

GWU Data Science Datathon Guidelines

1. DATES:

- The problem statement be released on Friday, April 2 at 8:00 a.m.
- The deliverables must be submitted before midnight on April 4.

2. DATA:

This [dataset](#) is derived from the Metropolitan Police Department's (MPD) crash data management system (COBALT) and represents DDOT's attempt to summarize some of the most requested elements of the crash data. Some of the features or information contained in the dataset are the type of participants (driver, occupant, bicyclist, pedestrian), age of participants, if injured, severity (minor, major, fatal), if person was issued a ticket, was person in vehicle where speeding was indicated, etc. The target feature is: "Fatal" (minor, major, fatal).

3. REPORT:

The report (in the form of Jupyter notebook) should include the following sections:

- Abstract: Summarize the key findings
- Pipeline: Implement the full pipeline of the project, including:
 - Data preprocessing
 - EDA
 - Hyperparameter tuning and / or model selection
 - Interpretation: e. g., feature's predictive power over the target
- Discussion: Suggest policy (to improve road safety) based on the findings, must be supported by report from the literature (preferably from top conference / journals)

4. VIDEO:

Record a 5 minute long presentation, which should include a focus on the interpretation and discussion described in discussion section of the report (above).

5. DELIVERABLES:

As this datathon is individual-based, each participant should submit a txt file named `firstname_lastname.txt`, [to the link](#) and include:

- The github repo link of the report (in the form of Jupyter notebook)
- The link of the video

6. PRIZE/AWARD:

- First prize - \$250 gift card
- Second prize - \$100 gift card
- Third prize - \$75 gift card