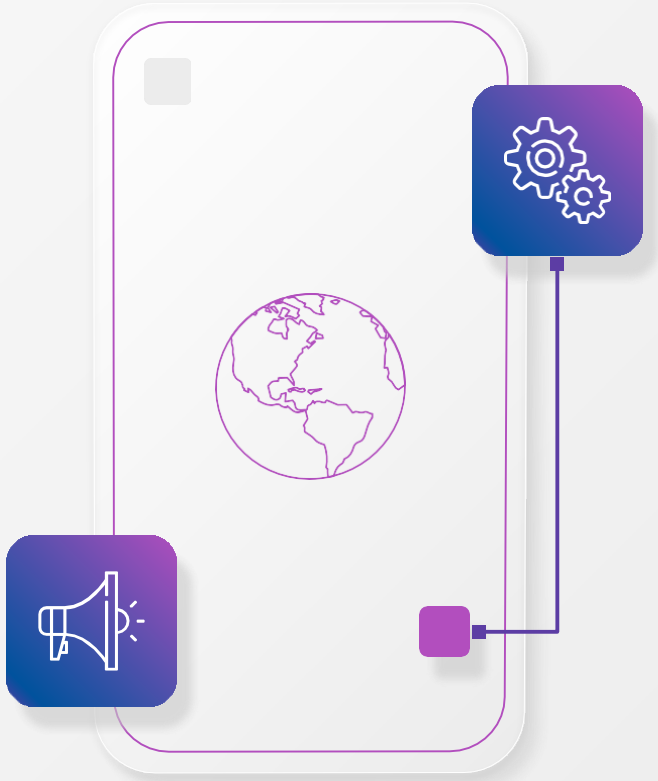
DataCo GLOBAL

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

- tableau
- PostgreSQL
- aws

Languages

- pandas
- python

Tools

- GitHub
- jupyter

Algorithms

- Logistic Regression

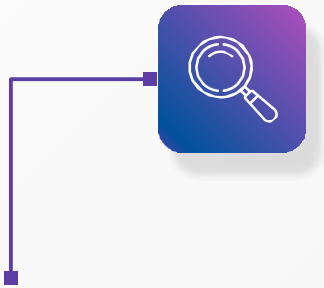


TABLE OF CONTENTS

01 DATA EXPLORATION

The use of ETL Process

Sourcing the Data

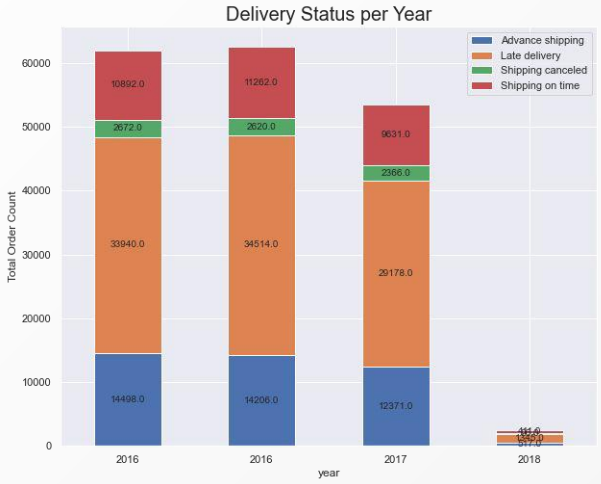
Our team was able to source the data in order to answer the analysis questions. The Data Exploration following the ETL Process.

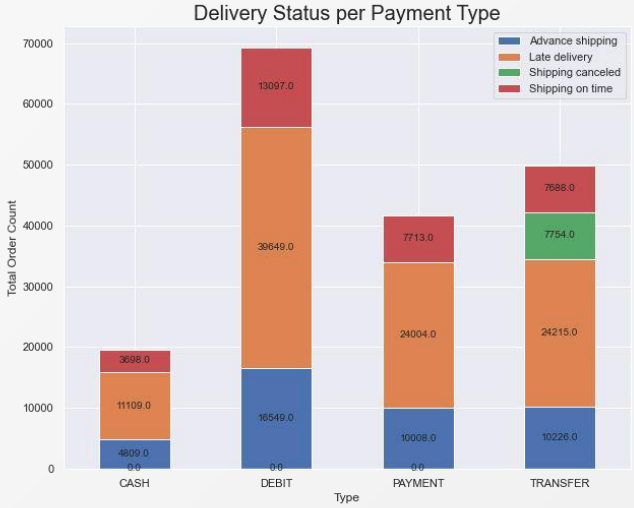
DataCo Source can be found **here**

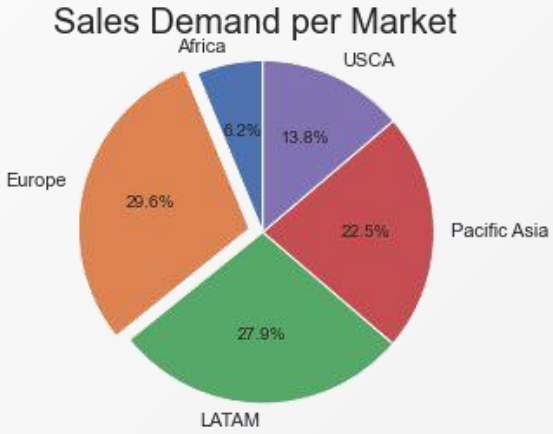
Extract and clean the data from the source, ensuring it has all necessary information

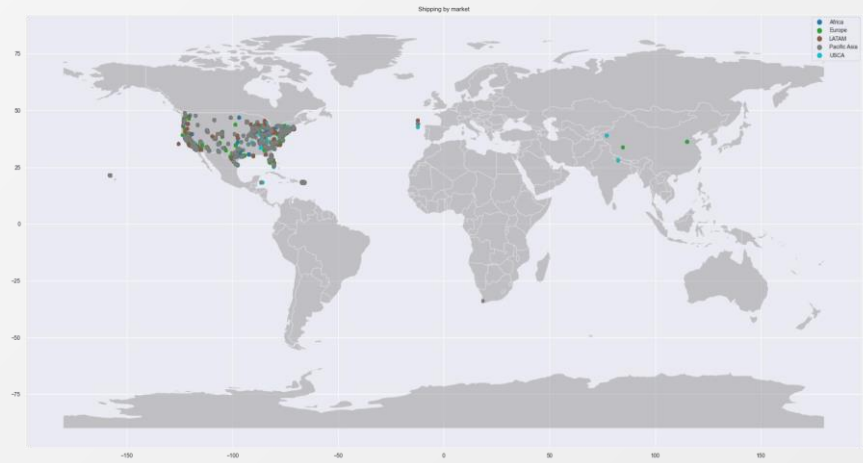
Transform the data, removing null values and unnecessary columns; creating separate dataframes and creating visualizations

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


Delivery Status per Year
Advance shipping
Late delivery
Shipping canceled
Shipping on time
Total Order Count
year
2016
2016
2017
2018
10892.0
2672.0
33940.0
14498.0
11262.0
2620.0
34514.0
14206.0
9631.0
2366.0
29178.0
12371.0


Delivery Status per Payment Type
Advance shipping
Late delivery
Shipping canceled
Shipping on time
Total Order Count
Type
CASH
DEBIT
PAYMENT
TRANSFER
3698.0
11109.0
4809.0
13097.0
39649.0
16549.0
7713.0
24004.0
10008.0
7688.0
7754.0
24215.0
10226.0

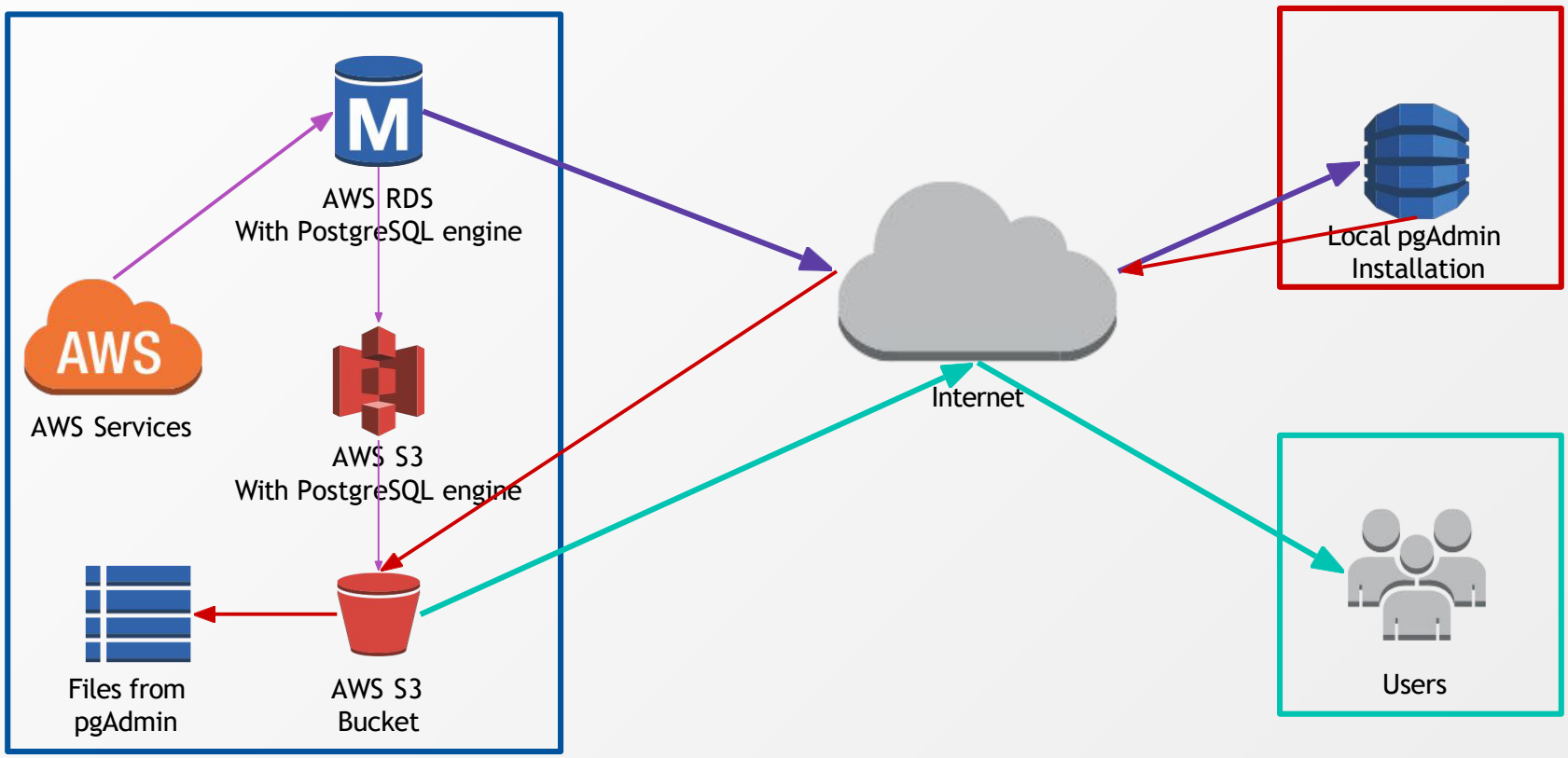

Sales Demand per Market
Africa
USCA
Europe
Pacific Asia
LATAM
6.2%
13.8%
29.6%
22.5%
27.9%



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

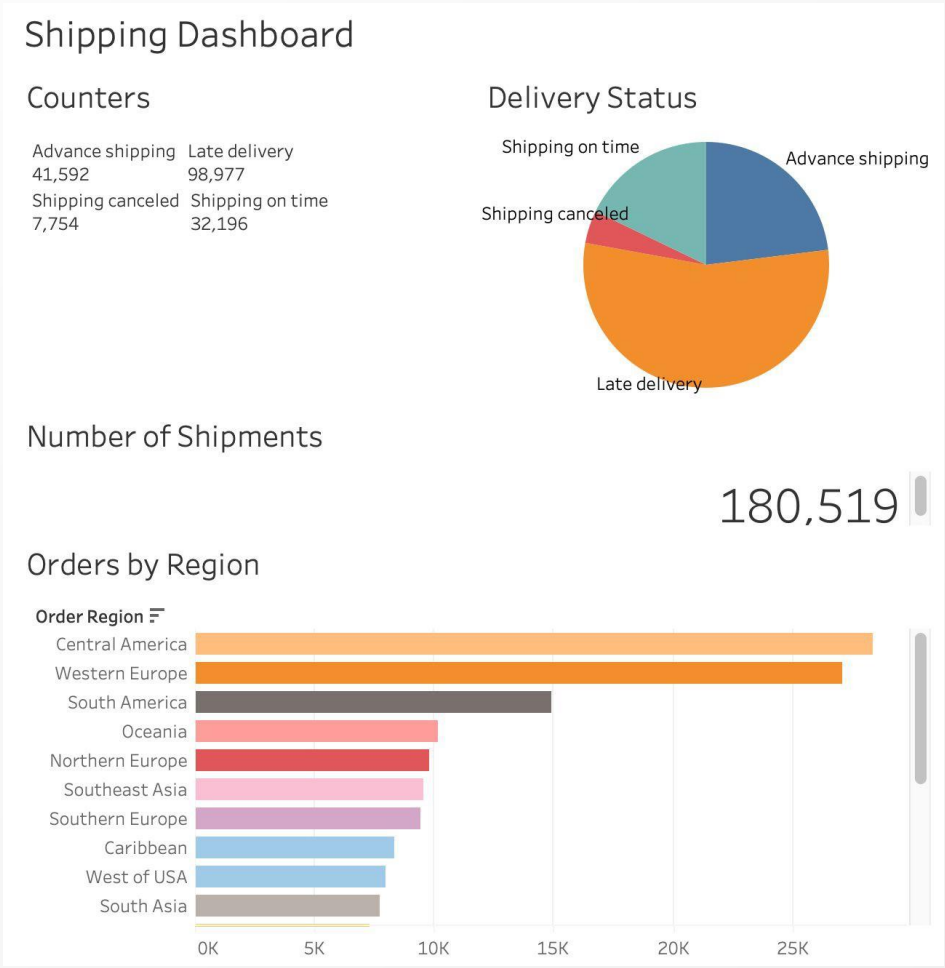
Dashboard Elements

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



To access the live Demo, click **here**

THANKS!

Visit our GitHub to access the full Report