AIT 614

Final project

Team 1

Step 1: Open "google_colab_team1.ipynb" using Google colab.

Step 2: upload the clean data set "car_crashdata_pysparkclean (1).csv" to google colab. This is the processed dataset.

Step 3: Run the code.

Note:
- The colab file has step by step process of the workflow.
- The file "Crash_Reporting_-_Drivers_Data-1.csv" is the original dataset.
- Refer to "Mongodb connections.pdf" to connect Mongodb compass to Mongodb Atlas.
- Refer to "connecting databricks and mongodb.pdf" to connect Mongodb Atlas to Databricks.
- The file "Car_Crash_Processing (1).ipynb" is used to clean the dataset and feature engineering.
- The cleaned dataset is names as "car_crashdata_pysparkclean (1).csv".
- The Geospatial analysis is done in Tableau, the related file is "AIT614_finalProject_Tableau_results.twbx".
- The EDA and machine learning part is included to google colab file.