DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023

# Milestone 2 – Sales Forecasting and Late Delivery Prediction

## Introduction – Background and Business Problem

Supply Chain Analytics is a necessary role in today's world. Many organizations struggle with supply management issues and require a better understanding of their current struggles. Applied data science toward supply chain business problems allows companies to recognize useful insights to overcome these challenges. According to AI Monks (2023), "Supply chain forms the backbone of most organizations, especially the ones which are operations heavy. It is essential for such organizations to run their supply chain processes efficiently and smoothly" (para. 1). Not all organizations will run their supply chain processes in the same manner. Organizations will need to develop their own processes based on their business needs. Some common organizational functions where Supply Chain Analytics can be applied are Demand Planning, Procurement, Inventory Management, Logistics & Transportation, and Vendor Management (AI Monks, 2023). Machine learning is a great way to address challenges in these areas. The article from Ariwala (2022) outlines the use of machine learning for predictive analytics with demand forecasting, fraud prevention, warehouse management, and quality inspections.

A supplier that understands their customers' needs and provides material (or services) in a timely fashion becomes invaluable. Customers tend to retain suppliers with high performance and this project will focus on two main aspects. The first objective is to develop a model for sales forecasting. A foundational understanding of customer demand allows suppliers to better manage their own internal supply chain processes. The second focus is to determine a method for predicting late deliveries. This allows the supplier to correct ongoing issues with late deliveries or communicate with customers early in the process. The relationships between customers and suppliers within a system often get compared to a stream. One supplier may report to a particular customer, then that customer serves as a supplier for an alternate customer downstream and so on... Organizations can relay these strategies to their suppliers "upstream" to continue to meet customer expectations. The research questions are outlined below:

Case Study 1: Prediction of Sales

- Which categories had the highest sales?
- Which customers bought the most?
- Are there any variables strongly correlated with sales?
- Which model provides the best accuracy for forecasting?

Case Study 2: Prediction of Late Deliveries

- Which categories had the highest number of late deliveries?
- Does a particular product tend to be late?
- Do customers in a particular geographic area tend to receive late shipments?
- Are there any variables strongly correlated with late deliveries?
- Which model provides the best accuracy for predicting late deliveries?

## Data Explanation, Methods, and Assumptions

The dataset for this analysis was from Mendeley Data and represents supply chain data from the company DataCo Global. Only the structured data from the DataCoSupplyChainDataset.csv was utilized. The link to the data can be found below:

DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023

[DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS - Mendeley Data]

The shape of the dataset initially included 53 variables and over 180,500 records. The source for the data included a quick description of each variable which can be referenced in the Appendix. All data cleaning steps were clearly outlined for transparency.

The Cross Industry Standard Process for Data Mining (CRISP-DM) was followed for this project. The high-level phases included Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The Business Understanding phase was previously outlined in the prior section. To recap, this project aimed to improve supply chain performance through sales forecasting and prediction of late delivery. The intent was to find actionable insights that surface from this analysis. The Data Understanding step involved Exploratory Data Analysis. This included univariate and bivariate analysis of the variables. The Data Preparation step was performed to prepare the data for each respective model. The forecasting sales model is a regression problem whereas the late delivery model is a classification problem. Supervised learning algorithms were utilized for each. For forecasting sales, Linear, Lasso, Decision Trees and Random Forest models were considered. For predicting late deliveries, Logistic Regression, Decision Trees, and Random Forest models were considered. Since there were two types of problems being addressed, the evaluation metrics were different for the models predicting future sales ($R^2$ or RMSE) compared to prediction of a late delivery (accuracy, precision, recall, F1 score). Cross-validation was used to better understand model performance. Lastly, a recommendation will be communicated whether to deploy the models. This was an iterative process and steps were revisited often throughout the analysis.

There were a few assumptions to highlight prior to moving to the analysis section. The insights found from the analysis are not a "one size fit all" approach. Organizations can utilize their own data with a similar methodology outlined in this analysis to realize their own challenges and required actions for improvements. The assumption here is that organizations, or individuals, will use this analysis as a template for the methodology to follow for supply chain improvements. Another assumption for this analysis is that the data was collected fairly without bias. The chosen data set already had the records and variables collected, so this was not part of the focus of this analysis. The variables were evaluated during the Data Understanding phase to review for any data quality issues.

There was an in-depth Data Understanding phase. There were 19 columns removed from the data set initially for multiple reasons. If most records were missing from the column within the data set, then the column was removed. If the column provided sensitive information related to privacy, then the column was removed. If the column represented duplicate information, then the column was also removed prior to the analysis. The list below summarizes the dropped features prior to Exploratory Data Analysis (EDA).

Dropped Features prior to EDA: Product Description, Order Zipcode, Customer Lname, Customer Zipcode, Product Image, Customer Fname, Customer Email, Customer Password, Product Status, Category Id, Department Id, Order Id, Order Item Cardprod Id, Product Card Id, Product Category Id, Order Item Discount Rate, Longitude, Latitude, Order Customer Id

DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023

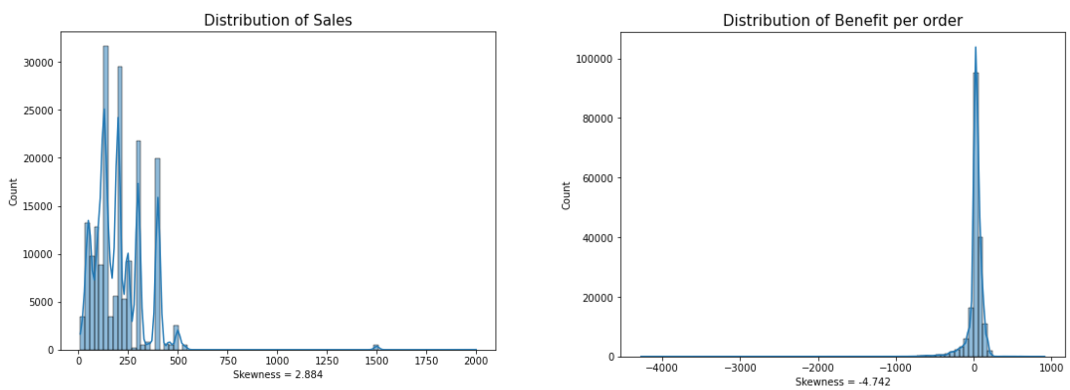
## Analysis

The analysis for this project was performed in Jupyter Notebook via Python. Although there were many interesting insights, the focus will remain on sales forecasting and late delivery prediction relevance. The remainder of this section will overview the main findings for the research questions posed for each respective case study.

Case Study 1: Prediction of Sales

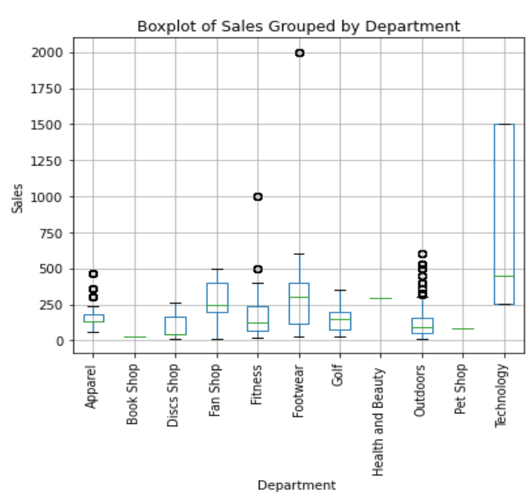
To begin, Exploratory Data Analysis was performed on the data to better understand the features. The distribution of Sales and Benefit per Order are shown in Figure 1.

*Figure 1: Distribution of Sales and Benefit per Order*



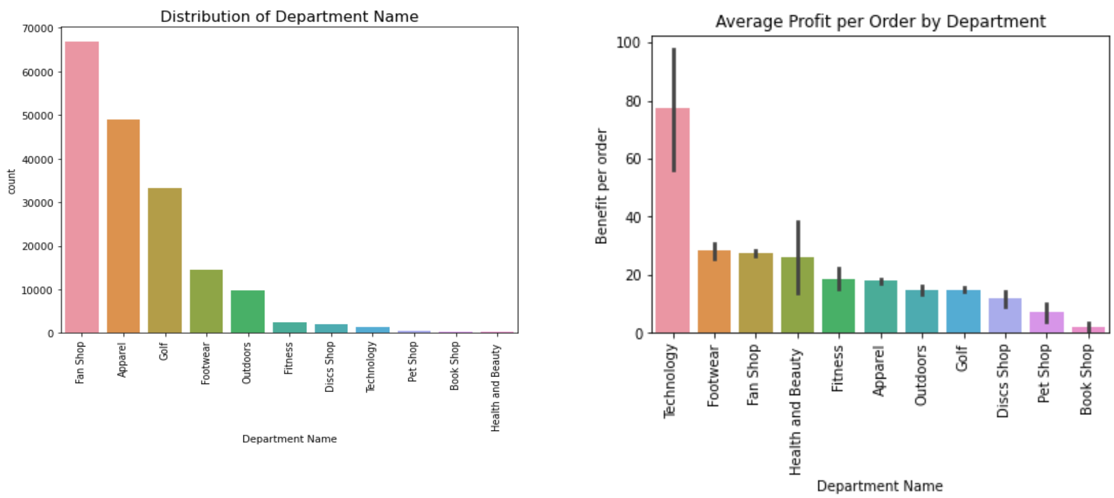

Most sales are worth less than $500, although there are a few exceptions around the $1500 mark. This tends to match the Product Price distribution very closely. A profit is not made on all orders. In fact, the distribution of Benefit per order inclines to skew towards losing money on over half of the orders. Figure 2 shows the sales distributions broken down by department.

*Figure 2: Sales by Department*



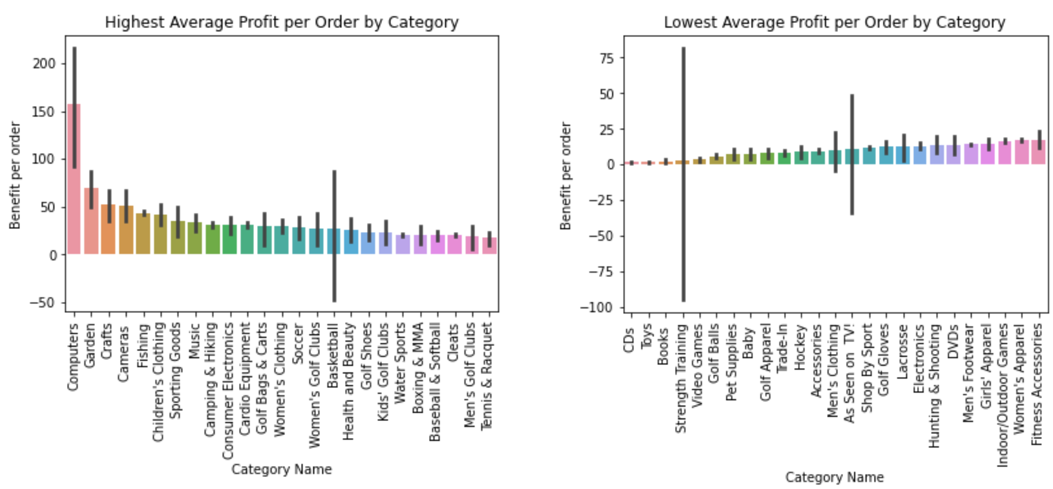
DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023
Technology sales tend to be the highest followed by Footwear, Fan Shop, Health and Beauty, and Fitness. The lowest sales are from the Book Shop, Pet Shop, and Disc Shop. Most orders came from the Fan shop, Apparel, Golf, Footwear, and Outdoors departments. However, the same findings hold true when reviewing the Average Profit per Order by Department (see Figure 3).

*Figure 3: Count of Orders by Department (left) and Average Profit per Order by Department (right)*



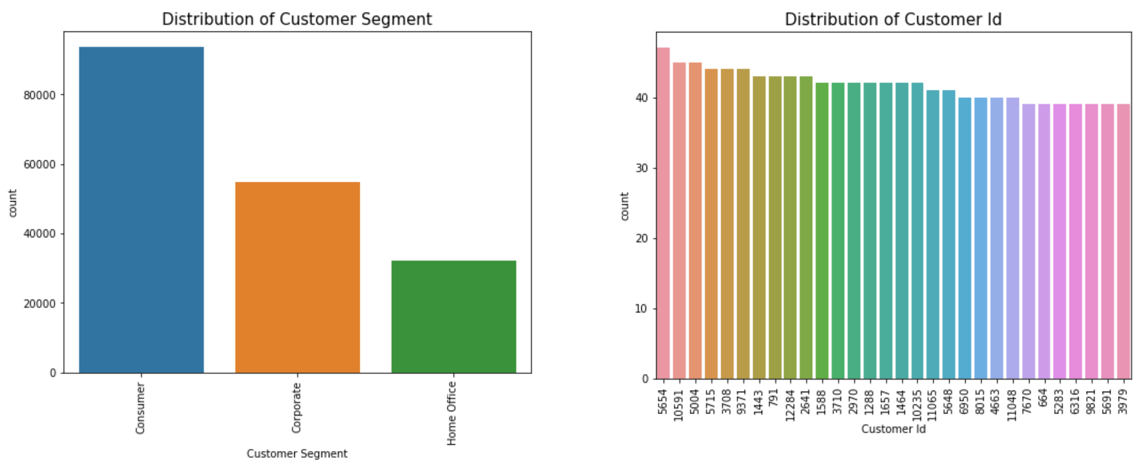

The highest Average Profit per Order by Category were Computers, Garden, Crafts, Cameras, Fishing, and Children's Clothing. The lowest average profit per order by category were CD's, Toys, Books, Strength Training, and Video Game as shown in Figure 4.

*Figure 4: Highest (right) and Lowest (left) Average Profit per Order by Category*



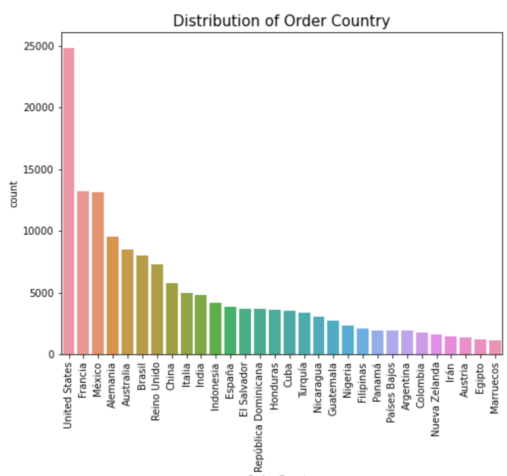

It was important to understand where the orders came from for the Sales. This was approached through Customer Segmentation, Customer ID's, and Order Location. Most orders came from Consumers and Corporate customers. Several customers submitted over 40 purchase orders as seen in  Figure 5.

*Figure 5: Order Counts by Customer Segment (left) and Customer Id (right)*



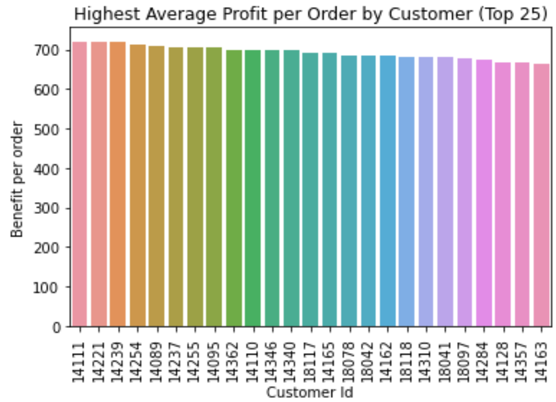

Most orders came from the United States with around 250,000 orders between 2015-2018. France and Mexico were the next highest order totals but were closer to 140,000. This is shown below in Figure 6.

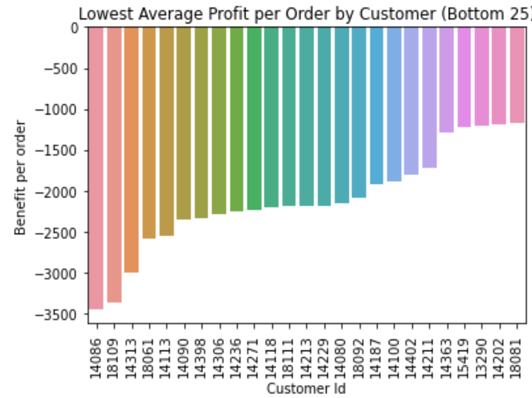
*Figure 6: Count of Orders by Country*




The highest average profit per order from customers came in just over $700. The lowest average profit per order by customer was as much as a $3500 loss. Recommend looking into the higher profit purchases by customers for additional insights. Also, it would be useful to review the purchases from the customers with the largest profit losses for additional insights.

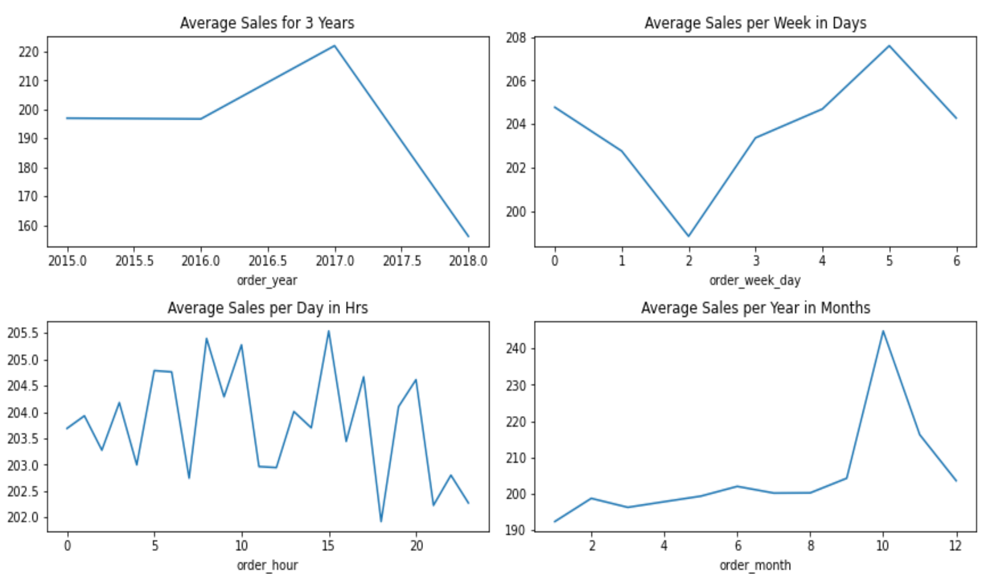
*Figure 7: Highest (left) and Lowest (right) Average Profit per Order by Customer*



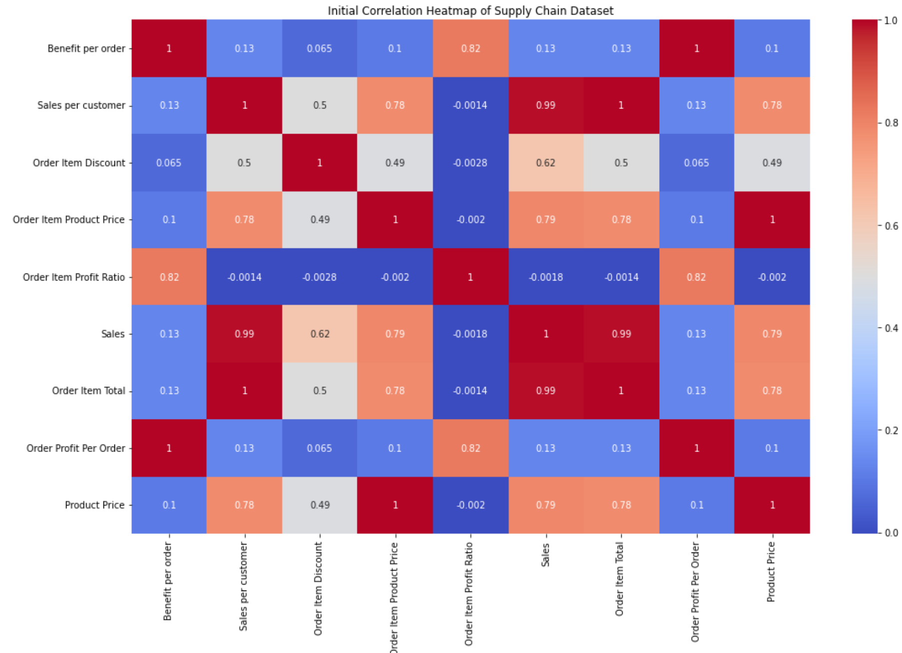

The next aspect of the analysis was whether there were any changes in Sales over time. The best years for Sales were 2015-2017. From 2017-2018, there was a sharp decline in Sales which should be investigated further. In addition, orders tend to come in later in the week, later in the year, and at any hour of the day. Figure 8 shows the general trend for Sales from the different time perspectives. This was reviewed during EDA but was not considered during the modeling phase.

*Figure 8: Time Analysis for Sales*




There were a few major findings for features correlated with Sales. Figure 9 shows the correlation heatmap of continuous, numeric features included in the supply chain data set. There were several features with a strong correlation with Sales. Those features included Product Price, Order Item Total, Order Item Product Price, Order Item Discount, and Sales per Customer.

*Figure 9: Correlation Heatmap of Supply Chain Dataset*




It was noted that there was multicollinearity present between features. As a result, features with similar record details were removed during model preparation. The columns retained for model preparation were Benefit per order, Order Item Discount, Order Item Profit Ratio, Product Price, and Sales. The records with suspected fraud or pending payment status were removed from the data prior to training as well. The data was split into 80% training and 20% testing for Linear Regression, Lasso Regression, Decision Tree, and Random Forest models. The main metric for evaluation of model performance was $R^2$, but Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were considered as well. Table 1 summarizes the results for each respective model.

*Table 1: Sales Forecasting Model Performance*

| | MAE | MSE | RMSE | R2_Score_Train | R2_Score_Test | Model |
|---|---|---|---|---|---|---|
| **RandomForest** | 1.292949 | 56.631167 | 7.525368 | 0.999181 | 0.996738 | RandomForest |
| **DecisionTree** | 1.256761 | 92.787967 | 9.632651 | 0.999530 | 0.994656 | DecisionTree |
| **Lasso** | 55.203839 | 5235.633555 | 72.357678 | 0.703851 | 0.698444 | Lasso |
| **LinearRegression** | 54.914229 | 5237.498800 | 72.370566 | 0.704701 | 0.698337 | LinearRegression |

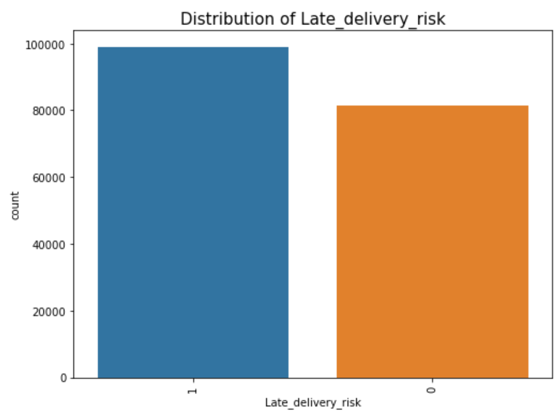
The Random Forest Regressor performed the best on the test data with an $R^2$ value of 99.67%. Use caution when deploying the model as Sales seem to fluctuate depending on industry/market trends. The two models that performed the best are susceptible to overfitting. Additional business-savvy expertise is advised in addition to the deployment of this model. The training for this model is recommended to occur once a month to ensure the model keeps up with recent data. The performance shall be continuously monitored once deployed.

DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023

<u>Case Study 2: Prediction of Late Deliveries</u>

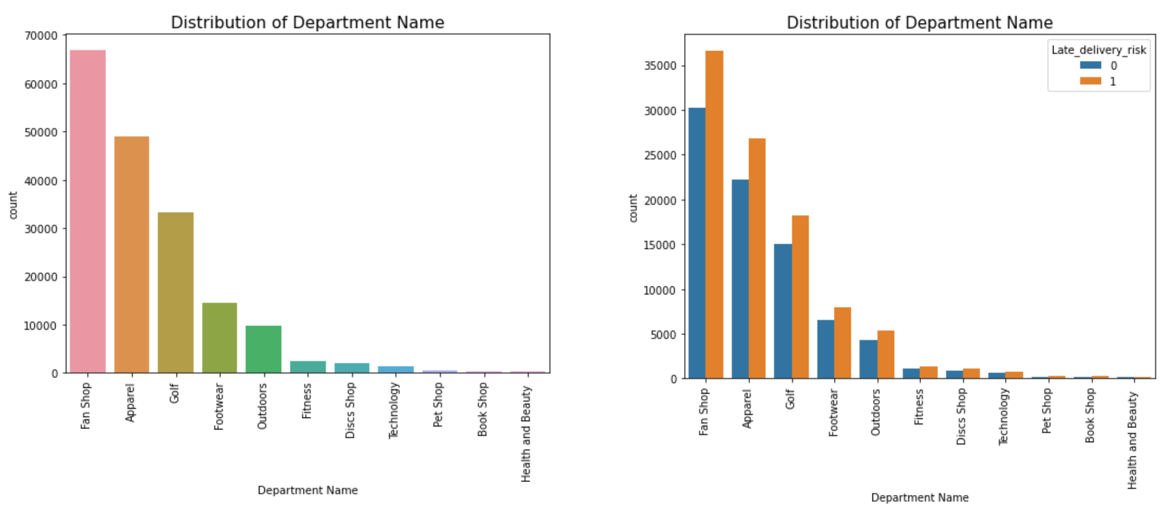
Exploratory Data Analysis was also utilized to understand the features for this case study, but with an emphasis on late deliveries. As seen in Figure 10, there was decent balance for each category in the data set.

*Figure 10: Count of Late Delivery (1-Late, 0-On-Time)*



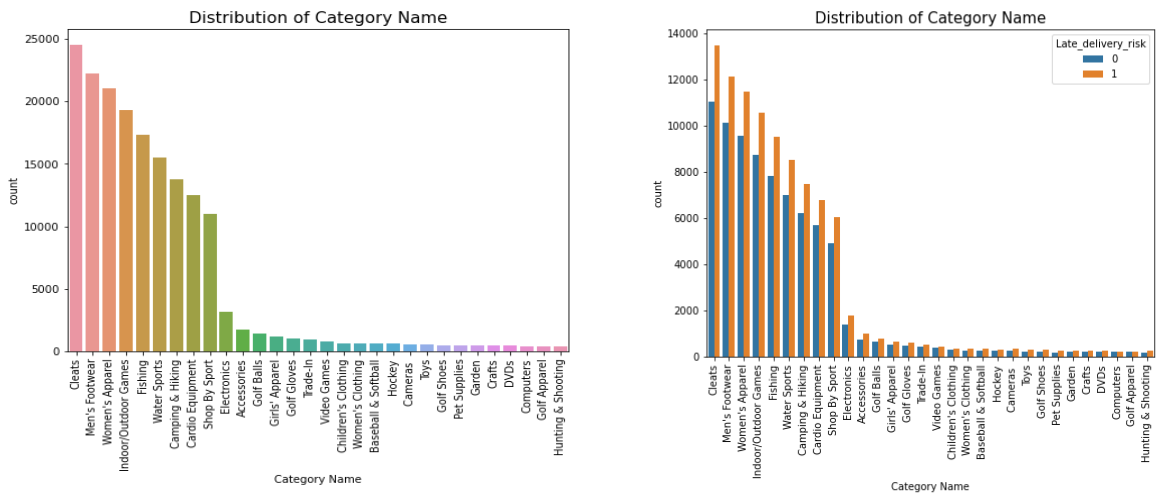

Late deliveries appeared to be a systematic issue present across the board. The issue was explored from a few different perspectives. First, late deliveries were present in all departments and were related to the number of incoming orders. The department with the highest number of late deliveries was the Fan Shop. Figure 11 shows the translation between the number of orders by department (left) and the breakdown of late deliveries (right).

*Figure 11: Late Deliveries by Department*



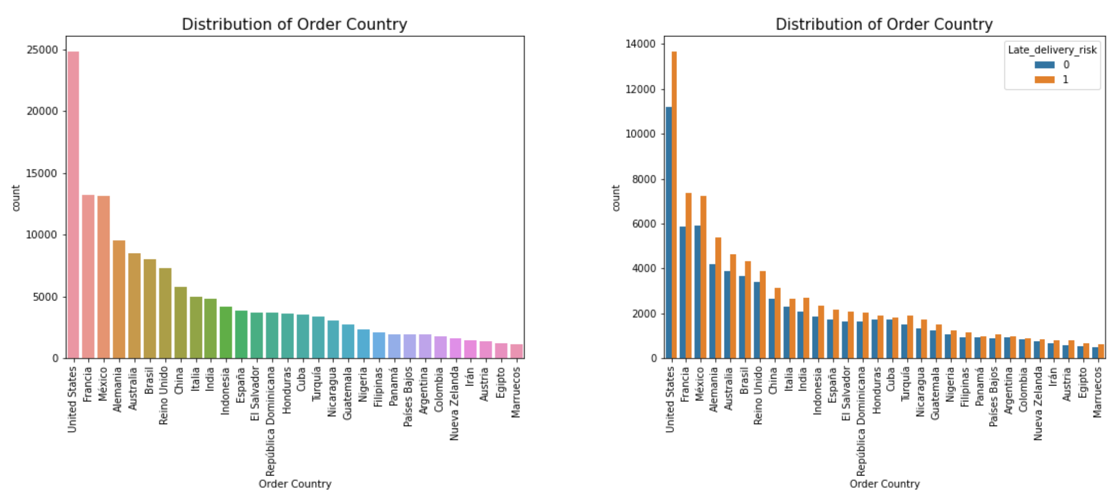

An alternative approach chosen for investigating the late delivery issue was by product. The top 30 products ordered within the timespan were reviewed for any trends with late deliveries. The same trend held true for higher number of late orders for products with more orders. Figure 12 shows the translation from the number of product orders (left) broken down by the number of late deliveries (right). Late deliveries are present in all the top product orders.

*Figure 12: Late Deliveries by Product Category*



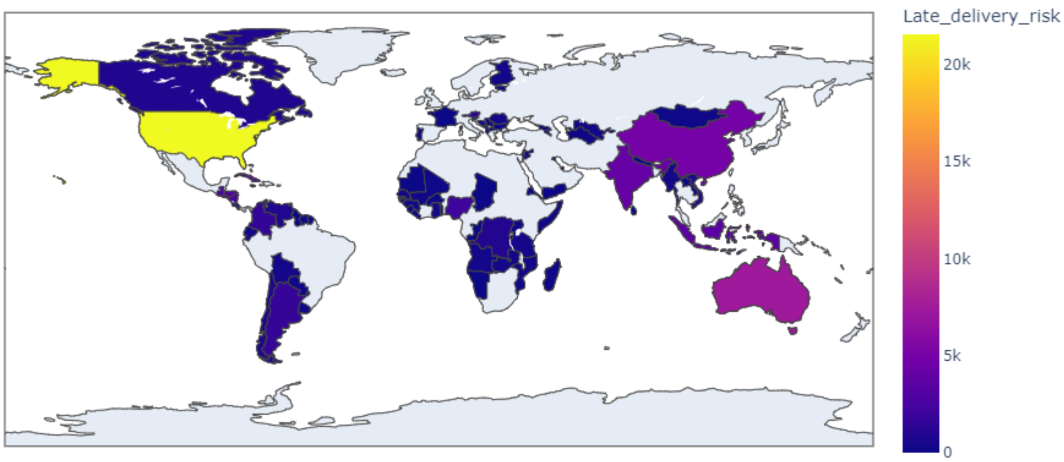

The geographic location of the orders presented an intriguing investigation point. Like the previous findings, the late deliveries were present in orders regardless of which Country. The quantity of late deliveries seemed to match the trend for the number of orders received. These findings are highlighted in Figure 13.

*Figure 13: Late Deliveries by Order Country*



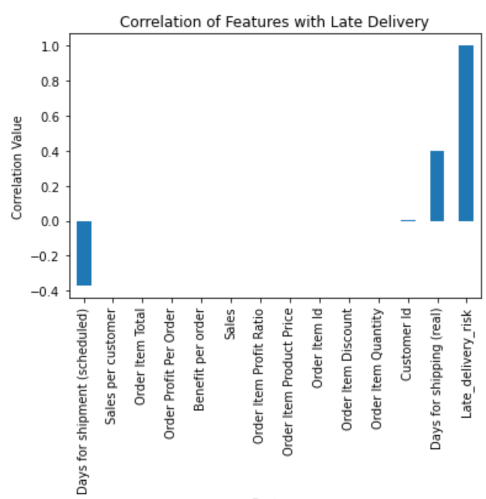

A map was generated to review the late deliveries in the various countries as shown in Figure 14. Although the United States had the most late delivery risk orders, this could be misleading since most orders were placed by the U.S. as well. The trend for late deliveries tends to increase based on the quantity of orders rather than geographic location.

*Figure 14: Map of Late Delivery Risk Count*



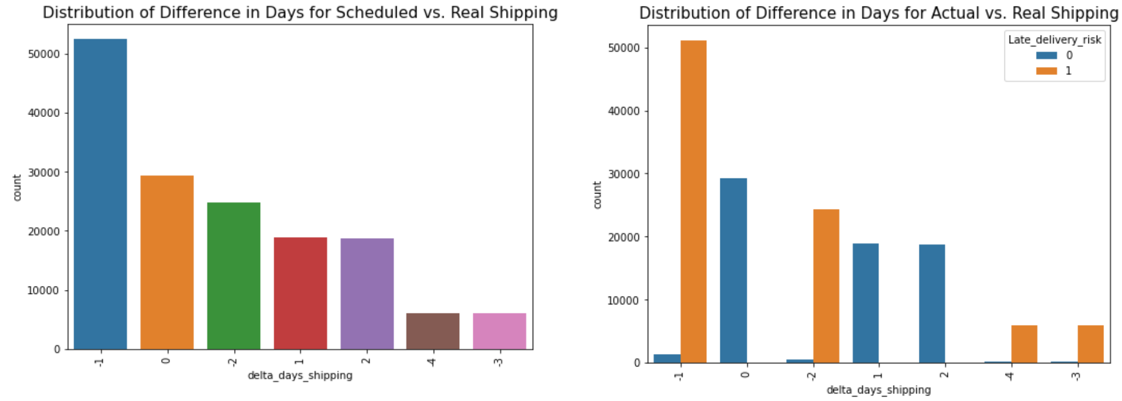

There were only a few variables strongly correlated with late deliveries. Those features were Days for shipment (scheduled) and “Days for shipping (real)”. See Figure 15 below for reference.

*Figure 15: Correlation of Features with Late Delivery Risk*



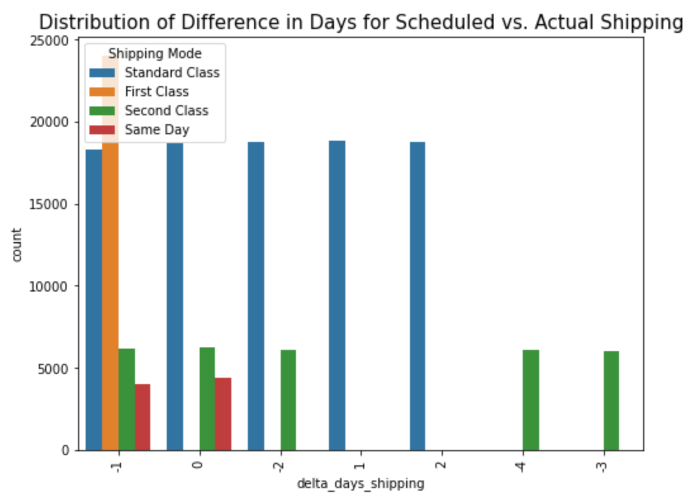

Most shipments arrive one day late or on time. The shipping mode and count of difference in days between scheduled and shipping is shown in Figure 16.

*Figure 16: Difference in Days Scheduled vs. Actual Count (left) and by Late Delivery (right)*



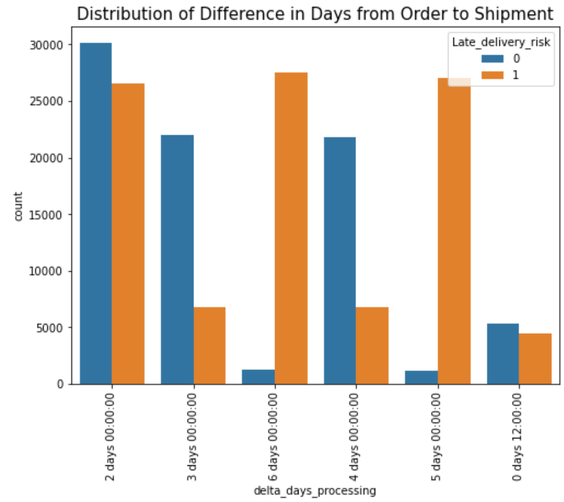

The main insight here is that the elimination of the day late (or 2 days late) deliveries will drastically reduce the on-time delivery metric. It was important to review the shipping mode for the delta days analysis as well. The day late deliveries were not one shipping mode, but rather an assortment of Standard Class, First Class, Second Class, and Same Day. Majority of shipments for a day late were from Standard Class, Second Class, or Same Day indicating they were running late initially.

*Figure 17: Count of Delta Days for Scheduled vs. Actual Shipping*



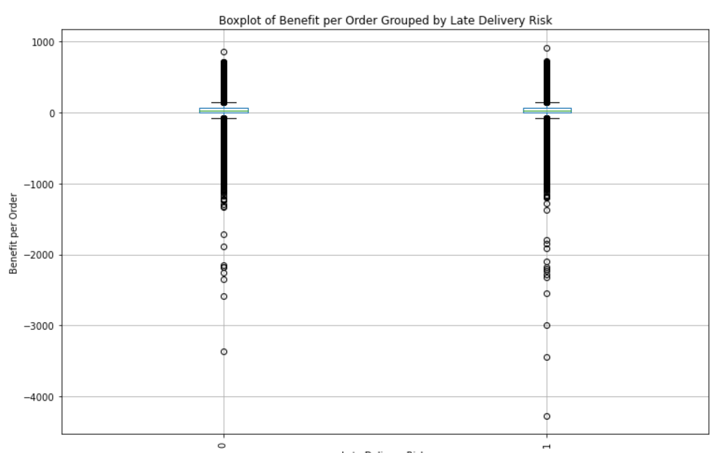

The next step was to investigate the difference (in days) from when an order was placed to when an order was shipped. Essentially, it was being reviewed whether there was too much time processing between when the order was placed and when the product shipped. Figure 18 shows that most orders took 2 days, but a large quantity of orders could take between 3-6 days. Over 50,000 orders were late when the processing time took over 5 days from when an order was placed!

*Figure 18: Difference in Days from Order Date to Ship Date Count (left) and by Late Delivery (right)*


Although a lot of orders were late for delivery, there did not appear to be a major impact on the profit made per order. Figure 19 shows a similar distribution for profit when grouped by on-time and late deliveries.

*Figure 19: Profit per Order by Late Delivery (1-Late, 0-On-Time)*



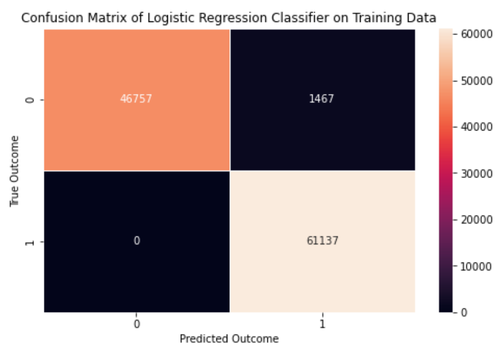

The columns included in the data frame for late delivery prediction were Late_delivery_risk, Days for shipping (real), Days for shipment (scheduled), Order Region, Department Name, and Category Name. The data was split into 30% Training and 70% Testing subsets in preparation for the predictive models. The models chosen for this analysis were Logistic Regression, Decision Tree Classifier, and Random Forest Classifier. For the Logistic Regression model, no additional preparation was performed. For the Decision Tree and Random Forest Classifiers, the features were transformed into standard scalers. The metrics used for the evaluation of the models were Accuracy, Precision, Recall, and F1-score. A summary of the performance for each model is shown in Table 2.

*Table 2: Late Delivery Model Prediction Performance*

| | Model | Model_Accuracy_Test | Model_Accuracy_Training | Model_Precision_Score | Model_Recall_Score | Model_F1_Score |
|---|---|---|---|---|---|---|
| **Logistic_Regression** | Logistic_Regression | 0.986345 | 0.986586 | 0.976091 | 1.000000 | 0.987901 |
| **Random_Forest** | Random_Forest | 0.985918 | 0.986787 | 0.976109 | 0.999196 | 0.987517 |
| **Decision_Tree** | Decision_Tree | 0.985491 | 0.986805 | 0.976162 | 0.998354 | 0.987133 |

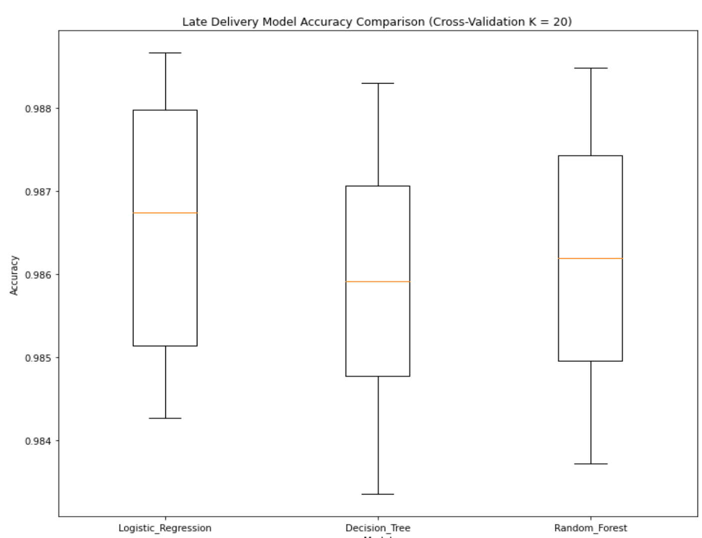
Logistic Regression performed the best with an accuracy of 98.64% on the test data. The Confusion Matrix for the Logistic Regression model is shown below.

*Figure 20: Logistic Regression Confusion Matrix*




As seen in the matrix (and Recall Score), there were no predicted on-time deliveries that arrived late. However, there were 1,467 deliveries predicted late, but arrived on time. Cross-validation was also utilized to show model performance repeatability. Logistic Regression tended to perform the best.

*Figure 21: Cross-Validation (kfold = 20) Model Performance*



The Logistic Regression model appears to be suitable for deployment. It is recommended to provide some current data prior to fully deploying it. Also, continuously monitor the model to ensure the performance is up to date.

## Conclusion, and Recommendations

In conclusion, the CRISP-DM methodology was followed to analyze the DataCo Global Supply Chain data for relevant insights related to sales forecasting and late delivery prediction. Several insights were found during the Exploratory Data Analysis step for each case study. The findings are highlighted below:

Case Study 1: Prediction of Sales

- The Technology department had the highest sales and provided the largest profit per order on average (~$77.25). Note: Fan Shop had the most quantity of orders.
- Customer ID 5654 bought the most often, but several customers have placed over 40 orders. The company made the most profit from Customer ID 14111 with an average of $720.30 per order.
- The features with the strongest correlation with Sales were Product Price, Order Item Total, Order Item Product Price, Order Item Discount, and Sales per Customer. However, there was multicollinearity present between some features in the data. The main features considered for the modeling phase were Benefit per order, Order Item Discount, Order Item Profit Ratio, Sales, and Product Price.
- The model with the best accuracy for Sales Forecasting was Random Forest Regressor with an $R^2$ value on test data of 99.67%.

Case Study 2: Prediction of Late Deliveries

- The Fan Shop had the highest number of late deliveries, but this department also had the highest number of orders.
- Similarly, cleats had the highest number of late deliveries for a product, but also had the highest number of orders compared to other products.
- Rather than geographic location, late deliveries appeared to be a systematic issue independent of location. The higher number of orders resulted in a higher number of late deliveries.
- The two features strongly correlated with late deliveries were scheduled shipping days (inverse) and actual shipping days (direct).
- The Logistic Regression model performed the best for predicting late deliveries with an accuracy of 98.64% on test data.

Prior to deploying the model, recommend consulting technical experts within the supply chain group to confirm/discuss the findings. In addition, the data utilized in this analysis was from 2015-2018. I recommend utilizing more up-to-date data for training and testing. Lastly, I would recommend trying some alternative models for each case study to see if there are any additional improvements to the current model performance. Additional actions from this analysis include looking into the decline in sales from 2017-2018, investigating high profit orders from customers, and adjusting the scheduled delivery days currently used (or improve the lead time for days between orders processed and orders shipped).

DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023

## Future Uses, Limitations, and Challenges

This analysis provides a methodology for other organizations to follow to better understand their own supply chain issues. The supply chain data will vary for each organization since supply chain processes are customized to the business needs. The challenges can be approached in a similar manner with data-driven insights that will improve the current situation. This analysis is not a one-size-fit-all solution. The methodology can be applied across different organizations to recognize specific improvement areas custom to that company. The main challenge for this analysis was narrowing down the analysis to fit both case studies. Next time, recommend only selecting one specific case study and treating the other as an independent project. At times it became convoluted during the analysis of the supply chain data set.

## Implementation Plan

The Random Forest Regressor for prediction of sales and the Logistic Regression Classifier for late delivery prediction appear to be ready for deployment. Prior to releasing the model, a few recommendations need to be made. First, it is recommended to consult the supply chain team with the findings from the analysis. Second, train the model with data that is more current than 2018. Work with the supply chain team to determine a frequency to train and monitor the models.

## Ethical Assessment

Ethics play an essential role in data analysis projects. This analysis was no different. To maintain data privacy, the features with customer information in the data set were removed (names, emails, passwords, etc…). Although the data set was made public, the consent of these individuals and organizations would be required to disclose the personal information in the analysis. Each step was outlined in the process to clearly, and transparently, show the audience the choices made in the analysis. There was no intent to deceive with the charts or results. The intent for the analysis was to present the findings in an honest manner that clearly shows the story with the data. Lastly, the method used for this analysis can be replicated within additional organizations to recognize improvement opportunities within the supply chain.

## References

AI Monks (n.d.). *Top 10 Analytics Projects in Operations and Supply Chain*. Resources. Retrieved June 10, 2023, from https://aimonks.com/supply-chain-analytics-top-10-analytics-projects/

Ariwala, P. (2022, September 30). *9 Ways Machine Learning Can Transform Supply Chain Management*. Artificial Intelligence and Machine Learning. Retrieved June 10, 2023, from https://marutitech.com/machine-learning-in-supply-chain/

Constante, F., Silva, F., & Pereira, A. (2019, March 12). *Datasets*. DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS. Retrieved June 10, 2023, from https://data.mendeley.com/datasets/8gx2fvg2k6/5

Diaz, J., Marcolis, C., Washburn, R. (2023, March 31). *CDW*. How the Modern Data Platform Fuels Success. Retrieved June 19, 2023, from https://www.cdw.com/content/cdw/en/articles/dataanalytics/how-the-modern-data-platform-fuels-success.html?cm_ven=acquirgy&cm_cat=bing&cm_pla=SLG+Analytics&cm_iteBundle+Modern+Data+Platform+Broad&s_kwcid=AL!4223!10!73392704939126!73392726583203&ef_id=f97c2ba4bbf8166f67094f346210261b:G:s&msclkid=f97c2ba4bbf8166f67094f346210261b

Hotz, N. (2023, January 19). *What is CRISP DM?*. Data Science Process Alliance. Retrieved June 16, 2023, from https://www.datascience-pm.com/crisp-dm-2/

IBM (2021, August 17). *CRISP-DM Help Overview*. Documentation. Retrieved June 10, 2023, from https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview

Tiwari, S. (2019, March 12). *DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS*. About Dataset. Retrieved June 10, 2023, from https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis?resource=download&select=DataCoSupplyChainDataset.csv

Ying, K. (2021, September 16). *Data Science Ethics – What Could Go Wrong and How to Avoid It*. Data Science. Retrieved June 10, 2023, from https://www.freecodecamp.org/news/the-ethics-of-data-science/

DSC 680 Applied Data Science (2237-1)
Bellevue University
Professor Williams
Milestone 2 – Draft White Paper
Jake Meyer
06/25/2023

# Appendix A: Milestone Template Criteria

Each Milestone must follow the following template

**Milestone 1 - Proposal**

Submit a proposal with the following topics covered as a PDF

- Topic - Describe and name your project in 1-2 sentences max
- Business Problem - Describe the business problem your project is trying to solve and/or the research questions you will explore
- Datasets - where are you getting your data? Describe the data that you will use to solve the problem
- Methods - What analysis methods will you use to complete this project? Note this is just a proposal, your project can adapt as you work on it
- Ethical Considerations - What are some potential ethical concerns of this topic or analyzing the data?
- Challenges/Issues - What are some issues and challenges do you think you might face?
- References - What sources will you use to validate your results and support your project topic?

**Milestone 2 - Draft White Paper**

Submit a draft of your white paper with the following topics covered as a PDF

- Business Problem
- Background/History
- Data Explanation (Data Prep/Data Dictionary/etc)
- Methods
- Analysis
- Conclusion
- Assumptions
- Limitations
- Challenges
- Future Uses/Additional Applications
- Recommendations
- Implementation Plan
- Ethical Assessment

## Appendix B: Variable Descriptions for Data Set

| FIELDS | DESCRIPTION |
|---|---|
| Type | Type of transaction made |
| Days for shipping (real) | Actual shipping days of the purchased product |
| Days for shipment (scheduled) | Days of scheduled delivery of the purchased product |
| Benefit per order | Earnings per order placed |
| Sales per customer | Total sales per customer made per customer |
| Delivery Status | Delivery status of orders Advance shipping, Late delivery, Shipping canceled, Shipping on time |
| Late_delivery_risk | Categorical variable that indicates if sending is late (1), it is not late (0). |
| Category Id | Product category code |
| Category Name | Description of the product category |
| Customer City | City where the customer made the purchase |
| Customer Country | Country where the customer made the purchase |
| Customer Email | Customer's email |
| Customer Fname | Customer name |
| Customer Id | Customer ID |
| Customer Lname | Customer lastname |
| Customer Password | Masked customer key |
| Customer Segment | Types of Customers Consumer, Corporate, Home Office |
| Customer State | State to which the store where the purchase is registered belongs |
| Customer Street | Street to which the store where the purchase is registered belongs |
| Customer Zipcode | Customer Zipcode |
| Department Id | Department code of store |
| Department Name | Department name of store |
| Latitude | Latitude corresponding to location of store |
| Longitude | Longitude corresponding to location of store |
| Market | Market to where the order is delivered Africa, Europe, LATAM, Pacific Asia, USCA |
| Order City | Destination city of the order |
| Order Country | Destination country of the order |
| Order Customer Id | Customer order code |
| order date (DateOrders) | Date on which the order is made |
| Order Id | Order code |
| Order Item Cardprod Id | Product code generated through the RFID reader |
| Order Item Discount | Order item discount value |
| Order Item Discount Rate | Order item discount percentage |
| Order Item Id | Order item code |

| | |
|---|---|
| Order Item Product Price | Price of products without discount |
| Order Item Profit Ratio | Order Item Profit Ratio |
| Order Item Quantity | Number of products per order |
| Sales | Value in sales |
| Order Item Total | Total amount per order |
| Order Profit Per Order | Order Profit Per Order |
| Order Region | Region of the world where the order is delivered Southeast Asia ,South Asia ,Oceania ,Eastern Asia, West Asia , West of USA , US Center , West Africa, Central Africa ,North Africa ,Western Europe ,Northern , Caribbean , South America ,East Africa ,Southern Europe , East of USA ,Canada ,Southern Africa , Central Asia , Europe , Central America, Eastern Europe , South of USA |
| Order State | State of the region where the order is delivered |
| Order Status | Order Status COMPLETE, PENDING, CLOSED, PENDING_PAYMENT, CANCELED, PROCESSING , SUSPECTED_FRAUD , ON_HOLD , PAYMENT_REVIEW |
| Product Card Id | Product code |
| Product Category Id | Product category code |
| Product Description | Product Description |
| Product Image | Link of visit and purchase of the product |
| Product Name | Product Name |
| Product Price | Product Price |
| Product Status | Status of the product stock If it is 1 not available, 0 the product is available |
| Shipping date (DateOrders) | Exact date and time of shipment |
| Shipping Mode | The following shipping modes are presented Standard Class, First Class, Second Class, Same Day |