I would like to thank your team for providing me this opportunity to work on the time-based assignment where I could demonstrate my proficiency in data wrangling, manipulation and visualisation.

For task 1, I analysed each CSV file on MS Excel and Tableau and then used python via Jupyter Notebook to further analyse the data, dump the required table/fields into the dataframe and perform various manipulations on the data. Once I was done extracting the required information, I put the data into a dataframe, performed few sanity checks and exported the dataframe to an Excel file. Then, I imported that Excel file into Tableau to do visualisation activities. I tried to make use of all the fields and show my findings and analysis through different kinds of visualisations.

For task 2, I used the MS SQL Server 2014 Management Studio to create the tables with the non-null, primary key and foreign key constraints. Then, I dumped the dummy data as shown in the sample inputs for respective tables and worked up on the questions asked. First question required the joining concept where as second question required group by concept.

I've enclosed the respective task files in sub folder. Below are the details:

1. Task 1 folder
   a. ACCIDENT.csv -- input file
   b. ACCIDENT_LOCATION.csv – input file
   c. Findings.pdf -- My findings and analysis in PDF format
   d. NODE.csv – input file
   e. Task 1.ipynb – Python code in Jupyter Notebook (Anaconda Navigator)
   f. Task 1.pdf – PDF version of the python code from Jupyter Notebook
   g. Task 1.twb – Consists of all the visualisations of the excel file output from Task 1.ipynb

2. Task 2 folder
   a. PWC SQL Test.SQL – Consists of the SQL script and the answers in SQL file format
   b. PWC SQL Test.txt – Consists of the SQL script and the answers in txt format (notepad)