DataCo GLOBAL

BUSINESS CASE

UofT Data Analytics Bootcamp 2022

Summary

Overview:
DataCo Global is a company dedicated to provisioning, production, sales and commercial distribution.
DataCo is in the Supply chain business with three different products:
- Clothing
- Sports
- Electronic Supplies

During 4 years, DataCo Global experienced a high rate on Late Deliveries that had impact in sales.

Goal:
By analyzing the DataCo data and trends between 2015 to 2018, we will be able to answer the two main questions:

- **How to predict the supply chain disruption?**
- **What type of features are impacting those disruptions?**

OUR TEAM

AUDREY MONJARAS

BRENT NEWMAN

BRANDON WILLIAMS

TOOLS AND TECHNOLOGIES

- Logistic Regression

GitHub

Technologies   Languages   Tools   Algorithms

TABLE OF CONTENTS

01
DATA EXPLORATION
The use of ETL Process

02
DATABASE
Creating Connections

03
MACHINE LEARNING
Making Predictions

04
DASHBOARD
Interactive view

01
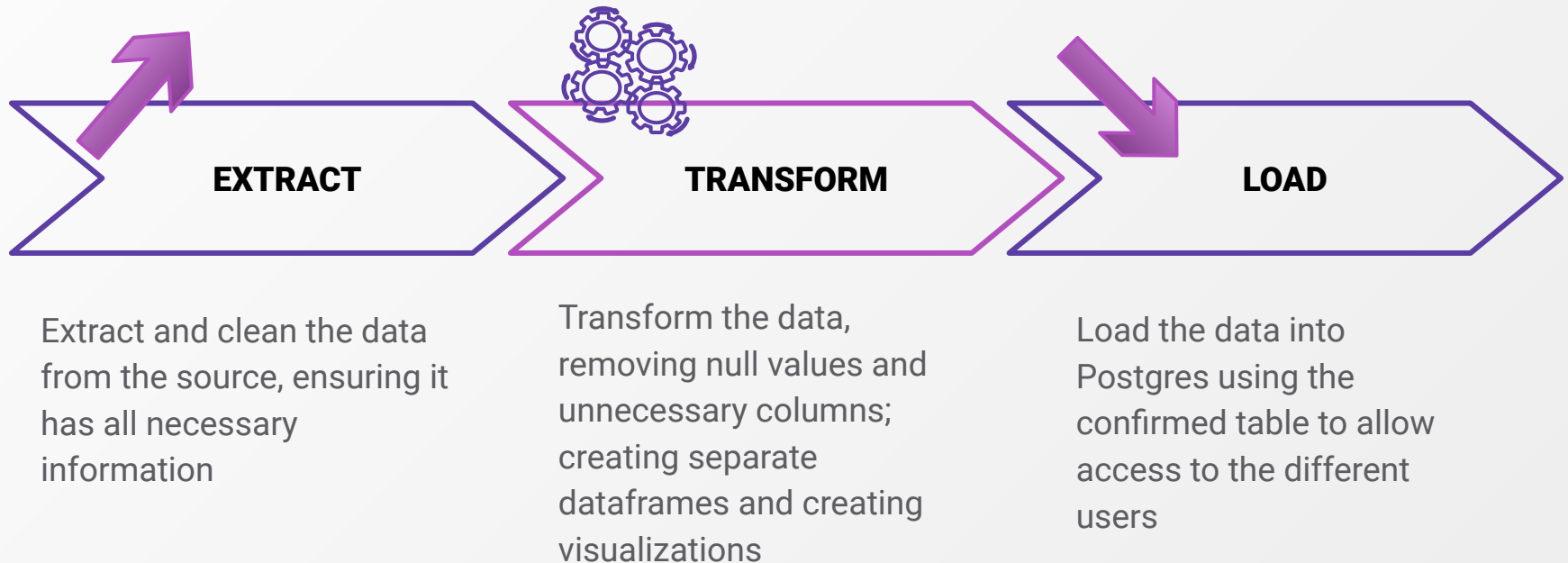
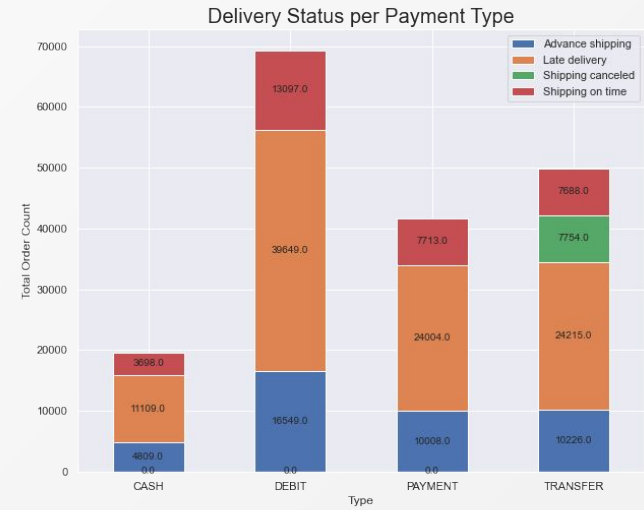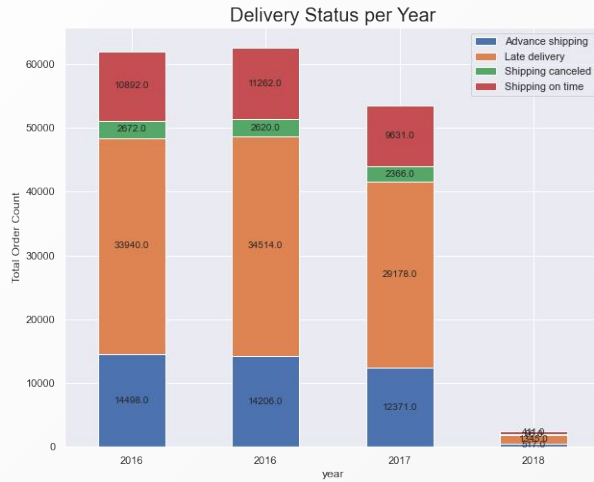# DATA EXPLORATION

The use of ETL Process
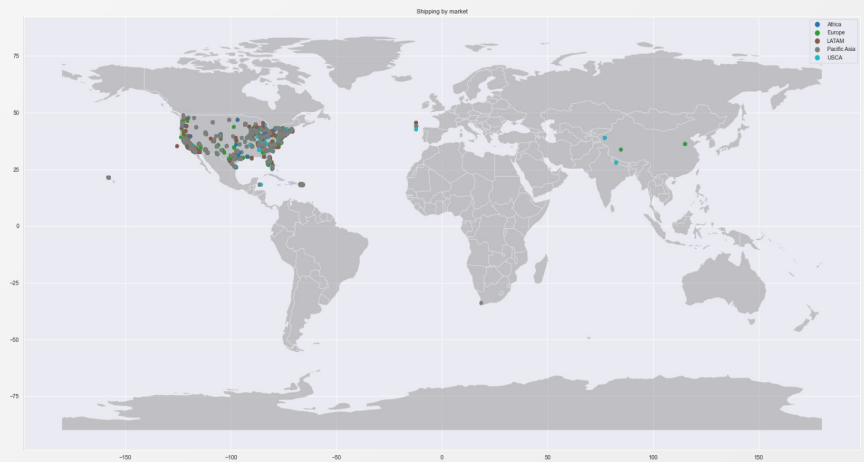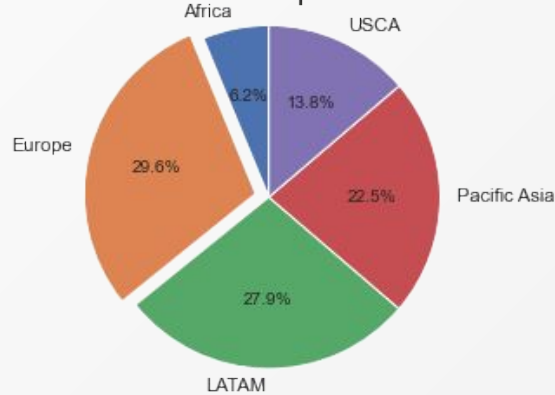
Sourcing the Data

Our team was able to source the data in order to answer the analysis questions. The Data Exploration followed the ETL Process.

DataCo Source can be found **here**

EXTRACT

TRANSFORM

LOAD

Extract and clean the data from the source, ensuring it has all necessary information

Transform the data, removing null values and unnecessary columns; creating separate dataframes and creating visualizations

Load the data into Postgres using the confirmed table to allow access to the different users

Delivery Status per Year

Delivery Status per Payment Type

Sales Demand per Market

Shipping by market

02
DATABASE

Creating connections

Database Process

AWS RDS
With PostgreSQL engine

AWS Services

AWS S3
With PostgreSQL engine

Files from
pgAdmin

AWS S3
Bucket

Internet

Local pgAdmin
Installation

Users

03
MACHINE LEARNING
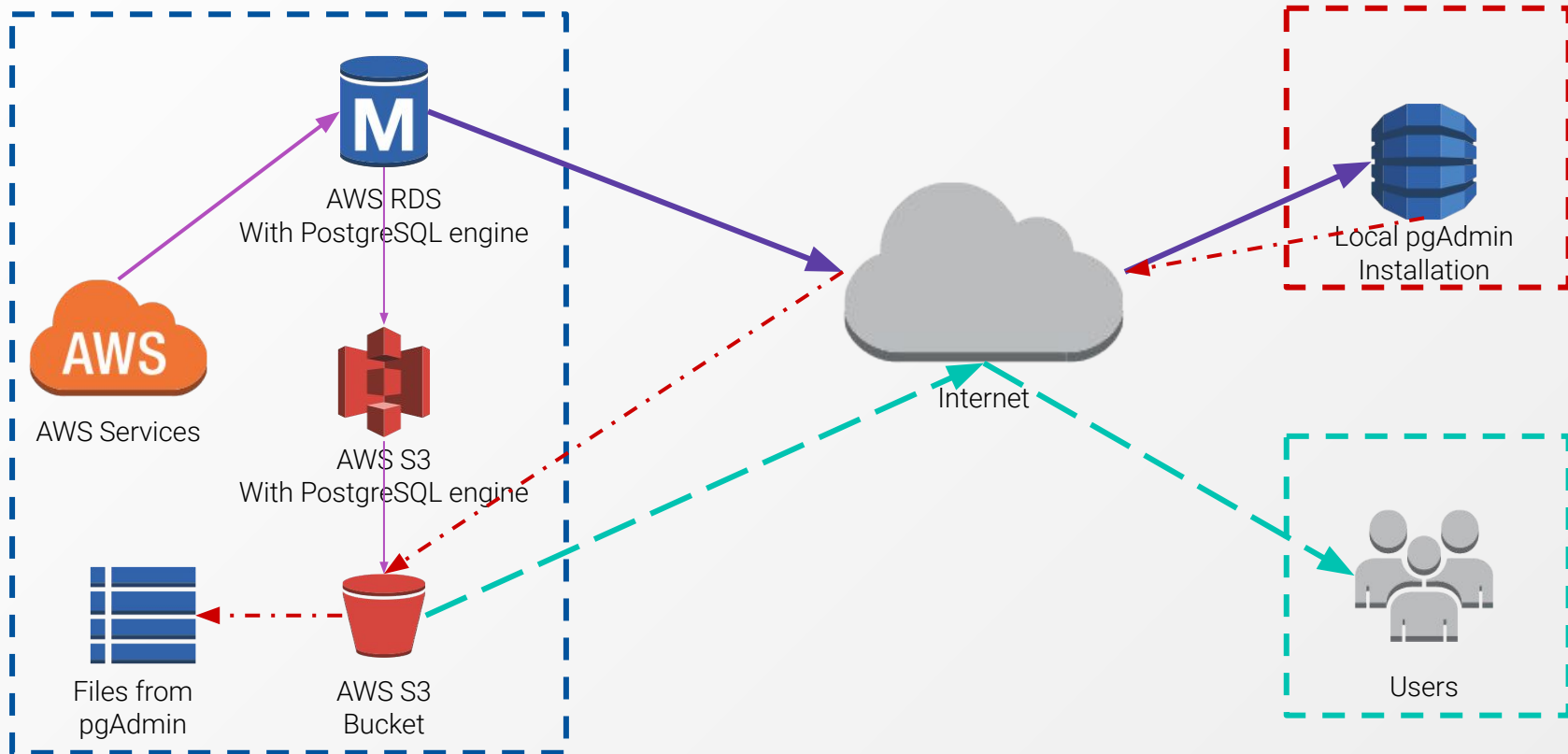
Making predictions

Preliminary Data Processing

We selected logistic regression to answer our questions defined in the **summary**. Logistic Regression is an example of supervised learning. Used to calculate / predict the probability of a binary event occurring.

Prepare Data for Logistic Regression

- **Binary Output Variable:** logistic regression is intended for binary classification problems (0 or 1 classification)
- **Remove Noise:** consider removing outliers and possibly misclassified instances from the training data
- **Remove Correlated Inputs:** like linear regression, the model can overfit if we have multiple highly-correlated inputs.

Model Testing

MODEL NAME	DATA PROCESSING	MODEL PARAMETERS	ACCURACY
LogisticRegression()	50/50 train test split	(solver='lbfgs', random_state=1)	50%
	AFTER REMOVING CORRELATED INPUTS	(solver='lbfgs', random_state=1)	70%

Conclusion

Analysis

After completing the project and reviewing the model prediction, it is not clear what are the features impacting the disruption. From the ETL process we can determine one strong possible cause and we can recommend to DataCo Group to open more stores on their different Markets (Africa, Europe, Pacific Asia)

What could be done different

- Explore more models to improve fit such as Random Forest
- Include a model evaluation such as Root Mean Square Error (RMSE)

04
DASHBOARD

Interactive View

Dashboard Elements

Interactive Elements

- The country will be selected from a drop down menu and will be able to display the status of shipments

Other Elements

- The remaining elements in the dashboard are static , but designed to showcase the different skills learned through the bootcamp
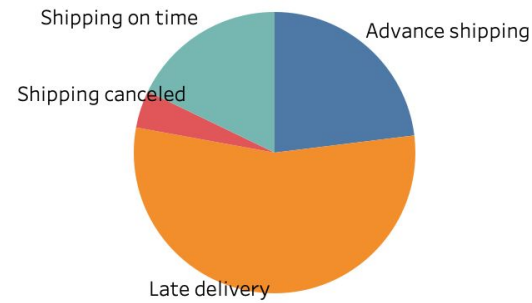- All the tables and charts reflect the information from our dataset

Shipping Dashboard

Counters

Advance shipping 41,592	Late delivery 98,977
Shipping canceled 7,754	Shipping on time 32,196

Delivery Status

- Shipping on time
- Advance shipping
- Shipping canceled
- Late delivery

Number of Shipments

180,519

Orders by Region

Order Region

Region	
Central America	
Western Europe	
South America	
Oceania	
Northern Europe	
Southeast Asia	
Southern Europe	
Caribbean	
West of USA	
South Asia	

0K 5K 10K 15K 20K 25K
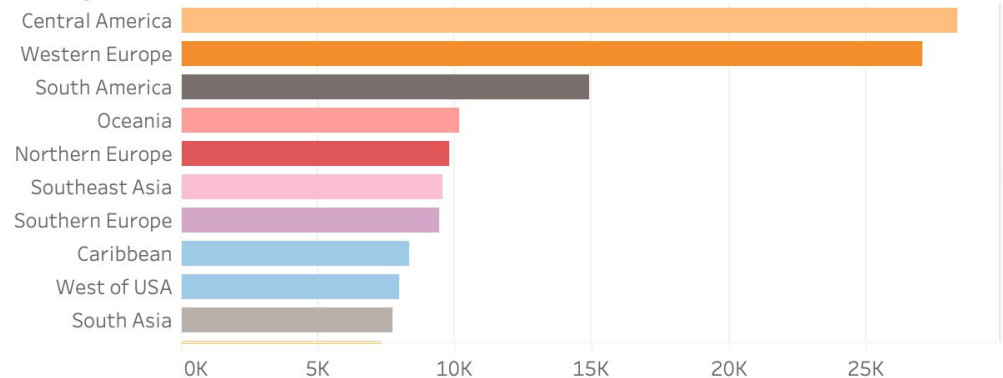
THANKS!

Visit our GitHub to access the full Report