Business Problem | Data Inputs | Data Discovery | Prediction Process | User Output

# IMPROVE FORECAST ACCURACY OF INBOUND MERCHANDISE SHIPMENTS

- PROBLEM: Inaccurate projections of number of containers required on inbound vessels leads to unexpected costs, shipment delays, and reduced trust with business partners.
  - **Under-forecast**: steamship lines will not have enough room onboard for our containers.
    - Option 1: Pay spot market rates to find space and ship product on time.
    - Option 2: Delay product shipment until following week = late delivery to distribution centers & stores = missed sales opportunities.
  - **Over-forecast**: results in charge for unused vessel space and damages relationships with carriers.
    - Lost revenue for carriers if they could have given the unused space to another shipper.
- Cross-functional teams involved
  - Steamship lines & freight forwarders (external)
  - Inbound Logistics (internal)
  - DC Operations (internal)
  - Supply Chain Data Science (internal)

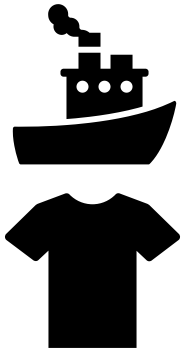
Business Problem → **Data Inputs** → Data Discovery → Prediction Process → User Output

## INBOUND SHIPMENT HISTORY

- 1 year of weekly inbound purchase order shipments
- 21 export shipping ports
- 6 inbound global distribution centers
- SKU-level purchase order data

## PRODUCT ATTRIBUTES

- Purchase Order & SKU information
  - Ordered unit quantities
  - Unit dimensions (length, width, height)
  - SKU Category (i.e. jeans, shirts, dresses)
  - Factory information
  - Carton standard
  - **Cubic meters shipped (target variable for prediction)**



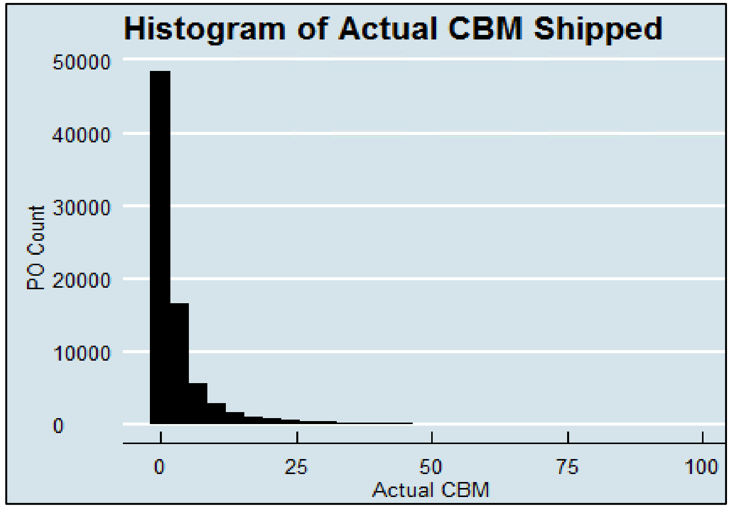
**IF WE KNOW WHERE FUTURE ORDERS ARE SHIPPING FROM/TO, HOW BIG THE UNITS ARE, ORDER QUANTITIES, CAN WE PREDICT HOW MANY CUBIC METERS THE ORDER WILL COMPRISE?**

Business Problem | Data Inputs | **Data Discovery** | Prediction Process | User Output

# DATA DISCOVERY & EXPLORATION PRIOR TO PREDICTIVE MODELING

## Target Variable Exploration

- Discovered that nearly half of all orders had shipment volume of **exactly 1.0 cubic meters**.
  - Data Issue!
  - Collaboration with logistics team to raise awareness of incorrect data.
  - Fixed data issue for historical data to train model with.
  - Cost save discovered by uncovering discrepancy.
    - Overpayments to transload providers.




| cbm_value | po_count | pct_total |
|---|---|---|
| cbm != 1 | 40798 | 51.34 |
| cbm == 1 | 38667 | 48.66 |

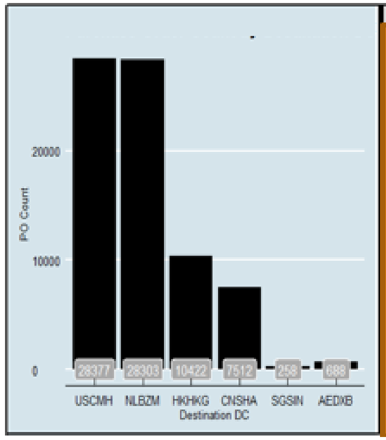
Business Problem | Data Inputs | **Data Discovery** | Prediction Process | User Output

# DATA DISCOVERY & EXPLORATION PRIOR TO PREDICTIVE MODELING

## Equal shipments to all distribution centers?

- US & Netherlands DC's received the most shipments.
  - Dictated prioritization
- All export ports do not ship to all locations.
  - Some vendors may only ship to certain DC's.
- Order size differs by destination DC.
  - Not only do the most orders ship to US & Netherlands DC's, but the size of the orders to those destinations tend to be larger.

### ORDERS SHIPPED TO DESTINATION DC

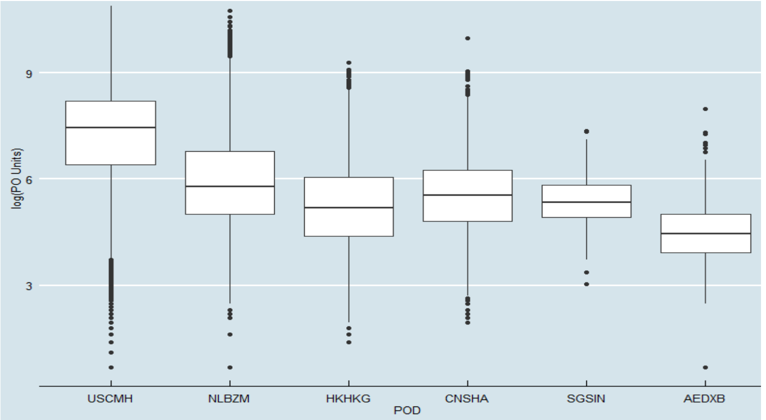



21 export ports

| port_of_export | USCMH | NLBZM | HKHKG | CNSHA | SGSIN | AEDXB |
|---|---|---|---|---|---|---|
| | 0 | 183 | 88 | 4 | 0 | 0 |
| | 668 | 609 | 362 | 353 | 6 | 6 |
| | 3420 | 2850 | 1846 | 0 | 18 | 83 |
| | 4080 | 5137 | 2063 | 1616 | 44 | 89 |
| | 1714 | 1390 | 830 | 662 | 14 | 80 |
| | 2122 | 2388 | 1049 | 0 | 48 | 42 |
| | 4530 | 3688 | 0 | 1714 | 35 | 124 |
| | 1948 | 1751 | 744 | 569 | 9 | 121 |
| | 1026 | 1016 | 390 | 296 | 4 | 23 |
| | 5338 | 5379 | 2124 | 1654 | 30 | 57 |
| | 953 | 1050 | 22 | 0 | 13 | 0 |
| | 28 | 17 | 7 | 0 | 1 | 0 |
| | 1007 | 1011 | 418 | 330 | 20 | 30 |
| | 1365 | 891 | 421 | 314 | 11 | 28 |
| | 97 | 91 | 58 | 0 | 5 | 5 |
| | 0 | 348 | 0 | 0 | 0 | 0 |
| | 0 | 328 | 0 | 0 | 0 | 0 |
| | 0 | 296 | 0 | 0 | 0 | 0 |
| | 12 | 0 | 0 | 0 | 0 | 0 |
| | 35 | 0 | 0 | 0 | 0 | 0 |
| | 34 | 0 | 0 | 0 | 0 | 0 |

*Hidden for confidentiality

### UNITS PER PURCHASE ORDER BY DESTINATION DC

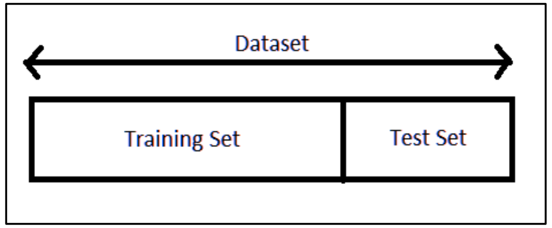


Business Problem | Data Inputs | Data Discovery | **Prediction Process** | User Output

# PREDICT CUBIC METERS PER SHIPMENT

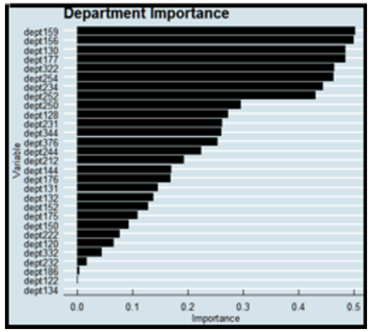
## 1. DATA SPLITTING

Divided labeled data in 75/25 train/test splits.




## 2. FEATURE EVALUATION

Conducted tests for variable importance to understand which variables are prominent predictors.




## 3. FEATURE SELECTION

Used feature evaluations to select variables to use as predictors.

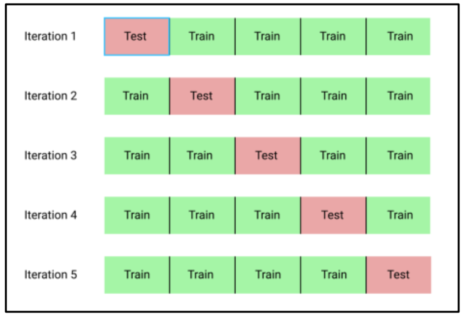
<u>Selected Predictors</u>

- Units Ordered
- Export location
- Factory Name
- Category (i.e. jeans, sweaters)
- Carton Standard (suggested units to pack per box)
- Unit dimensions
  - Unit Length
  - Unit Width
  - Unit Height

## 4. MODEL DEVELOPMENT

Build and test multiple models to evaluate prediction accuracy.

Train models with 5-fold cross validation.



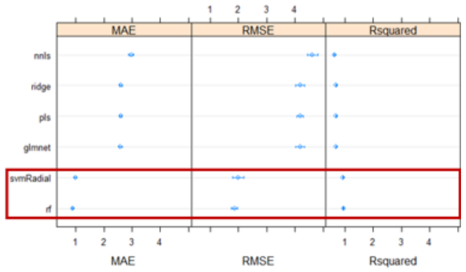
Business Problem | Data Inputs | Data Discovery | **Prediction Process** | User Output

# PREDICT CUBIC METERS PER SHIPMENT

## 5. MODEL SELECTION

Select model with best test-set RMSE and R-Squared.

**Selected Random Forest model.**




## 6. EVALUATE PREDICTIONS

Compare predictive model to existing forecast method.

| Bucket | New Model | Status Quo |
|---|---|---|
| Under 1* | 77.9% | 56.7% |
| 1 to 2 | 89.7% | 73.3% |
| 2 to 5 | 97.1% | 89.5% |
| 5 to 10 | 99.4% | 95.8% |

*Under 1: 77.9% of predicted values are within 1 cubic meter (CBM) of the actual CBM on the shipment, compared to only 56.7% within 1 CBM using the existing method.

## 7. DEPLOY MODEL AND MAKE PREDICTIONS

Use models to make volume predictions on future purchase orders.

1. Obtain SKU information and order to ship in next 8 weeks (SQL).
2. Use script to transform data into required input format for prediction (R).
3. Make predictions on unseen data (R).
4. Aggregate predictions into weekly shipments (R).
5. Visualize results in easy to use format for Logistics team (Tableau).

Business Problem | Data Inputs | Data Discovery | Prediction Process | User Output

# AGGREGATE INDIVIDUAL PURCHASE ORDER VOLUME PREDICTIONS INTO WEEKLY SHIPMENT PROJECTIONS

## SAMPLE PREDICTION OUTPUT

| PO | POD (Destination) | POE (Export) | Ship Week | Predicted CBM | Predicted Containers |
|---|---|---|---|---|---|
| 1234 | USCMH | CNTAO | 201921 | 35 | 1.0 |
| 5678 | USCMH | CNTAO | 201921 | 30 | 1.0 |
| 9012 | USCMH | CNTAO | 201921 | 10 | 0.5 |
| 3456 | USCMH | VNSGN | 201922 | 400 | 6.5 |
| 7890 | USCMH | VNSGN | 201922 | 150 | 2.5 |

## FINAL OUTPUT ON TABLEAU SERVER (USER INTERACTIVE TOOL)

| | | | Pred GAC Cal Wk / Pred Sail Cal Wk | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Carrier | POD | POE | 201920 201921 | 201921 201922 | 201922 201923 | 201923 201924 | 201924 201925 | 201925 201926 | 201926 201927 | 201927 201928 | Grand Total |
| TOLL | USCMH | CNDLC | | | | | 1.0 | | | | 1.0 |
| | | CNTAO | 2.5 | 3.0 | 1.5 | 6.0 | 5.0 | 2.0 | 6.5 | 3.5 | 30.0 |
| | | HKHKG | 4.5 | 9.0 | 5.5 | 5.0 | 5.0 | 2.0 | 4.5 | 6.5 | 42.0 |
| | | KHPNH | 1.5 | 3.0 | 12.5 | 4.5 | 4.5 | 2.5 | 13.0 | 2.0 | 43.5 |
| | | LKCMB | 0.5 | 0.5 | 1.0 | 0.5 | | 2.0 | | 3.0 | 7.5 |
| | | PHMNL | 2.0 | 3.5 | 2.5 | 4.0 | 2.0 | 2.5 | 3.0 | 1.5 | 21.0 |
| | | TRIST | 0.5 | | 0.5 | | 0.5 | 0.5 | 0.5 | | 2.5 |
| | | VNHPH | 16.5 | 12.0 | 12.5 | 15.5 | 17.5 | 7.0 | 17.0 | 8.0 | 106.0 |
| | | VNSGN | 9.0 | 15.0 | 9.5 | 21.0 | 12.0 | 5.0 | 5.5 | 6.0 | 83.0 |
| | NLBZM | CNSHA | 1.0 | 1.0 | 1.0 | 1.0 | 3.5 | 0.5 | 2.0 | | 10.0 |
| | | CNTAO | 0.5 | 1.5 | 2.0 | 3.5 | 0.5 | 2.0 | 2.5 | | 12.5 |
| | | GTSTC | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | | | 3.0 |
| | | HKHKG | 1.5 | 2.5 | 2.0 | 2.0 | 2.0 | 1.5 | 2.0 | 0.5 | 14.0 |
| | | PHMNL | 0.5 | 0.5 | 0.5 | 1.5 | 1.0 | 1.0 | 1.0 | 0.5 | 6.5 |
| | | USEZA | | | | | 1.0 | | | 0.5 | 1.5 |
| | | VNSGN | 4.0 | 3.0 | 4.5 | 5.0 | 3.0 | 1.5 | 2.5 | 1.0 | 24.5 |