Supply-Chain Analysis

- Devansh Purwar

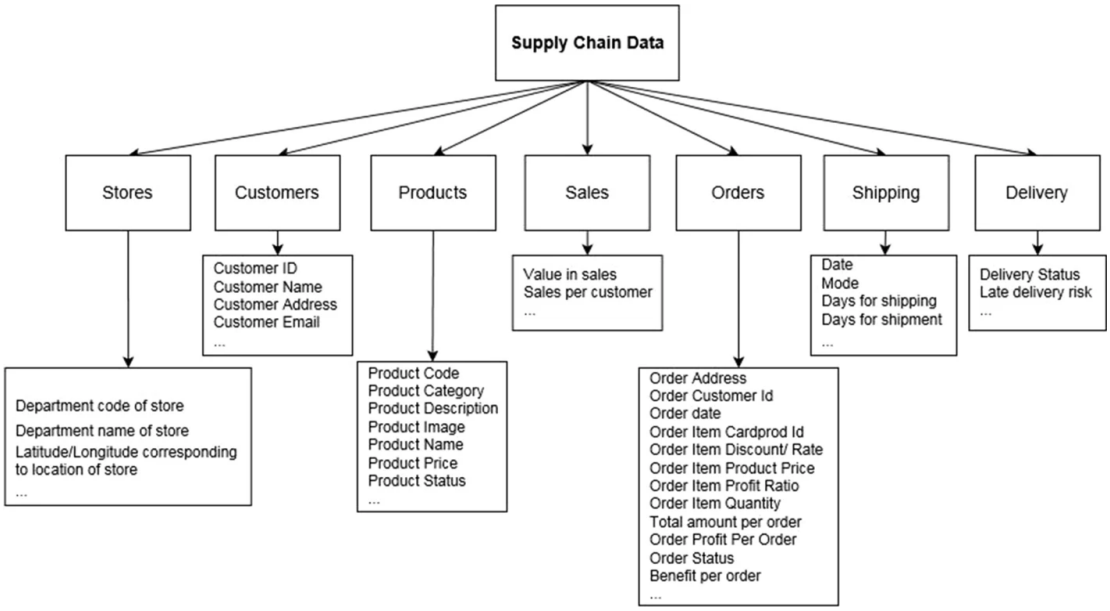
Abstract

Big data analytics (BDA) in supply chain management (SCM) is receiving growing attention. This is due to the fact that BDA has a wide range of applications in SCM, including customer behavior analysis, trend analysis, and demand prediction. In this report, I investigate the predictive BDA applications in supply chain demand forecasting to propose a classification of these applications, identify the gaps, and provide insights for future research.

Introduction

As a valuable asset for decision-making, Data Analytics can play a pivotal role in transforming and improving the functions of the supply chain. In this changing business environment, business leaders prefer to take decisions bearing in mind the data-driven insights rather than relying on their intuitions (Davenport 2006). Due to the perceived benefits of BDA, organizations are highly motivated to develop their technical and organizational capabilities to extract value from data.A DataSet of Supply Chains used by the company DataCo Global was used for the analysis

Dataset



Taxonomy of supply chain data

- Dataset contain 180519 data points and 53 fields

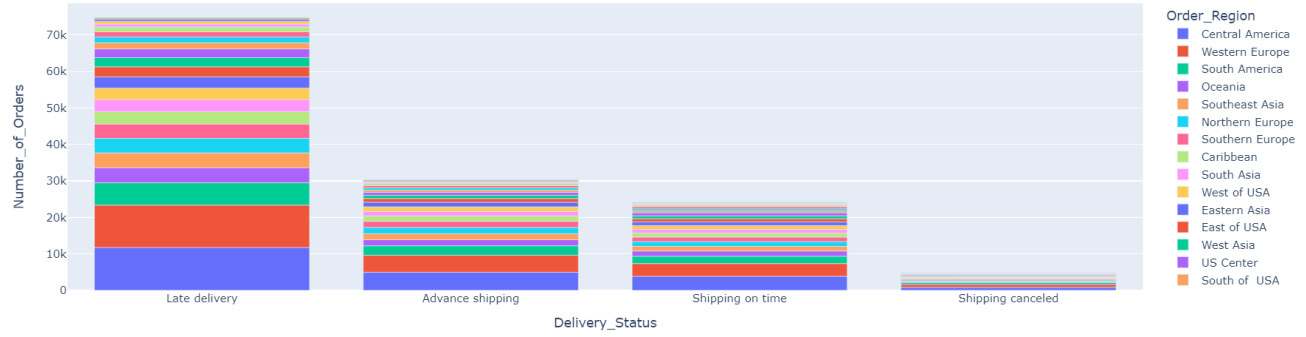
Acknowledgements & Source
Fabian Constante,
Instituto Politecnico de Leiria Escola Superior de Tecnologia e Gestao
Contribution: Master interested in topics related to Big Data
Fernando Silva, António Pereira

https://data.mendeley.com/datasets/8gx2fvg2k6/5

Data Cleaning

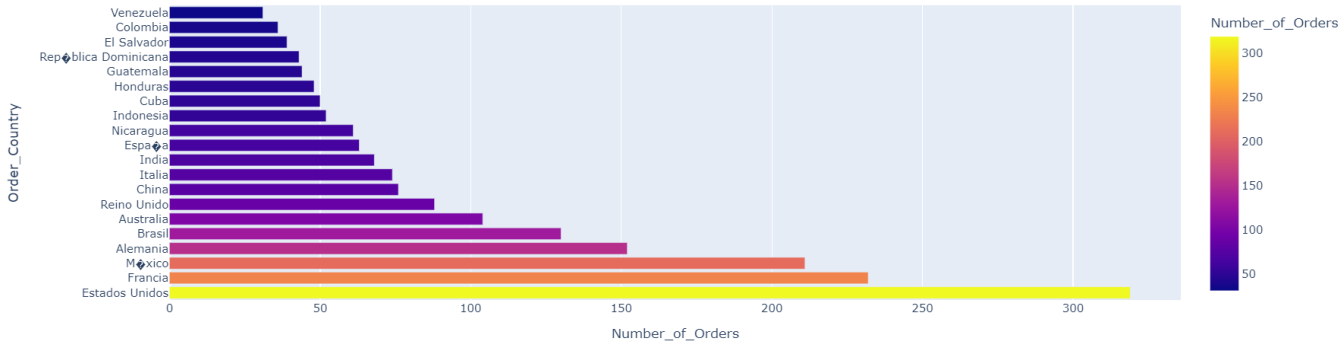
- Checked NaN Values
- Checked empty fields
- Checked variance in a column
- Checked redundant columns

Exploratory Data Analysis



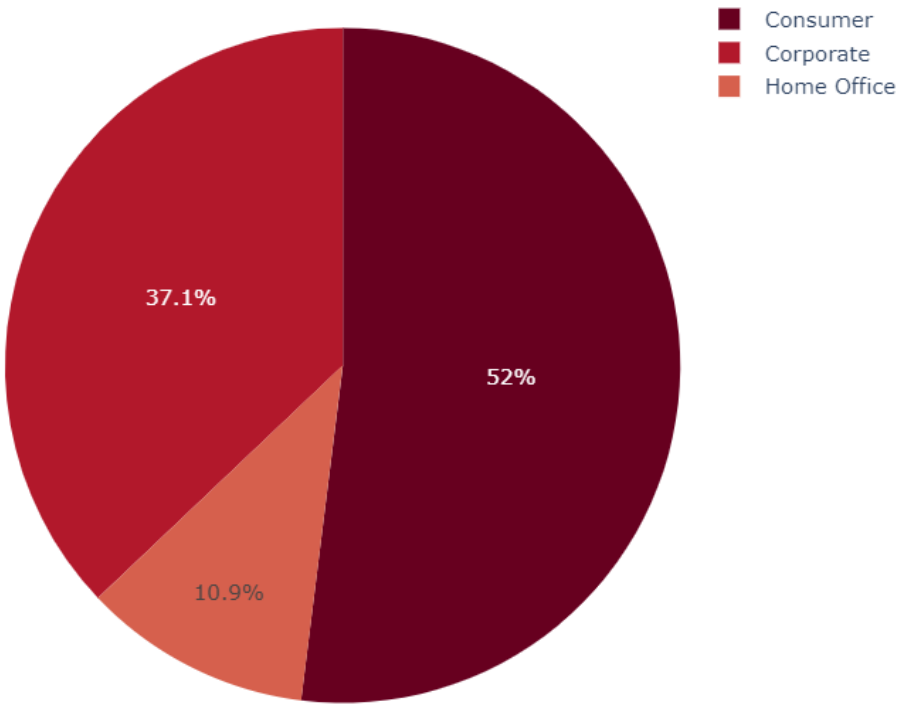
Numbers of orders over delivery status region wise

- Late Delivery is over 50% of total orders.

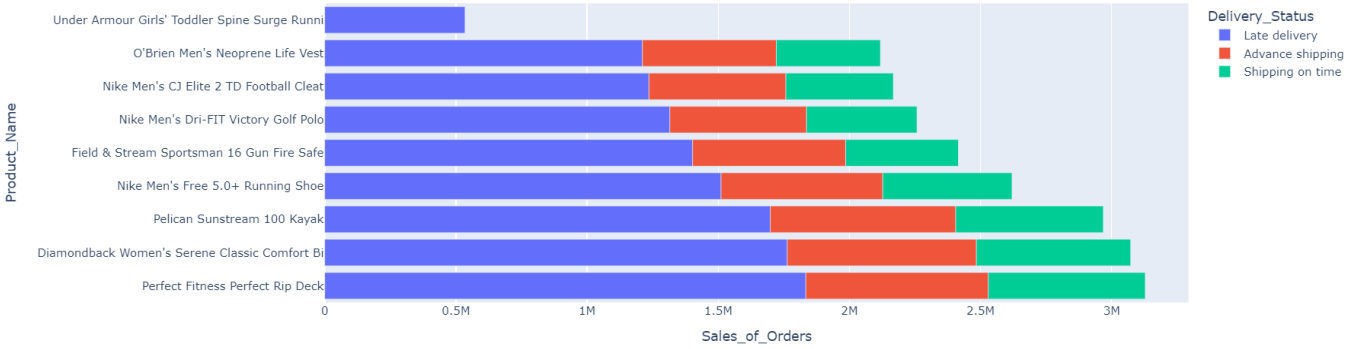


Number of suspected fraud orders over different countries

- Median of suspected fraud orders from each country is 80.

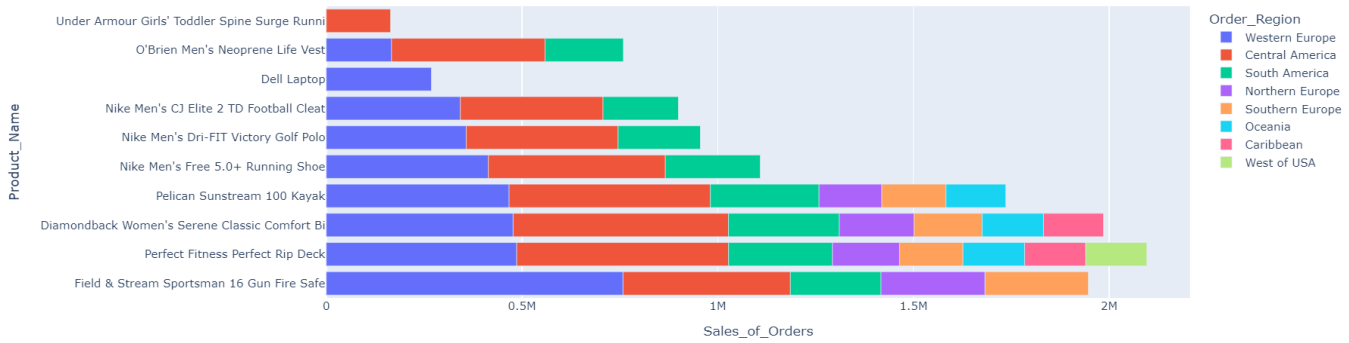


Customer Segment



Product Sales orders classification over Delivery Status

- Some products have a record of delivering late every time .

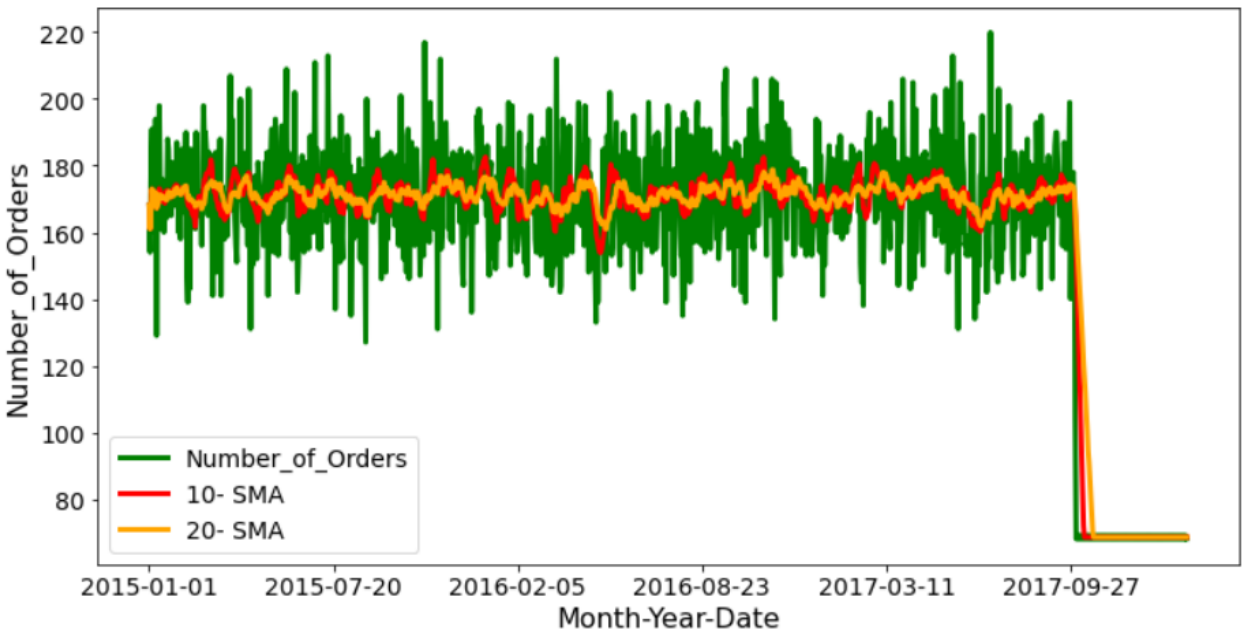


Product Sales orders classification over Order Region

- Some products have demand only in particular region.

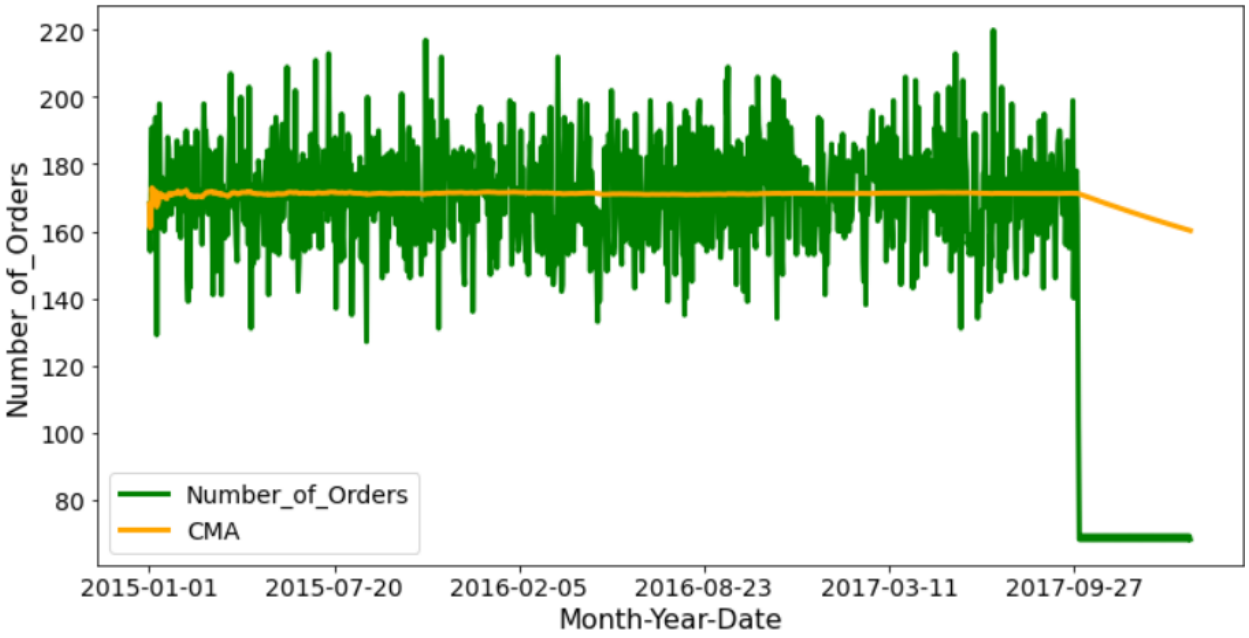
Time Series Analysis

It is used for supply chain management because it's constantly fluctuating over time or is affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing which major role for these types of industries.



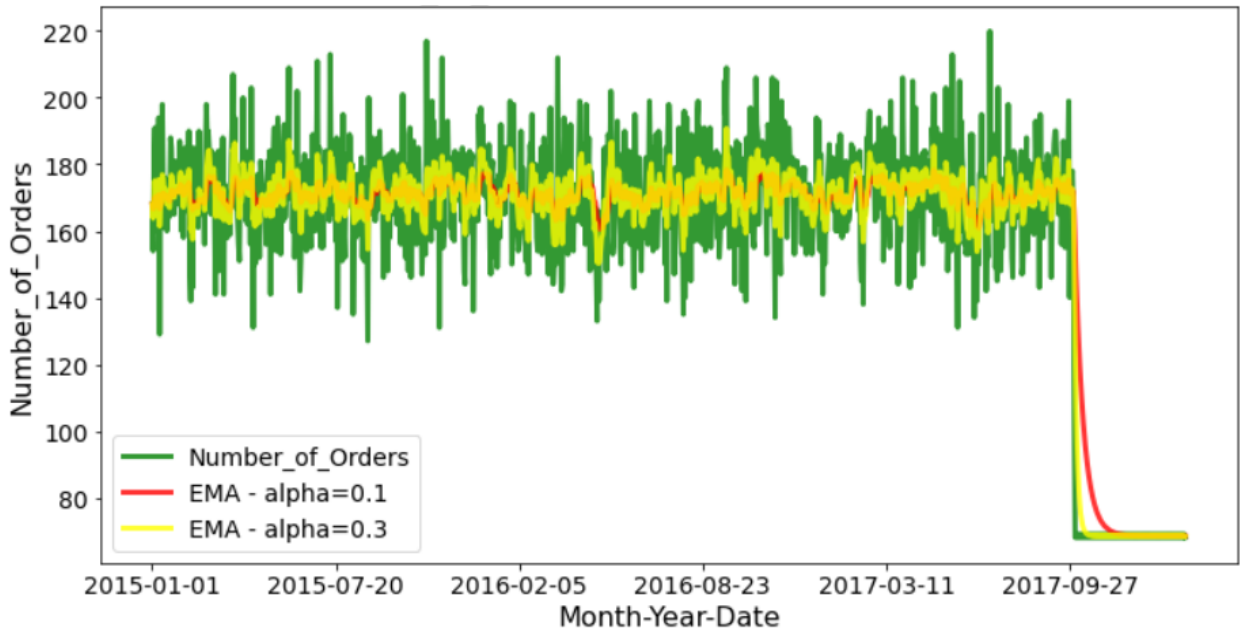
Simple Moving Average

A simple moving average (SMA) calculates the average of a selected range of prices, usually closing prices, by the number of periods in that range.



Cumulative Moving Average

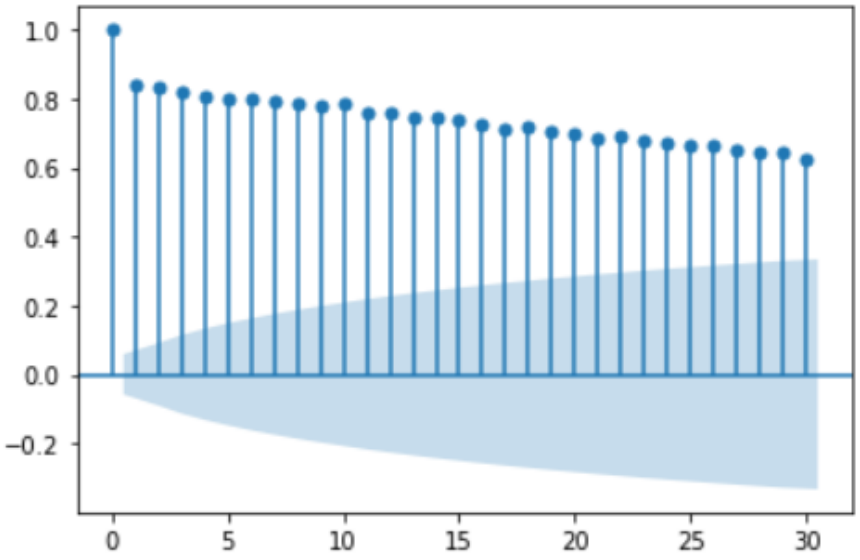
The Cumulative Moving Average is the unweighted mean of the previous values up to the current time t.



Exponential Moving Average

An exponential moving average (EMA) is a type of moving average (MA) that places a greater weight and significance on the most recent data points. The exponential moving average is also referred to as the exponentially weighted moving average. An exponentially weighted moving average reacts more significantly to recent price changes than a simple moving average simple moving average (SMA), which applies

an equal weight to all observations in the period.



Autocorrelation with 30 days lag

Autocorrelation is the relationship between two values in a time series. To put it another way, the time series data are correlated, hence the word. “Lags” are the term for these kinds of connections.
The number of intervals between two measurements is known as the lag.

Late Delivery Prediction

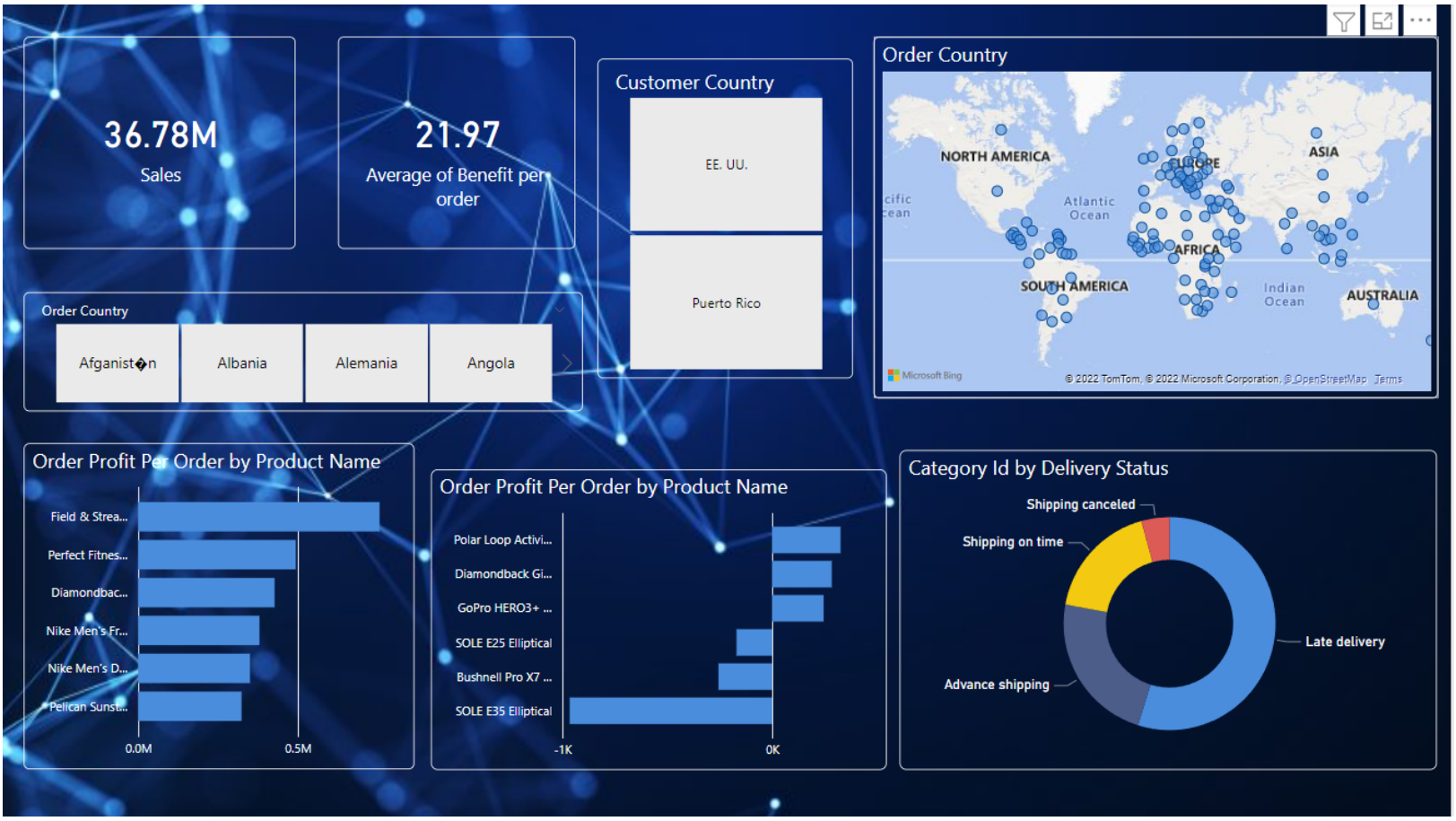
For Late delivery prediction , first I had done the binary classification of late delivery column from shipping date (scheduled) and shipping date(real). After that selected features from a dataset for late delivery model ('Type','Days_for_shipment_(scheduled)','Category_Name','Customer_Country','Customer_Segment','Order_Country','Order_Item_Quantity','Shipping_Mode','Order_State','Customer_State','Late_Delivery_Label')
Separated categorical and numerical features and followed Ordinal Encoding then merge both dataframes and splited train and test dataset with 8:2 for model.
By using Grid search CV, finding the best hyperparameters for each model.
Trained models :

- DecisionTreeClassifier
- KNeighborsClassifier
- LogisticRegression
- SupportVectorMachine
- RandomForestClassifier
- XGBClassifier
- CatBoostClassifier

```
ACCURACY SCORE
Accuracy score of the Decision Tree model is 0.6994515843119876
Accuracy score of the KNN model is 0.6418956348327055
Accuracy score of the Logistic Regression model is 0.6982051850210503
Accuracy score of the SVM model is 0.5719033902060714
Accuracy score of the Random Forest Tree model is 0.6982051850210503
Accuracy score of the XGBoost model is 0.7059051628628407
Accuracy score of the CatBoost model is 0.7181475736760 47
```

```
F1 SCORE
F1 score of the Decision Tree model is 0.7112483035738045
F1 score of the KNN model is 0.5815993009934953
F1 score of the Logistic Regression model is 0.7057679844458847
F1 score of the SVM model is 0.0
F1 score of the Random Forest Tree model is 0.7057679844458847
F1 score of the XGBoost model is 0.7190559348044663
F1 score of the CatBoost model is 0.7246753246753246
```

Dashboard



Conclusion

After analyzing the DataCo Company dataset it has been discovered that both Western Europe and Central America are the regions with the highest number of sales but also the company lost most revenue from these regions only. And both these regions are suspected to the highest number of fraudulent transactions and orders with more late deliveries. The total sales for the company were consistent until the 2017 Quarter 3 and 10% increase in total sales by quarter and then suddenly dipped by almost 65% in 2018 quarter 1. October and November are the months with most sales in the total year. Most people preferred to do payment through debit card and all the fraud transactions are happening with wire transfer so the company should be careful when customers are using wire transfer as the company was scammed with more than 100k by a single customer. All the orders with the risk of late delivery are delivered late every time. Most of the orders with Cleats, Men's Footwear, and Women's Apparel category products are causing late delivery also these products are suspected to fraud the most.