



Bikeshop Shipping Delay Analysis

With: PostgreSQL & Tableau

Ifal Nurfalah Supriatna



— — —

Background

Bikeshop is an offline retail company that sells bicycles. In recent years, the company has frequently faced shipping delays, resulting in customer complaints and operational inefficiencies.

Data

The dataset, obtained from the sales and production databases and processed using data mining techniques to extract relevant records, represents Bikeshop sales data 2017 with **20 columns** and **4,722 rows**.

Objective

To address this issue, an **analysis was conducted to identify the scale and segmentation of delivery delays based on factors such as day, month, store, and destination city**.

Hypotheses

The analysis assumes that shipping delays at Bikeshop are not random but are influenced by certain patterns, such as specific days, months, store locations, or destination cities.



— — —

Data Preparation

The dataset was already well-structured, with consistent formats and valid values across all columns. Therefore, the data preparation process focused mainly on removing duplicate records and identifying any null or missing values to ensure data accuracy and reliability.

Removing duplicate records

```
WITH cte_dup AS(
    SELECT *,
        ROW_NUMBER() OVER(
            PARTITION BY order_id, brand_name, product_name, order_date, list_price) AS dup_row
    FROM set.bike_shop_dataset
)
SELECT * FROM cte_dup WHERE dup_row > 1
```

Total rows: 33

After checking the dataset, it was found that there were **33 duplicate records**.

Identifying any null or missing values

```
SELECT * FROM set.bikeshop_data_nodup
WHERE order_id IS NULL
OR customer_id IS NULL
OR product_name IS NULL
OR quantity IS NULL
OR list_price IS NULL
OR revenue IS NULL
OR profit IS NULL
```

Total rows: 0

After checking the dataset, **no missing values were found**.



EDA Process

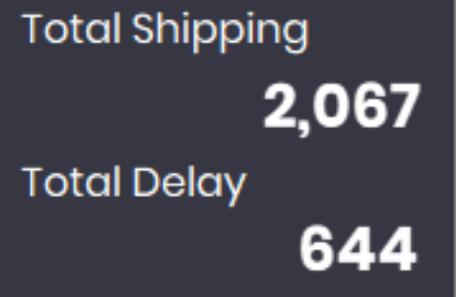
Shipping Performance Overview

The analysis examines the total number of shipments in 2017, the number of delayed shipments during that year, and the percentage of delays in 2017.

```
SELECT
    COUNT(order_id) AS total_ord,
    COUNT(CASE WHEN shipping_status = 'Delay' THEN order_id END) AS delay_ord,
    ROUND(
        COUNT(CASE WHEN shipping_status = 'Delay' THEN order_id END) * 100.0 / COUNT(order_id),
        2) AS delay_percentage
FROM set.bikeshop_data_nodup
WHERE EXTRACT (YEAR FROM order_date) = 2017
```

Query

Result

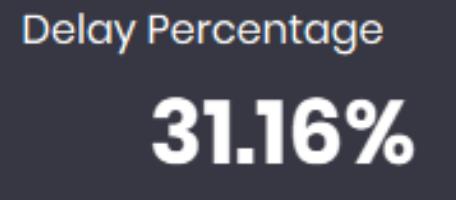


In 2017, a total of **2,067 shipments** were recorded, showing a high volume of delivery activity.

Out of all shipments, 644 experienced delays, indicating that delivery issues were relatively common.

The delay rate of 31.16% suggests a need for process improvement in logistics or distribution to enhance delivery efficiency.

Insight





EDA Process

Delay by Day

The purpose of this exploration is to analyze shipping delays by day of the week compared to the total shipments in 2017.

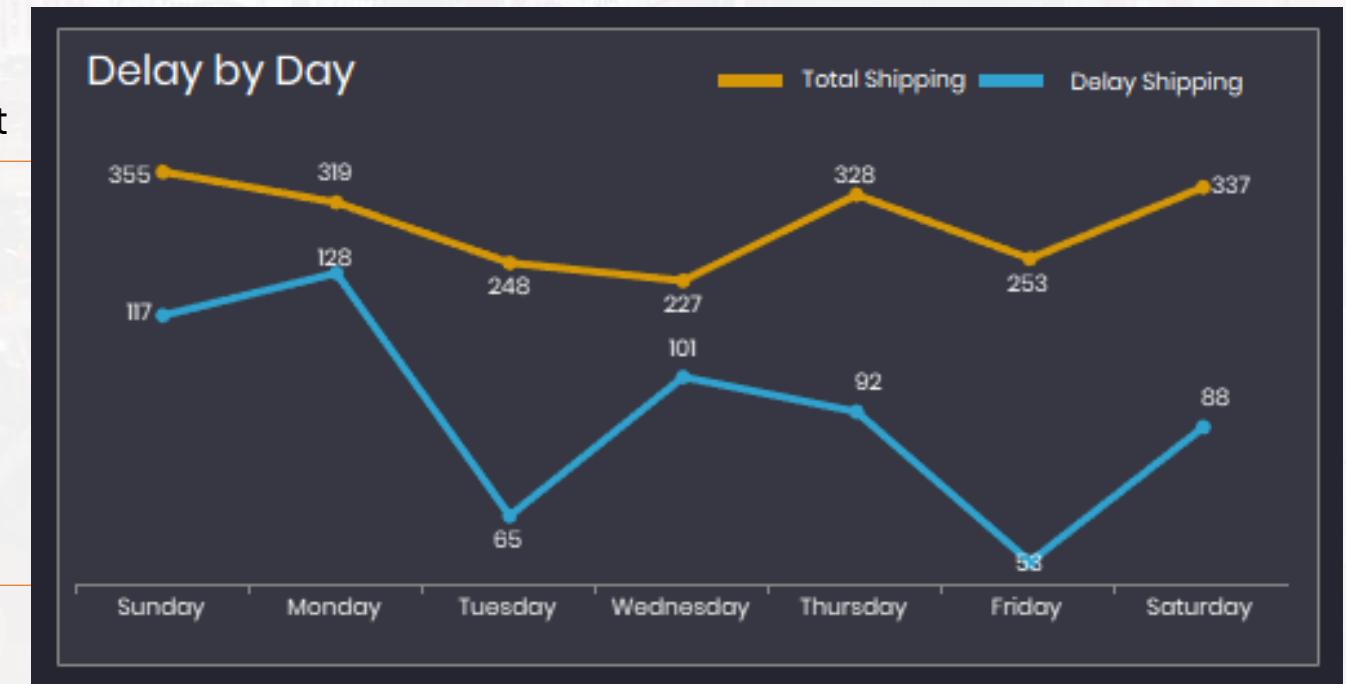
```
SELECT
    TO_CHAR(order_date, 'Day') AS delay_by_day,
    COUNT(order_id) AS total_delay
FROM set.bikeshop_data_nodup
WHERE EXTRACT (YEAR FROM order_date) = 2017
    AND shipping_status = 'Delay'
GROUP BY delay_by_day
ORDER BY total_delay DESC
```

Query

Result

There appears to be a **partial correlation between order volume and shipping delays**, as higher shipments on Sunday and Monday are followed by more delays. However, the pattern is not consistent across all days, indicating that factors such as operational efficiency or resource allocation may also influence delivery performance.

Insight



In summary, there is some correlation, but it's not purely linear, meaning order volume alone doesn't fully explain the delay pattern.



EDA Process

Delay by Month

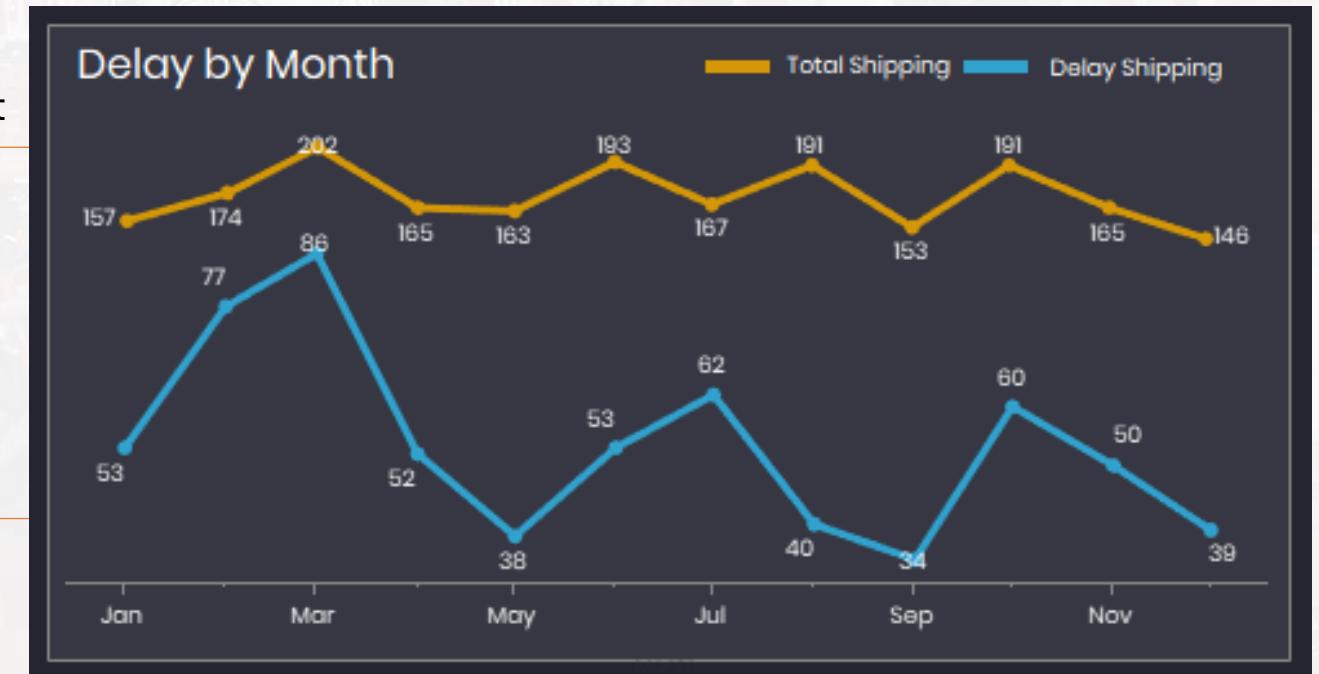
The purpose of this exploration is to analyze shipping delays by month compared to the total shipments in 2017.

```
SELECT
    TO_CHAR(order_date, 'Month') AS delay_by_month,
    COUNT(order_id) AS total_delay
FROM set.bikeshop_data_nodup
WHERE EXTRACT (YEAR FROM order_date) = 2017
    AND shipping_status = 'Delay'
GROUP BY delay_by_month
ORDER BY total_delay DESC
```

Query

Result

Insight



The analysis suggests a **clear correlation** between the total shipping volume and the number of delayed shipments, but with varying degrees of strain.

March stands out as the most challenging month for the bike shop's logistics, experiencing the absolute peak in both total volume and number of delays, as well as the highest relative delay rate.



EDA Process

Delay by Store

The purpose of this exploration is to analyze shipping delays by store.

```
SELECT
    store_name,
    COUNT(order_id) AS total_delay
FROM set.bikeshop_data_nodup
WHERE EXTRACT (YEAR FROM order_date) = 2017
    AND shipping_status = 'Delay'
GROUP BY store_name
ORDER BY total_delay DESC
```

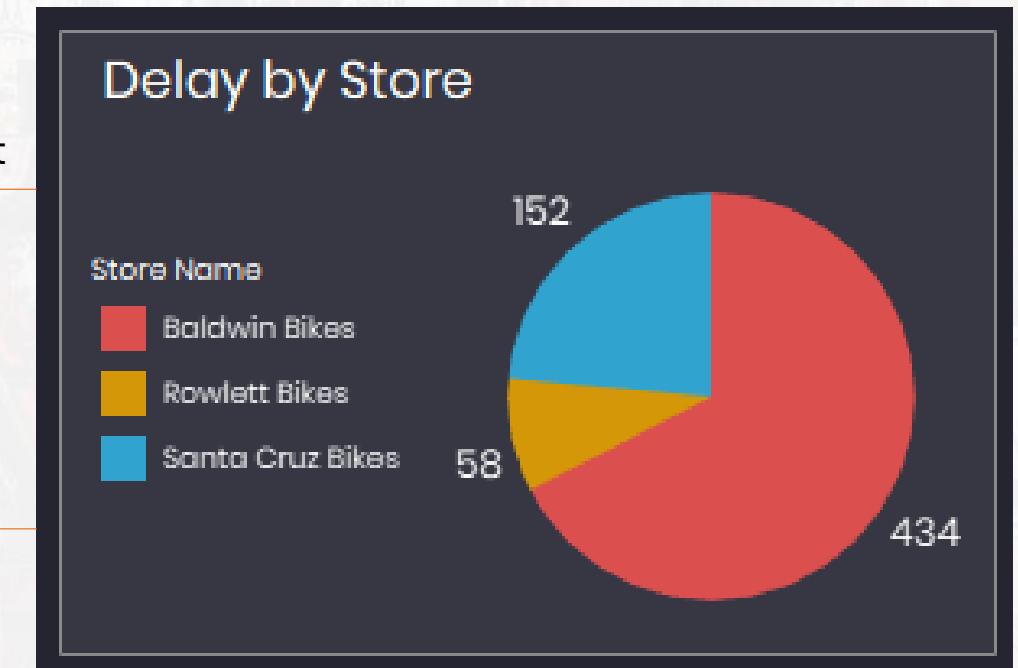
Query

Result

Insight

65% of delays occurred at Baldwin Bikes, indicating that this store faces significant operational or logistical challenges compared to others.

This could be due to higher order volume, limited staff capacity, or inefficiencies in the delivery process that need further investigation to improve overall shipping performance.





EDA Process

Delay by Destination City

The purpose of this exploration is to analyze shipping delays by destination city.

```
SELECT
    cust_city,
    COUNT(order_id) AS total_delay
FROM set.bikeshop_data_nodup
WHERE EXTRACT (YEAR FROM order_date) = 2017
    AND shipping_status = 'Delay'
GROUP BY cust_city
ORDER BY total_delay DESC
```

Query

Top 5 City with Highest Delay

3 19

Smithtown

New Rochelle

San Angelo

Sunnyside

Ballston Spa

The **Baldwin** shipping station is the primary source of delays contributing to the highest counts in multiple major destinations including **19 to Smithtown** and **15 to Spring Valley**. Fixing Baldwin's internal process will yield the largest overall delay reduction.

Insight



— — —

Conclusions

- **Temporal Pattern (Day of Week):** The highest order day is Sunday, while the highest delay day is Monday.
- **Temporal Pattern (Month):** March is the month with the highest level of both total orders and shipping delays throughout 2017.
- **Origin/Source Bottleneck:** Baldwin Bikes in the city of Baldwin is the store and origin city with the highest total number of shipping delays in 2017.

Recommendations

- **Address the Primary Source:** Immediately audit the **Baldwin** shipping station to correct internal operational inefficiencies, such as staff shortages or systemic faults, which are causing the largest overall number of route delays.
- **Manage Peak Strain:** Increase staffing and logistics capacity on **Sunday and Monday** and during the **March** peak to prevent the most significant daily and monthly backlogs.
- **Optimize Critical Routes:** Focus targeted carrier and infrastructure reviews on the worst-performing corridors, specifically **Baldwin to Smithtown**.



— — —

Thank You!

Ifal Nurfalah Supriatna