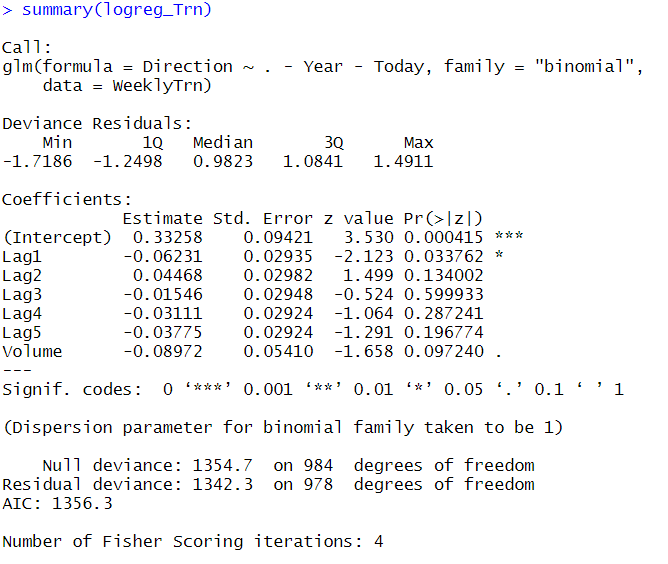
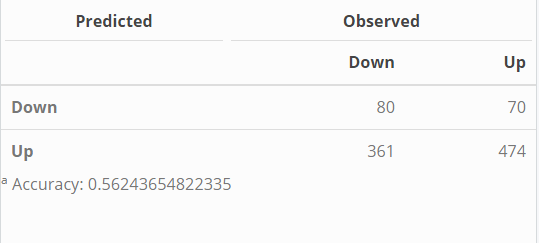
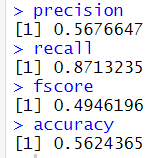
**Logistic Regression using all the variables except “Year” and “Today”:**

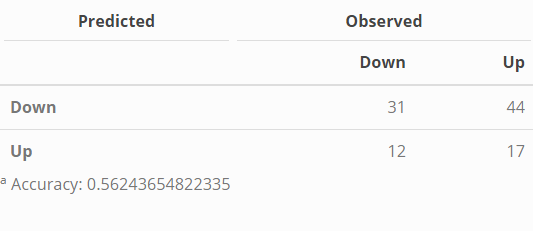
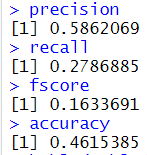
**Summary:**



**Confusion Matrix of Training Data:**

**** ****

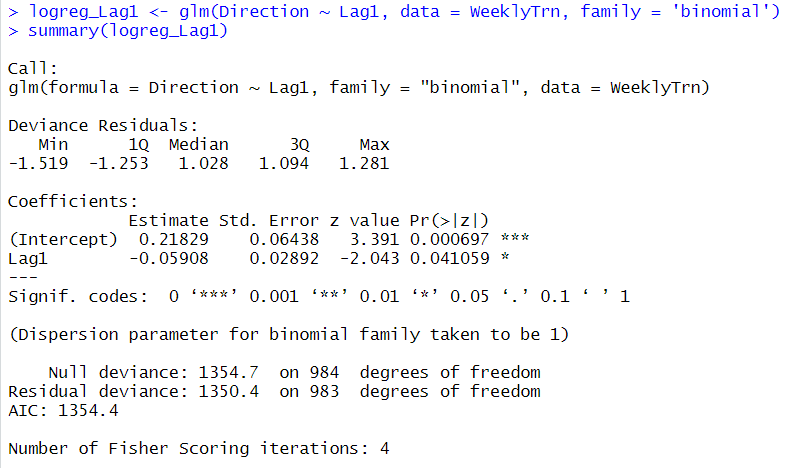
**Confusion Matrix of Test Data:**

**** ****

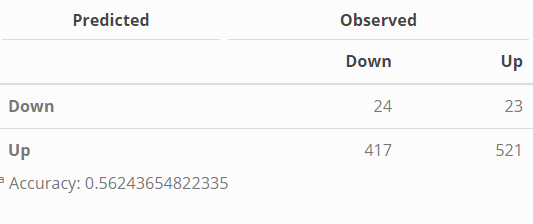
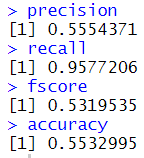
**Area under curve:** 0.5177278

**Logistic Regression using Lag1 Variable**

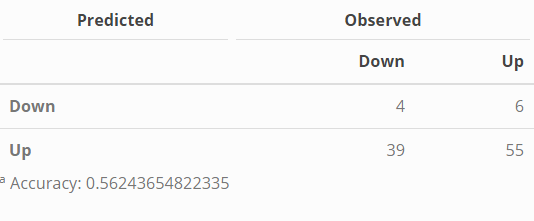
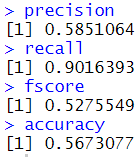
**Summary:**

****

**Confusion Matrix of Training Data:**

**** ****

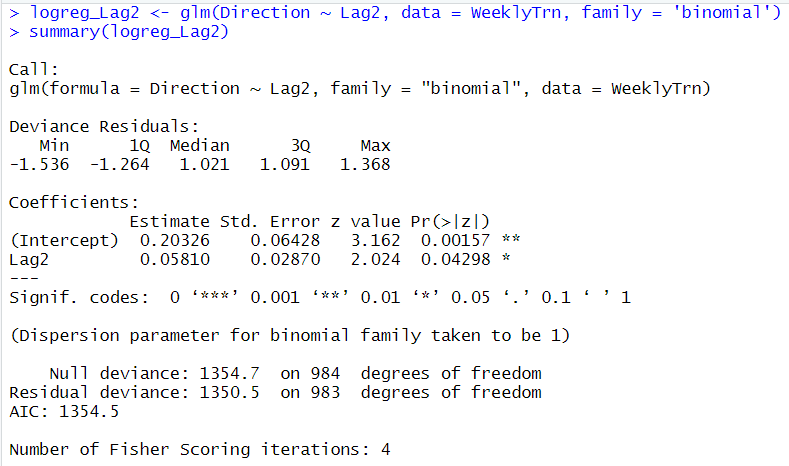
**Confusion Matrix of Test Data:**

**** ****

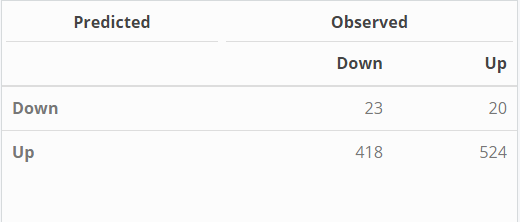
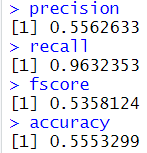
**Area under curve =** 0.4864659

**Logistic Regression using Lag2 Variable**

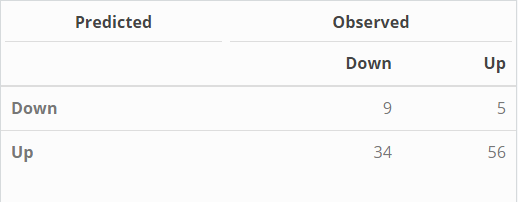
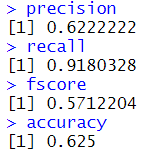
**Summary:**

****

**Confusion Matrix of Training Data:**

**** ****

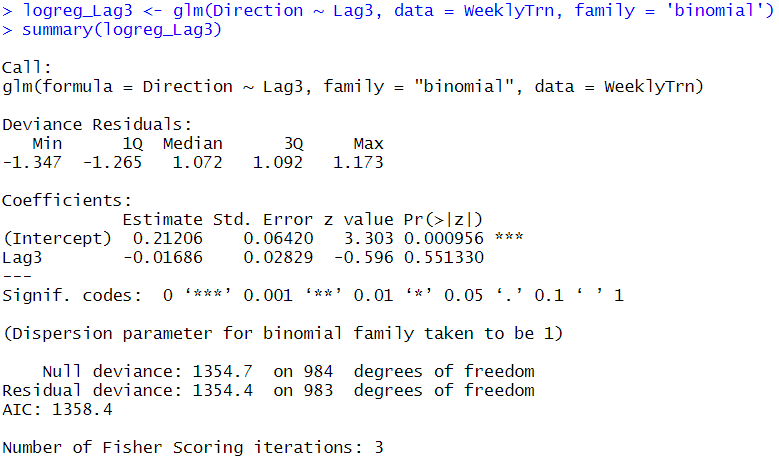
**Confusion Matrix of Test Data:**

**** ****

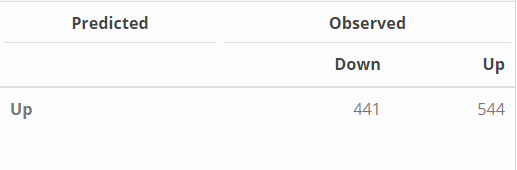
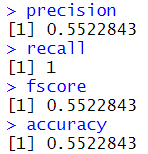
**Area under curve = 0.546321**

**Logistic Regression using Lag3 Variable**

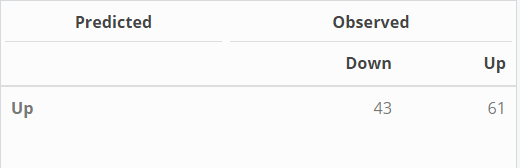
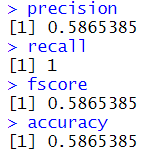
**Summary:**

****

**Confusion Matrix of Training data:**

**** ****

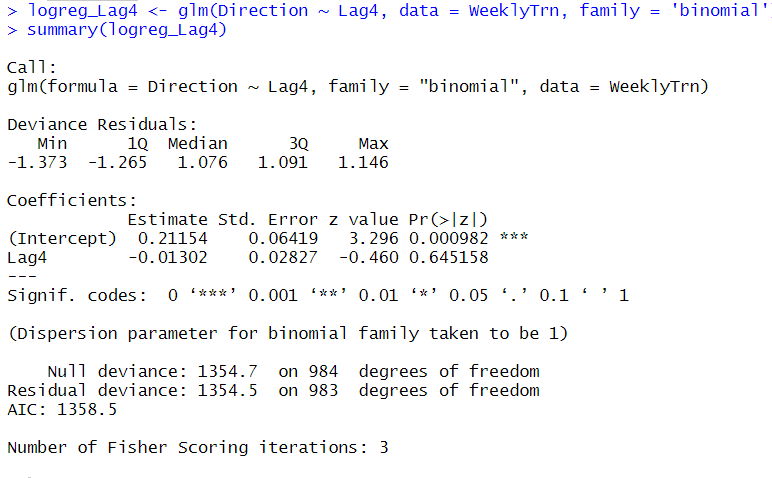
**Confusion Matrix of Test data:**

**** ****

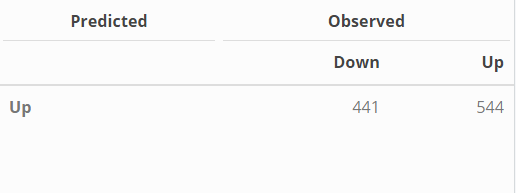
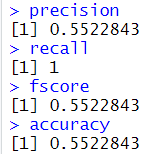
**Area under curve =** 0.5242089

**Logistic Regression using Lag4 Variable**

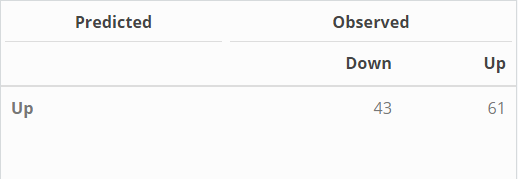
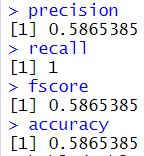
**Summary:**

****

**Confusion Matrix of Training data:**

**** ****

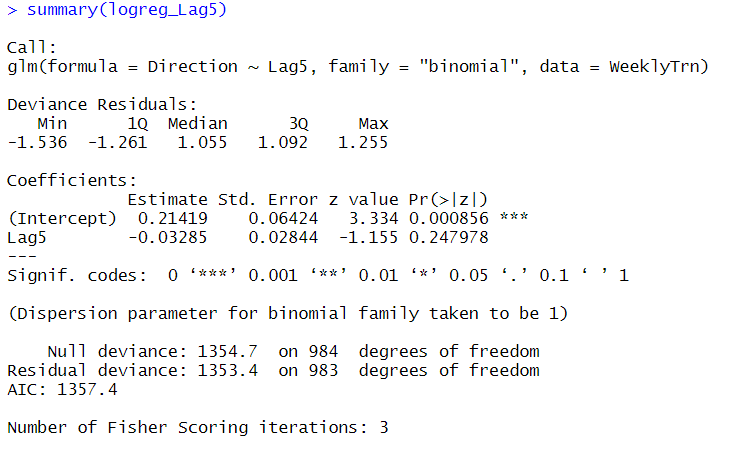
**Confusion Matrix of Test data:**

**** ****

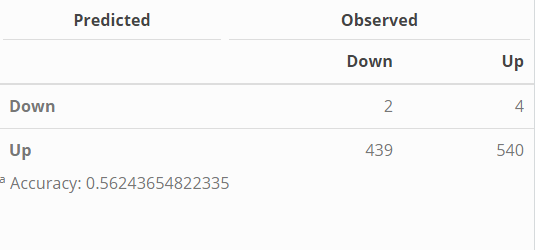
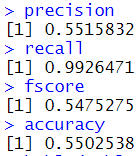
**Area under curve =** 0.5257339

**Logistic Regression using Lag5**

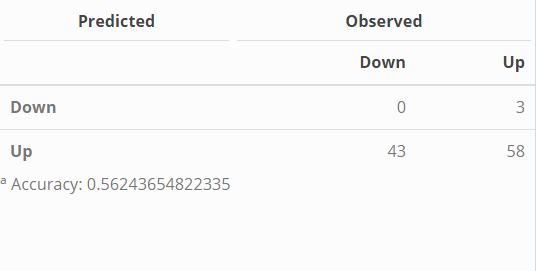
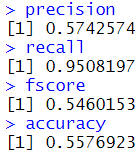
**Summary:**

****

**Confusion Matrix of Training data:**

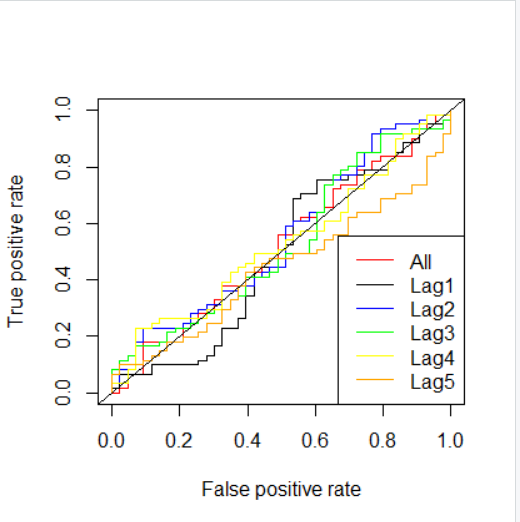
**** ****

**Confusion Matrix of Test data:**

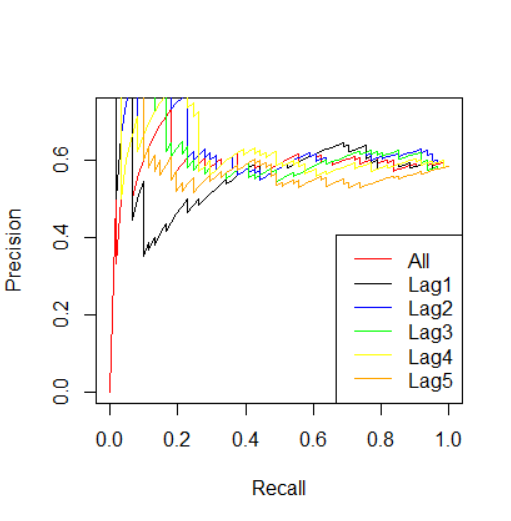
**** ****

**Area of Curve =** 0.4422417

* **ROC Curves for all the classifiers:**

****

**PR Curve for all classifiers:**

****

1. 0.38
2. 50 hours
3. Confusion Matrix and Accuracy of each model is specified above.
4. The model created using variable “Lag2” is the best model as the accuracy = 0.625 (62.5%) and F-score = 0.571 (57.1%). No, the best model does not achieve the best accuracy or F-score on the training data.
5. The best model in terms of the Area Under Curve (AUC) is the model created using only Lag 2 variable 0.546321 (54.7%)
6. The best model in terms of the Area Under Curve (AUC) is the model created using only Lag 2 variable 0.546321 (54.7%)
7. Based on the performance evaluation metrics like: Precision = 62.2%, Recall = 91.8%, Accuracy = 62.5%, AUC = 54.63%, F-Score = 57.12%

The model created using “Lag2” variable is the best model. The only statistically significant variable is “Lag2” which can be understood by the stochasticity of the data. Hence, Model created using only “Lag2” is the best model for this problem.

Solution for the 1st and 2nd question:

