**COVID-19 Impacts and Pandemic Planning for NCDOT Construction and Maintenance Operations**

In order to get the final input for life regression, the selected parameters should be filtered and evaluated step by step with multiple statistical tool and logistics. With initial investigation, the parameters we select is:

|  |  |
| --- | --- |
| Revenue Related | Gasoline\_Gallons |
| Diesel\_Gallons |
| Kerosene\_Gallons |
| Alternative\_Fuels\_Gallons |
| Highway\_Fuel\_Use |
| Total\_Gallons\_ST\_Road\_Tax |
| Registration\_Fee |
| Driver\_license\_Fee |
| Emission\_inspection |
| Overweight |
| Title\_fee |
| Economy | Quarterly\_GDP\_of\_NC |
| Labor\_Force |
| Employment |
| Unemployment |
| CPI\_U |
| Monthly\_Retail\_and\_Food\_Services |
| Climate | Ave\_Temperature\_F |
| Ave\_Precipitation |
| Ave\_Snowfall |
| Ave\_snow\_depth |
| Ave\_heating\_degree\_day |
| Ave\_Cooling\_Degree\_Day |
| Hurrican\_times\_occurred |
| Daily\_Dew\_Point |
| Sea\_Level\_Pressure |
| Daily\_Visibility |
| Daily\_Wind\_Speed |
| Daily\_Humidity |

**Step1**: Univariate test (proc univ), eliminate abnormal data (or incomplete ones), after evaluating data, the result is

|  |  |
| --- | --- |
| Revenue Related | Gasoline\_Gallons |
| Diesel\_Gallons |
| ~~Kerosene\_Gallons~~ |
| ~~Alternative\_Fuels\_Gallons~~ |
| Highway\_Fuel\_Use |
| Total\_Gallons\_ST\_Road\_Tax |
| Registration\_Fee |
| Driver\_license\_Fee |
| Emission\_inspection |
| Overweight |
| Title\_fee |
| Economy | Quarterly\_GDP\_of\_NC |
| Labor\_Force |
| Employment |
| Unemployment |
| CPI\_U |
| Monthly\_Retail\_and\_Food\_Services |
| Climate | Ave\_Temperature\_F |
| Ave\_Precipitation |
| Ave\_Snowfall |
| ~~Ave\_snow\_depth~~ |
| Ave\_heating\_degree\_day |
| Ave\_Cooling\_Degree\_Day |
| Hurrican\_times\_occurred |
| ~~Daily\_Dew\_Point~~ |
| ~~Sea\_Level\_Pressure~~ |
| ~~Daily\_Visibility~~ |
| Daily\_Wind\_Speed |
| ~~Daily\_Humidity~~ |

**Step2**: Correlation (proc corr), using Pearson correlation test, we eliminated the variables which did not pass any of the significant test with other variables, the result is

|  |  |
| --- | --- |
| Revenue Related | Gasoline\_Gallons |
| Diesel\_Gallons |
| ~~Highway\_Fuel\_Use~~ |
| Total\_Gallons\_ST\_Road\_Tax |
| Registration\_Fee |
| Driver\_license\_Fee |
| Emission\_inspection |
| Overweight |
| Title\_fee |
| Economy | Quarterly\_GDP\_of\_NC |
| Labor\_Force |
| Employment |
| Unemployment |
| CPI\_U |
| Monthly\_Retail\_and\_Food\_Services |
| Climate | Ave\_Temperature\_F |
| ~~Ave\_Precipitation~~ |
| Ave\_Snowfall |
| Ave\_heating\_degree\_day |
| Ave\_Cooling\_Degree\_Day |
| Hurrican\_times\_occurred |
| Daily\_Wind\_Speed |

**Step3**: from filtered results of correlation, using Principle Component Analysis (PCA) to find vectors that could represent “Revenue related”, “Economy”, “Climate” best, using each variable in vector which is greater 0.35 as threshold as filter criteria, the result is

|  |  |
| --- | --- |
| Revenue Related | Gasoline\_Gallons |
| Diesel\_Gallons |
| Total\_Gallons\_ST\_Road\_Tax |
| Registration\_Fee |
| ~~Driver\_license\_Fee~~ |
| ~~Emission\_inspection~~ |
| Overweight |
| Title\_fee |
| Economy | Quarterly\_GDP\_of\_NC |
| Labor\_Force |
| Employment |
| Unemployment |
| CPI\_U |
| Monthly\_Retail\_and\_Food\_Services |
| Climate | Ave\_Temperature\_F |
| ~~Ave\_Snowfall~~ |
| Ave\_heating\_degree\_day |
| Ave\_Cooling\_Degree\_Day |
| ~~Hurrican\_times\_occurred~~ |
| ~~Daily\_Wind\_Speed~~ |

**Step4**: from filtered results of correlation, using Canonical Correlation analysis to find most important variables in each set that has great impacts on the other sets, this step did not filter out any of the variables from step3. The direct interpretation will be that, “Gasoline\_Gallons”, “Diesel\_Gallons”, “Total\_Gallons\_ST\_Road\_Tax”, “Registration\_Fee”, “Overweight”, “Title\_fee” can represent “Revenue Related” most and also have the strongest correlation with the “Economy”, “Climate”. Same for the “Economy” and “Climate”.

|  |  |
| --- | --- |
| Revenue Related | Gasoline\_Gallons |
| Diesel\_Gallons |
| Total\_Gallons\_ST\_Road\_Tax |
| Registration\_Fee |
| Overweight |
| Title\_fee |
| Economy | Quarterly\_GDP\_of\_NC |
| Labor\_Force |
| Employment |
| Unemployment |
| CPI\_U |
| Monthly\_Retail\_and\_Food\_Services |
| Climate | Ave\_Temperature\_F |
| Ave\_heating\_degree\_day |
| Ave\_Cooling\_Degree\_Day |

**Step5**: for the variables left in the step4, using time series analysis, ARIMA to find time series correlation. ARIMA can intuitively find if the variable is influenced by itself or by other unmeasurable factors (noise). Initially, we find some non-stationary factors, which means they have growth pattern in the time series. We manipulated the data to eliminate the non-stationary issues (use ). Here we identified the pattern of AR, MA, and ARMA. the result is

|  |  |
| --- | --- |
| Revenue Related | Gasoline\_Gallons (AR) |
| Diesel\_Gallons (MA) |
| Total\_Gallons\_ST\_Road\_Tax (MA/ARMA) |
| Registration\_Fee (MA) |
| Overweight (MA) |
| Title\_fee (ARMA) |
| Economy | Quarterly\_GDP\_of\_NC (MA) |
| Labor\_Force (AR) |
| Employment (AR) |
| Unemployment (AR) |
| CPI\_U (ARMA) |
| Monthly\_Retail\_and\_Food\_Services (MA) |
| Climate | Ave\_Temperature\_F (MA) |
| Ave\_heating\_degree\_day (MA) |
| Ave\_Cooling\_Degree\_Day (MA) |

**Step6**: using the VERMAX to find the most relevant parameters in each set (Revenue, Economy, and Climate) in a time series analysis. which we can eliminate some parameters that might not have significant influence on other variables within the set during a time series.

|  |  |
| --- | --- |
| Revenue Related | ~~Gasoline\_Gallons (AR)~~ |
| ~~Diesel\_Gallons (MA)~~ |
| ~~Total\_Gallons\_ST\_Road\_Tax (MA/ARMA)~~ |
| Registration\_Fee (MA) |
| Overweight (MA) |
| Title\_fee (ARMA) |
| Economy | ~~Quarterly\_GDP\_of\_NC (MA)~~ |
| Labor\_Force (AR) |
| Employment (AR) |
| ~~Unemployment (AR)~~ |
| CPI\_U (ARMA) |
| ~~Monthly\_Retail\_and\_Food\_Services (MA)~~ |
| Climate | Ave\_Temperature\_F (MA) |
| Ave\_heating\_degree\_day (MA) |
| Ave\_Cooling\_Degree\_Day (MA) |

**Step7**: “Economy” and “Climate” will treat as independent variables in the survival analysis. We are assuming *Event1*, *Event2*, and *Event3* where revenues decline, how many time periods (months) they experienced before the decline. And listed the corresponding values for those identified independent variables. The input for life regression model is concluded in the Excel that we sent.