**Connor Bruce – Project 2 Proposal**

For project 2, I would like to use the same data that I used for project 1 which included two separate tables with pavement condition data from FHWA’s Long Term Pavement Program. These tables fit the requested requirements of being under 100MB, over 1000 rows, and over 10 columns. ANALYSIS IRI has roughly 150,000 rows and roughly 20 columns. ANALYSIS\_RUTTING has roughly 25000 rows and 30 columns. These tables are obtained using the following steps:

1. Go to <https://infopave.fhwa.dot.gov/>
2. Under the ‘Data’ tab, click ‘Analysis Ready Datasets’
3. Under the Performance tab in the middle, expand the panel and then check the box for ‘Analysis Ready Rutting Dataset’ and ‘Analysis Ready IRI Dataset’.
4. Click ‘Add to Selection’, then ‘Add to Data Bucket’ when it appears below.
5. At the top right, there will be a drop-down menu when you hover over ‘Data Bucket (2)’. Click ‘Data (2)’ under this menu.
6. Enter the information, select Microsoft Excel as the Export File Format, and then click ‘Submit for Data Extraction’
7. You will receive an email to confirm the data has been sent for retrieval, then another once it is ready. Click the link in the second email to then be directed to the data bucket where you can download.
8. You will download a .zip file. When extracted, there will be an excel spreadsheet with multiple sheets including ANALYSIS\_IRI and ANALYSIS\_RUTTING.

In addition to the two mentioned tables, there will be multiple metadata table included. These tables can be used to tie the different sections of pavement to their construction information, the experiment type it is under, and other data in the database such as traffic and weather data. As I bring in these tables, I will provide information on how to access them as well as some quick EDA to give perspective. By tying additional data using these tables, including traffic and weather data, I intend to address the following hypotheses:

1. In my Project 1 presentation, I mentioned that IRI and Rutting are loosely positively correlated because they are impacted by the same factors, but these factors are weighted differently. By combining traffic data to rutting data and IRI data, I would like to model the rate at which IRI and rutting increase as a function of traffic. I hypothesize that traffic would have a stronger effect on rutting and aim to provide evidence of this by showing that there are higher deterioration rates of rutting in areas with higher traffic.
2. The second hypothesis is that areas with higher precipitation and higher temperatures will typically have faster deterioration rates for rutting and IRI. I will model IRI and Rutting rates as a function of both precipitation and temperatures (separately and together) to show this.
3. (Maybe) I will look at IRI deterioration rates between asphalt surfaces and concrete surfaces in terms of precipitation and traffic. I hypothesize that the deterioration rates will be higher for concrete as traffic increases, but higher for asphalt as precipitation increases.

Both of these hypotheses are expected but not obvious as there are additional factors that tie into the deterioration rates such as material and age of the pavement. They are both also answerable by comparing deterioration rates as functions of their respective factors (temperature, precipitation, and traffic).