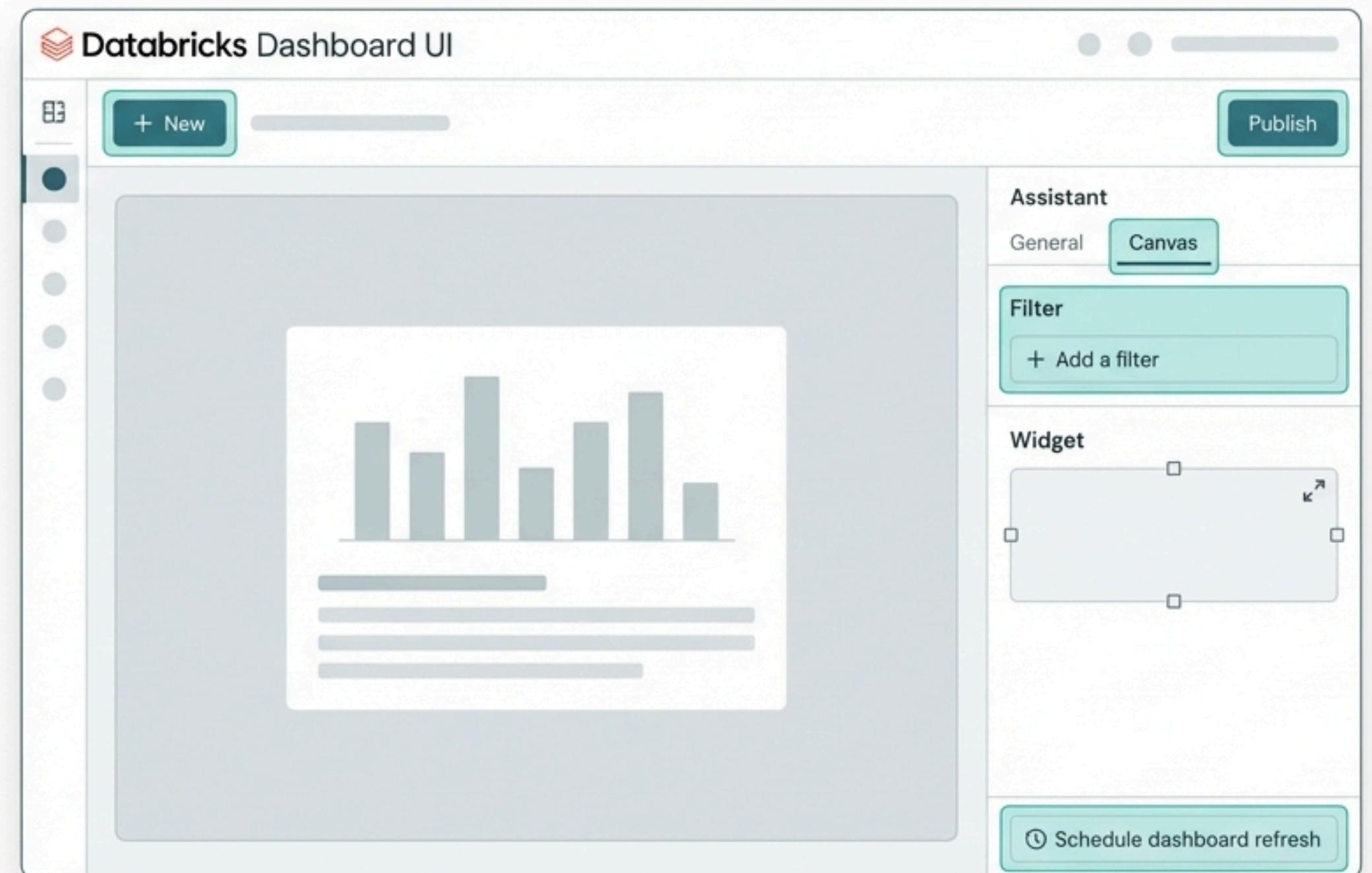
Drag and drop to resize widgets.

5. Publish

Share with stakeholders.

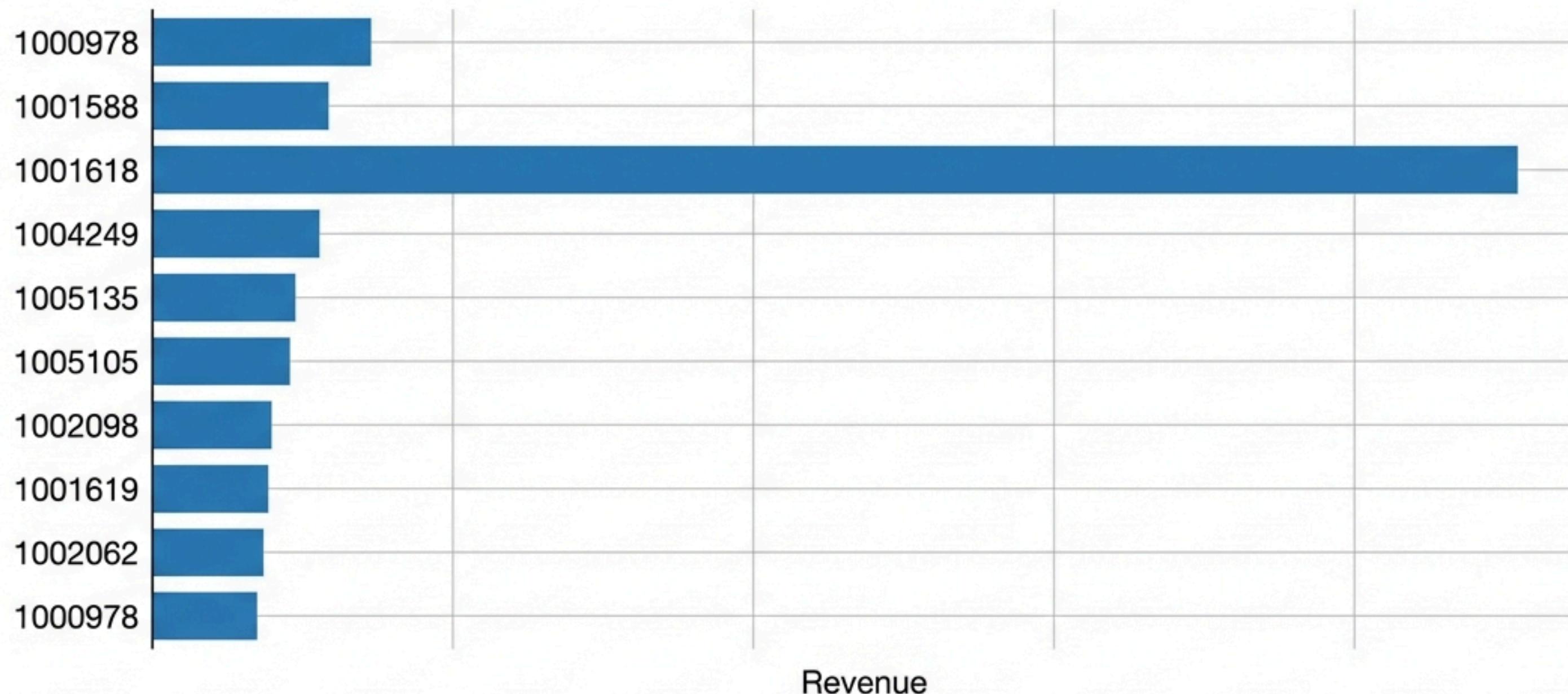
6. Refresh

Schedule automated updates.



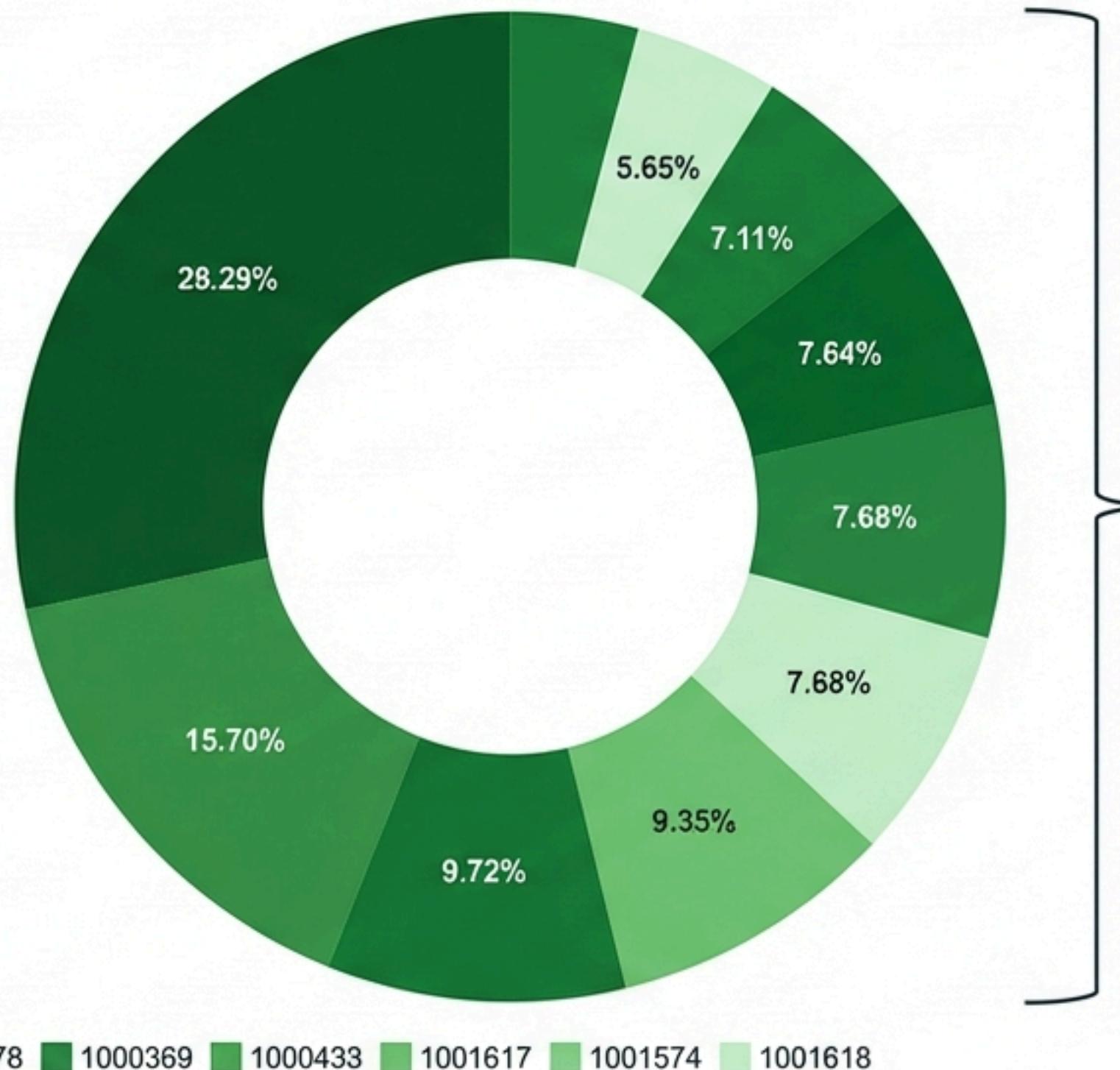
Visualizing Product Performance

Revenue by Product (Top 10)



Visualizing Composition

Revenue Share per Product



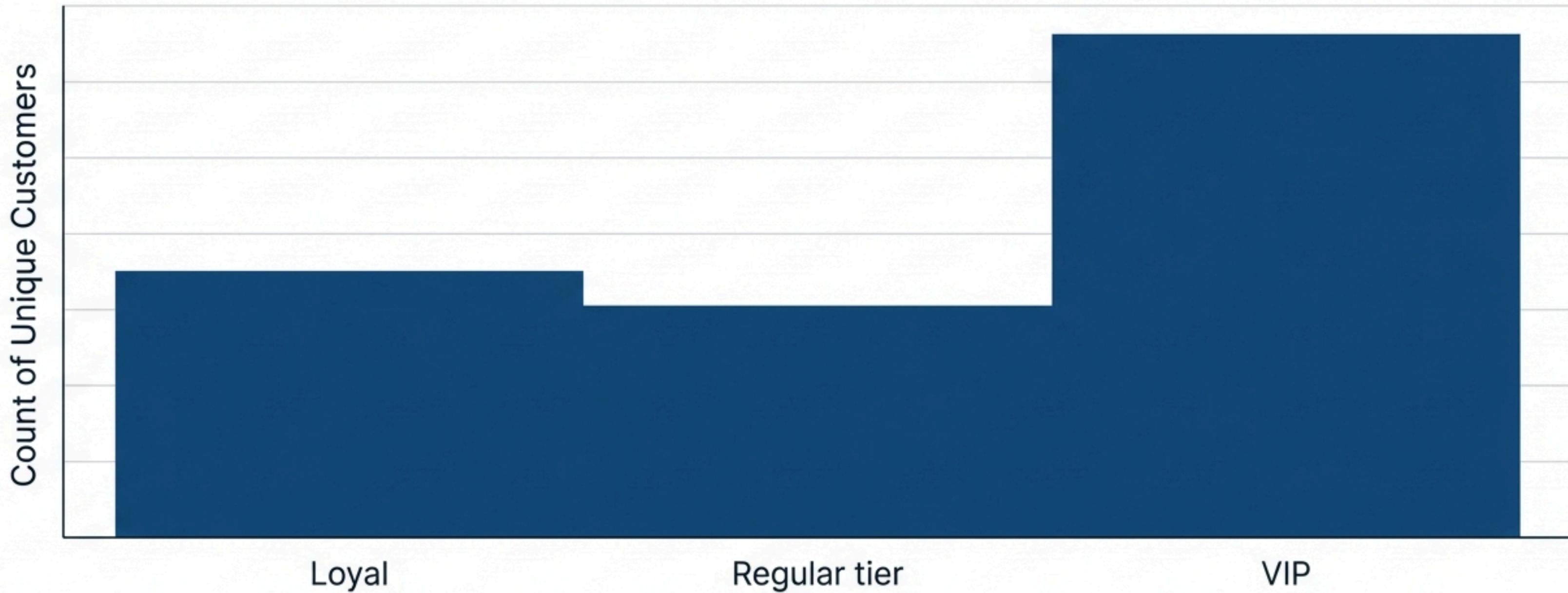
44% of total revenue
from top 2 products.

product_id

■ 1001618 ■ 1000978 ■ 1000978 ■ 1000369 ■ 1000433 ■ 1001617 ■ 1001574 ■ 1001618

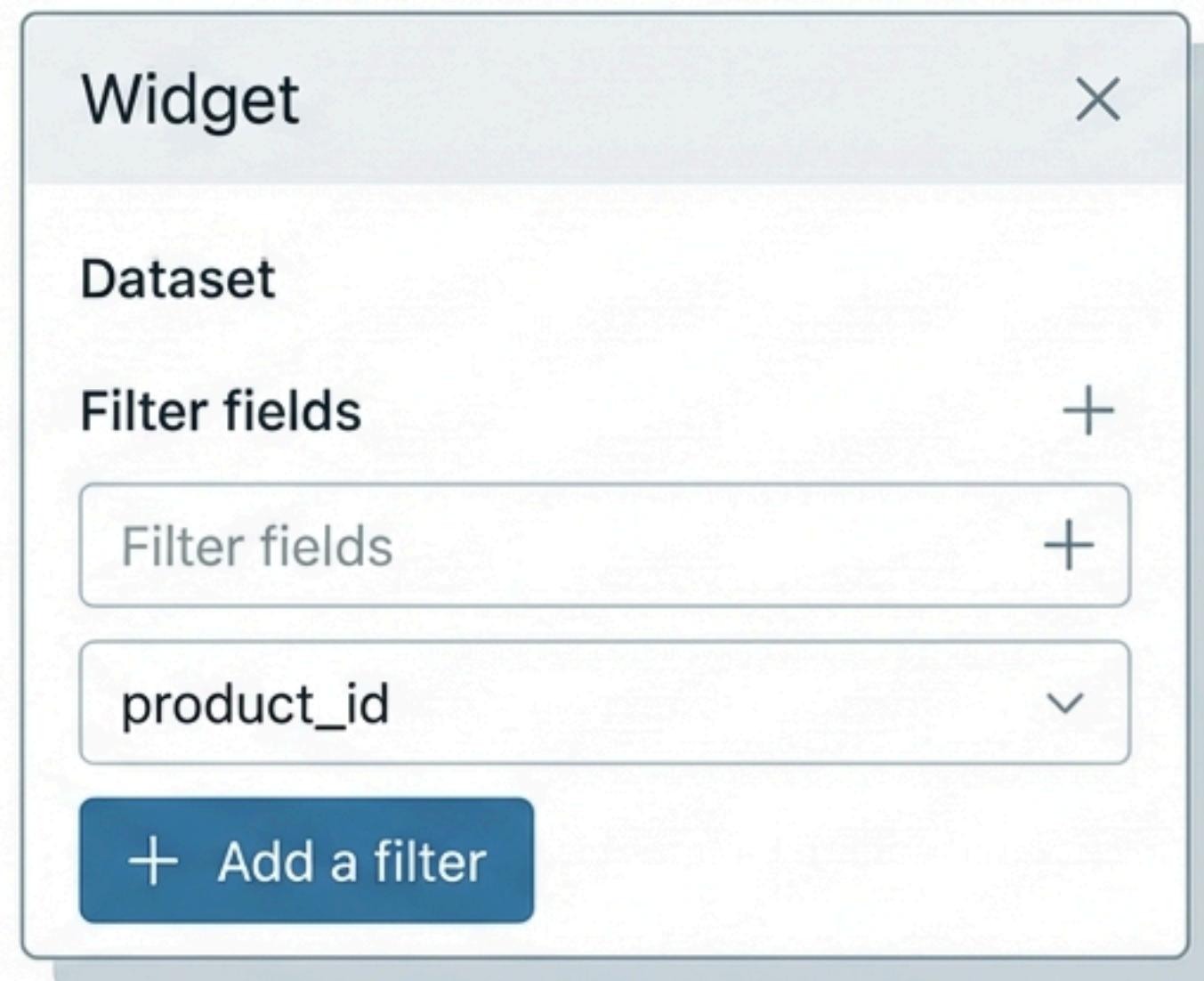
Visualizing Segments

Unique Customers by Tier

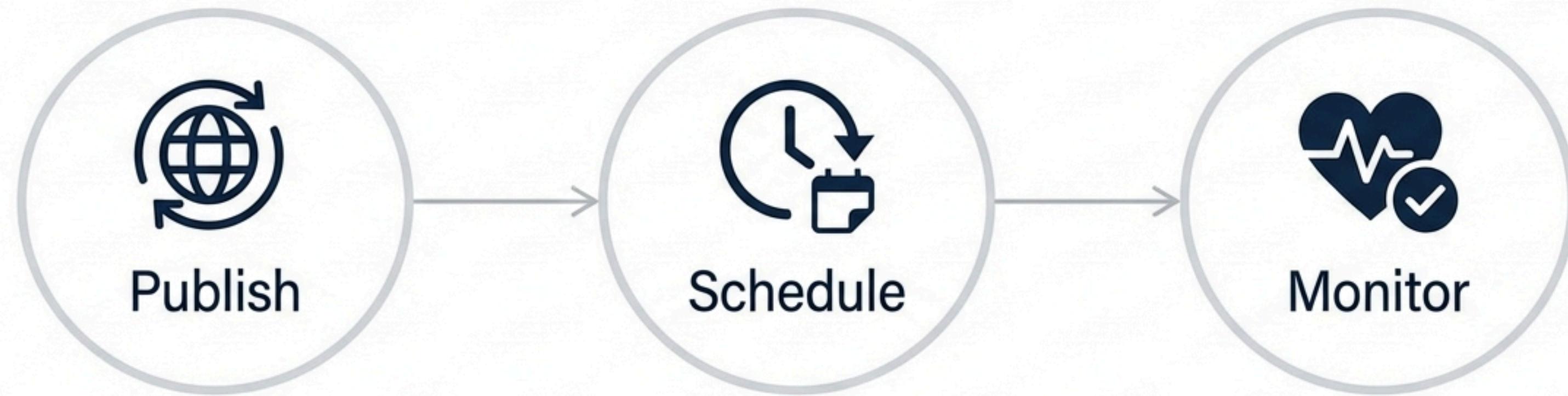


Adding Interactivity

Filters allow business users to slice data without writing code.



Production: Schedule & Refresh



Lock version for
viewers.

Define refresh
cadence (e.g., Daily).

Ensure data
currency.

Set up a refresh schedule in the dashboard settings to keep your data up to date.

Mission Accomplished



- **Ingested** Gold-level commerce data.
- **Calculated** trends and segmentation logic.
- **Visualized** insights for impact.
- **Deployed** a live, interactive data product.