|  |
| --- |
| On-Time Performance Factors |
| Analyzing the Effects of Absenteeism and Other Environmental Factors on the Timeliness of RTS Buses in Monroe County, NY |
|  |

Jeff Palmerino, Analytics Intern

**Abstract**

The purpose of this analysis is to explore the effects of absenteeism, the day of the week, and other environmental factors on on-time performance within Monroe County RTS buses. Through aforementioned analysis, it can be concluded that absenteeism, other environmental factors, and the day of the week have a statistically significant correlation with OTP.

Data used in the analysis was restricted to the time period December 30th, 2013, to October 13th, 2016. Due to a window change that occurred on April 1st, 2016, the definition of a bus being on-time changed. Before the change, buses were given a time period of -2:59 to +5:59 around their scheduled arrival time to be considered on-time. After the change, the time period shrunk to -2:00 to +5:00. In order to adjust for this in the analysis, OTP raw percentages were standardized so the data could be pooled together. Otherwise, the data would have had an inconsistent dependent variable in the model.

Other outside factors, measured at daily occurrences, included: temperature, whether school was in session, if it snowed, number of change offs, the amount of revenue service buses, total amount of precipitation, boarding’s per bus, and lift count. It is expected that OTP will be higher as temperature increases and to be lower as temperature decreases. If school was in session, OTP was expected to drop because there would be more buses and cars in general on the road, increasing traffic, and increasing the risk of buses being late. ‘If it snowed’ and the ‘total amount of precipitation’ were also factors that were expected to lower OTP. Snow and rain tend to make people drive cautiously, taking this into consideration, OTP is expected to be lower when it snows, and when total precipitation increases. If a bus has to be removed from service due to a problem, it is recorded as a ‘change-off’ in RTS Operations Data. If the total number of change offs increases, OTP is also expected to decrease. Finally, it is also expected that the average number of boardings per pus and lift count to decrease OTP. Both variables deal with more spent time at a stop. The more people that get on, the more people that have to get off, which means more time spent at a stop. On top of that, ‘lift count’ refers to if a rider needs to use the wheelchair lift. Taking the time to make sure someone gets on and off the bus safely will increase the risk of the bus being late to its next stop.

Through the analysis, absenteeism was found to be statistically significant. With the highest rates of absenteeism coming on Fridays and Saturdays, along with the lowest rates of OTP also occurring on the same days, it did not come as a shock when Friday and Saturday were the only statistically significant days of the week.

From the analysis it can be concluded that fluctuations in OTP percentages are significantly affected by what day of the week it is, multiple environmental factors (i.e. temperature and snow), whether or not school is in session, the rate of absenteeism, and the total number of change-offs in a day.

**Building the Models**

**Absenteeism on OTP**

Above is the relationship between absenteeism and OTP over roughly the last three years. The trend-line created within the data points shows a negative effect of higher absenteeism rates on OTP, and further supports the initial hypothesis that higher rates of absenteeism would cause OTP to drop.

**Daily On-Time Performance Rates**

Over the last few years RTS has experienced relatively steady OTP percentages, with only a slight decline. The effect of the window change that RTS implemented on April 1st, 2016 (denoted by the vertical black line) is quite visible. Once implemented, the range of on-time buses shifted down. This was expected because buses that would have been considered on-time may now be considered early or late due to the smaller window.

|  |  |
| --- | --- |
| After Window Change |  |
| Mean | 89.653 |
| Std. Deviation | 2.888 |
| Standard Error | 0.2063 |
| Median | 89.74 |
| Range | 18.22 |
| Minimum | 77.59 |
| Maximum | 95.81 |
| Count | 196 |

|  |  |
| --- | --- |
| Before Window Change |  |
| Mean | 93.11 |
| Std. Deviation | 3.833 |
| Standard Error | 0.1336 |
| Median | 93.86 |
| Range | 37.33 |
| Minimum | 61.88 |
| Maximum | 99.21 |
| Count | 823 |

The summary statistics above show the effect of the window change so far. OTP has dropped nearly 4%, and the maximum OTP value after the window change is nearly 4% less than the maximum value before the change. The difference in standard deviations is partially due to ‘after window change’ not including any winter data. Winter tends to cause more variation in OTP, and since the new window change doesn’t have any, it makes sense that it is lower. It would not come as a surprise if the standard deviation of the new window, however, became higher and close to the old window as it starts to accumulate winter data. Although the new window has not been implemented for long and its sample size is smaller than that of the previous window, because the data was standardized and pooled, this allows informative analysis to be conducted.

**Variables**

Multiple dummy and continuous variables were chosen for the analysis, and their sources can be found in the footnotes of this page. The following variables were used as dummies in the models:

* Days of the week0
* Snow1
* School2

The following variables were continuous or discrete in the analysis:

* Temperature3, Temperature.23 (the quadratic temperature variable)
* Absenteeism Rate4
* Total Number of Change-Offs5
* Buses in Revenue Service6
* Total Precipitation7
* Boarding’s per Bus8
* Lift count9

Although subtle, the above graph shows the slight quadratic relationship between temperature and OTP and for that, ‘Temperature.2’ was added as a quadratic term to the model. This is not quite what was expected, however, when analyzing the two as a linear functionA, a trend-line was created that had an R-squared value of 0.006. This helped to justify adding a quadratic term to the model.

A Linear relationship trend line can be found in appendix 0 Monday-Saturday, one category must be omitted when using dummy variables, which leaves something with the value of zero with which to compare all other categories, Sunday was chosen to omit 1 Weather Underground archived weather history files 2 RCSD Academic Calendar 3 Weather Underground archived weather history files 4 RTS Monroe Operator Absence Data 5-9 RTS Operations Data

**Correlation**

**Non-Reduced Model Regression Results**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | T - Value | P - Value |
| (Intercept) | 0.733 | 0.287 | 2.558 | 0.010 |
| Monday | -0.221 | 0.185 | -1.198 | 0.231 |
| Tuesday | -0.240 | 0.189 | -1.271 | 0.204 |
| Wednesday | -0.337 | 0.191 | -1.764 | 0.078 |
| Thursday | -0.309 | 0.187 | -1.648 | 0.0997 |
| Friday | -0.562 | 0.194 | -2.892 | 0.004 |
| Saturday | -1.179 | 0.108 | -10.92 | 2.61E-26 |
| Temperature | 0.028 | 0.008 | 3.493 | 0.00049 |
| Temperature.2 | -0.00036 | 7.80E-05 | -4.708 | 2.85E-06 |
| Absenteeism | -1.748 | 0.630 | -2.774 | 0.00564 |
| School | -0.209 | 0.105 | -1.985 | 0.04759 |
| Snow | -0.210 | 0.103 | -2.041 | 0.0413 |
| Change-Offs | -0.0226 | 0.012 | -1.924 | 0.0443 |
| Revenue Service Buses | -0.00208 | 0.001 | -1.935 | 0.0531 |
| Total Precipitation | -0.185 | 0.122 | -1.527 | 0.127 |
| Boardings per Bus | 0.00027 | 0.00056 | 0.481 | 0.6303 |
| Lift Count | 0.00038 | 0.00043 | 0.889 | 0.374 |

The non-reduced model shown above is the first model we created to try and explain variation in OTP. Each variable’s p-value10, indicated above, tells us whether or not the variable was statistically significant. As stated before, the variables Friday, Saturday, Temperature, Temperature.2, Absenteeism, School, Snow and Change-Offs resulted in statistical significance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Adjusted R - Squared | Residual Std. Error | F - Stat | P - Value | AIC |
| 0.193 | 0.8660 | 16.1775  (16 and 1018 DF) | < 0.0001 | 2629.40 |

The non-reduced model, although it included a lot of variables, only explained roughly 20% of the variation in OTP of RTS buses. The model as a whole, however, resulted in a statistically significant p-value when an F-Test11 was conducted.

10A p-value is defined as the probability of obtaining a result equal to or more extreme than what was observed, when in fact the null hypothesis is true, if it is less than the alpha level, the p-value is statistically significant  11 The F-test of the model’s overall significance. It compares a model with no predictors (intercept only model) to the model that you specify.

**Correlation**

Using Pearson’s correlation coefficients12 , one positive, but mostly fairly weak correlation, and multiple weak, negative correlations between the explanatory variables and OTP were found. A positive correlation would mean that as one variable increased, so did OTP and as the variable decreased, OTP decreased also. On the other hand, negative correlations mean that as the variable increased, OTP declined and as the variable declines as OTP increases.

|  |  |  |
| --- | --- | --- |
| Variable | Correlation with OTP | P-Value |
| Temperature | -0.078 | 0.00049 |
| Absenteeism | -0.18 | 0.00520 |
| Change-Offs | -.120 | 0.0443 |
| Revenue Service Buses | -.150 | 0.05315 |
| Boardings per Bus | 0.063 | 0.6303 |
| Lift Count | -.120 | 0.3740 |
| Total Precipitation | -0.050 | 0.12710 |

Of the above correlations, temperature, absenteeism, and change-offs were found to be statistically significant, using an alpha level of .05 as the threshold. Out of the variables that were statistically significant, absenteeism and change-offs both produced negative correlations like expected. However, one factor produced results that were unexpected. Temperature was expected to be a positive correlation, yet yielded a negative one.

|  |  |  |
| --- | --- | --- |
| Variable | CorrelationSummer | CorrelationWinter |
| Temperature | -.115 | .222 |

When investigating further into the possible cause(s) of the unexpected result with the temperature factor, the correlations between temperature and OTP were found to be severely different when temperature was split up into two seasonal sub-categories. The new window only spans from April 1st, 2016 to October 13th, 2016 in the data set, so that was defined as ‘summer’ (‘winter’ included all other days). These new correlations tell us that as temperatures rose in the summer, and as temperatures rose in the winter, OTP also rose. This trend further justifies the use of a quadratic temperature factor in the model.

12 Pearson’s correlation coefficient, a measure of the strength and direction of the linear relationship between two variables

**Building the Final Reduced Model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | T - Value | P - Value |
| (Intercept) | 0.721 | 0.236 | 3.053 | 0.00231 |
| Friday | -0.295 | 0.0814 | -3.633 | 0.00029 |
| Saturday | -0.9816 | 0.0901 | -10.883 | 3.699E-26 |
| Temperature | 0.0315 | 0.00769 | 4.104 | 4.374E-05 |
| Temperature2 | -0.000401 | 7.676E-05 | -5.228 | 2.073E-07 |
| Absence Rate | -2.944 | 0.5667 | -5.195 | 2.4672E-07 |
| School | -0.5532 | 0.0677 | -8.165 | 9.483E-16 |
| Snow | 8.431e-07 | 7.501e-06 | -2.326 | 0.0201 |
| Change Offs | -0.0324 | 0.01175 | -2.76395 | 0.00581 |

Once the factors that were significant and those that were not were identified, the model was reduced to the factors above that were deemed statistically significant. Making that distinction allows us to create a model with the factors that are most influential on OTP.

The above bar-plot further justifies Friday and Saturday being the only days of the week to turn out as statistically significant. One could expect from this graph that Saturday would have a larger effect on OTP than Friday does; which is confirmed as the coefficient for Saturday is more influential than that of Friday.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Adjusted R - Squared | Residual Std. Error | F - Stat | P - Value | AIC13 |
| 0.177 | 0.8744 | 28.347  (8 and 1018 DF) | < 0.0001 | 2617.66 |

The final reduced model took a slight drop in the amount of variation it could explain. Comparing it to the non-reduced model, which had an adjusted r-squared value of 19.3%, the reduced model can only explain 17.7% of the variation in OTP. This could simply be due to the fact that there are fewer variables included in the reduced model. The slight drop, however, does not mean the reduced model is worse. In fact, with the AIC value dropping nearly 12 points, it can be concluded that the reduced model is significantly better at minimizing information loss than the non-reduced model. Finally, the model as a whole still resulted in a statistically significant p-value when an F-Test was conducted.

13 Akaike Information Criterion (AIC), the lower the score the better, helps decide what models are better than others