**Model Building**

The objective here is to build a model, using historical data that will determine an optimum weight of the product to be shipped each time from the respective warehouse. This is a regression problem. Multiple models are trained and tested. The various types of regression models used are:

1. Linear Regression
2. Ridge Regression
3. Lasso Regression
4. Decision Tree Regression
5. Random Forest Regression
6. KNN Regression
7. AdaBoost Regression

**1. Linear Regression**

* This model produced an R-squared and adjusted R squared value 0.978.
* From the model summary it is understood that features transport\_issue\_l1y, distributor\_num, wh\_est\_year, storage\_issue\_reported\_l3m, wh\_breakdown\_l3m, Location\_type, temp\_reg\_mach, approved\_wh\_govt\_certificate can be considered as relevant features.
* Test and Train score percentage are 97.70% and 97.75% which indicate that no overfitting and underfitting has occurred.
* MAE = 1298.8109065670133
* MSE value = 3079574.5660483395
* RMSE = 1754.871666546685

**2. Ridge Regression**

* The train and test score percentage are 97.75% and 97.70% which indicate that no overfitting and underfitting has occurred.
* MAE = 1298.7457301465547
* MSE = 3079409.245126389
* RMSE = 1754.8245624923277

**3. Lasso Regression**

* The train and test score percentage are 97.75% and 97.70% which indicate that no overfitting and underfitting has occurred.
* MAE = 1296.8250053171755
* MSE = 3076376.8879492655
* RMSE = 1753.9603438930042

**4. Decision Tree Regression**

* The train and test score percentage are 100% and 98.80% which indicate that no overfitting and underfitting has occurred.
* MAE = 837.1171903214157
* MSE = 1602122.000902853
* RMSE = 1265.7495806449444

**5. Random Forest Regression**

* The train and test score percentage are 99.9% and 99.3 % which indicate that no overfitting and underfitting has occurred.
* MAE = 681.096743409173
* MSE = 828661.1152029947
* RMSE = 910.3082528478991

**6. KNN Regression**

* The train and test score percentage are 73.6% and 58.9%. It has a low test score and a good training score. This shows a condition of overfitting.
* MAE = 5799.377248104008
* MSE = 55066187.09821597
* RMSE = 7420.6594786592905

**7. AdaBoost Regression**

* The train and test score percentage are 97.1 % and 96.9% which indicate that no overfitting and underfitting has occurred.
* MAE = 1570.2045720023452
* MSE = 4130340.9728428284
* RMSE = 2032.3240324423732

From the above 7 models the best 3 models are produced by Random Forest Regression, Decision Tree Regression and Ridge Regression. This is determined based on the RMSE value.