



# 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
- Sporting Goods
- 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**

GitHub / ETL /  
Presentation



**BRENT NEWMAN**

Tools / Machine Learning /  
Dashboard

**BRANDON-SCOTT  
WILLIAMS**

Database / Machine  
Learning





# TOOLS AND TECHNOLOGIES



**Technologies**



**Languages**

**GitHub**



**Tools**

- Logistic Regression

**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

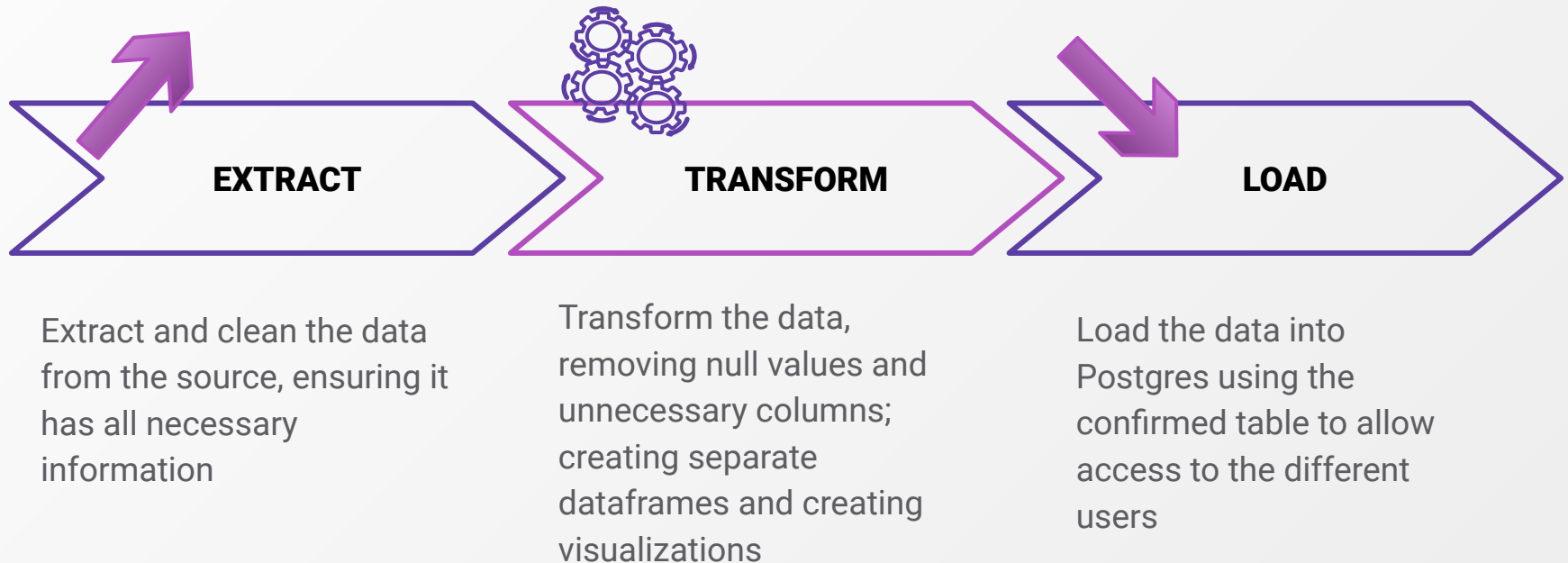


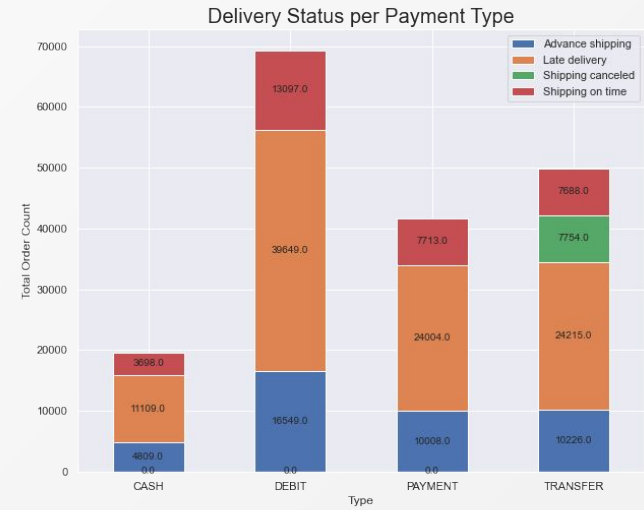
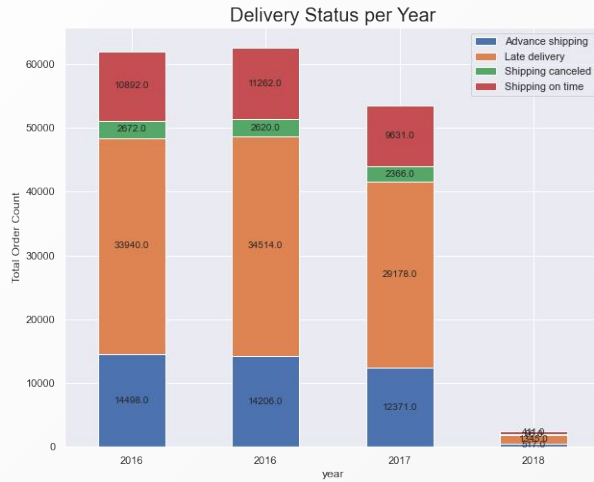


# Sourcing the Data

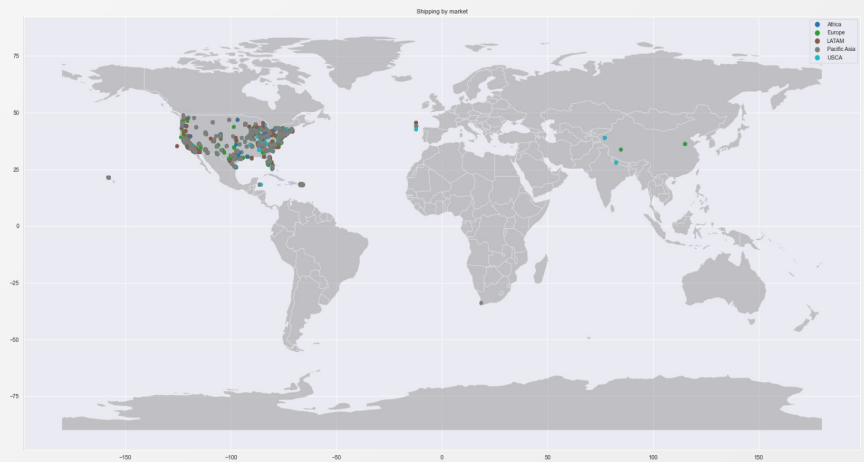
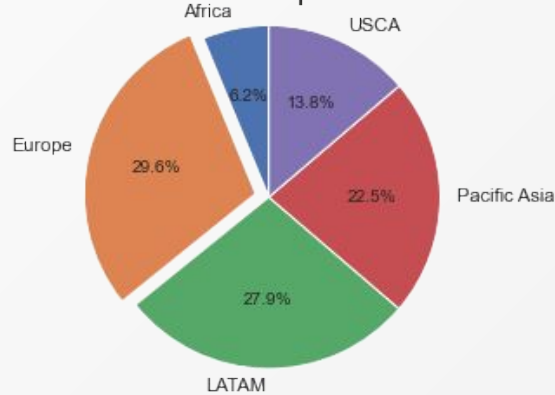
Our team was able to source the data in order to answer the analysis questions. The Data Exploration followed the ETL Process.

DataCo Source can be found [here](#)





Sales Demand per Market







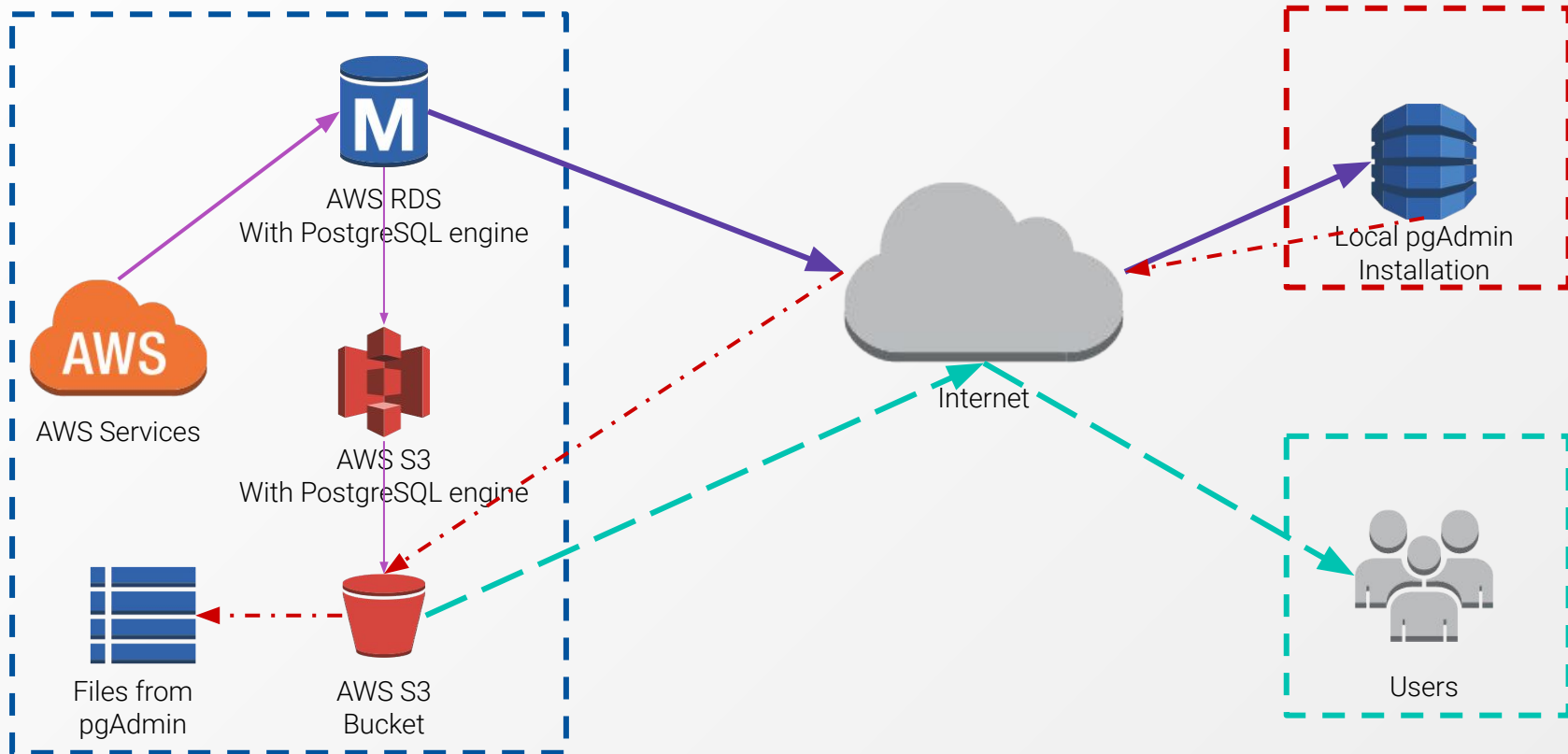
# 02 DATABASE

Creating connections





# Database Process





# 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.

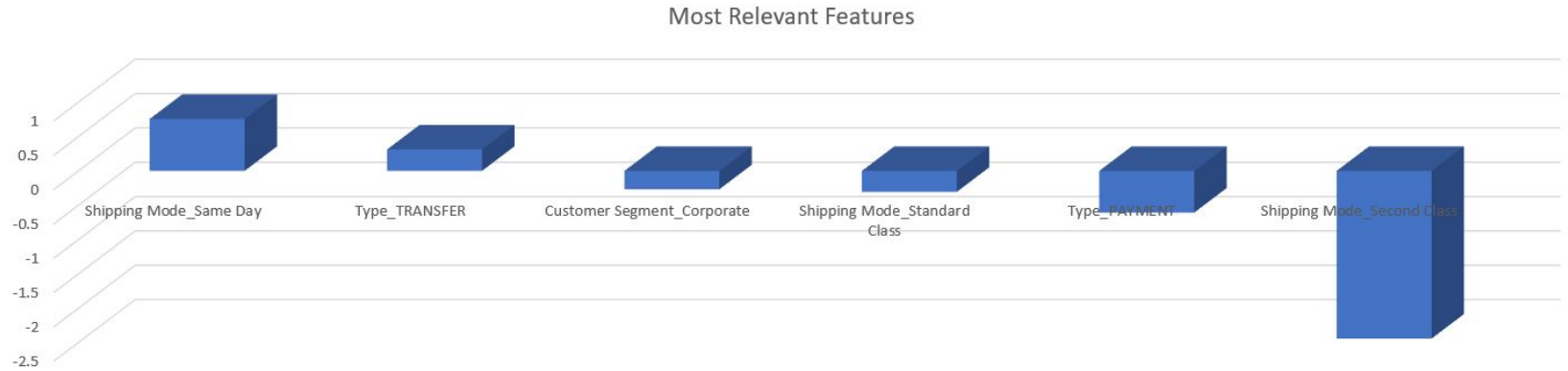




# Model Testing

MODEL NAME	DATA PROCESSING	MODEL PARAMETERS	ACCURACY
LogisticRegression()	80/20 train test split	(solver='lbfgs', random_state=1)	50%
	AFTER REMOVING CORRELATED INPUTS	(solver='lbfgs', random_state=1)	70%





After the feature analysis it is noticeable that the Second Class shipping is creating the largest negative impact on deliveries.





# Conclusion

## Analysis

After completing the project and reviewing the model prediction, we prove that the biggest feature impacting the deliveries is "Second Class Shipping".

### Recommendations

- Investing in the variables affecting the Second Class Shipping
- Adjusting expected shipping time for that class

## 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



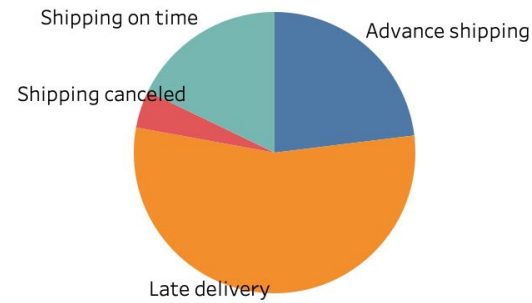


## Shipping Dashboard

### Counters

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

### Delivery Status

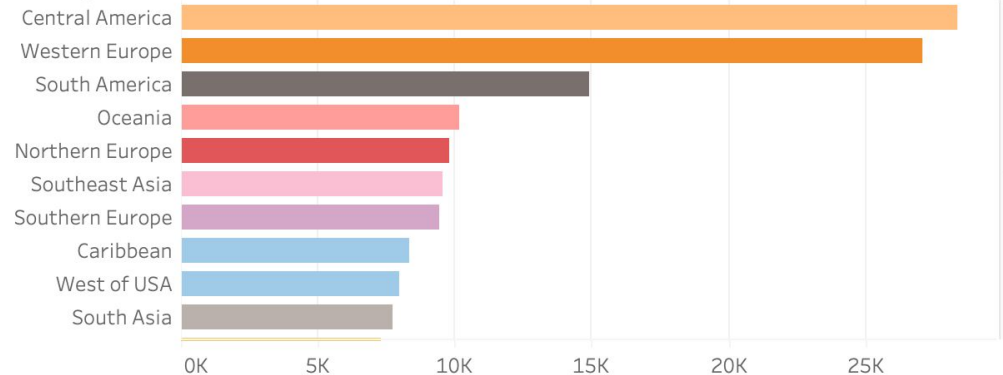


### Number of Shipments

180,519

### Orders by Region


Order Region



Two light gray squares stacked vertically on the left side of the slide.

# THANKS!

Visit our GitHub to access the full  
Report

Two light gray squares stacked vertically on the right side of the slide.

CREDITS: This presentation template was created by **Slidesgo**, including  
icons by **Flaticon** and infographics & images by **Freepik**

Please keep this slide for attribution