

E-Commerce Project Presentation

Presented by Ziad Asal, Ibrahim Refaey
and Ibrahim Izz ElDeen

Supervised by Eng. Essam Ammar

Data Engineering Mentorship 2023

Introduction

We have different sets of information about E-commerce, each focusing on specific aspects.

Key Categories:

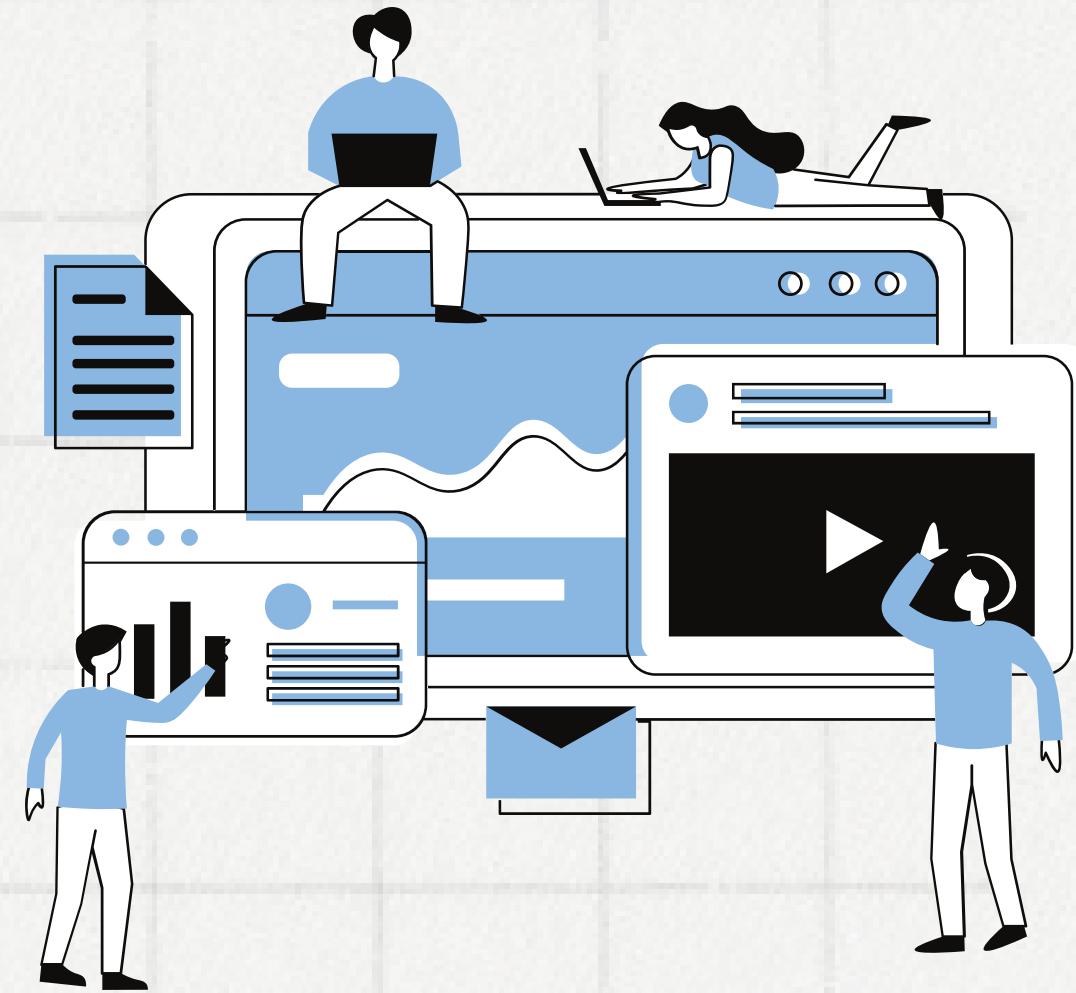
- Products available
- Users involved
- Orders placed
- Items within orders
- Sellers involved
- Feedback received

Project Goal: Creating a system to organize this data, helping to answer important business queries.

Project Idea

This project aimed to develop Datawarehouse that organized all different sets of data to adapt with business requirements also to support decision making to improve business as possible as.

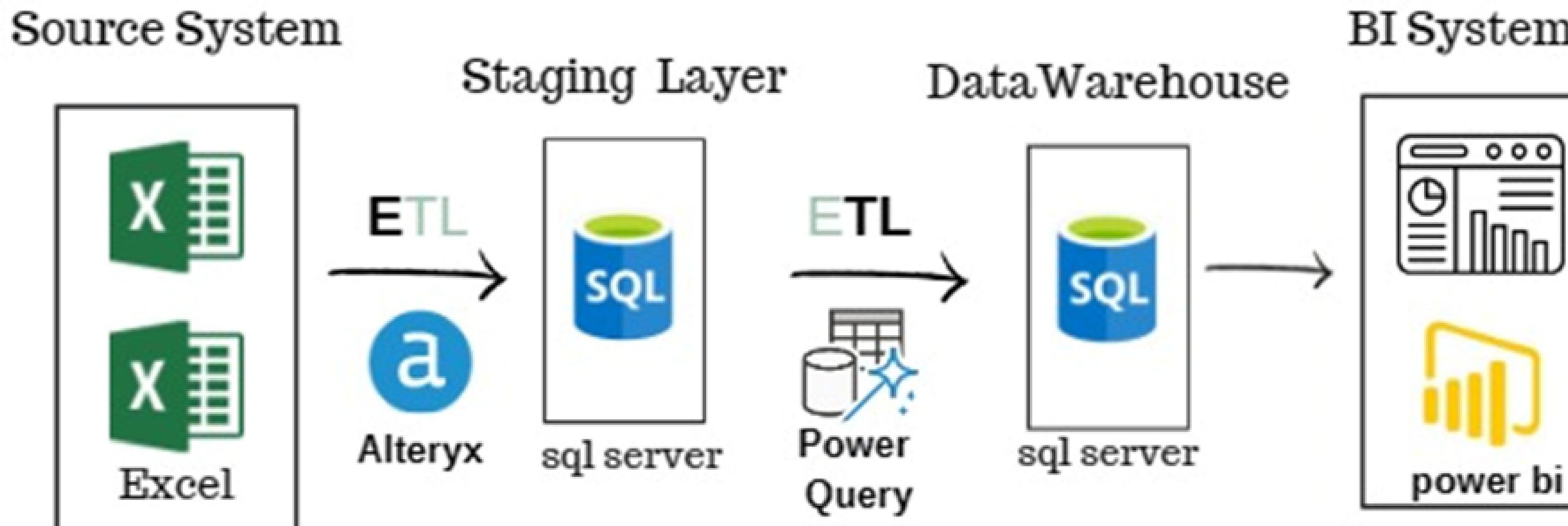
Then Create ETL pipeline to apply Extract, Transform, Load into the Datawarehouse to answer Business Questions.



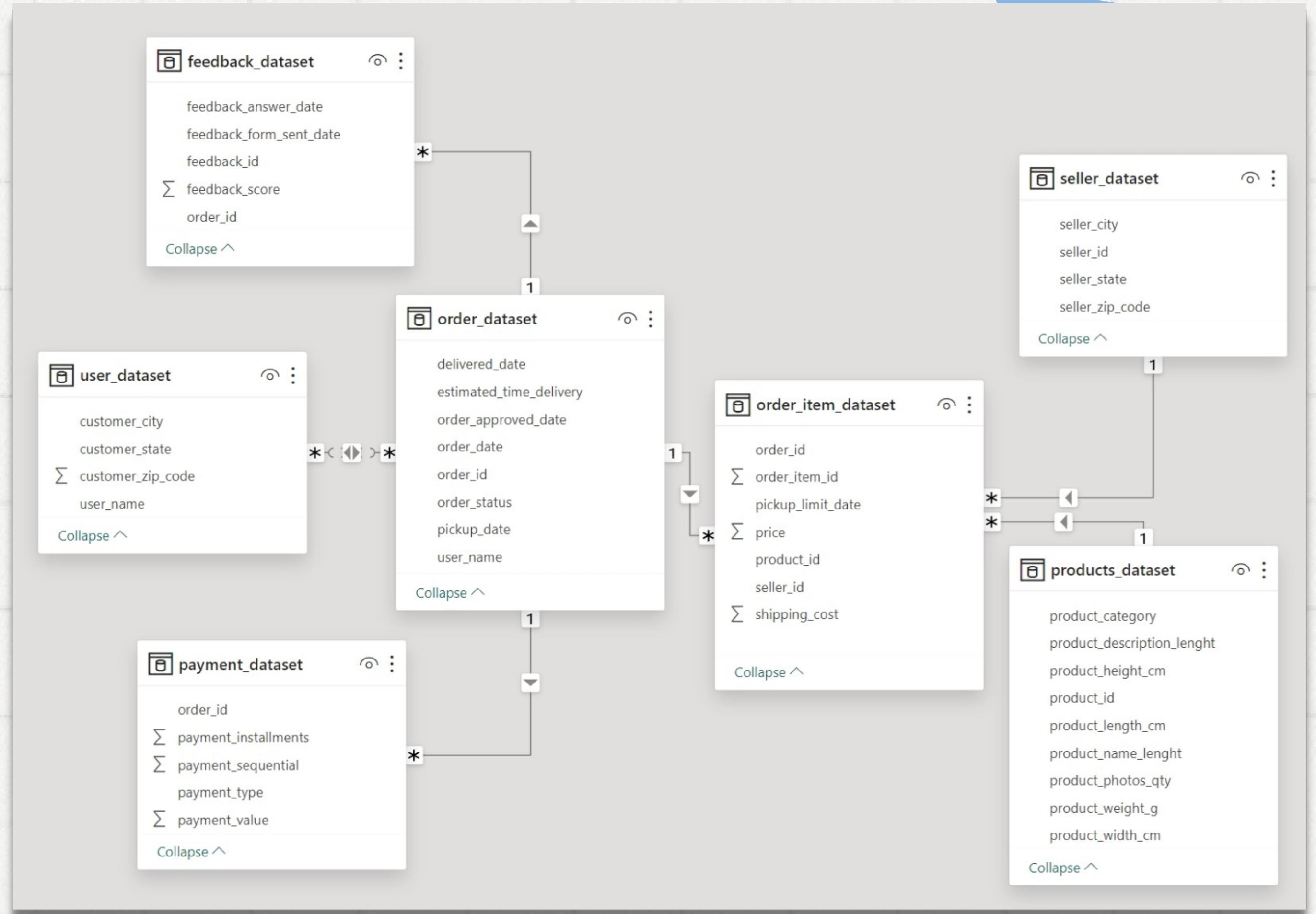
Process

1.	Integrate
2.	Analyze
3.	Reporting

Data Engineer Ecommerce



Staging Layer (OLTP Model)



**OLAP Model Needed, But
First, Our Data Needs
Some Transformations!**

ETL & Data Modeling Processes



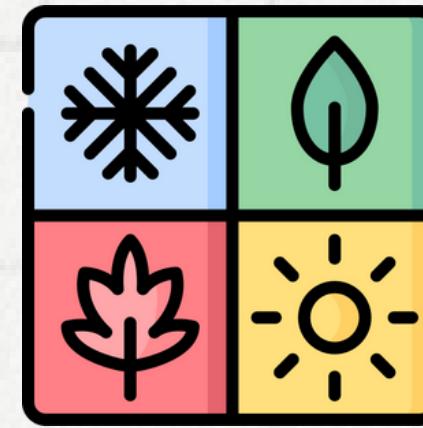
SCD in Users and Feedbacks Tables. Only Keep Most Recent Records



Group by Order ID and Product ID in Order Items Table to Reduce Records and Calculate Product Items Quantity



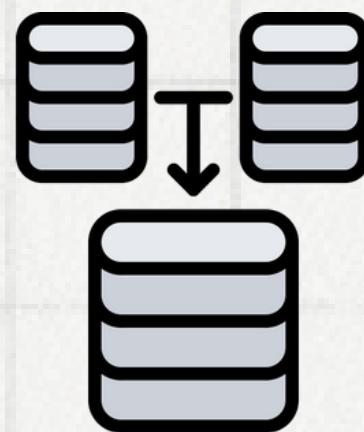
Split Date Columns to Separated Dates and Times Columns



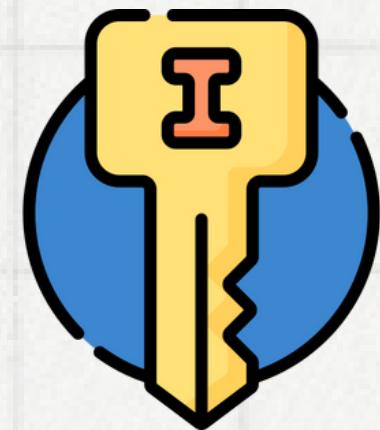
Create Dates Dimension with Surrogate Key Index, Date, Year, Month, Day and Season (Spring, Summer, Fall and Winter)



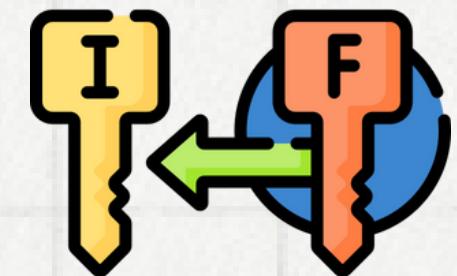
Create Times Dimension with Surrogate Key Index, Time, Hour, Minute, Second, AM/PM and Time of Day (Morning, Afternoon, Evening and Night)



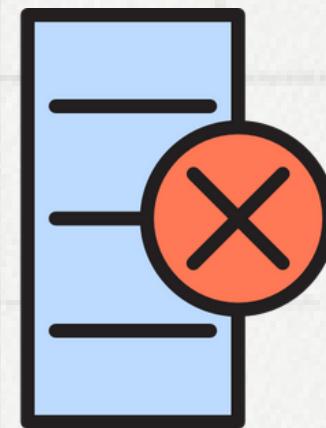
Merge Orders and Order Items Tables as One Fact Table



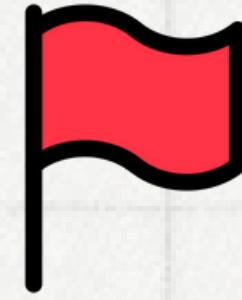
Create New Surrogate Key Indexing as New PK For Other Dimensions



Link All FKs in Fact Table with PKs in Dimensions



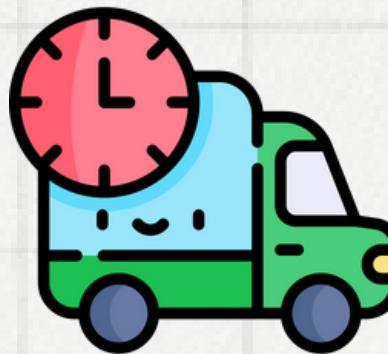
Delete Previous Keys and Unused Columns



Derive a Delay Check Flag Column, "Delayed" for Estimated Delivery Date < Delivery Date



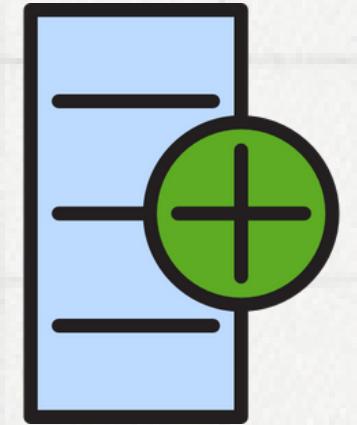
Derive Delivery Delay Days Column, Which Returns Delay Duration in Days



Derive Shipping Days Column, Which Returns Shipping Duration in Days



Clean, Format & Trim Data

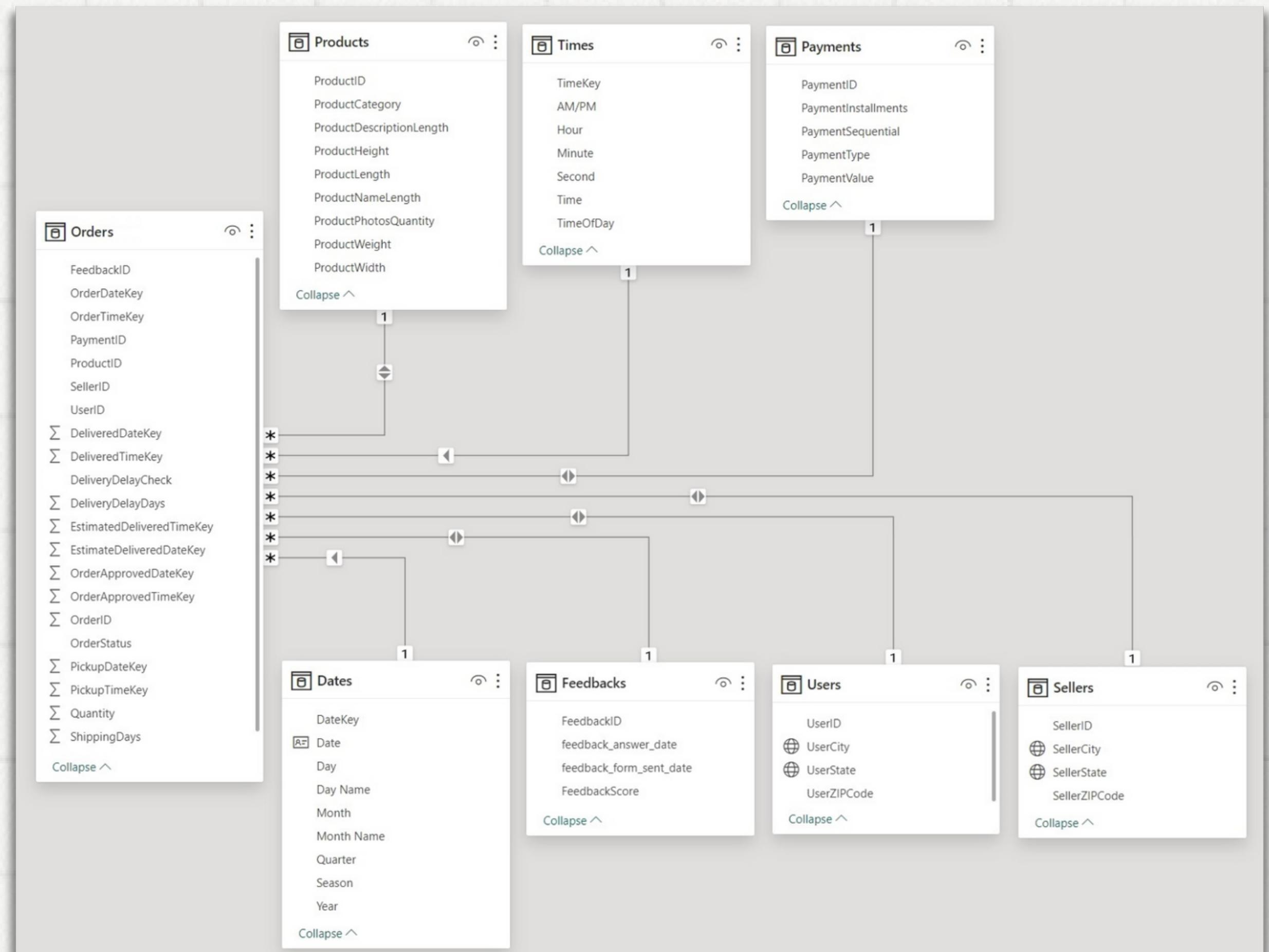


Optional: Add Valuable Attributes to Fact Table for Better Retrieval Instead of Joining Tables Every Time



We Now Have A Star-Schema Data Model!

Final Data Model



**Data Model is Now
OLAP-Ready! Let's Dive
Into Some Business
Analysis**

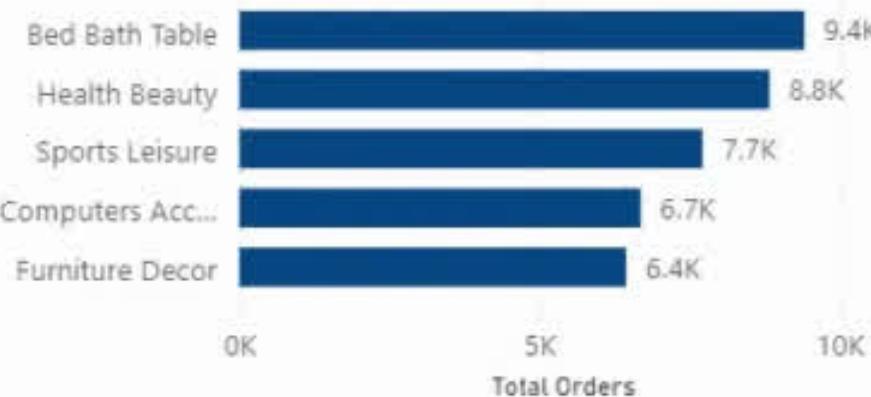
E-Commerce Dashboard

Data Engineering Mentorship 2023

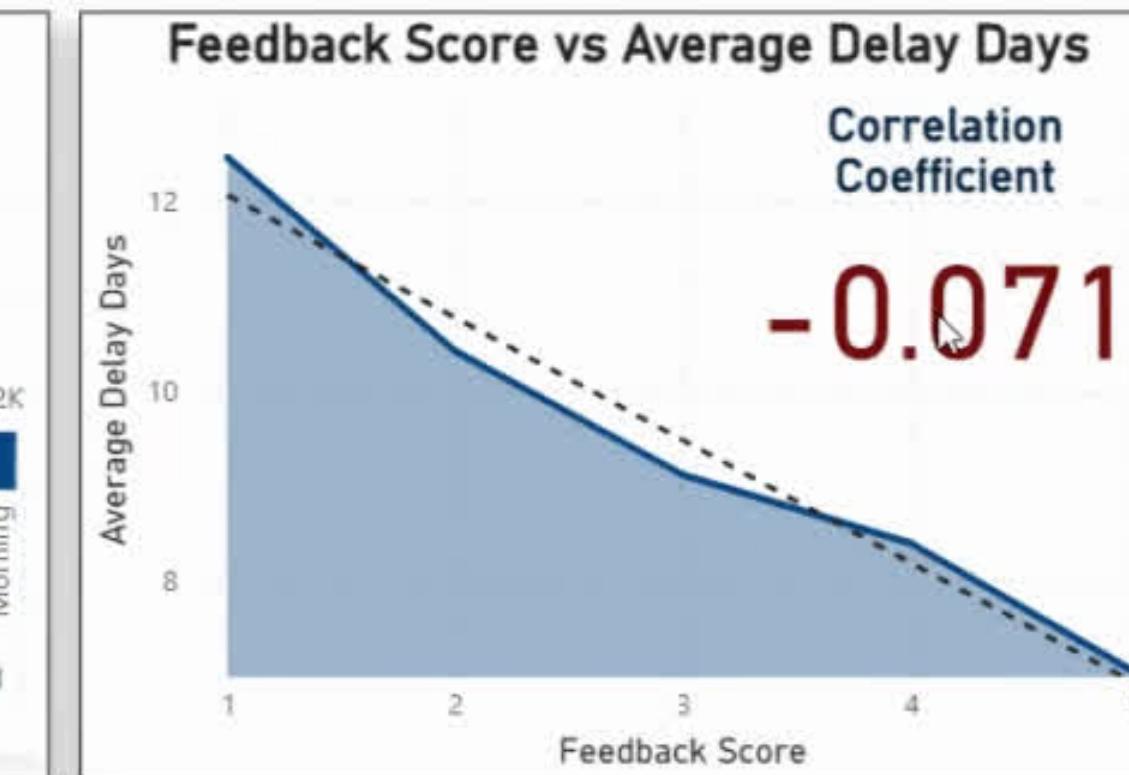
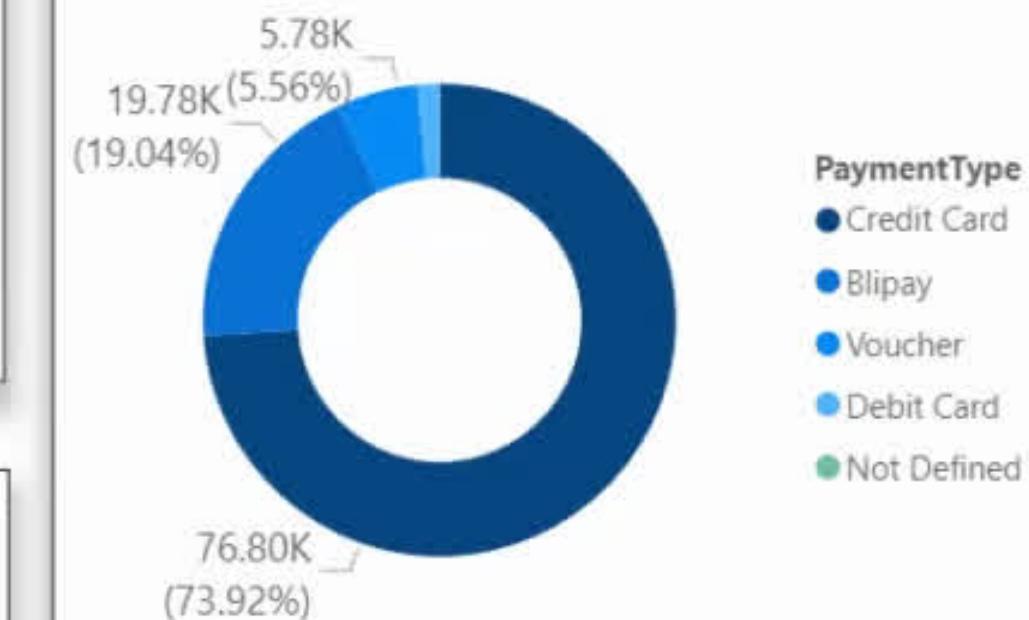
Total Orders
99.44K

Total Revenue
Rp16.0bn

Total Orders by Product Category



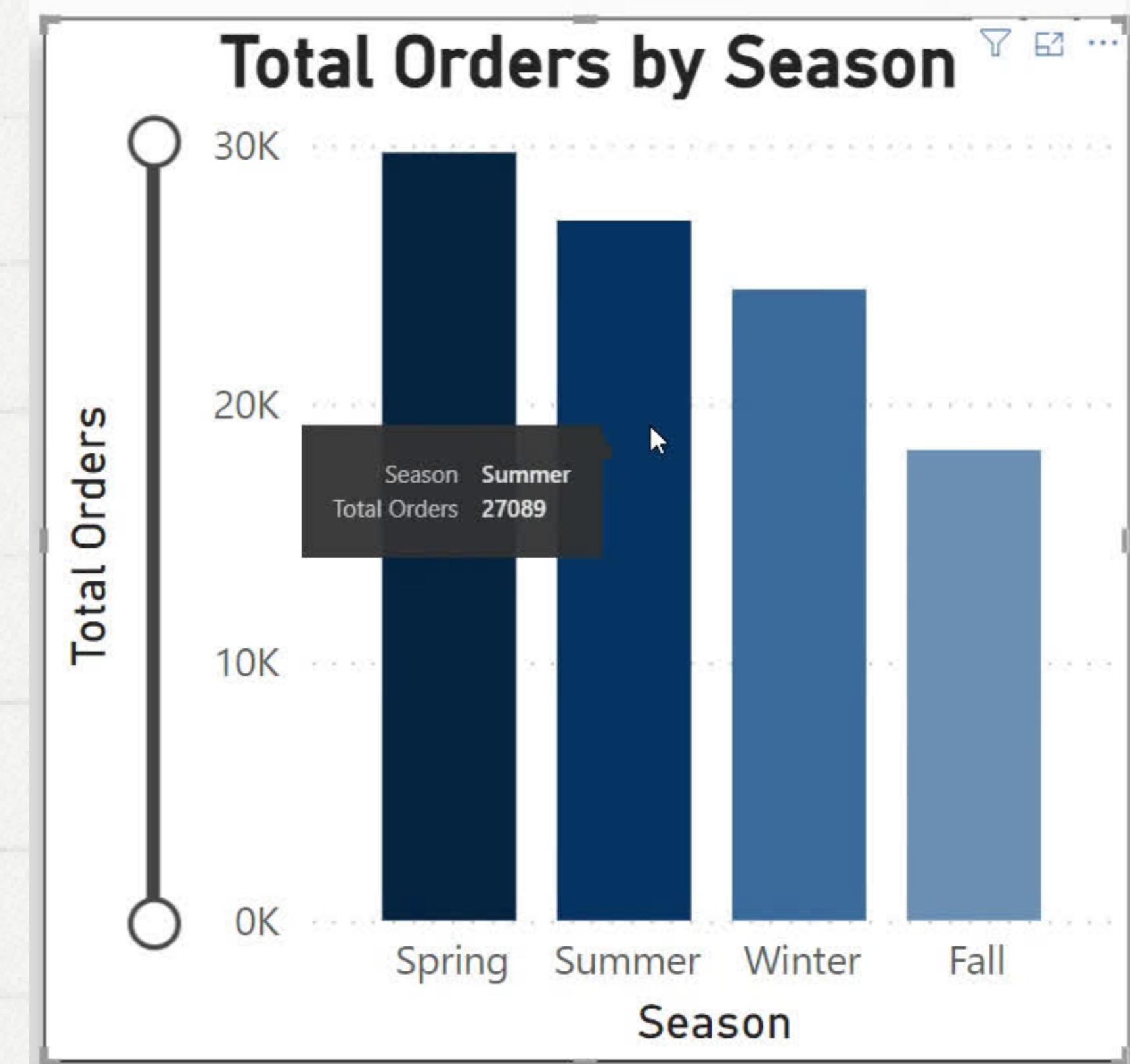
Total Payments by Payment Type



- What is the Peak Season of our E-Commerce?

The highest number of orders is in **Spring**, reaching close to 29.72K. This suggests that Spring is the busiest season for orders.

Fall, on the other hand, has the lowest number of orders, around 18.21K. This indicates a significant drop in orders during this season.



- **What Time Users Are Most Likely Making An Order Or Using our E-Commerce App?**

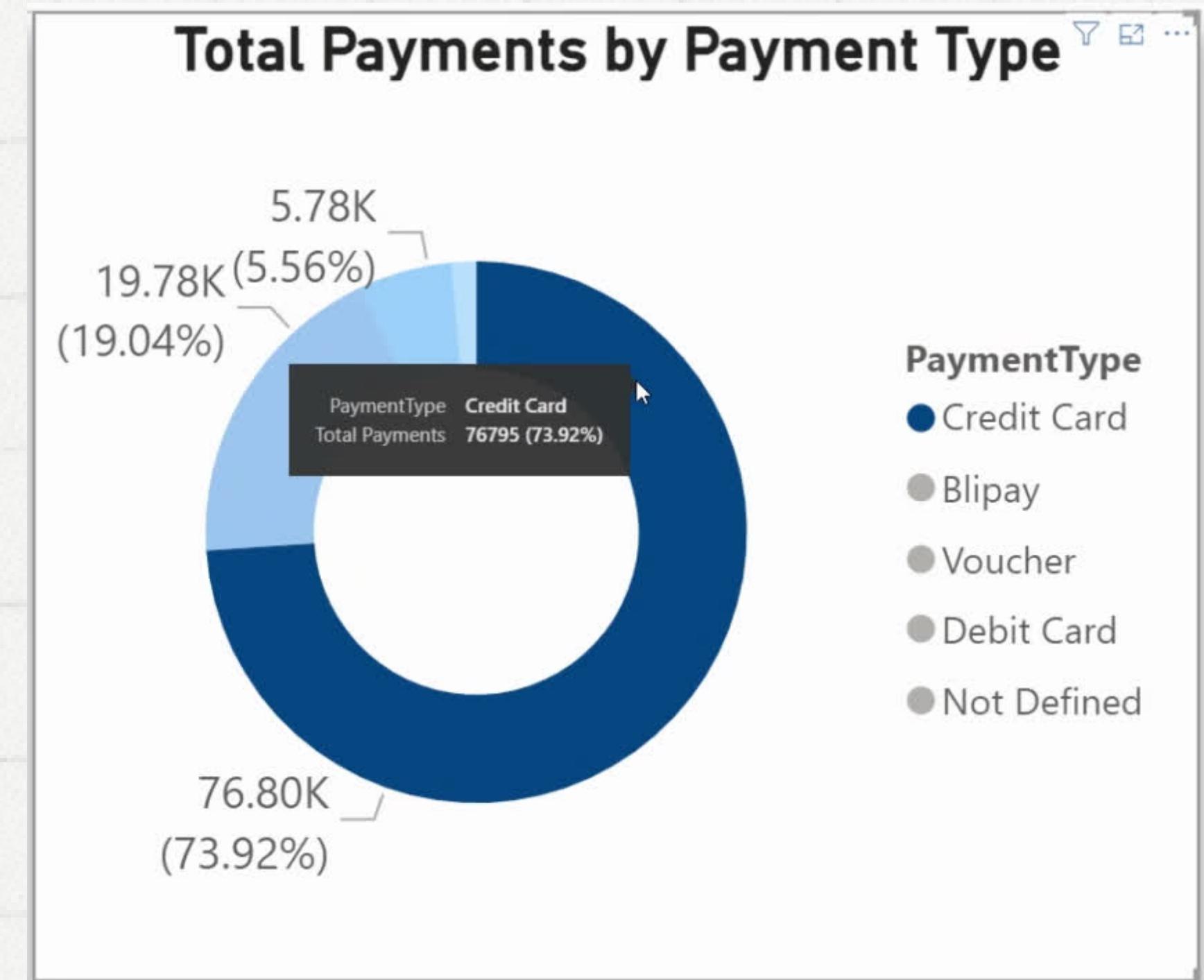
The highest number of orders are received during the **Afternoon** and **Evening** hours, peaking at 6.7K orders around **16:00 (4 PM)**.

There is a significant drop in orders during the late night to early morning hours (4-5 AM).



- What Is The Preferred Way To Pay In The E-Commerce?

Credit Card is the most used payment method. Followed by Billpay, Voucher, Debit Card and Undefined Payment Methods.



- How Many Installments are usually done when paying in the E-Commerce?

There are total of 99441 orders, with total revenue of Rp16bn, and average payment installments of 2.85 per order.

Total Orders	Total Revenue	Average Installments	Delayed Orders
99.44K	Rp 16.0bn	2.93	6535

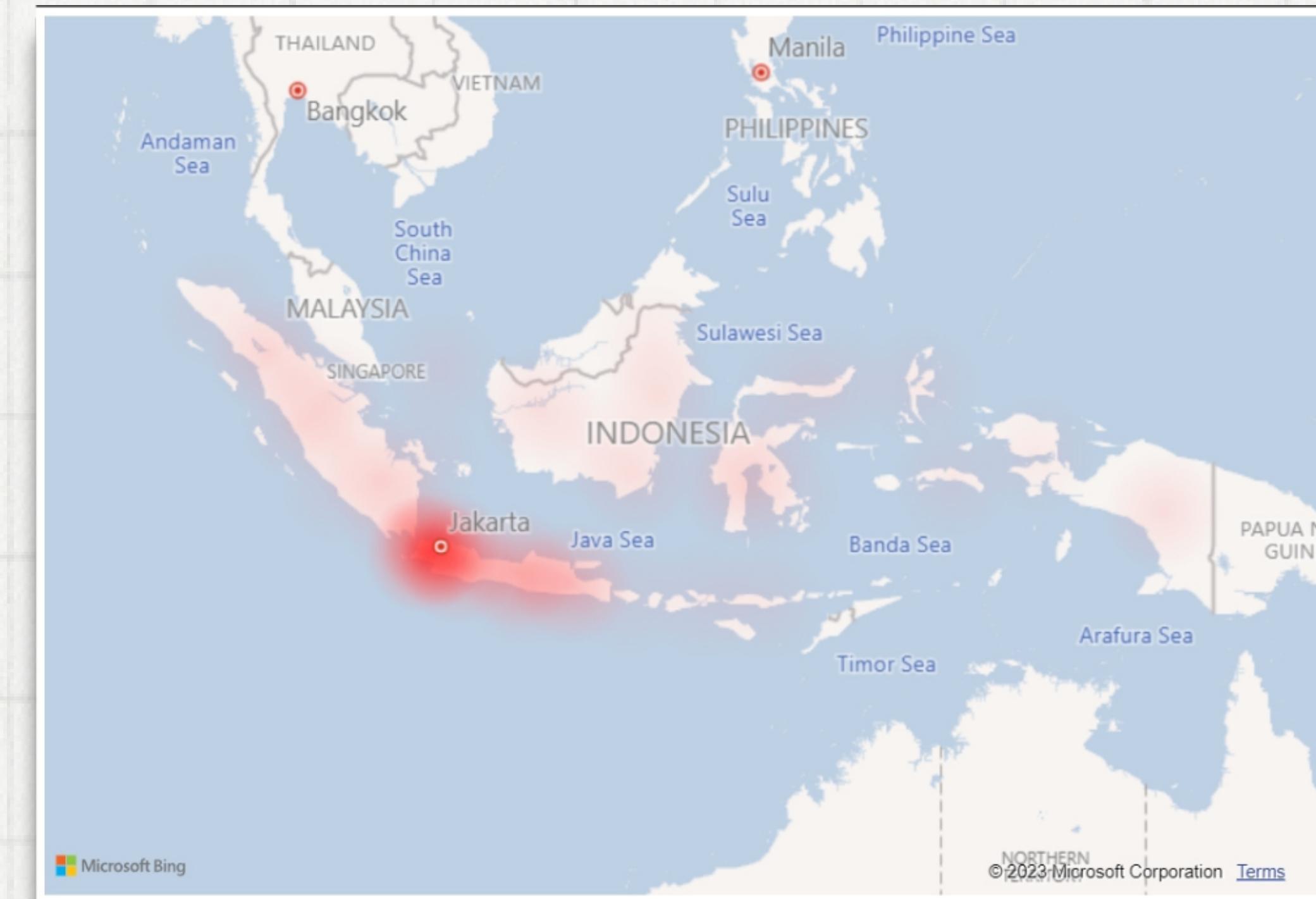
- What Is The Frequency Of Purchase In Each State?

Banten has the highest total orders of 21,167, indicating a high demand in this region, followed by other states.

The Purchase Frequency, however, is almost consistent across all states, ranging from 1.02 to 1.05.

User State	Total Orders	Orders/Customer
Banten	21167	1.04
Jawa Barat	12844	1.04
Dki Jakarta	12526	1.04
Jawa Tengah	8583	1.03
Jawa Timur	8448	1.03
Sumatera Utara	3928	1.03
Sulawesi Selatan	2383	1.04
Sumatera Selatan	2148	1.03
Sumatera Barat	1890	1.04
Papua	1801	1.03
Di Yogyakarta	1758	1.03
Kalimantan Timur	1687	1.04
Lampung	1677	1.04
Kalimantan Barat	1569	1.03

The Heat Map shows that The Capital Jakarta and states around it have the highest number of orders.



• Which Logistic Route Has Heavy Web Traffic In Our Ecommerce?

When it comes to web traffic, **Kota Tangerang** has the highest number of Seller-Customer Transactions, indicating a strong local commerce.

General			Total Orders	Total Delayed Orders
Avg Delivery Delay Days	Delivery Delay Days/Order	Avg Shipping Days		
Logistic Route				
Seller City	User City	Total Orders		
Kota Tangerang	Kota Tangerang	4182		
Kota Tangerang	Kota Jakarta Barat	1432		
Kabupaten Berau	Kota Tangerang	1113		
Kota Tangerang	Kabupaten Tangerang	605		
Kabupaten Berau	Kota Jakarta Barat	504		
Kota Tangerang	Kabupaten Bekasi	485		
Kabupaten Bogor	Kota Tangerang	437		
Kota Tangerang	Kota Jakarta Selatan	387		
Kota Tangerang	Kota Jakarta Timur	378		
Kota Jakarta Selatan	Kota Tangerang	371		
Total		99441		

- Which Logistic Route Has Heavy Traffic In Our Ecommerce?

The route from Kabupaten Tangerang to Kabupaten Bintan experiences the longest average delay of 181 days.

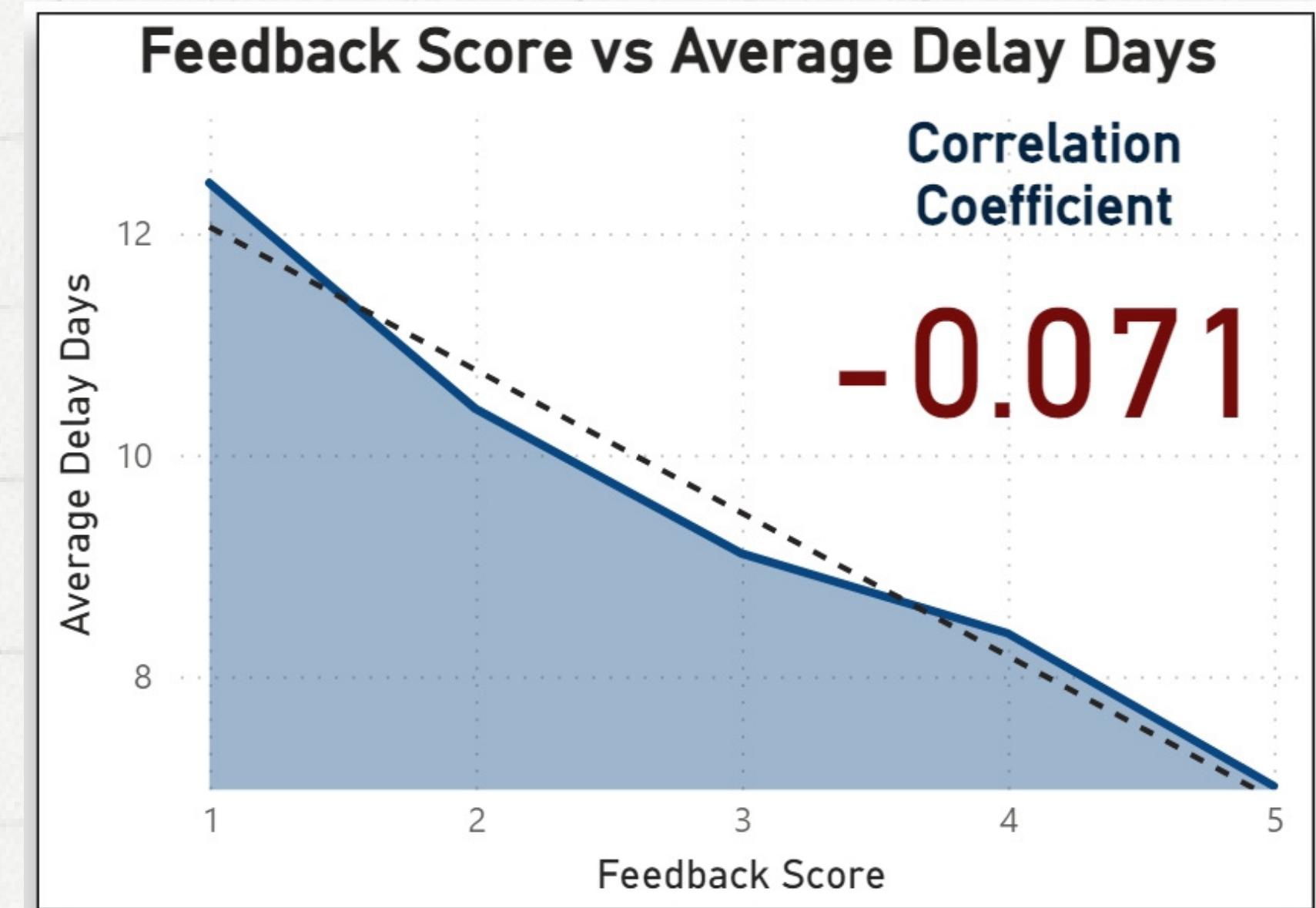
General		Total Orders	Total Delayed Orders
Avg Delivery Delay Days	Delivery Delay Days/Order	Avg Shipping Days	
Logistic Route			
Kabupaten Tangerang	Kabupaten Bintan		181.00000
Kota Tangerang	Kabupaten Lampung Selatan		175.00000
Kabupaten Nias Selatan	Kabupaten Tasikmalaya		167.00000
Kabupaten Kendal	Kabupaten Bangka Tengah		166.00000
Kabupaten Simeulue	Kabupaten Bungo		162.00000
Kota Balikpapan	Kota Bekasi		161.00000
Kota Malang	Kabupaten Jepara		161.00000
Kota Mojokerto	Kota Tasikmalaya		159.00000
Kota Bandung	Kota Manado		155.00000
Kabupaten Jombang	Kota Bukittinggi		153.00000
Total			0.00163

- How Many Late Delivered Orders In Our Ecommerce? Are Late Orders Affecting Customer Satisfaction?

There are **6535** delayed orders in our E-Commerce.

The graph shows a **very weak negative correlation** between Feedback Score and Average Delay Days, as indicated by a Correlation Coefficient of -0.071.

Delayed Orders
6535



- How Long Is The Delay For Delivery/Shipping Process In Each State?

Kepulauan Bangka Belitung has the highest average shipping days of 12.97, indicating that deliveries take the longest time in this region.

User State	Avg. Shipping Days	Avg. Delivery Delay Days
Kepulauan Bangka Belitung	12.97	16.43
Bali	12.30	13.34
Sulawesi Barat	11.96	14.71
Kepulauan Riau	11.83	14.16
Sulawesi Tenggara	11.83	12.16
Nusa Tenggara Timur	11.81	10.52
Sulawesi Utara	11.69	12.44
Papua	11.60	8.51
Lampung	11.56	10.54
Aceh	11.48	10.03
Nusa Tenggara Barat	11.44	16.95
Papua Barat	11.29	11.61
Kalimantan Tengah	11.19	10.65
Jambi	11.12	14.56

- How Long Is The Difference Between Estimated Delivery Time And Actual Delivery Time In Each State?

Nusa Tenggara Barat has the highest average delivery delay of days, indicating that deliveries take the longest time in this region.

User State	Avg. Shipping Days	Avg. Delivery Delay Days
Kepulauan Bangka Belitung	12.97	16.43
Bali	12.30	13.34
Sulawesi Barat	11.96	14.71
Kepulauan Riau	11.83	14.16
Sulawesi Tenggara	11.83	12.16
Nusa Tenggara Timur	11.81	10.52
Sulawesi Utara	11.69	12.44
Papua	11.60	8.51
Lampung	11.56	10.54
Aceh	11.48	10.03
Nusa Tenggara Barat	11.44	16.95
Papua Barat	11.29	11.61
Kalimantan Tengah	11.19	10.65
Jambi	11.12	14.56

E-Commerce Dashboard

Data Engineering Mentorship 2023

Total Orders

99.44K

Total Revenue

Rp16.0bn

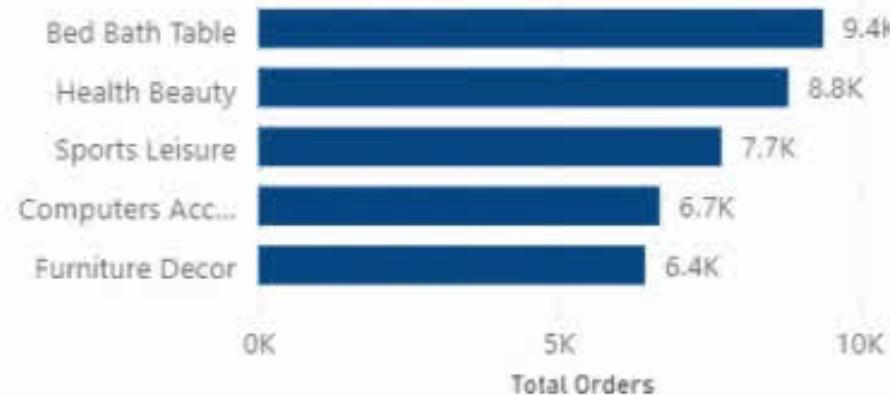
Average Installments

2.93

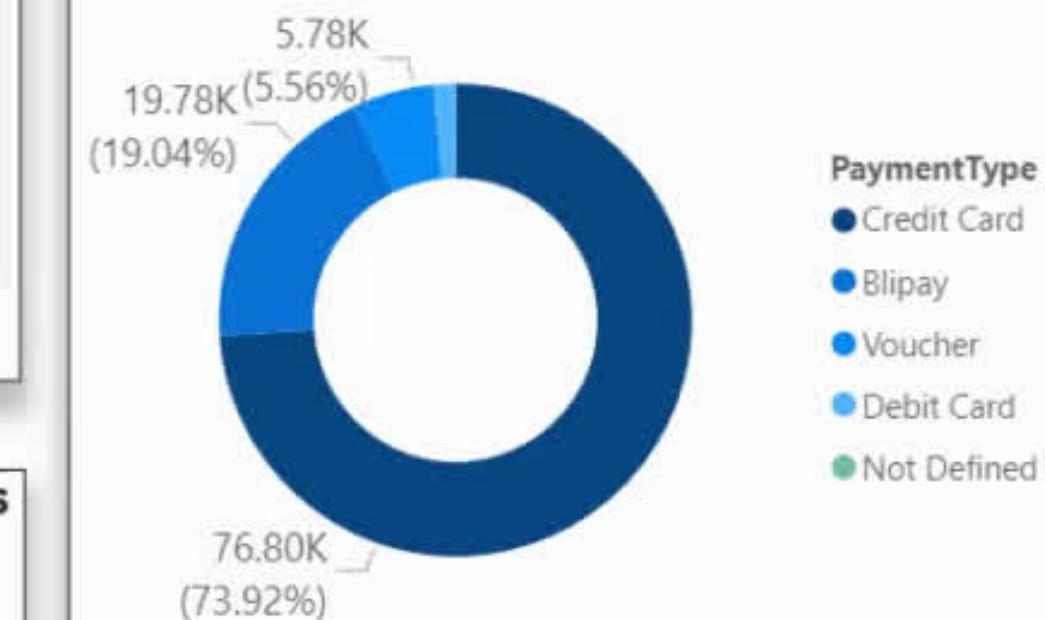
Delayed Orders

6535

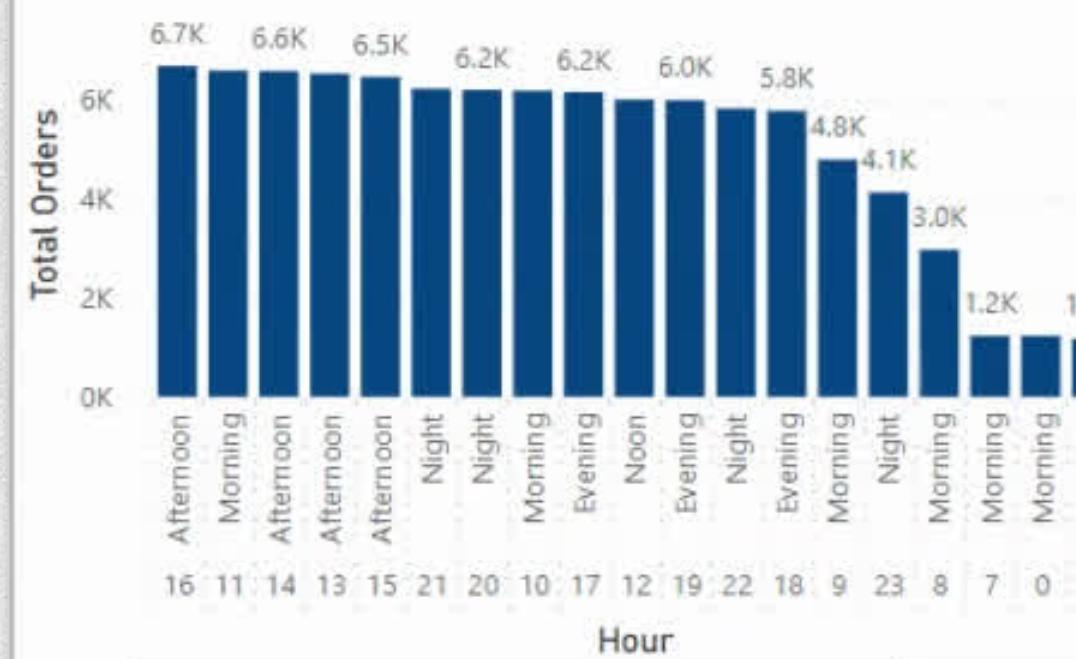
Total Orders by Product Category



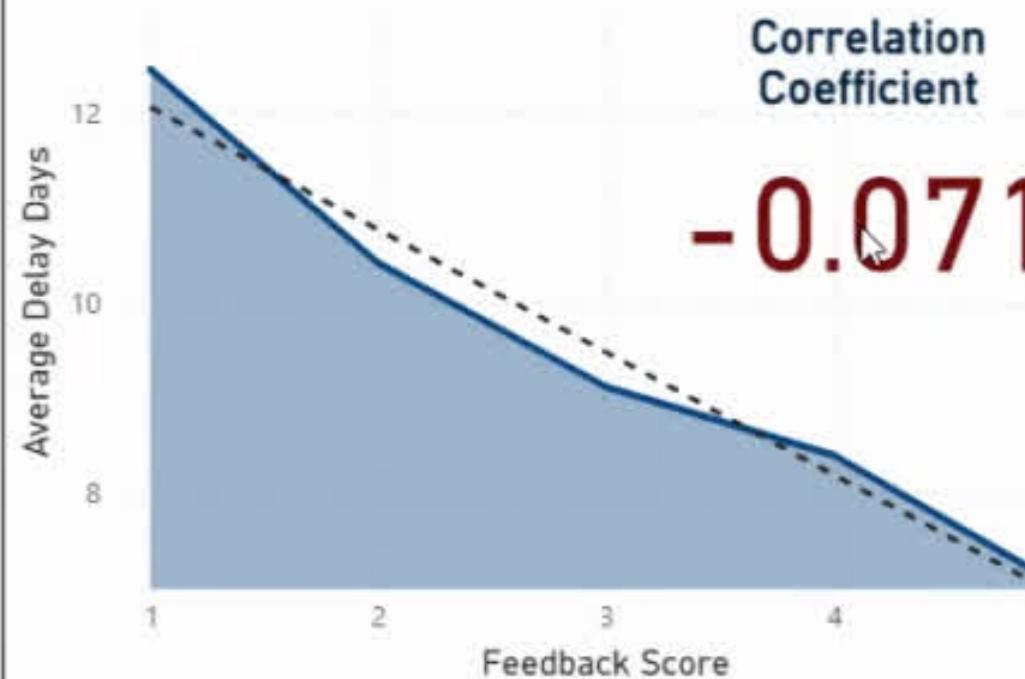
Total Payments by Payment Type



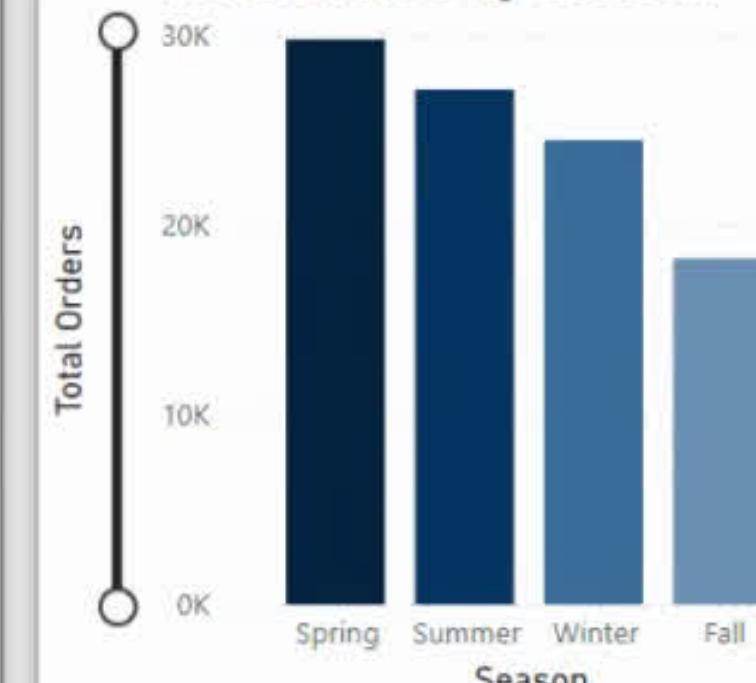
Total Orders by Time



Feedback Score vs Average Delay Days



Total Orders by Season



Some Insights

Outliers in Sales

24/11/2017 → 1283 Orders

Avg Price of Product 8613 in
24/11/2017

Rp 103K

Avg Price of Product 8613
except 24/11/2017

Rp 158K



November 24, 2017 was a Friday in Indonesia.

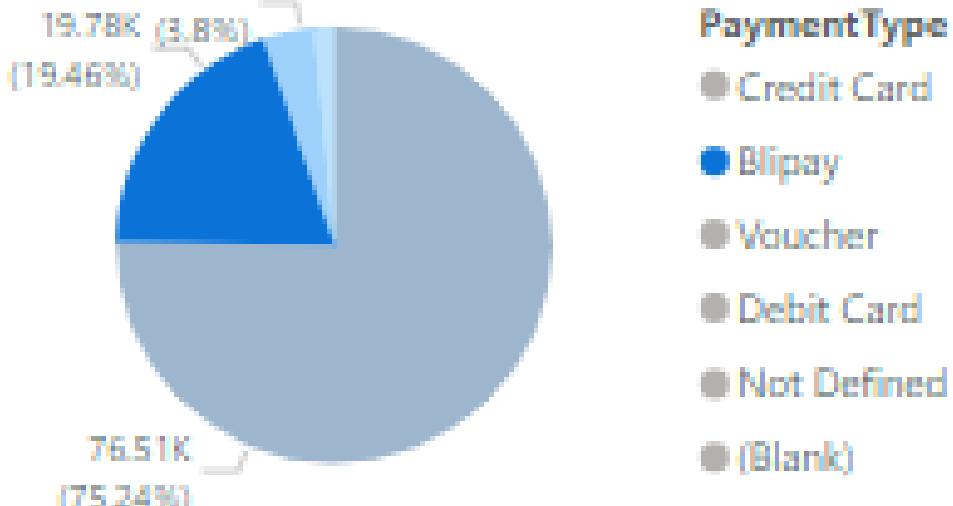
It was a public holiday in Indonesia to celebrate the 72nd Anniversary of the country's independence from the Netherlands.

Using BillPay

1.00

Average Installments

Payment Type



Sales per Category



More Details



<https://github.com/ziadasal/End-to-end-data-engineer-project-E-commerce>

Thank you!