**DANGEROUS INTERSECTIONS**

**Slide 11:** Convert Hour and Year to an Integer

**Slide 12:** Show me the first 20 lines of my data

**Slide 12:** In the location column, delete any data in ()

**Slide 13:** in the location column, alphabetize the intersections delimited by a /

**Slide 16:** I would like to do the same with the locations that are formatted with four numbers followed by a W or an S, then four numbers followed by a W or an S. I want the lower number first, but the W or S needs to go with it.

**Slide 18:** Insert a new column called score - base the value on the Primary Offense column. If Primary Offense contains "FATALITY" OR "HOMICIDE", put a 3 in the score column. If the Primary Offense column contains "INJURY", put a 2 in the score column, else 1.

**Slide 19:** Based on the score column, rank the top 15 locations descending

**Slide 19:** Display the top 15 locations, grouped by score (if 3 call it Fatality, if 2 call it Injury, if 1 call it Non-Injury)

**Slide 21:** on the total dataset, create a matrix with DOW in the rows and Hour in the columns, use purple.

**Slide 23:** graphs for day of week and year

**Slide 24:** Can we look at the distribution of crash categories for 2025? Add a count label.

**Slide 25:** Comparison charts by category across the years

**Slide 26:** Export final dataframe and all the graphs as jpgs

**Officer Availability**

**Slide 35:** # Sort by Unit then Date/Time

**Slide 36:** # create a new column called Time that subtracts the current Date/Time from the previous Date/Time

**Slide 37:** # remove any rows that have a value in the Time column over 5:00:00

**Slide 38:** # create a new column called Date that includes only the date from the Date/Time column

**Slide 39:** # convert Date column to a Date format

**Slide 40:** Filter on the Status column, only rows with ENRT if it follows a row with CMPLT

**Slide 41:** Create a new column called SUM, which sums the Time column, group by Unit and Date, Create a new column called AVG, which averages the Time column, group by Unit and Date, Create a new column called SumAVG, which averages the SUM column, group by Unit, Create a new column called AVGAvail, which averages the AVG column, group by Unit

**Slide 42:** # remove duplicate Units

**Slide 43:** # run descriptive statistics on AVGAvail

**Slide 44:** # export DFF and descriptive statistics of AVGAvail in Excel

**FORECASTING**

**Slide 50:** In google colab, I have 12 Excel files that I want to import and merge, but before I merge the files, I want to insert a column named Zone into each file and use the name of the file as the value.

**Slide 52:** # convert Report Date/Time and Occurred On to Date/Time format

**Slide 53:** # insert a column called DOW for the Day of the Week name of Report Date/Time

**Slide 53:** # insert a column called Hour for the hour of day of Report Date/Time

**Slide 54:** # sort by Zone then Report Date/Time ascending

**Slide 55:** # Insert a new column called DBH which calculates the date difference from the line above it, group by Zone

**Slide 56:** The result of the code above is only giving me one decimal place, for example 4.0 on the days, I am getting 0.0 on some that are one day apart so I want to see that it is likely rounding down.

**Slide 58:**

# prompt: descriptive statistics on DBH by zone

# Group data by ‘Zone’ and get descriptive statistics for ‘DBH’

**Slide 59:** In google colab, I have descriptive statistics on column ‘DBH’ called zone\_dbh\_stats that are grouped by column ‘Zone’. I want to add the following columns: the last Report Date/Time for each zone (called Last Crime), a calculation of mean + std (called Mean+STD), a date column which adds the last two added column (last date of each zone + mean+std (called Forecast), todays day-day report is run (called Today) and date diff between Today and Forecast (called Due).