The Road to Better Deliveries: Pruning Lead Time

GROUP : 9 Sunny Khade, Nidhi Verhani, Gunjan Kale, Dev Kotian, Parv Maru

DATASET: LOGISTICS

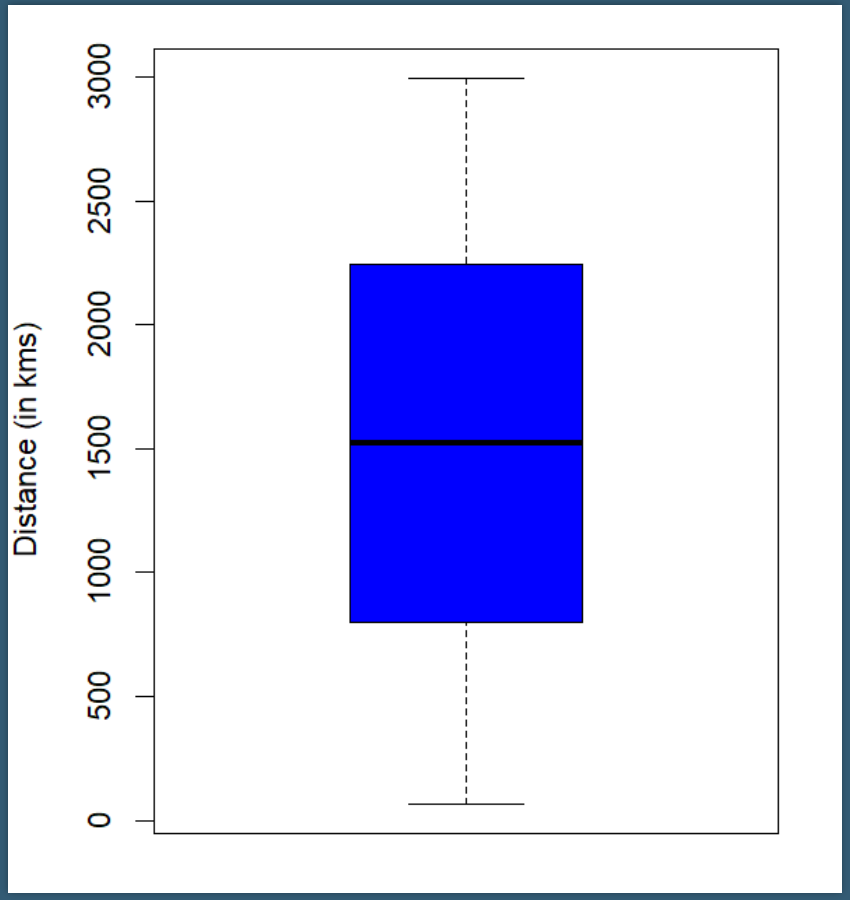
Research Problem

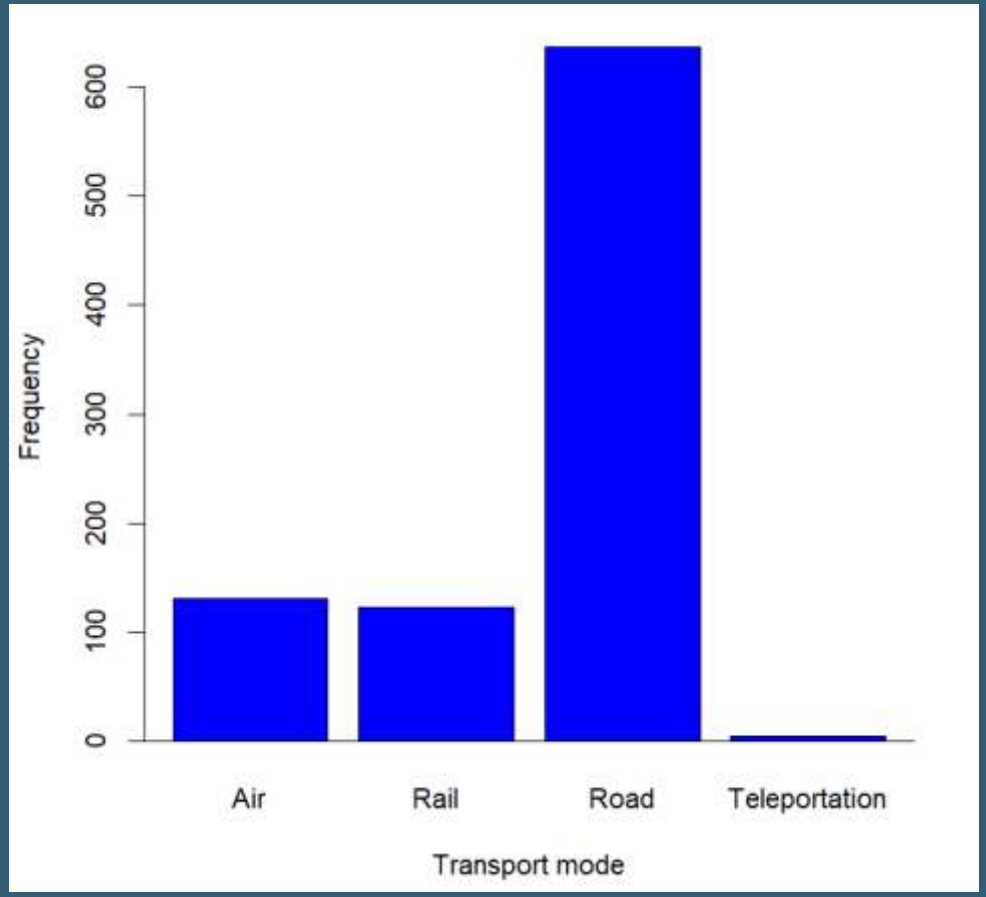
Can we predict lead time of deliveries in order to minimize it ?

Data Preprocessing

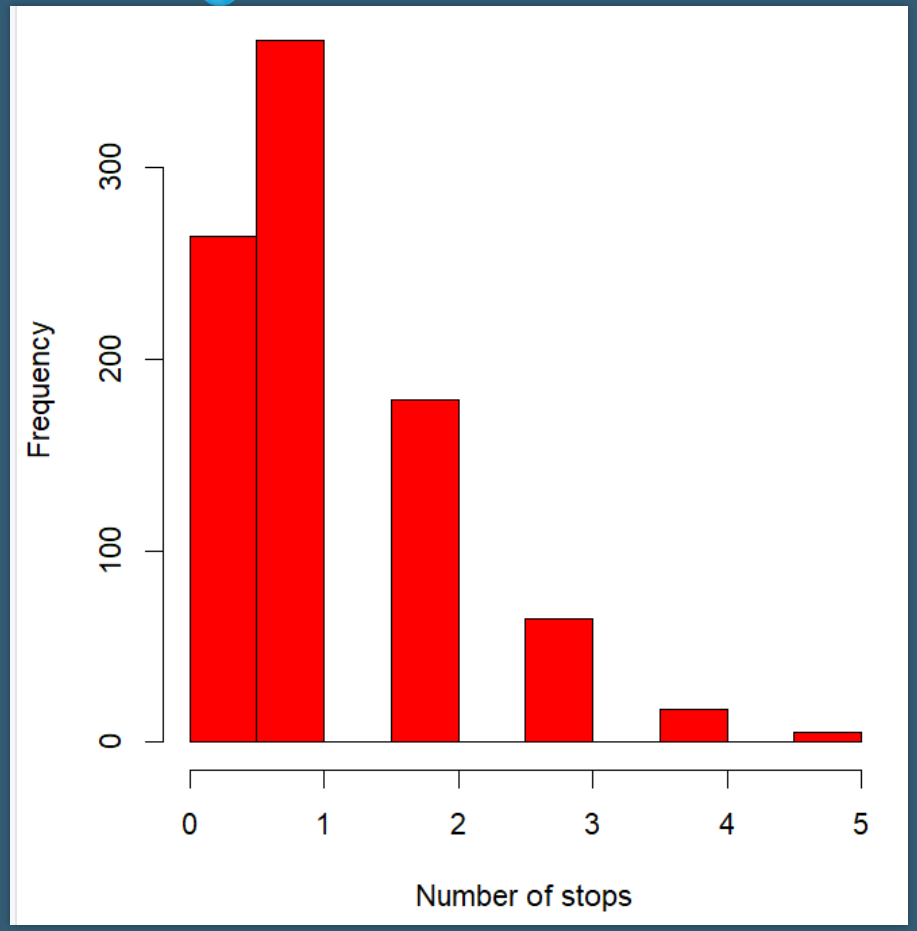
- Removing NULL values.
- Removing outliers (using IQR).
- Removing negative “delivery_lead_time” values.
- Removing unwanted columns: “id”, “factory_location”, “destination”.

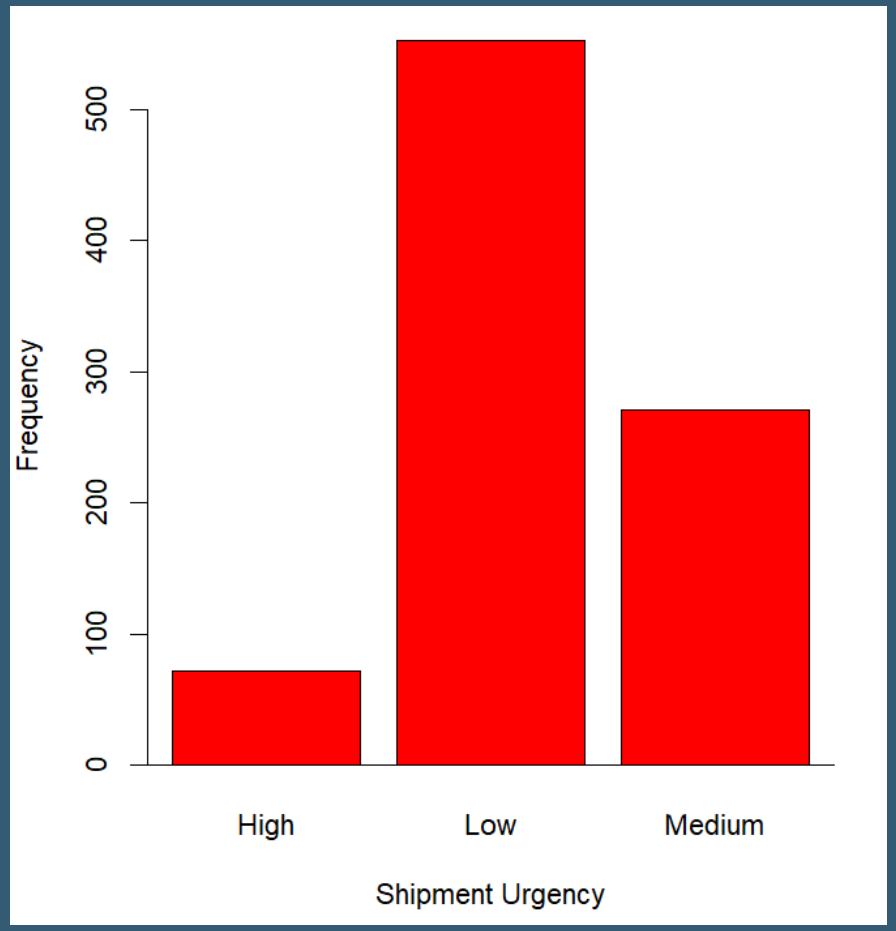
The Data (after preprocessing)



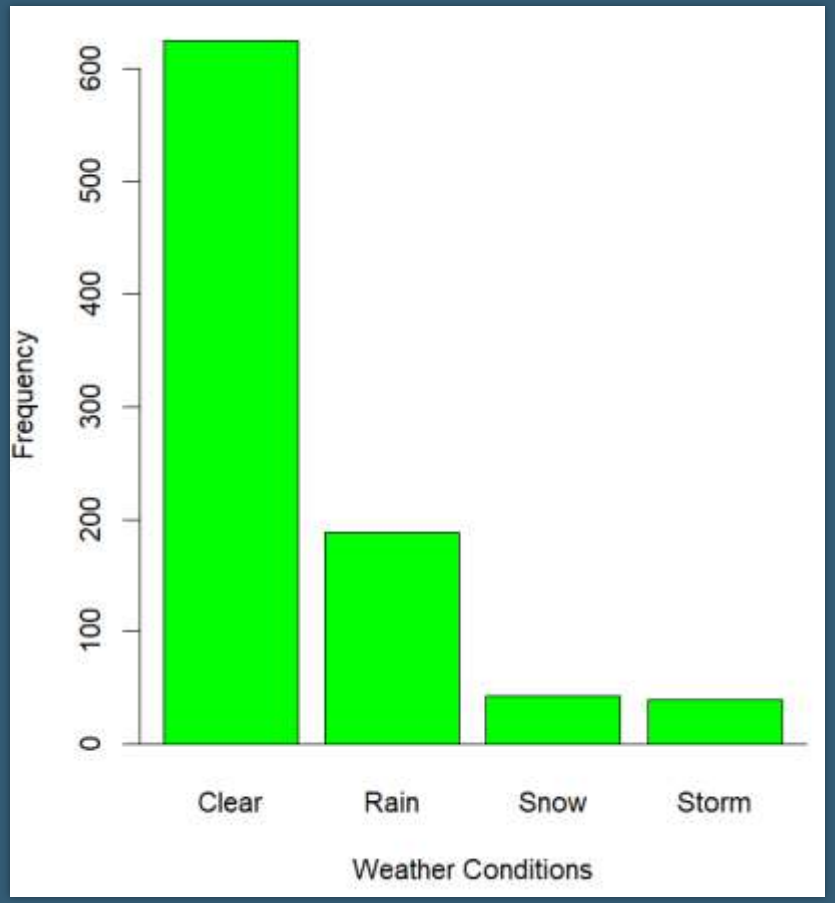


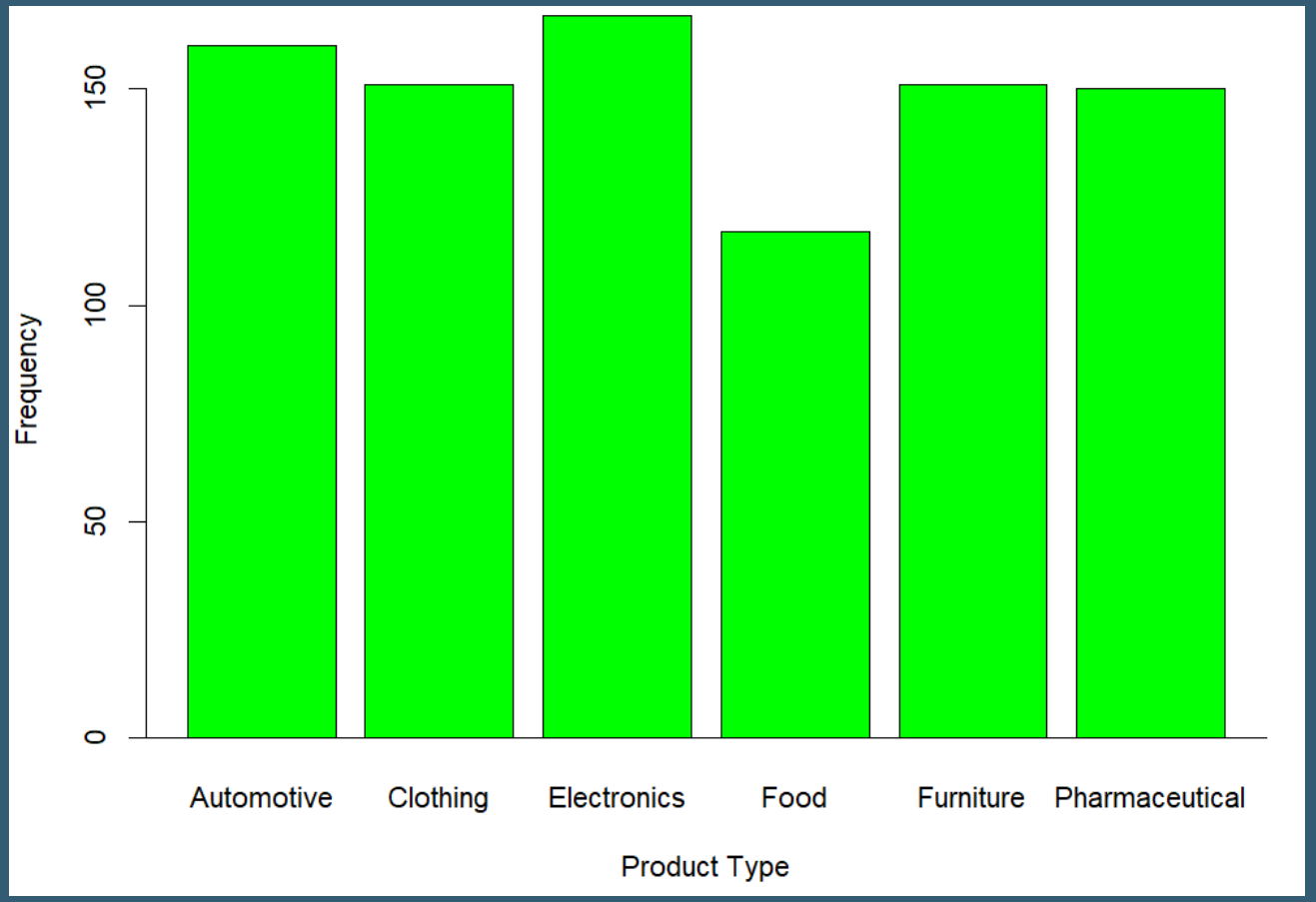
The Data (after preprocessing)



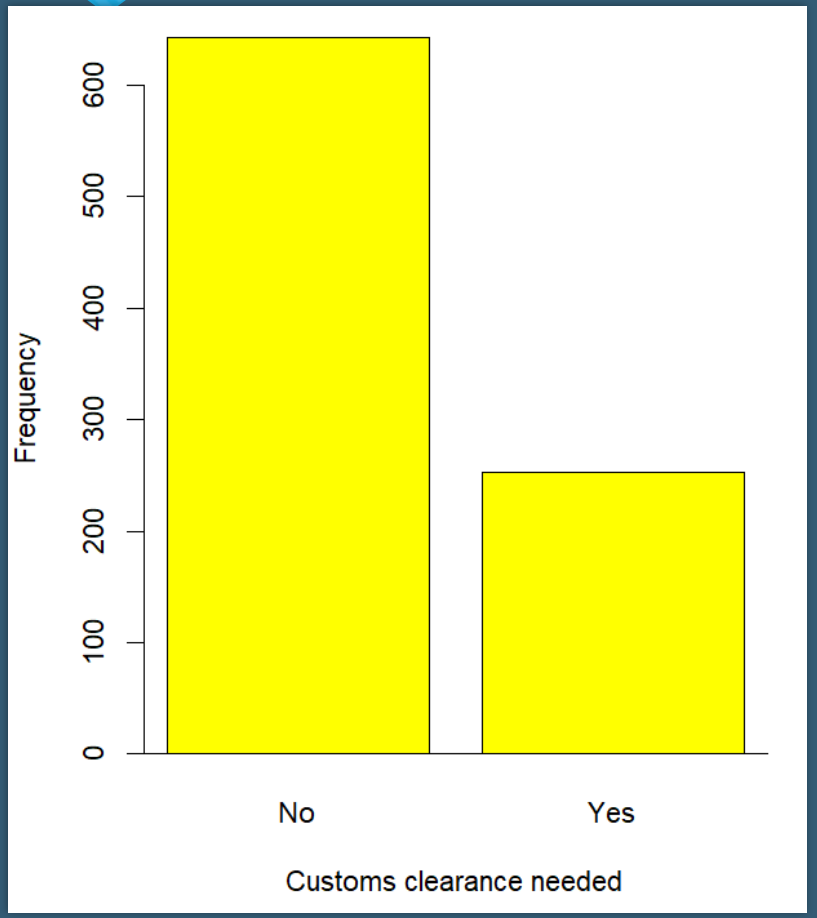


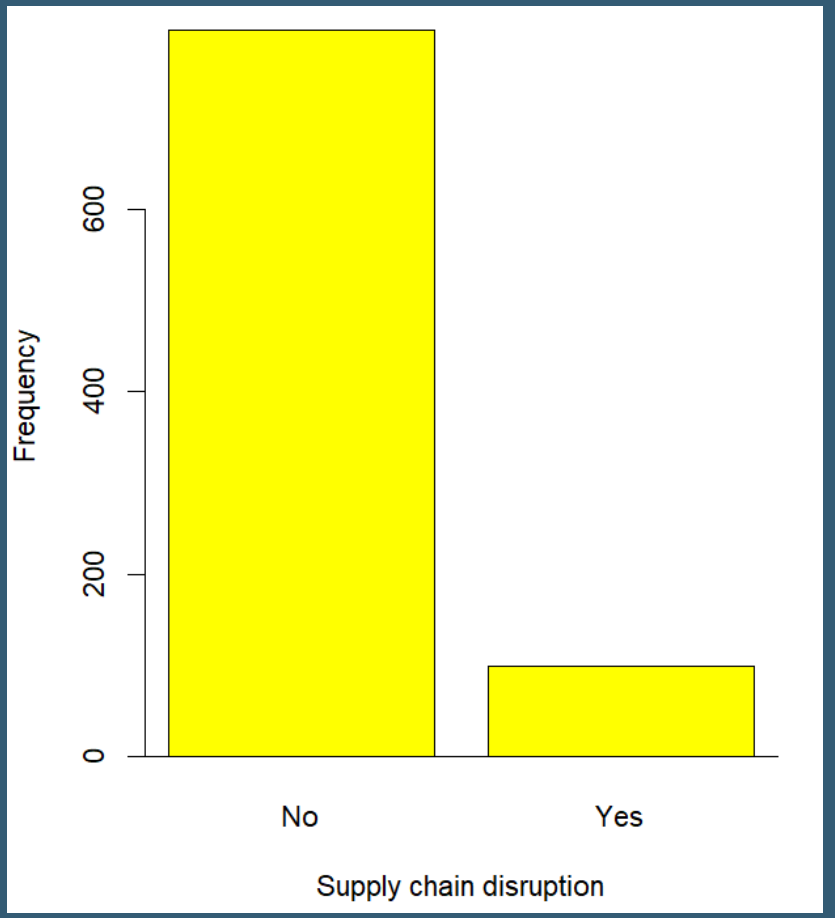
The Data (after preprocessing)





The Data (after preprocessing)





Dummy Variables

- Categorical Variables: transport_mode, shipment_urgency, weather_conditions, product_type, packaging_type, special_handling, customs_clearance, supply_chain_disruption, seasonality, temp_control
- Created dummy variables for categorical variables using “as.factor()” method

Data Partitioning

- Created a 60-40 split.
- 60% Data – Training dataset
- 40% Data – Validation dataset

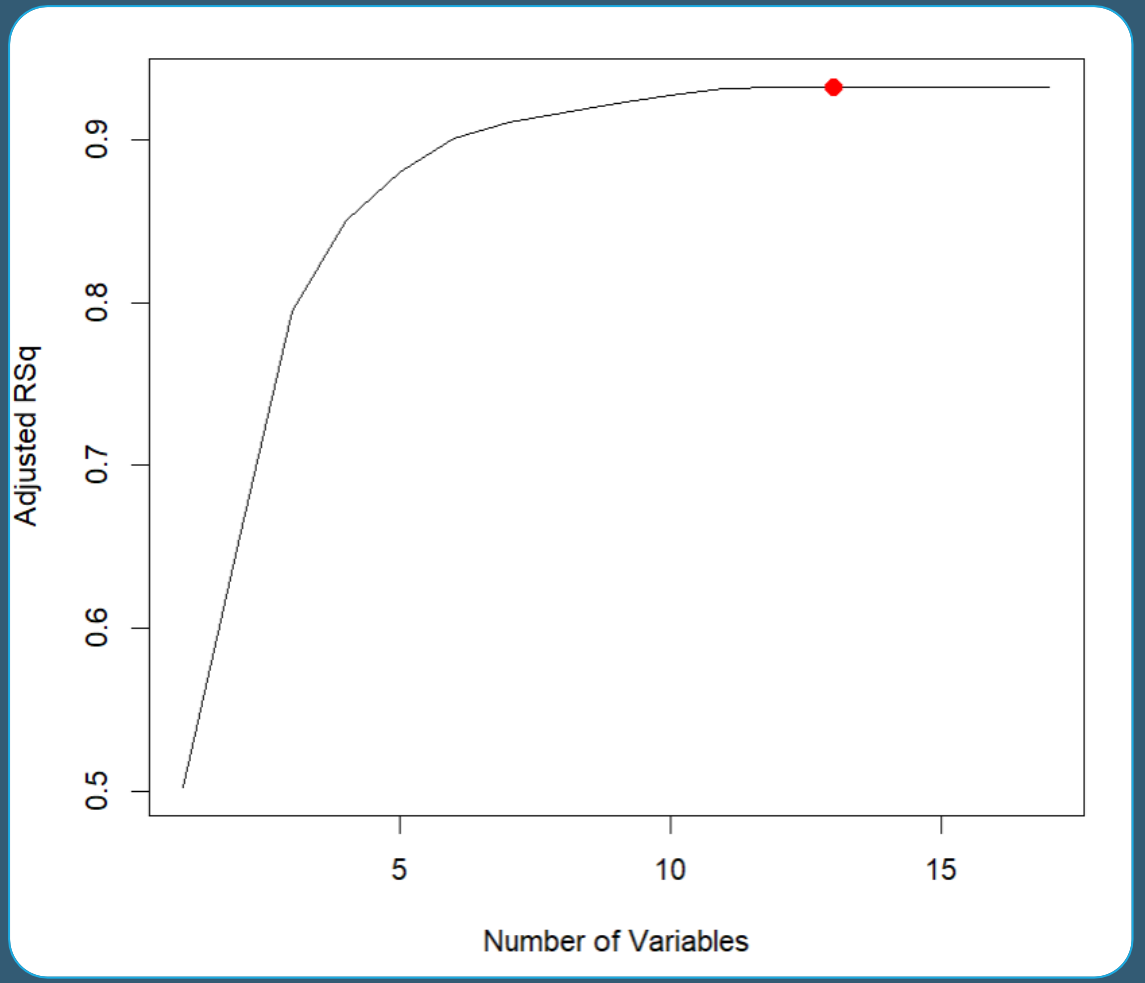
```
set.seed(1)
train_indices <- sample(1:nrow(logistics_cleaned),
                        size=nrow(logistics_cleaned)*0.6,
                        replace=F)

logistics_train <- logistics_cleaned[train_indices,]
logistics_validation <- logistics_cleaned[-train_indices,]
```

Variable Selection: Best Subset Method

Maximum Adjusted R Squared :

13 Variables



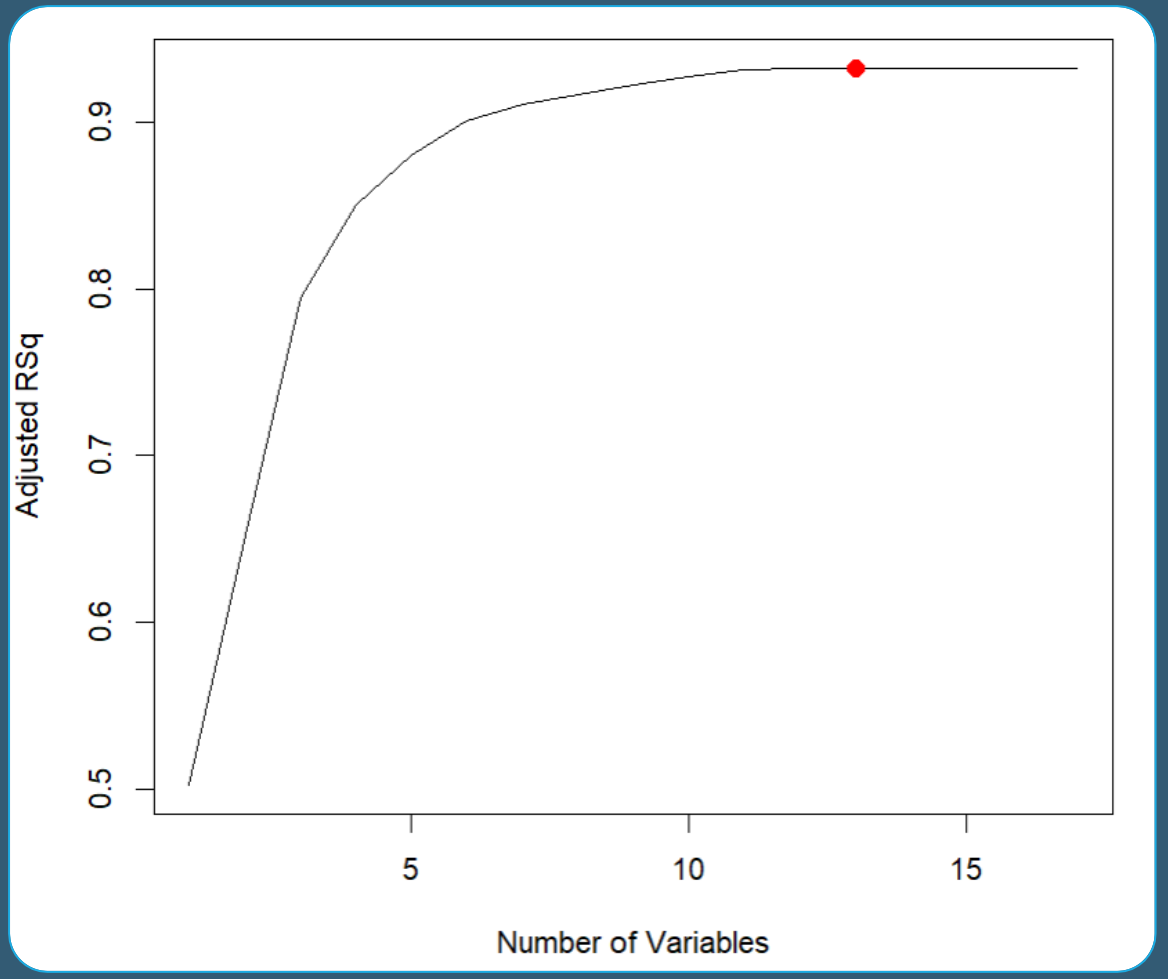
Variable Selection: Best Subset Method

- Variables selected by best subset method

```
> coef(regfit.best,13)
                (Intercept)                 distance_km       transport_modeRail
              -1.664384317                0.001775937               0.621624334
       transport_modeRoad transport_modeTeleportation                 num_stops
              1.091980170                0.822854677               0.473804044
      shipment_urgencyLow     shipment_urgencyMedium  weather_conditionsRain
              1.035542412                0.566004875               0.399715943
   weather_conditionsSnow     weather_conditionsStorm product_typeElectronics
              0.854918920                1.200982231               0.085952356
     customs_clearanceYes  supply_chain_disruptionYes
              1.879928436                2.825262539
>
```

Variable Selection: Forward Selection

- Same results as best subset selection method



Linear Regression Model

- Linear regression model using the variables selected from the best subset method.
- We got a RMSE of 0.542

```
> forecast::accuracy(predicted.best, logistics_validation$delivery_lead_time)
                     ME      RMSE       MAE          MPE    MAPE
Test set -0.01969293 0.542498 0.4224151 -0.6243155 13.705
```

Linear Regression Model

- Model Summary:
- We got an Adjusted R-Squared value of 0.9323

```
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                -1.666e+00  1.206e-01 -13.808  < 2e-16 ***
distance_km                 1.775e-03  2.792e-05  63.593  < 2e-16 ***
transport_modeRail          6.234e-01  9.154e-02   6.811 2.70e-11 ***
transport_modeRoad          1.093e+00  6.669e-02  16.387  < 2e-16 ***
transport_modeTeleportation 8.294e-01  2.809e-01   2.953  0.00329 **
num_stops                   4.722e-01  2.372e-02  19.907  < 2e-16 ***
shipment_urgencyLow         1.039e+00  8.821e-02  11.775  < 2e-16 ***
shipment_urgencyMedium      5.649e-01  9.371e-02   6.028 3.16e-09 ***
weather_conditionsRain      3.997e-01  5.793e-02   6.900 1.52e-11 ***
weather_conditionsSnow      8.580e-01  1.033e-01   8.307 8.58e-16 ***
weather_conditionsStorm     1.197e+00  1.207e-01   9.918  < 2e-16 ***
product_typeClothing        4.207e-02  7.890e-02   0.533  0.59415
product_typeElectronics     8.679e-02  7.666e-02   1.132  0.25812
product_typeFood            1.338e-02  8.424e-02   0.159  0.87384
product_typeFurniture      -2.800e-02  7.882e-02  -0.355  0.72259
product_typePharmaceutical -1.890e-02  7.987e-02  -0.237  0.81305
customs_clearanceYes        1.881e+00  5.256e-02  35.795  < 2e-16 ***
supply_chain_disruptionYes  2.825e+00  7.631e-02  37.018  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5397 on 519 degrees of freedom
Multiple R-squared:  0.9344,    Adjusted R-squared:  0.9323
F-statistic: 434.9 on 17 and 519 DF,  p-value: < 2.2e-16
```

Linear Regression Model

- Linear regression model using **all variables**:
- RMSE: 0.544
- Adjusted R-Squared: 0.9313

```
Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -1.590e+00  2.906e-01  -5.472 7.00e-08 ***
distance_km                  1.776e-03  2.878e-05  61.719  < 2e-16 ***
transport_modeRail           6.238e-01  9.296e-02   6.710 5.21e-11 ***
transport_modeRoad           1.101e+00  6.791e-02  16.216  < 2e-16 ***
transport_modeTeleportation  8.215e-01  2.860e-01   2.872  0.00424 **
fuel_price                   3.874e-02  5.620e-02   0.689  0.49090
shipment_weight              1.931e-06  4.247e-06   0.455  0.64957
shipment_volume              2.983e-04  8.670e-04   0.344  0.73093
num_stops                    4.726e-01  2.409e-02  19.619  < 2e-16 ***
shipment_urgencyLow          1.034e+00  8.978e-02  11.517  < 2e-16 ***
shipment_urgencyMedium       5.593e-01  9.634e-02   5.805 1.13e-08 ***
weather_conditionsRain       3.979e-01  5.885e-02   6.761 3.77e-11 ***
weather_conditionsSnow       8.738e-01  1.054e-01   8.292 1.00e-15 ***
weather_conditionsStorm      1.215e+00  1.226e-01   9.913  < 2e-16 ***
carrier_reliability         -2.301e-03  2.838e-03  -0.811  0.41794
order_volume                 5.209e-05  8.295e-05   0.628  0.53030
product_typeClothing         4.386e-02  7.972e-02   0.550  0.58243
product_typeElectronics      7.849e-02  7.811e-02   1.005  0.31544
product_typeFood             1.361e-02  8.587e-02   0.159  0.87412
product_typeFurniture       -3.413e-02  8.016e-02  -0.426  0.67042
product_typePharmaceutical  -2.556e-02  8.157e-02  -0.313  0.75412
packaging_typeOversized      4.295e-02  6.653e-02   0.646  0.51880
packaging_typeRefrigerated  -2.318e-02  6.491e-02  -0.357  0.72114
packaging_typeStandard      -1.887e-02  6.878e-02  -0.274  0.78392
special_handlingYes         -3.120e-02  6.096e-02  -0.512  0.60899
customs_clearanceYes         1.880e+00  5.362e-02  35.059  < 2e-16 ***
supply_chain_disruptionYes   2.822e+00  7.777e-02  36.284  < 2e-16 ***
seasonalityPeak              3.704e-02  5.266e-02   0.703  0.48210
temp_controlYes              4.477e-02  7.234e-02   0.619  0.53628
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

K-Nearest Neighbors

- Data Partitioning :
- 90% Data - Training validation dataset.
- 10% Data – Testing dataset.

```
train_indices_knn <- sample(1:nrow(logistics_cleaned),
                            size=nrow(logistics_cleaned)*0.9,
                            replace=F)

logistics_train_valid <- logistics_cleaned[train_indices_knn,]
logistics_test <- logistics_cleaned[-train_indices_knn,]
```

K-Nearest Neighbors

- Final value used for model: K=7

k	RMSE	Rsquared	MAE
1	1.430521	0.5556116	1.0616552
2	1.271151	0.6227520	0.9336230
3	1.250811	0.6352066	0.9498013
4	1.235130	0.6529772	0.9340141
5	1.222124	0.6699865	0.9293478
6	1.218744	0.6808743	0.9410978
7	1.215639	0.6991838	0.9439054
8	1.229281	0.7018544	0.9497455
9	1.239873	0.7082527	0.9662713
10	1.247913	0.7142950	0.9761372
11	1.265957	0.7136970	0.9914183
12	1.283541	0.7131971	1.0101693
13	1.299995	0.7087451	1.0235494
14	1.323000	0.7056871	1.0418352
15	1.331239	0.7087519	1.0483537

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was k = 7.

K-Nearest Neighbors

- We got a RMSE of 1.131

```
> forecast::accuracy(predicted_knn,logistics_test$delivery_lead_time)
                     ME      RMSE       MAE       MPE     MAPE
Test set -0.01462302 1.130595 0.9039881 -15.06652 29.51093
```

Model Selection

We select the Linear Regression model using the variables from best subset method.

RMSE = 0.542

Adjusted R-Squared = 0.9323

Thank You !