Logistic Regression Stats

Binary Independent Variable:

* Was there an injury? (0 = No, 1 = Yes)

Independent Input Variables:

* **Hour**: (0-23), numerical
* **Temperature**: numerical
  + Note: Some blank values from the weather data were replaced with -100 to differentiate the data and clearly show that they were outliers. This was done so we would not have to completely remove an entire entry due to one blank value.
* **Dewpoint**: numerical
* **Humidity**: numerical
* **Wind Speed**: numerical
* **Wind Gust Speed**: numerical
* **Wind Direction in Degrees**: numerical
* **Pressure**: numerical
* **Event**: (Clear, Rainy, Rainstorm, Cloudy, Foggy) 5 possibilities
* **Weekday**: (0 – 6), Monday – Sunday, numerical
* **Month**: (1 – 12), January – December, numerical

The Logistic Regression was performed using all of the values listed above using Python’s Logistic Regression function.

The following lines are the results from running a Logistic Regression Algorithm in Python:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [[10475 2]  [2678 3]] |  |  |  |  |
|  | Precision | Recall | F1-Score | Support |
| 0 | 0.80 | 1.00 | 0.89 | 10477 |
| 1 | 0.60 | 0.00 | 0.00 | 2681 |
| Avg/Total | 0.76 | 0.80 | 0.71 | 13158 |

A Precision and Recall Accuracy formula was used to determine the following:

* True Positives (TP): 10475
* True Negatives (TN): 3
* False Positives (FP): 2678
* False Negatives (FN): 2
* Formula: (TP + TN) / (TP + TN + FP + FN)
  + (10475 + 3) / (10475 + 3 + 2678 + 2) = 0.7963216294 (~80% Accuracy)

Running test over all variables:

LogisticRegression(C=1.0, class\_weight=None, dual=False, fit\_intercept=True,

intercept\_scaling=1, max\_iter=100, multi\_class='ovr', n\_jobs=1,

penalty='l2', random\_state=None, solver='liblinear', tol=0.0001,

verbose=0, warm\_start=False)

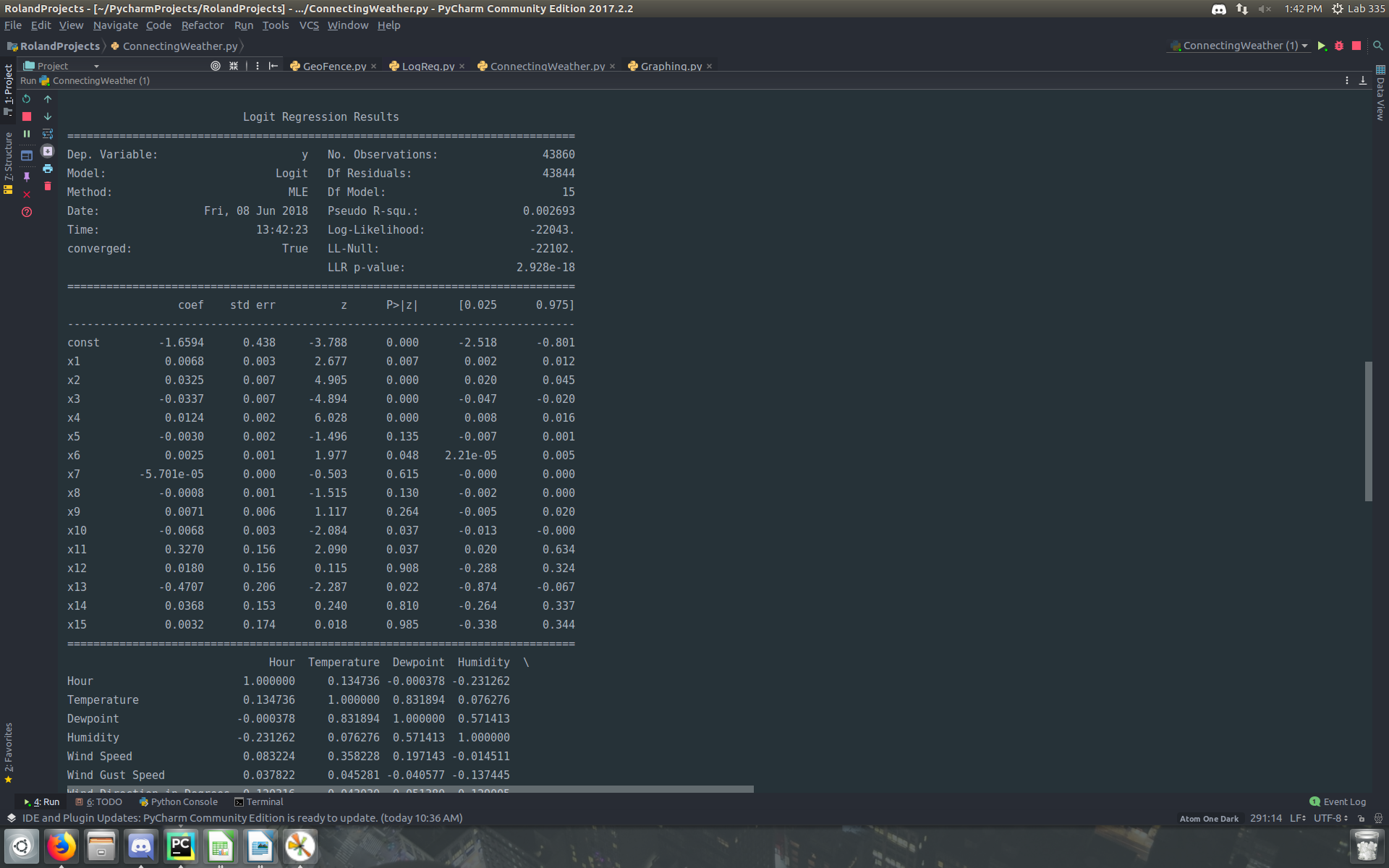
Accuracy Score: 0.796321629427, Recall Score: 0.796321629427

|  |  |  |
| --- | --- | --- |
| **Variables** | **Odds Ratios** | |
| Time Factors | | |
| * Hour | | 1.00816079 |
| * Weekday | | 1.04622437 |
| * Month | | 0.95594133 |
| Atmospheric Factors | | |
| * Temperature | | 1.01536094 |
| * Dewpoint | | 1.00075267 |
| * Humidity | | 1.00097968 |
| * Pressure | | 0.99995624 |
| Wind Factors | | |
| * Wind Speed | | 0.99727111 |
| * Wind Gust Speed | | 1.00777038 |
| * Wind Direction Degree | | 0.99883821 |
| Condition Factors | | |
| * Clear | | 1.29381204 |
| * Rainy | | .94293865 |
| * Rainstorm | | 0.89953121 |
| * Cloudy | | 0.95634182 |
| * Foggy | | 0.97018489 |

OR=1 Exposure does not affect odds of outcome

OR>1 Exposure associated with higher odds of outcome

OR<1 Exposure associated with lower odds of outcome



Warnings:  
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.  
[2] The condition number is large, 3.81e+04. This might indicate that there are  
strong multicollinearity or other numerical problems.