Section 6 Homework – Capstone development

Due: Tuesday before class (30 May, 6:30pm CT)

1. **Create a definitive problem statement to be solved with data analytics.** 
   * Have the amounts of wrecks increased significantly since the COVID lockdown here in San Antonio, TX?
   * What are the major contributing factors to the area’s wrecks?
2. **Provide your datasets**.

* Crashdata\_SA19,20,21:
  + Contains data for Crashes in San Antonio from 2019 to 2021.
  + Contains field picked my me from the website below.
* <https://cris.dot.state.tx.us/public/Query/app/home>
  + This site allowed me to query the exact fields before doing the data.

1. **Describe your plan** on how to use this data to answer your problem statement.

* I plan to use the Excel file to filter the data
  + Breaking it down by year. Once I have that, I will add each year to its own sheet in the file.
  + From there, I will analyze the data to determine the major contributing factors.
  + Once identified I will see if I trend existed throughout the years
  + I then plan to highlight a few other “good to know” areas in the data.
* Phyton or SQL (whichever I feel more comfortable with)
  + I plan on using PANDAS to create a data frame for each year and other fields I feel are appropriate.
  + Joins to combine tables if needed.
* Tableau to visualize the idea in my head.

**Submit your homework to git hub, to include pushing your dataset files to your repo.**