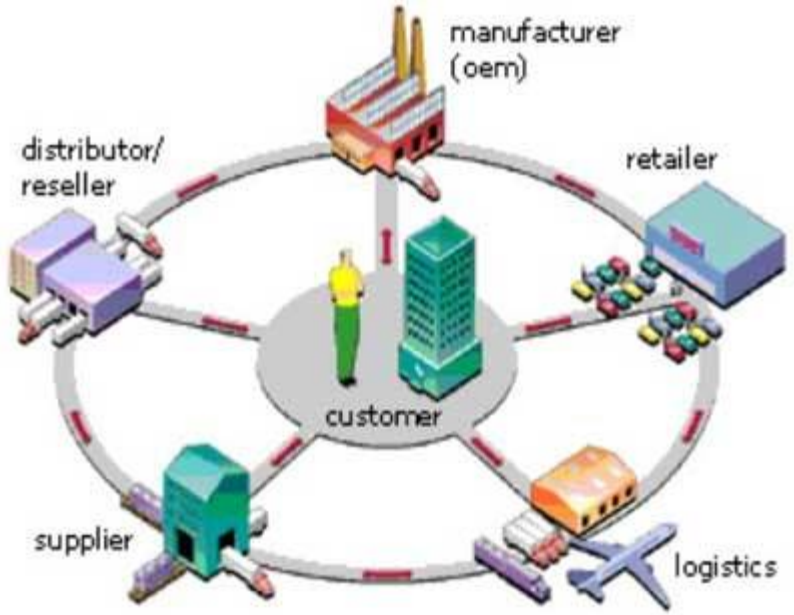
Supply Chain Management Analysis – Dataco Dataset



By B.Soundarya

INTRODUCTION

- **Supply chain management** involves optimizing product flow from suppliers to customers, ensuring efficiency and customer satisfaction.
- The project analyzes **DataCo’s supply chain data** to identify trends and inefficiencies.
- **Objective**: Improve decisions by understanding product performance and delivery insights.

PROBLEM STATEMENT

- DataCo faces challenges in managing its supply chain efficiently.
- Issues include:
    - Late deliveries causing customer dissatisfaction.
    - Fraudulent transactions affecting profit margins.
    - Inaccurate sales predictions leading to inventory issues.
- These challenges negatively impact overall business performance.

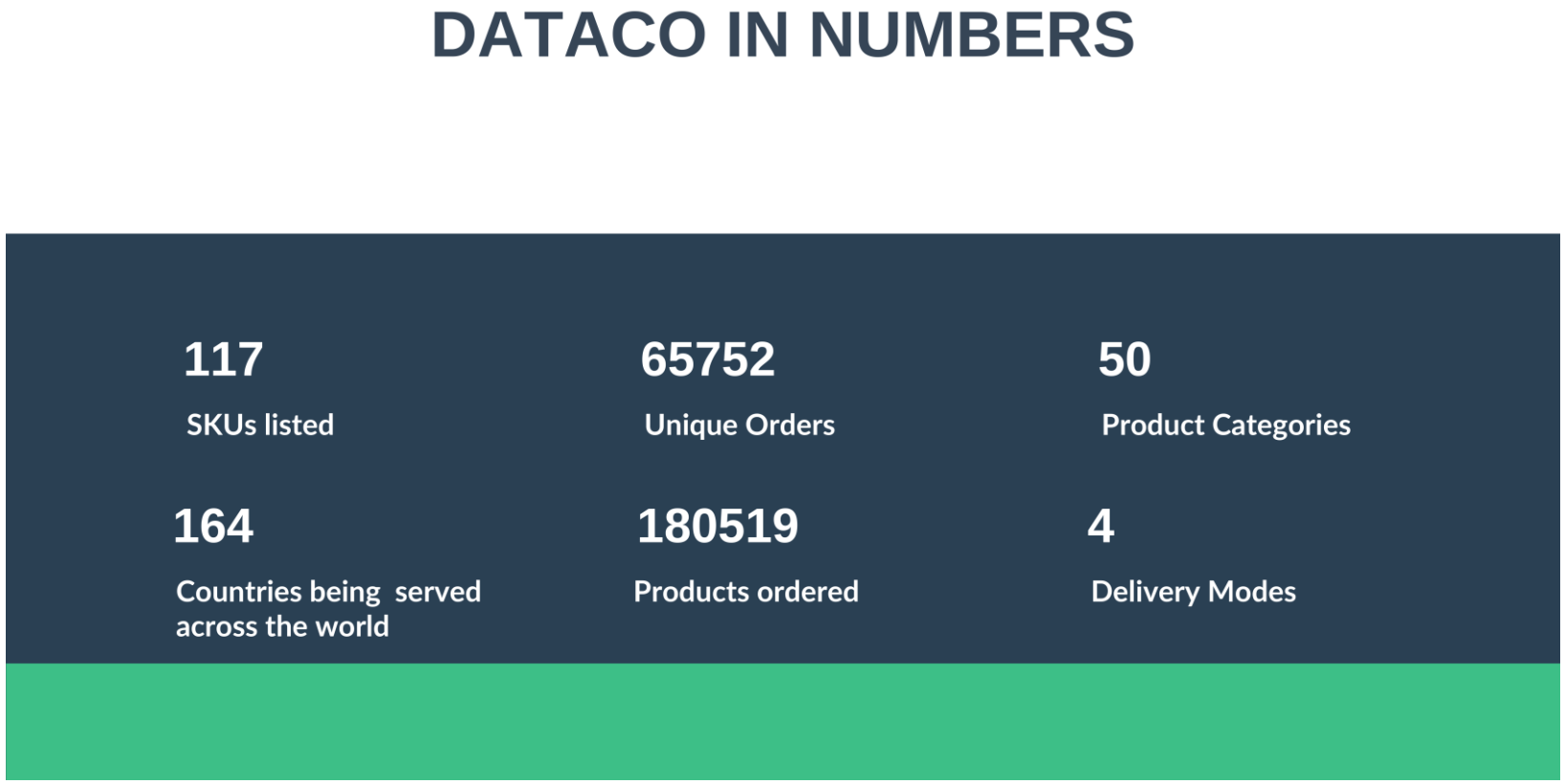
Goal:

- Improve supply chain efficiency by identifying key risks and optimizing operations.

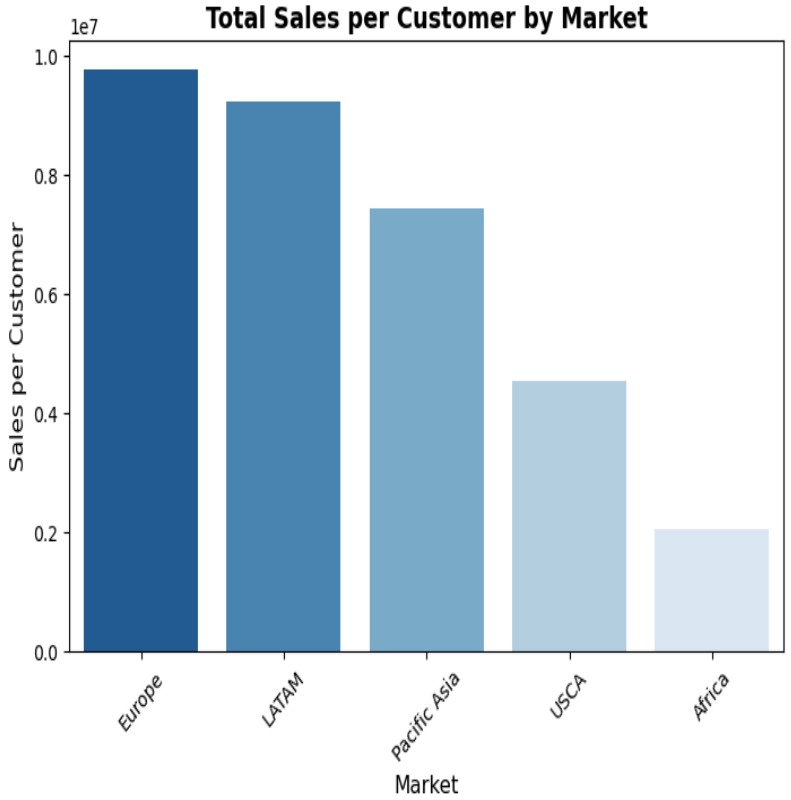
PROJECT OBJECTIVE

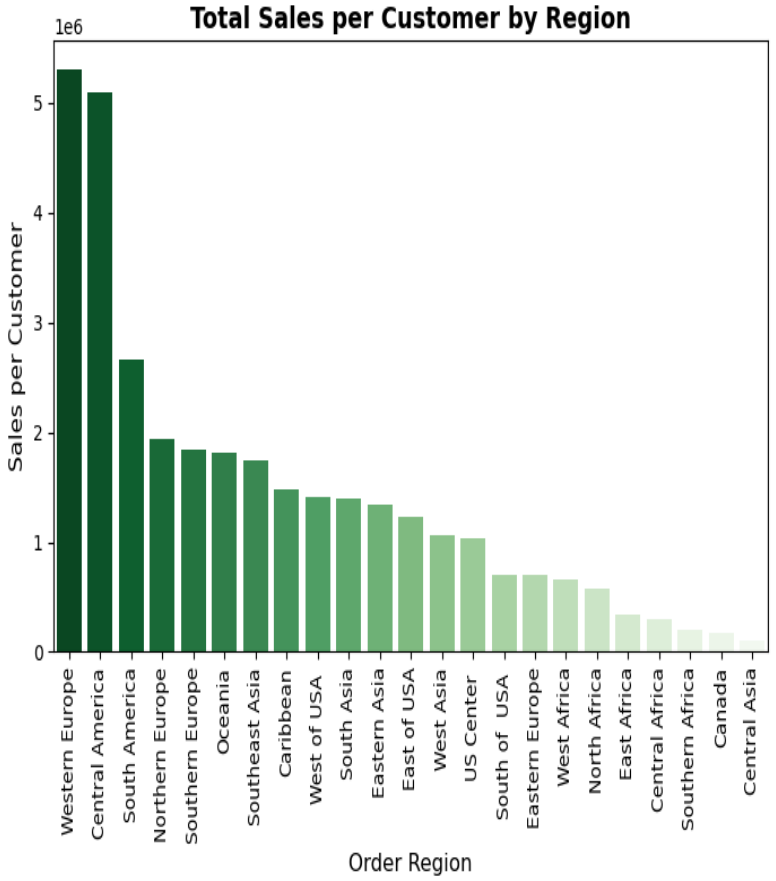
- Improve demand forecasting to optimize inventory levels.
- Detect fraudulent transactions and mitigate risks.
- Identify sales trends to enhance product and regional performance insights.
- Reduce delivery delays and improve customer satisfaction.
- Support data-driven decision-making for supply chain management.

DATASET OVERVIEW



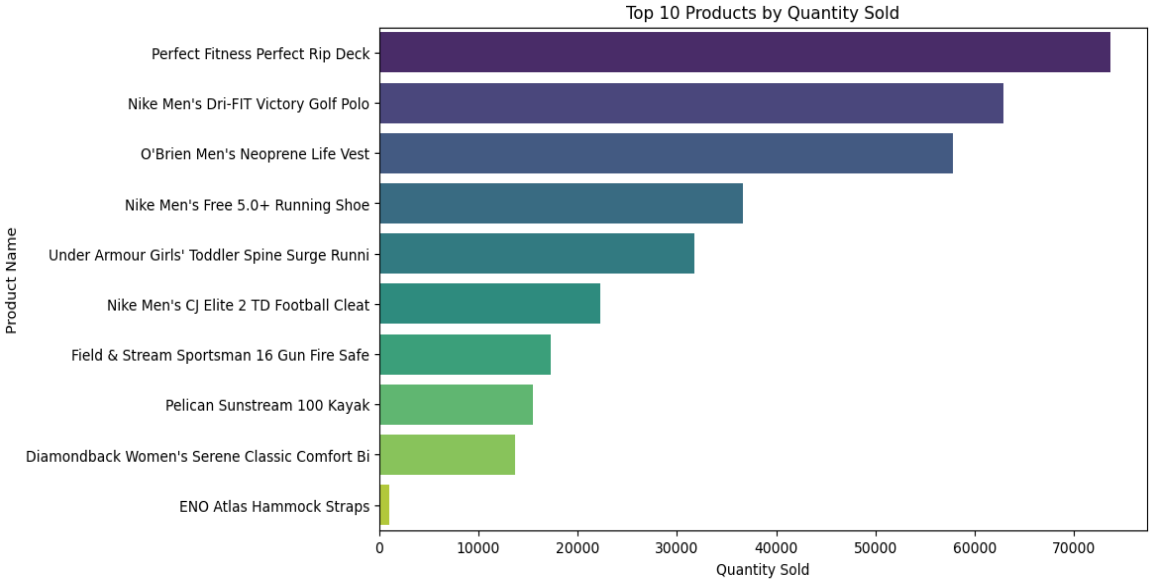
MARKET SHARE BY REGION



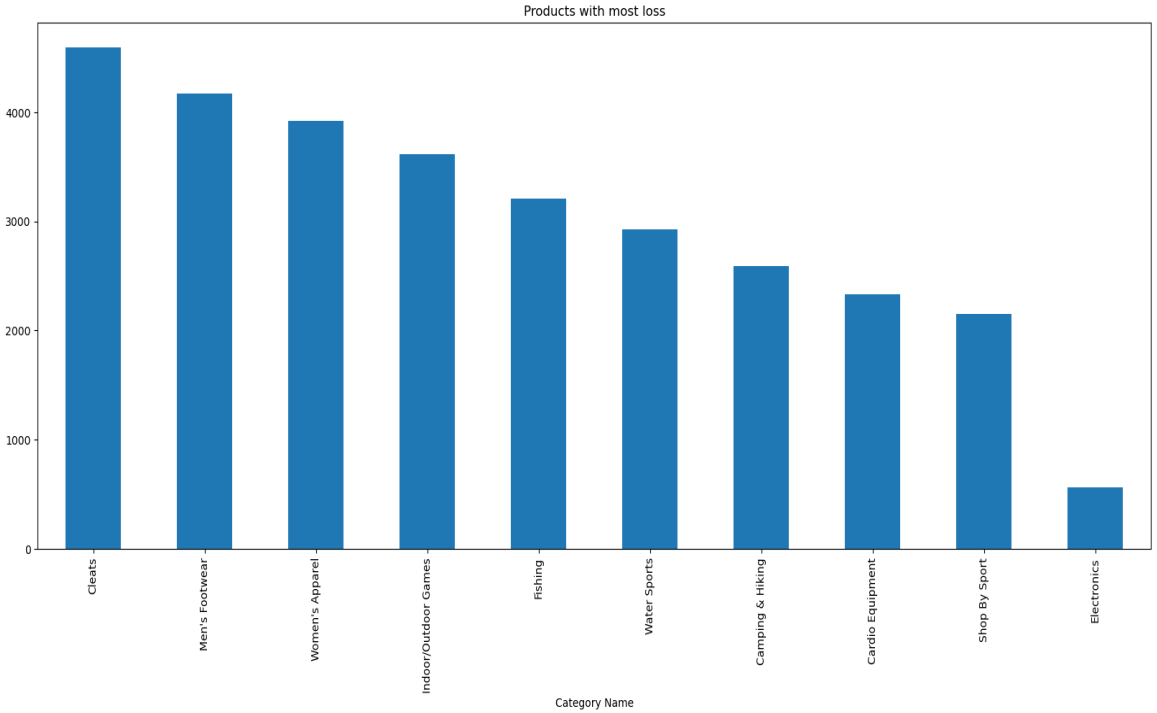


- **Europe** and **LATAM** lead in sales per customer.
- **Western Europe** and **Central America** dominate regional sales.
- **Africa** and **USCA** show lower sales, indicating potential for growth.

PRODUCTS WITH HIGHEST AND LOWEST PROFIT



- **Profitable Products:** Fitness, sports gear, and apparel lead in sales, especially Nike and Under Armour items.



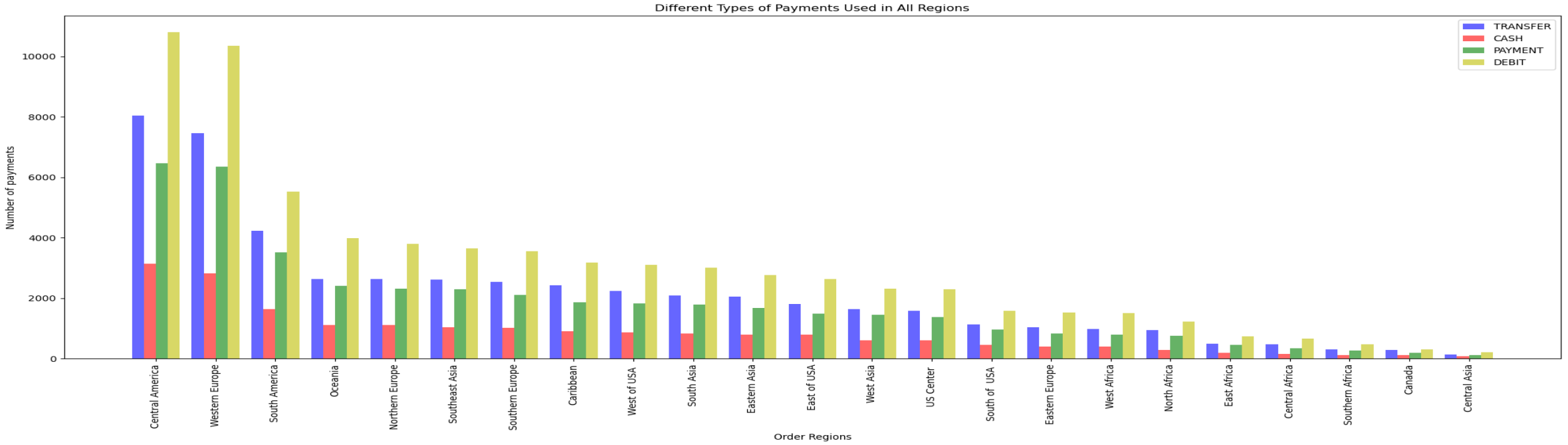
- **Loss-Making Categories:** Cleats, footwear, and apparel have the highest losses, signaling potential overstock or pricing issues.

SALES TREND BY QUARTER AND TIME

Actions: Focus marketing efforts on Fridays, peak hours, and seasonal months like November.

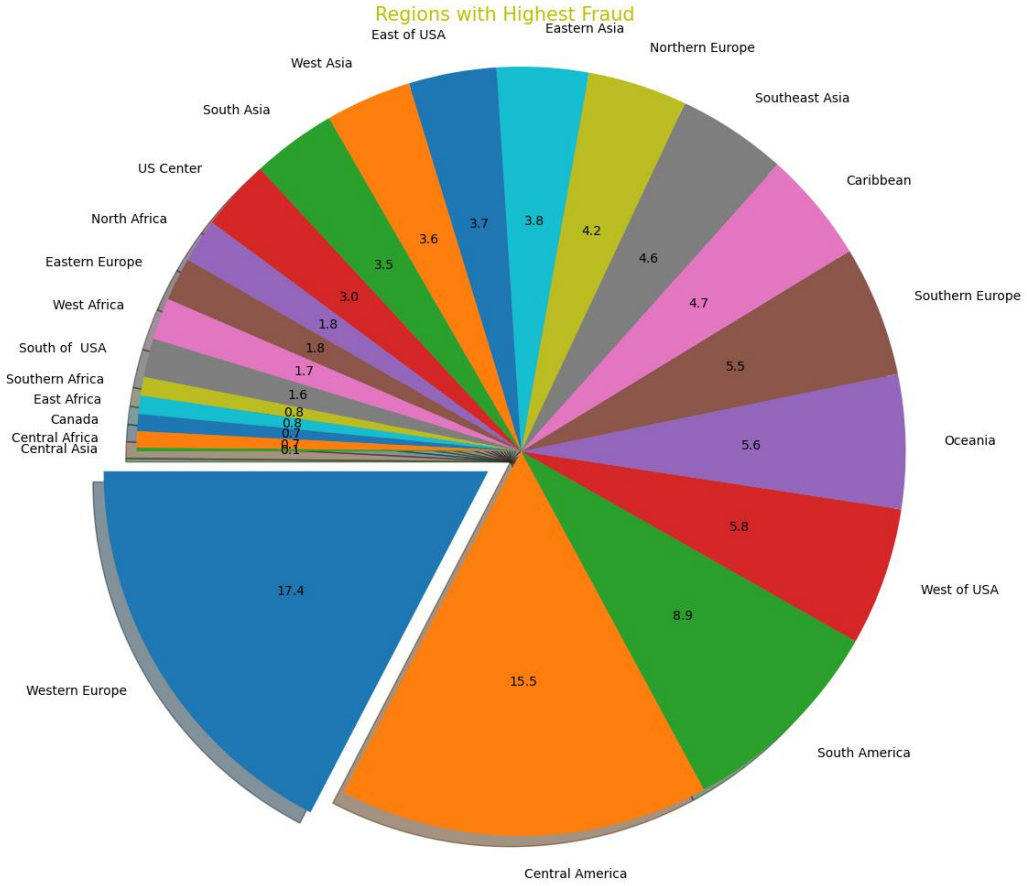
- **Yearly Trend:** Sales peaked in 2017, followed by a decline in 2018.
- **Weekly Pattern:** Highest sales occur on Fridays, while Tuesdays have the lowest sales.
- **Hourly Trends:** Early morning and late afternoon show higher sales activity.
- **Seasonal Peaks:** Sales spike in November, indicating holiday or promotional influence.

PAYMENT MODE



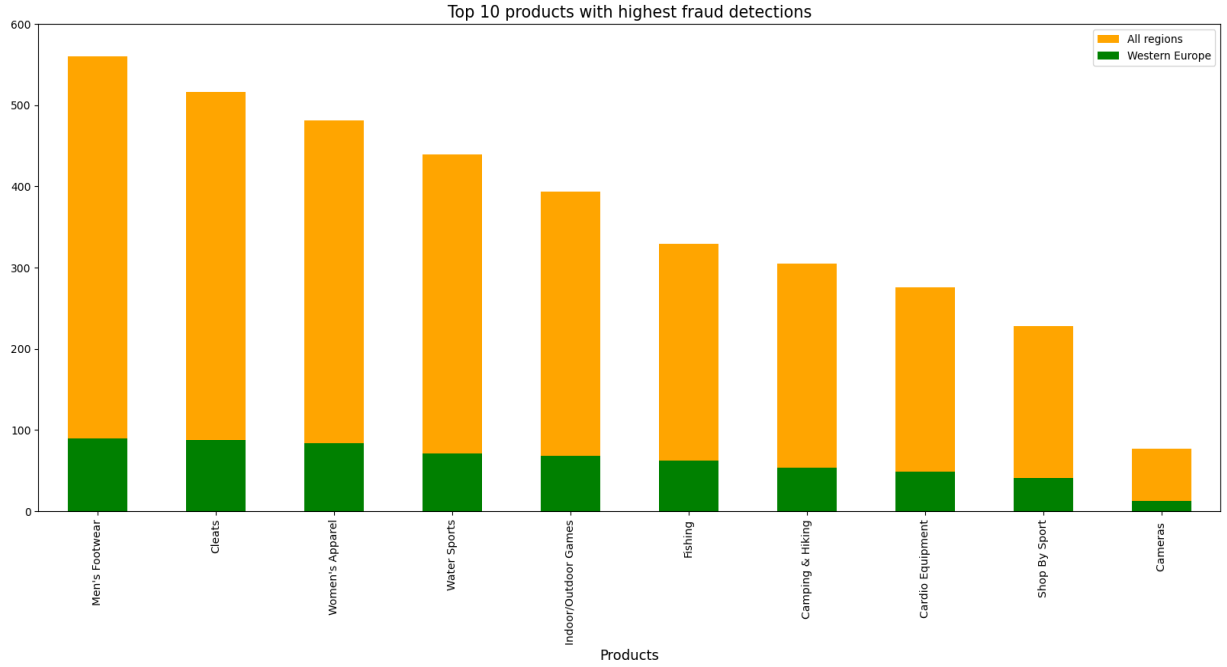
- **Debit** is the top payment mode across regions.
- **Transfers** are widely used in Central and Western Europe.
- **Cash** is least used overall, except in South Asia and Africa.
- Regional diversity indicates a need for tailored payment options.

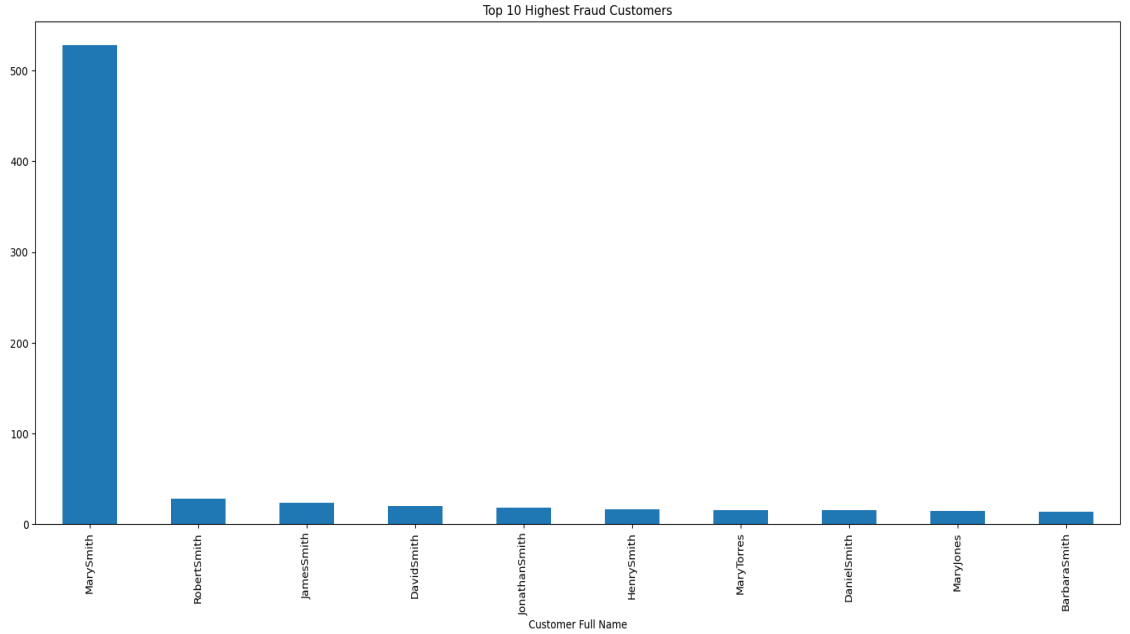
SUSPECTED FRAUD ANALYSIS



- Most fraud happens in **Western Europe, Central America,** and **South America**
- Fraud patterns vary across regions—focus on high-risk areas for fraud detection.
- Targeted strategies are needed to mitigate fraud in top-affected regions

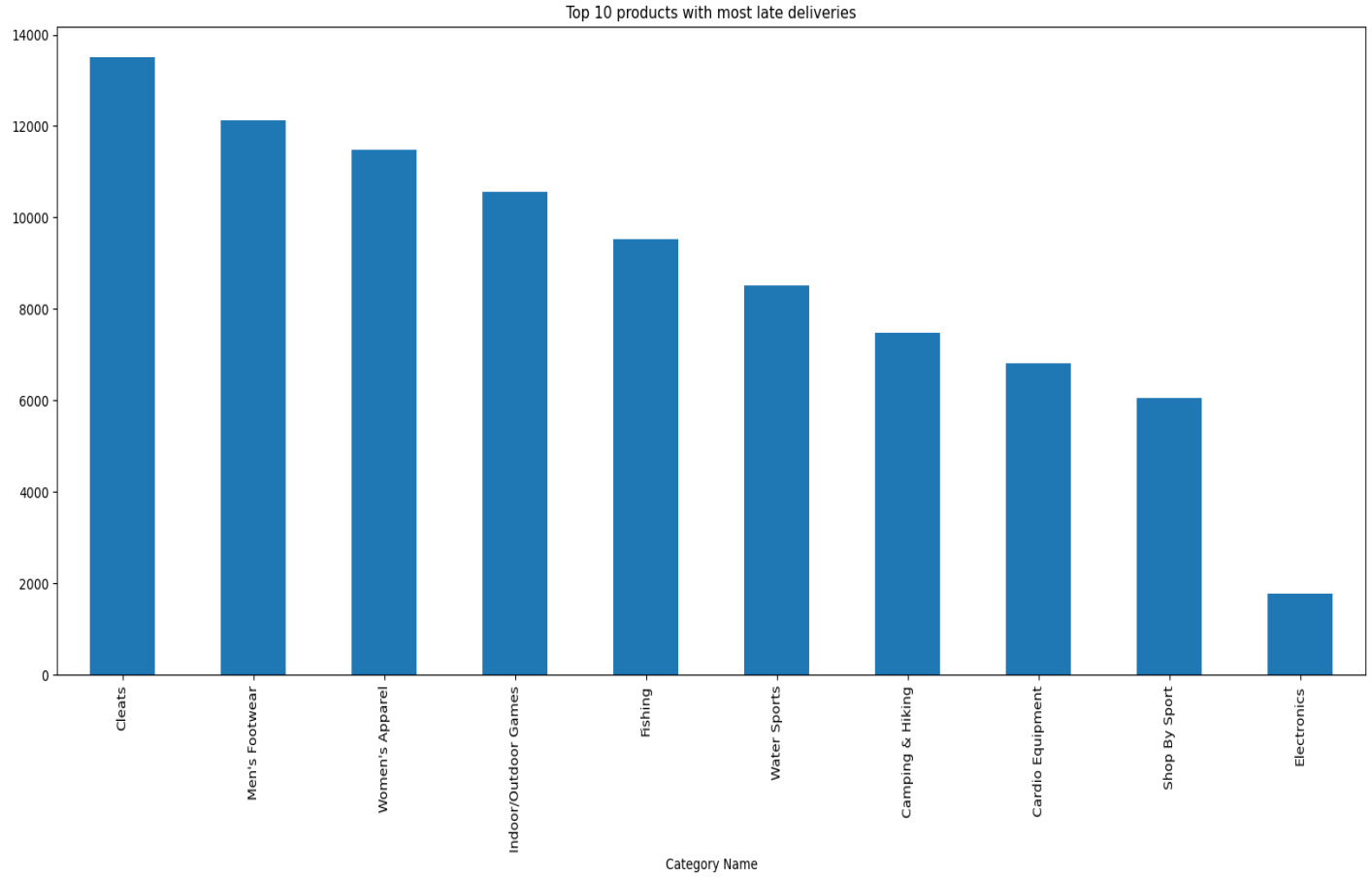
SUSPECTED FRAUD ANALYSIS





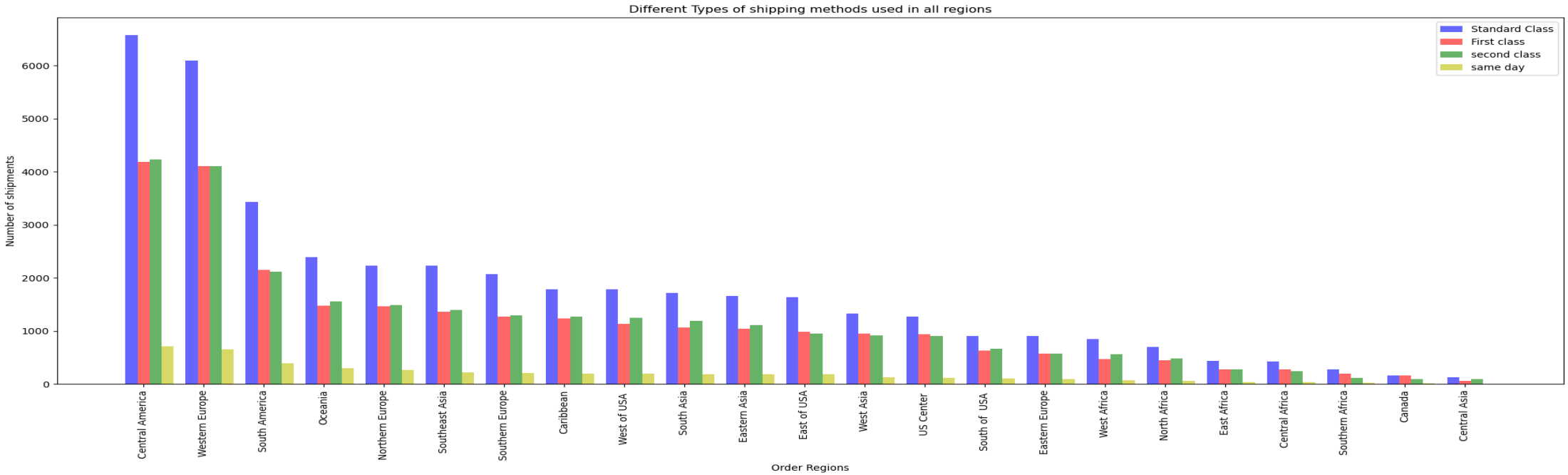
- **Product Fraud**: Men's footwear and cleats are the most targeted items, particularly in Western Europe, suggesting these areas require enhanced fraud detection measures.
- **Customer Fraud**: Mary Smith shows an unusually high number of fraud cases compared to others, indicating a need for closer monitoring.

LATE DELIVERIES BY PRODUCT CATEGORY



- **Cleats** experienced the highest number of late deliveries, followed by **Men’s Footwear** and **Women’s Apparel.**
- **Electronics** had the fewest late deliveries among the top 10 categories.

LATE DELIVERIES BY SHIPPING METHOD

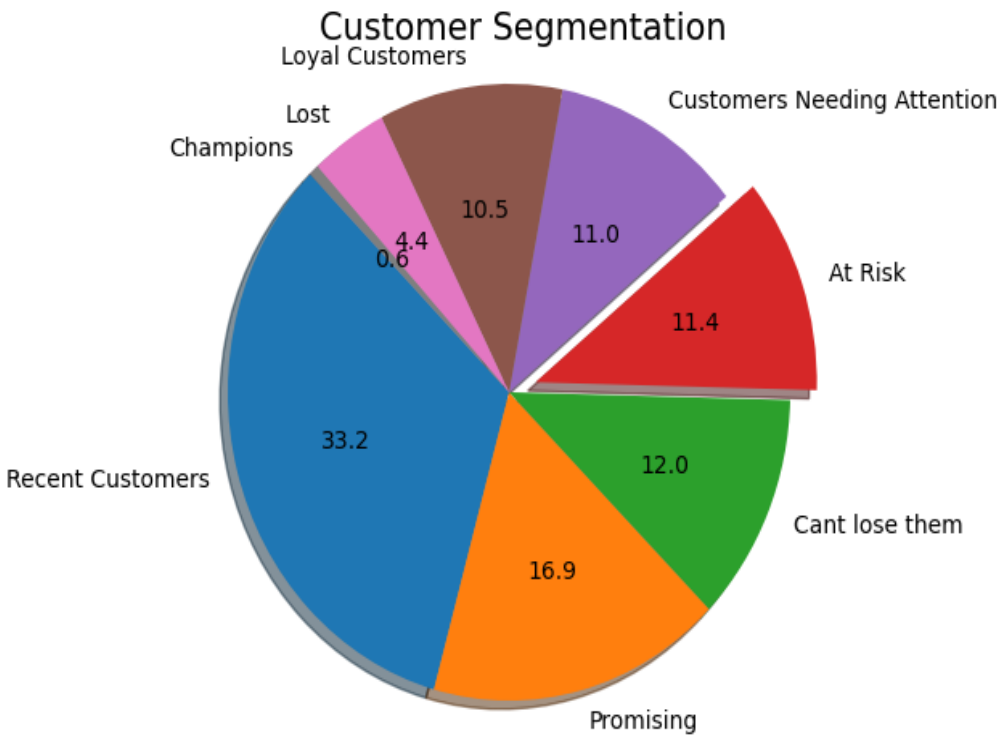


- **Standard Class** had the most late deliveries across all regions, while **Same-Day** shipping had the fewest.
- **First** and **Second Class** had similar late delivery counts

RFM ANALYSIS FOR CUSTOMER SEGMENTATION

- **RFM** (Recency, Frequency, Monetary) analysis is a customer segmentation technique used to evaluate customer behavior. It measures:
    1. **Recency:** How recently a customer made a purchase.
    2. **Frequency:** How often they make purchases.
    3. **Monetary:** How much they spend.

CUSTOMER SEGMENTATION



Actionable Insights:

1. **Loyal Customers (10.5%)**: Strengthen relationships with targeted loyalty programs and exclusive offers.
2. **Champions (0.6%)**: Reward with VIP treatment and involve them in referral programs.
3. **Promising (16.9%)**: Encourage more purchases through personalized offers and discounts.
4. **Can't Lose Them (12.0%)**: Take proactive measures like personalized re-engagement campaigns to retain.
5. **Customers Needing Attention (11.0%)**: Offer special deals and personalized messages to re-engage.
6. **At Risk (11.4%)**: Use urgent retention strategies, such as discounts or follow-up calls.
7. **Lost Customers (4.4%)**: Reactivate with one-time win-back offers or gather feedback to understand reasons for churn.

MODEL TRAINING OVERVIEW

- The machine learning models were trained to achieve two goals:
  - **Regression:** Predict sales and order quantities.
  - **Classification:** Detect fraud and identify late deliveries.

- Models Used:
  - **Regression**: Linear Regression, Ridge, Lasso, Random Forest, etc.
  - **Classification**: Logistic Regression, Decision Tree, Random Forest, XGBoost, etc.

MODEL PERFORMANCE COMPARISON(REGRESSION)

	Regression Model	MAE Value for Sales	RMSE Value for Sales	MAE Value for Quantity	RMSE Value for Quantity
0	Lasso	1.5500	2.3300	0.9000	1.030
1	Ridge	0.7500	0.9700	0.3400	0.520
2	Random Forest	0.8700	3.6790	0.0040	0.040
3	eXtreme gradient boosting	0.1500	0.5100	0.0008	0.008
4	Decision tree	0.0130	0.8700	3.6900	0.006
5	Linear Regression	0.0005	0.0014	0.3400	0.520

BEST MODEL INSIGHT:

- **Sales Prediction:**
  - The best model is **Linear Regression** with an MAE of **0.0005** and an RMSE of **0.0014** for Sales.
- **Quantity Prediction:**
  - The best model is **Decision Tree** with an MAE of **0.0040** and an RMSE of **0.006** for Quantity.

MODEL PERFORMANCE COMPARISON(CLASSIFICATION)

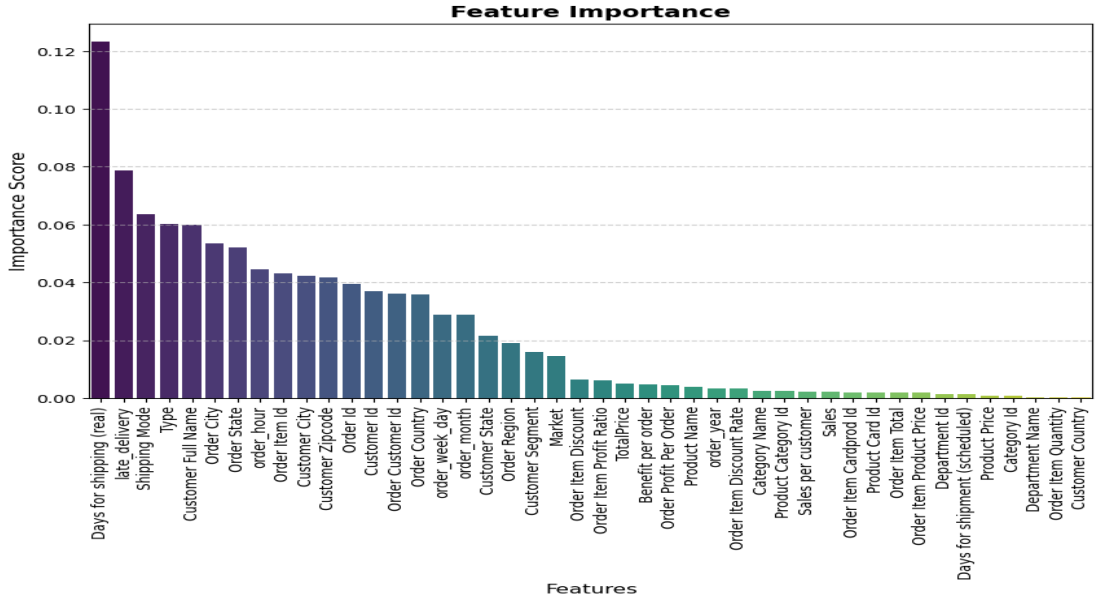
	Classification Model	Accuracy Score for Fraud Detection	Recall Score for Fraud Detection	F1 Score for Fraud Detection	Accuracy Score for Late Delivery	Recall Score for Late Delivery	F1 Score for Late Delivery
0	Logistic	97.80	59.40	31.22	98.84	97.94	98.96
1	Gausian Naive bayes	87.84	16.23	27.92	57.27	56.20	71.95
2	Support Vector Machines	97.75	56.89	28.42	98.84	97.94	98.96
3	K nearest Neighbour	97.36	41.90	35.67	80.82	83.45	82.26
4	Random Forest	98.66	98.93	60.79	99.60	98.52	99.74
5	eExtreme gradient boosting	99.03	92.39	75.86	99.13	98.45	99.21
6	Decision tree	99.08	81.43	80.16	99.37	99.41	99.42

BEST MODEL INSIGHT:

- **Fraud Detection:** Random Forest provides the best balance with high Recall (98.93%) and Accuracy (98.66%).
- **Late Delivery:** Decision Tree is the most effective, achieving the highest Accuracy (99.37%) and F1 Score (99.42%).

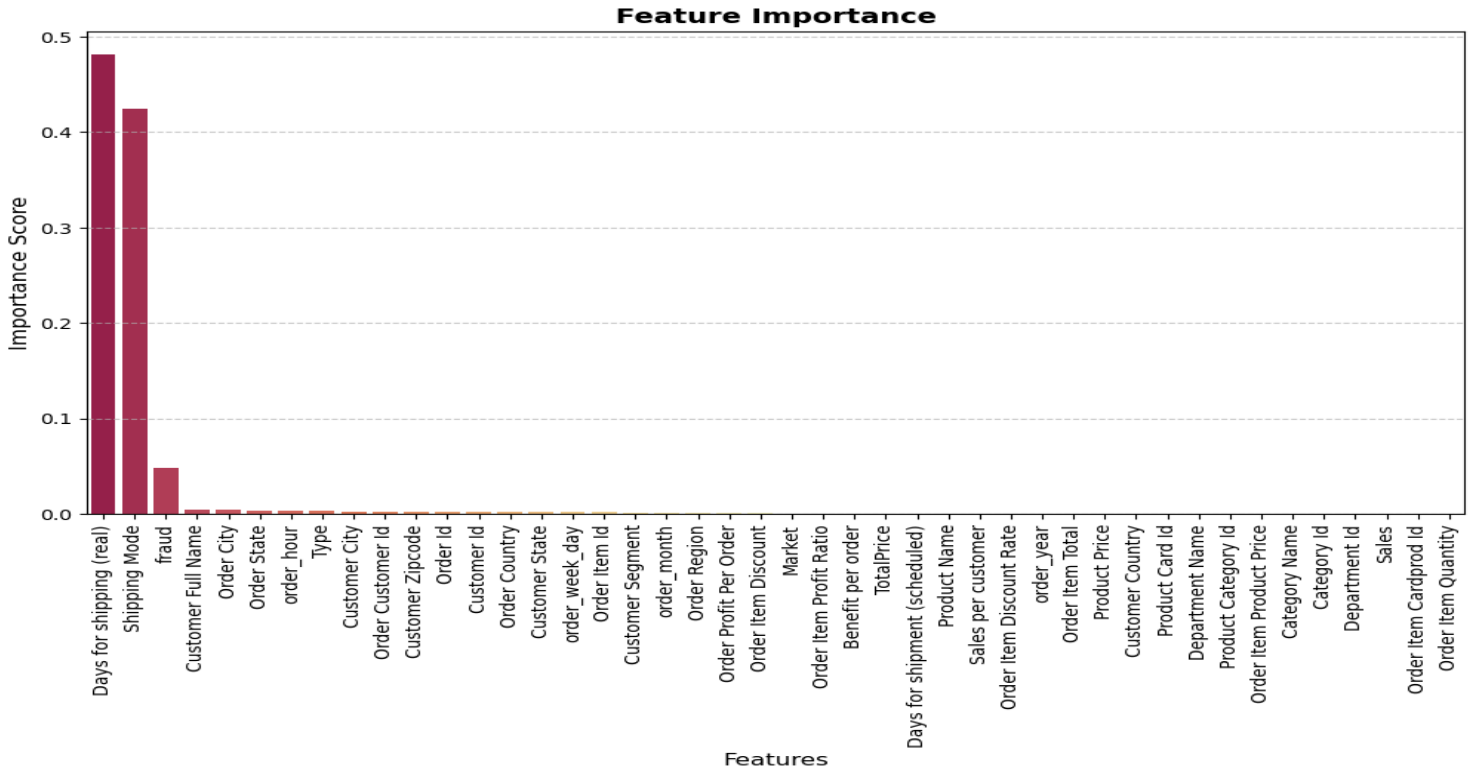
MODEL IMPROVEMENT TECHNIQUE

- **Cross-Validation:** Implemented to ensure the model generalizes well and is not overfitting.
- Feature Importance helps us identify which variables significantly impact the model's predictions.



- Days for Shipping (Real)" has 0.12 importance in fraud detection.
- Customer Name, Shipping Mode, and Payment Type (0.7) help detect repeat fraud.

MODEL IMPROVEMENT TECHNIQUE

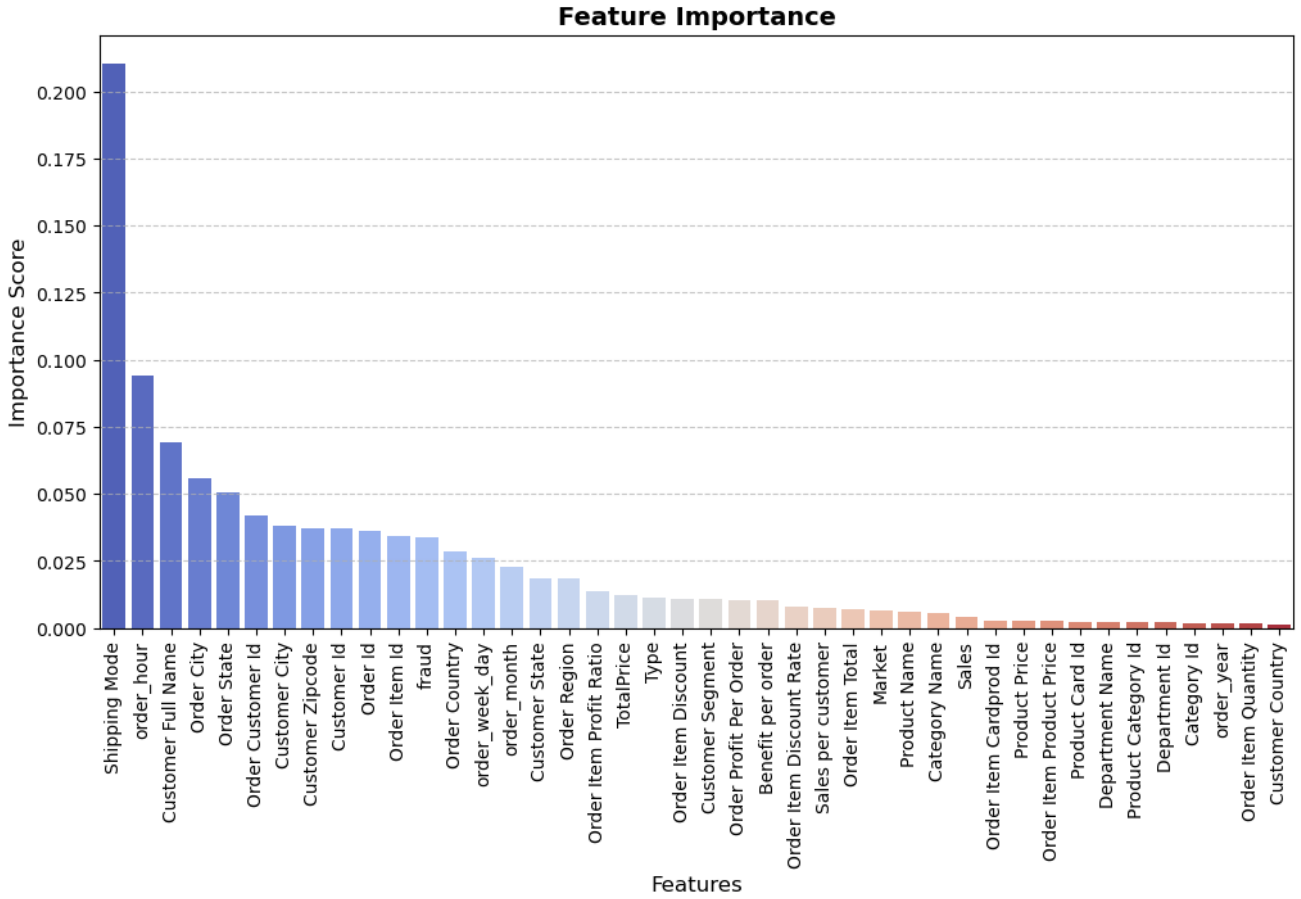


- Days for Shipping (Real) heavily influences late deliveries.
- Shipping Mode and Fraud impact delivery delays.

FINAL MODEL SELECTION

- **Best Model Identified: Decision Tree Classifier** outperformed other ML models for classification tasks.
- Selection Criteria:
  - **High Accuracy & F1-score (84%)** even after feature removal.
  - **Interpretability & Business Impact:** Clearly shows key factors affecting delivery efficiency
    - **Trade-offs Considered:** While complex models like XGBoost were tested, Decision Tree was preferred for its **transparency and real-world interpretability**.

FEATURE IMPORTANCE AFTER REMOVING SHIPPING DAYS



- **Model Performance:** Even without shipping days, accuracy remains at **84%**
- **Key Factors Now:** Shipping Mode, Order City, and Order State gain importance.
- **Business Impact:** Helps optimize faster delivery methods

CONCLUSION

- The dataset used in this analysis is from 2015-2018, before the COVID-19 pandemic. However, supply chain dynamics have changed significantly post-COVID, with increased disruptions, demand fluctuations, and digital transformation.
- To bridge this gap, I am comparing my findings with recent supply chain challenges and suggesting modern strategies to improve efficiency.

Supply Chain Comparison: 2015-2018 vs. Post-COVID

2015-2018 Findings	Post-Covid Changes	How to Improve?(Solutions)
Sales were highest in Europe & Central America. Africa & USCA had lower sales.	Demand changed due to economic instability and supply chain issues.	Use AI to track demand and adjust inventory based on trends.
Nike & Under Armour products made the most profit. Cleats & Footwear had the most losses	After COVID, people's shopping habits changed, affecting profits.	Set better pricing strategies and offer discounts based on demand.
Loyal customers (11%) spent the most, but 15% of customers were lost.	Customer buying behavior changed, making retention harder.	Use personalized offers, emails, and loyalty programs to keep customers.
Most fraud cases happened in Western Europe & South America. Men’s footwear was highly targeted	Fraud increased after COVID as more people used online payments.	Use AI to detect fraud patterns and secure payments with blockchain.

Supply Chain Comparison: 2015-2018 vs. Post-COVID

2015-2018 Findings	Post-Covid Changes	How to Improve?(Solutions)
Late deliveries were common with Standard Class shipping and certain products.	Global shipping delays became worse due to worker shortages and port congestion.	Work with multiple suppliers and track shipments in real-time to reduce delays.
Decision Tree model worked best for predicting fraud and late deliveries.	AI adoption increased after COVID to automate supply chains.	Use AI-powered tools to monitor supply chain performance and reduce risks.
Businesses depended on a few suppliers and centralized warehouses.	COVID exposed weak points in supply chains, leading to stock shortages.	Build supply chain resilience by using multiple suppliers, regional warehouses, and flexible sourcing.

BUSINESS RECOMMENDATIONS & FINAL CONCLUSION

- **Key Takeaways:**
  - AI-driven demand forecasting helps prevent stock shortages & excess inventory.
  - Fraud detection models reduce financial risks and improve security.
  - Optimized logistics strategies minimize late deliveries and improve customer satisfaction.
  - Resilience-building strategies ensure business continuity during disruptions.

BUSINESS RECOMMENDATIONS & FINAL CONCLUSION

- **Actionable Business Strategies:**
  - **Boost High-Growth Regions** → Increase inventory & marketing in Europe & Central America.
  - **Optimize Shipping Methods** → Reduce reliance on Standard Class; promote premium shipping.
  - **Fraud Prevention Strategy** → Strengthen fraud monitoring in high-risk regions & products.
  - **Customer Retention Programs** → Focus on at-risk customers with re-engagement offers.
  - **AI-Powered Decision Making** → Use ML models to forecast demand, detect fraud, and reduce losses.
  - **Enhance Supply Chain Resilience** → Adopt multi-sourcing, regional warehousing, and AI-driven risk management.

THANK YOU