

# SUPPLY CHAIN DATA ANALYSIS.

---

Under supervision of:

**Eng. Mennatallah Sleim El Saed**

PRESENTED BY:

Mariam Yasser - Zahra Mahmoud - Alaa Walid - Aya Ahmed - Rawan Ayman

# TABLE OF CONTENTS

01	INTRODUCTION	07	CHALLENGES AND SOLUTIONS
02	TOOLS AND TECHNOLOGIES USED	08	CONCLUSION
03	DATA COLLECTION AND CLEANING		
04	DATA MAPPING AND STRUCTURE		
05	EXPLORATORY DATA ANALYSIS (EDA)		
06	DASHBOARD AND QUERIES		

# 01

## INTRODUCTION

THIS PROJECT AIMED TO ANALYZE SUPPLY CHAIN DATA THROUGH 4 MAIN PHASES IN ORDER TO:



IMPROVE BUSINESS PERFORMANCE



REDUCE COSTS



ENHANCE DELIVERY EFFICIENCY



INCREASE CUSTOMER SATISFACTION

# 02

## TOOLS AND TECHNOLOGIES USED.



### EXCEL

Designing dimension and fact tables to establish a schema in power bi



### POWER BI.

Creating interactive dashboards for data visualization and answering stockholders questions



### PYTHON (PANDAS)

Data cleaning transformation , EDA.



### SQL

Querying, fixing country names, and integrating tables , Fixed error in country names.

POWER QUERY: PREPROCESSING AND SHAPING DATA WITHIN EXCEL ALSO WE USED IT IN DATA CLEANING

# DATA COLLECTION AND CLEANING

03

01

**COLLECT** data about supply chain (Data co for big data analysis) from kaggle and normalized it into orders, customers, products, and shipping (EXCEL)

02

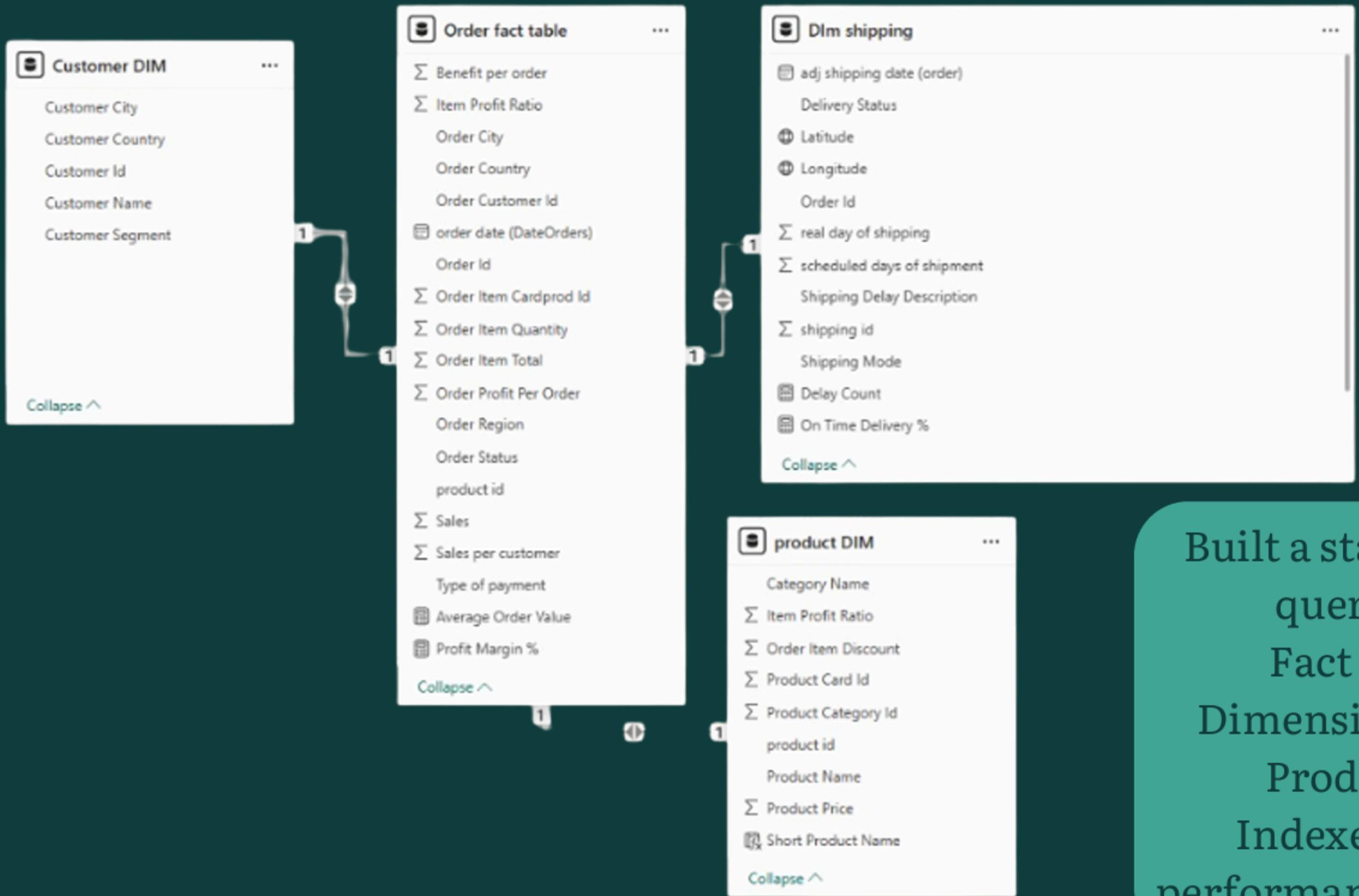
**REMOVE** unnecessary columns , missing values, and duplicates  
standardized formats.

03

**INTEGRATE** datasets to ensure proper relationships  
between tables.

# DATA MAPPING (SCHEMA)

04



Built a star schema for efficient querying and analysis  
Fact Table: order table  
Dimension Tables: Customer, Product, and Shipping  
Indexed tables to enhance performance in querying (PK,FK).



# EDA USING PYTHON

```
[ ] project['Order Status']=project['Order Status'].str.replace('_', ' ')
```

```
▶ project['Order Status']
```

```
→ Order Status
```

	Order Status
0	complete
1	pending
2	closed
3	complete
4	pending payment
...	...
995	pending payment
996	pending payment
997	canceled
998	pending payment
999	processing

01

Formatting the column order status by using replace function to remove the unintegrated symbols.

1000 rows x 1 columns

# 06

```
# Calculate Shipping Delay  
project["Shipping Delay"] = project["real day of shipping"] - project["scheduled days of shipment"]  
  
# Add a column to describe the shipping status  
project["Shipping Status"] = project["Shipping Delay"].apply(  
    lambda x: "Shipped Earlier" if x < 0 else ("On Time" if x == 0 else "Shipped Late")  
)  
  
# Add a column to show the absolute number of days early or late  
project["Shipping Delay Description"] = project["Shipping Delay"].apply(  
    lambda x: f"{abs(x)} day(s) early" if x < 0 else ("On Time" if x == 0 else f"{x} day(s) late")  
)  
  
# Display the updated DataFrame  
print(project[["real day of shipping", "scheduled days of shipment", "Shipping Delay", "Shipping Status", "Shipping Delay Description"]].head())
```

```
real day of shipping scheduled days of shipment Shipping Delay \\\n0 3 4 -1  
1 5 4 1  
2 4 4 0  
3 3 4 -1  
4 2 4 -2  
  
Shipping Status Shipping Delay Description  
0 Shipped Earlier 1 day(s) early  
1 Shipped Late 1 day(s) late  
2 On Time On Time  
3 Shipped Earlier 1 day(s) early  
4 Shipped Earlier 2 day(s) early
```

	real day of shipping	scheduled days of shipment	Shipping Delay	Shipping Status	Shipping Delay Description
0	3	4	-1	Shipped Earlier	1 day(s) early
1	5	4	1	Shipped Late	1 day(s) late
2	4	4	0	On Time	On Time
3	3	4	-1	Shipped Earlier	1 day(s) early
4	2	4	-2	Shipped Earlier	2 day(s) early
...	...	...	...	...	...
995	2	4	-2	Shipped Earlier	2 day(s) early
996	6	2	4	Shipped Late	4 day(s) late
997	5	2	3	Shipped Late	3 day(s) late
998	2	4	-2	Shipped Earlier	2 day(s) early
999	6	4	2	Shipped Late	2 day(s) late

1000 rows × 5 columns

## 02

Adding new calculated column to help in analyzing the shipment performance

```
[ ] project['adj shipping date (order)']=pd.to_datetime(project['adj shipping date (order)'],errors='coerce')

[ ] print(project['adj shipping date (order)'])

→ 0    2018-02-03
  1      NaT
  2      NaT
  3      NaT
  4      NaT
...
995   2018-01-12
996      NaT
997      NaT
998   2017-12-06
999   2017-12-10
Name: adj shipping date (order), Length: 1000, dtype: datetime64[ns]
```

03

Changed the data type

Changing the data type of the date/time to match the time zone.

Filling the null date by the date of the day by using function fillna and timestamp.today

# DASHBOARD AND QUERY

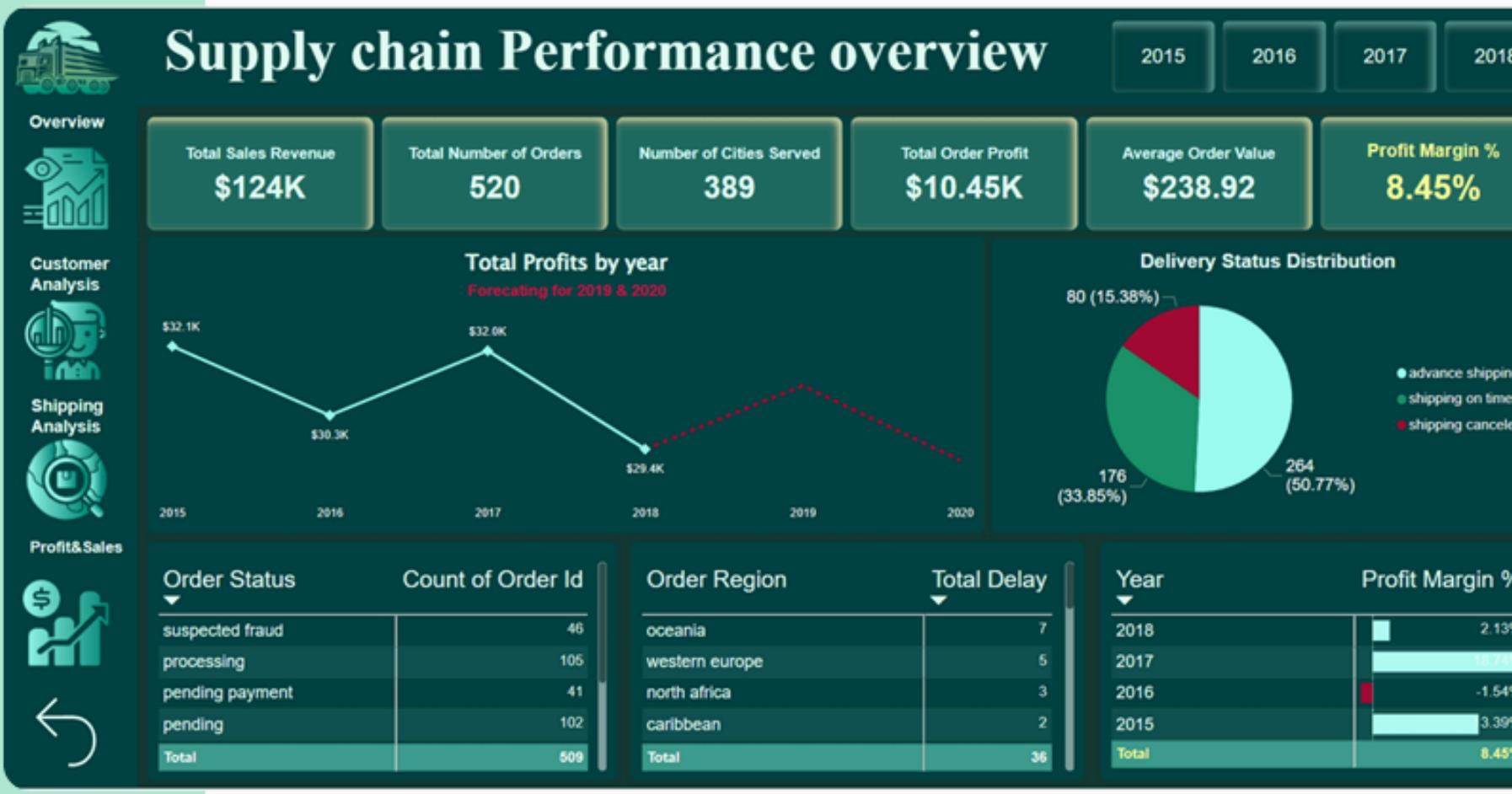


- Developed predictive questions based on historical data.
- Forecasted potential delivery delays and sales patterns.
- Used SQL and Power BI visuals to support forecasting.



# DASHBOARD AND VISUALS

09



# PAGE 1: OVERVIEW

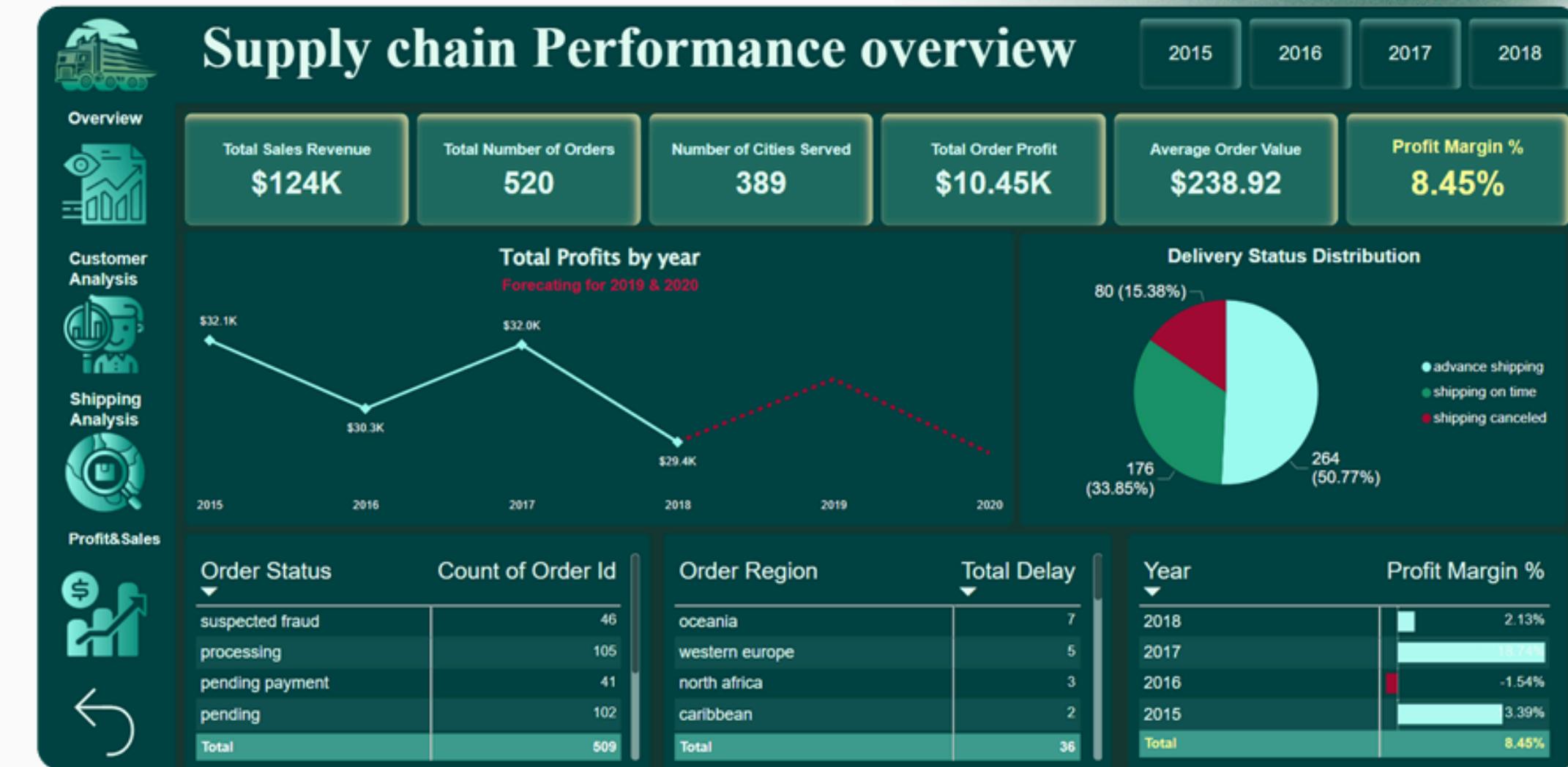
**PURPOSE:** SUMMARIZES OVERALL SUPPLY CHAIN PERFORMANCE FROM 2015–2018.

## Key Metrics:

- Total Sales Revenue: \$124K,
- Total Profit: \$10.45K .
- Total Orders: 520, Cities Served: 389 .
- Profit Margin: 8.45%, Average Order Value: \$238.92.

## Data Model:

Uses orders (sales, profit), shipping (delivery status, cities), and products (order details).



## Visuals:

- Line Chart: Profit trend (2015–2018) with 2019–2020 forecast.
- Pie Chart: Delivery status (33.85% on-time, 15.38% advanced, 50.77% canceled).
- Table: Order status (e.g., 102 pending, 41 payment pending).

## Insights:

- High cancellation rate (50.77%) signals shipping inefficiencies.
- Profit peaked in 2015 (\$29.44K), with a forecasted decline.
- 389 cities served, but delays in 36 regions need attention.

# PAGE 2: CUSTOMER ANALYSIS

**PURPOSE:** ANALYZES CUSTOMER REVENUE AND PREFERENCES BY SEGMENT AND CITY.

## Key Metrics:

- Revenue by City: Top cities (e.g., Caucasus \$1,340)
- Revenue by Segment: Consumer \$83K, Corporate \$18K, Home Office \$1K.
- Average Order Item Discount: 28.54% (Consumer), 20.51% (Corporate), 17.25% (Home Office).

## Data Model:

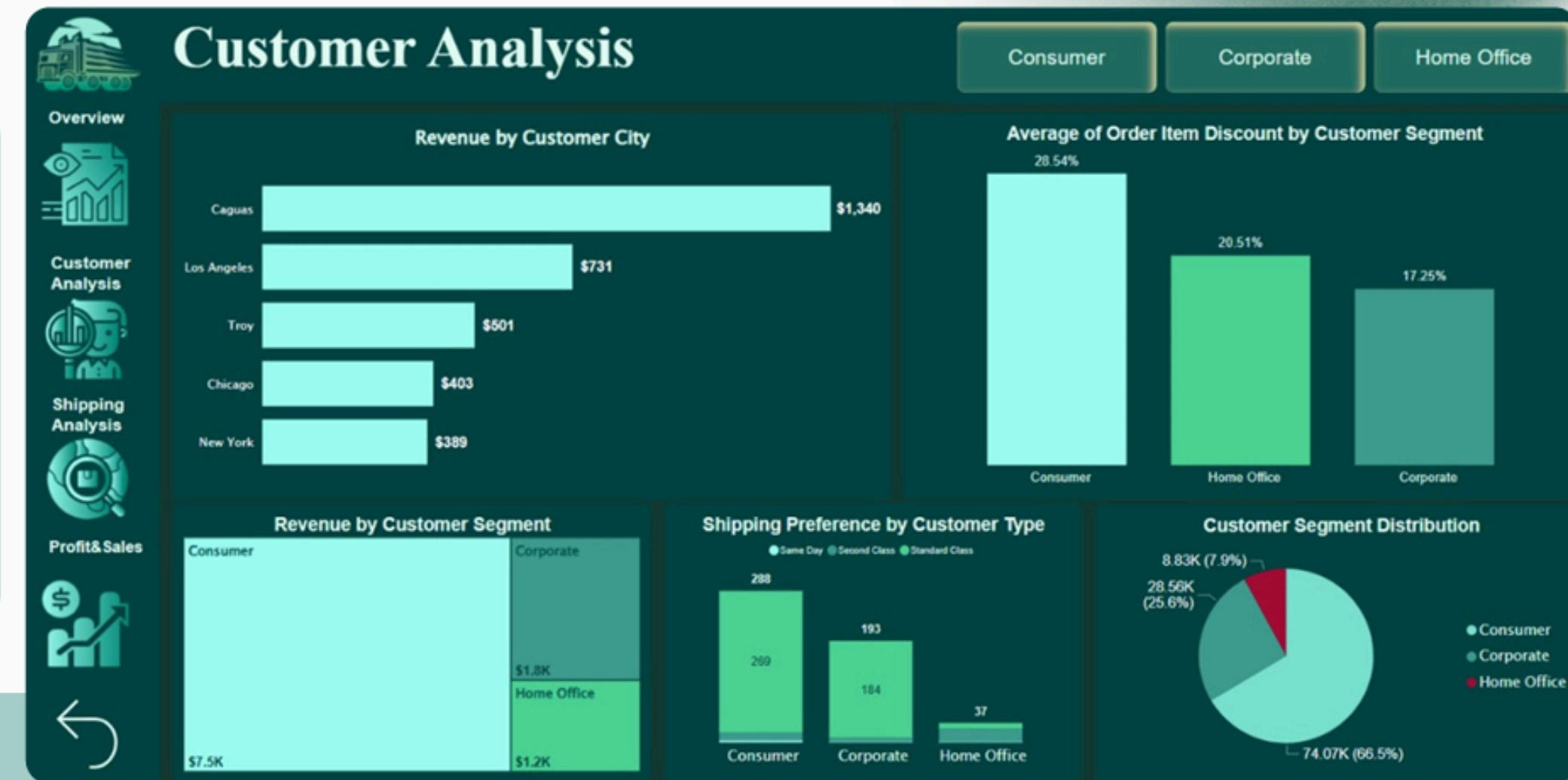
- Uses customers (segment), orders (revenue, discounts), and shipping (preferences).

## Visuals:

- Bar Chart: Revenue by customer city.
- Bar Chart: Shipping preference by customer type (e.g., Consumer \$32K, Corporate \$10K).
- Pie Chart: Customer segment distribution (Consumer 86%, Corporate 11%, Home Office 3%).

## Insights:

- Caucasus leads revenue, suggesting a key market focus.
- Consumer segment dominates (86%) with higher discounts, indicating pricing strategy impact.
- Shipping preference varies, with Consumer spending more than Corporate.



# PAGE 3: SHIPPING ANALYSIS

**PURPOSE:** EVALUATES SHIPPING EFFICIENCY AND DELAYS ACROSS REGIONS AND MODES.

## Key Metrics:

- Delivery Status Trends (2015–2018): Shipping on-time peaked at 80 in 2015, canceled at 58 in 2018.
- Shipping Delays by Region: Oceania (66 late), Western Europe (54 late).
- Delivery Status by Mode: Standard Class (243 canceled), Second Class (129 on-time).

## Data Model:

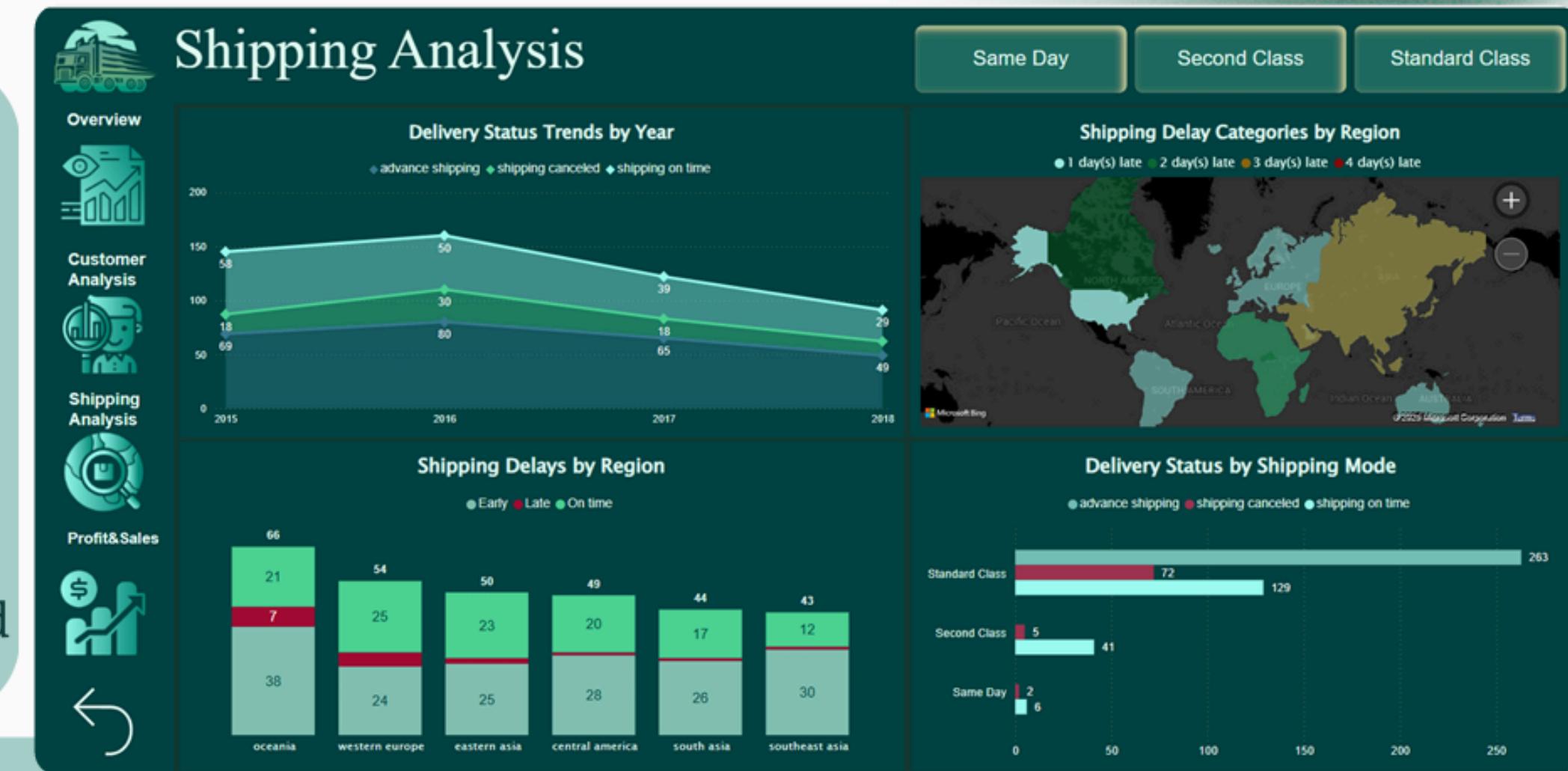
- Uses shipping (delivery status, region, mode) and orders (order details).

## Visuals:

- Line Chart: Delivery status trends over years.
- Map: Shipping delay categories by region (e.g., 1–4+ days late).
- Bar Chart: Delivery status by shipping mode (Same Day, Second Class, Standard Class).

## Insights:

- High cancellations in Standard Class (243) indicate potential inefficiencies.
- Oceania and Western Europe show significant delays, needing logistical improvements.
- On-time deliveries declined from 2015 to 2018, signaling a trend to address.



# PAGE 4: PROFIT AND SALES

**PURPOSE:** ANALYZES REVENUE, PROFIT, AND LOSSES BY CATEGORY AND PAYMENT METHOD.

## Key Metrics:

- Revenue by Category: Cardio Equipment (\$29K), Cleats (\$25K), Women's Apparel (\$15K).
- Profit by Category: Cleats (\$3.5K, 40% discount), Cardio Equipment (\$2K, 34% discount).
- Revenue Loss: \$82.9K due to cancellations, \$10K due to suspected fraud.

## Data Model:

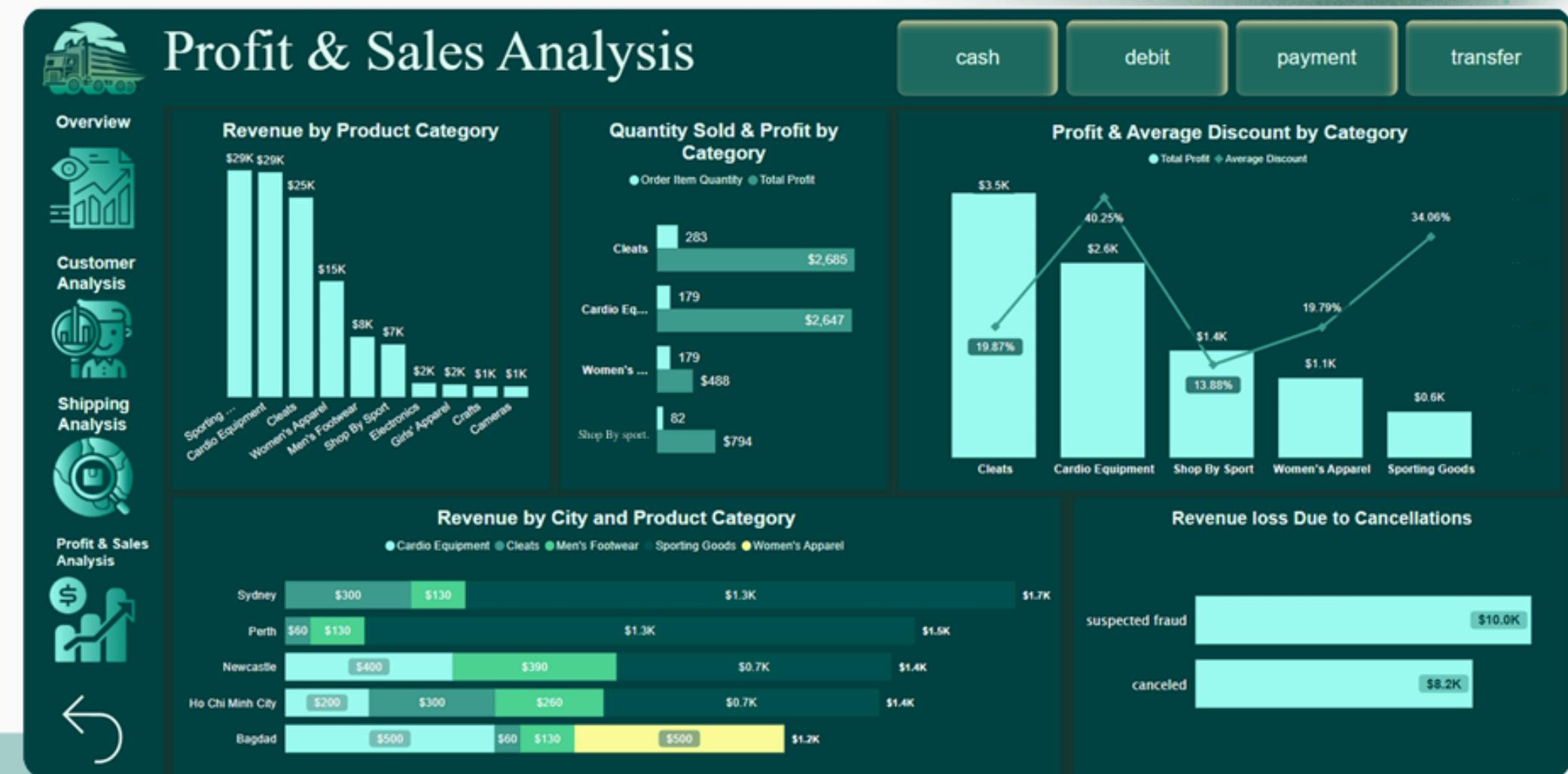
- Uses products (category), orders (revenue, profit, discounts), and shipping (cancellations).

## Visuals:

- Bar Chart: Revenue by product category.
- Combo Chart: Quantity sold and profit by category.
- Bar Chart: Revenue loss due to cancellations and fraud.

## Insights:

- Cardio Equipment leads revenue, but Cleats have the highest profit margin.
- High cancellation losses (\$82.9K) highlight the need for better order fulfillment.
- Discounts (e.g., 40% on Cleats) may be driving sales but impacting overall profit.



-----max profitable country---

```
select c.Customer_Country , max (o.sales) as 'max sales per country'
from [Customer DIM] c
join [Order fact table] o
on c.Customer_Id = o.Order_Customer_Id
group by c.Customer_Country
```

	Customer_Country	max sales per country
1	Puerto Rico	499.950012207031
2	USA	499.950012207031

-----Total sales-----

```
select SUM(Sales) as total_sales
from [Order fact table]
```

-----highest price in each cat-----

```
select Category_Name , MAX(Product_Price) as higest_price_product
from [product DIM]
group by Category_Name
```

-----non profit orders by the same customer-----

```
select Order_Id ,sum(Benefit_per_order) as non_profit_order
from [Order fact table]
where Benefit_per_order <0
group by Order_Id
```

## SQL QUERY FINANCIAL INSIGHT

	total_sales
1	123758.241622925

	Category_Name	higest_price_product
1	Accessories	24.9899997711182
2	Baby	59.0800018310547
3	Baseball & Sof...	59.9900016784668
4	Boxing & MMA	89.9899978637695
5	Cameras	452.040008544922
6	Cardio Equipm...	99.9899978637695
7	Children's Clot...	357.100006103516
8	Cleats	59.9900016784668

	Order_Id	non_profit_order
1	1756	-74.3899993896484
2	2203	-18.8099994659424
3	8636	-22.5
4	8678	-783.669982910156
5	9063	-190.649993896484
6	10014	-107.870002746582

-----processing time-----

```
select DATEDIFF(DAY, o.order_date_DateOrders,s.adj_shipping_date_order) as 'Processing time',o.Order_Id as 'order Id'
from [Order fact table] o
join [DIm shipping] s
on o.Order_Id =s.order_id
where DATEDIFF(DAY, o.order_date_DateOrders,s.adj_shipping_date_order) between 1 and 100
```

	Processing time	order Id
1	3	367
2	2	1991
3	4	2203
4	3	2332
5	4	2428
6	4	3828
7	3	4427
8	3	4487
9	4	5893
10	2	6176
11	4	6358
12	3	6522
13	2	6776
14	4	8123

# CALCULATING ORDER PROCESSING TIME AND LOSS

-----total loss-----

```
select sum(Benefit_per_order) as 'Total loss from orders'
from [Order fact table]
where Benefit_per_order <0
```

-----receipt-----

```
select p.Product_Name,p.product_id,p.Product_Price,c.Customer_Name as customer_name , c.customer_city,o.order_id,o.order_date_DateOrders
from [product DIM] p
join [Order fact table] o on p.product_id = o.product_id
join [Customer DIM] c on c.Customer_Id= o.Order_Customer_Id
join [DIm shipping] s on o.Order_Id=s.order_id
```

	Total loss from orders
1	-15900.7799220085

	Product_Name	product_id	Product_Price	customer_name	customer_city	order_id	order_date_DateOrders
1	Perfect Fitness Perfect Rip Deck	55831	59.9900016784668	Megan Smith	Lawrence	22297	2015-11-22
2	Nike Men's Dri-FIT Victory Golf Polo	131932	50	Mary Vincent	Caguas	52772	2017-02-09
3	Nike Men's Free 5.0+ Running Sh...	86490	99.9899978637695	Lori Fuller	Caguas	34631	2016-05-20
4	Nike Men's CJ Elite 2 TD Football ...	73253	129.990005493164	Mary Henry	Atlanta	29283	2016-03-03



-----product that no one ordered-----

```
select p.product_id,p.Product_Name,p.Category_Name,o.order_id
from [product DIM] p
left join [Order fact table] o
on o.product_id=p.product_id
where o.Order_Id is null
```

----- profit/loss per segment-----

```
select o.Order_Id,sum (o.Order_Profit_Per_Order) as 'Income per Segment',c.Customer_Segment,c.Customer_City
from [Order fact table] o
join [Customer DIM] c
on o.Order_Customer_Id=c.Customer_Id
group by c.Customer_Segment,c.Customer_City,o.Order_Id
```

-----total sales by customer country-----

```
select c.Customer_Country, p.Product_Name, SUM(o.Sales) as Total_Sales
from [Order fact table] o
join [Customer DIM] c on Order_Customer_Id = c.Customer_ID
join [product DIM] p on o.Product_ID = p.Product_ID
group by c.Customer_Country, p.Product_Name
order by c.Customer_Country, Total_Sales DESC
```

## ANALYZING SALES BY CUSTOMER & LOSS BY PRODUCT

	product_id	Product_Name	Category_Name	order_id
1	104577	Nike Men's CJ Elite 2 TD Football Cleat	Men's Footwear	NULL
2	113282	Team Golf Tennessee Volunteers Put...	Accessories	NULL

	Order_Id	Income per Segment	Customer_Segment	Customer_City
1	367	23.3899993896484	Corporate	Hayward
2	906	10.78999961853	Corporate	Sacramento
3	973	83.9800033569336	Consumer	Caguas
4	1386	13.260002288818	Consumer	Caguas
5	1756	-74.3899993896484	Corporate	Newark
6	1797	31.8799991607666	Consumer	Caguas
7	1991	65.5100021362305	Corporate	Mililani
8	2203	-18.8099994659424	Consumer	Caguas

	Customer_Country	Product_Name	Total_Sales
1	Puerto Rico	Smart watch	10160.25
2	Puerto Rico	Nike Men's Free 5.0+ Running Shoe	9099.09015655518
3	Puerto Rico	Nike Men's Dri-FIT Victory Golf Polo	7700
4	Puerto Rico	Perfect Fitness Perfect Rip Deck	5099.15018844604
5	Puerto Rico	Under Armour Girls' Toddler Spine ...	3119.21997070313
6	Puerto Rico	Web Camera	1356.12002563477
7	Puerto Rico	Nike Men's CJ Elite 2 TD Football ...	779.940032958984
8	Puerto Rico	adidas Youth Germany Black/Red	700



# CHALLENGES AND SOLUTIONS

## Issues:

- Faced issues with inconsistent data formats (especially date) and incorrect mappings , some problem with the file in code (UTF-8).

## Resolved using:

- Power Query python for date formatting.
- SQL for fixing country names..



وزارة الاتصالات  
وتكنولوجيا المعلومات



# THANK YOU!!

PRESENTED BY:

Mariam Yasser - Zahra Mahmoud - Alaa Walid - Aya Ahmed - Rawan Ayman