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 $y_t - y_{t-1}$). 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.