Data Analyst Intern - Take Home Exercise

Objective

Demonstrate required technical and professional skills for a Data Analyst. In this exercise, you will work with a sample data set and manipulate it into a workable format. The primary outputs will be a short video presentation and the code you used to produce your findings.

Timeline

Candidates will have 2 days to complete the exercise, submission should be made by email to connor@haasalert.com by midnight two full days after receipt of the exercise. Ie: if you receive the exercise Monday, you should submit by Wednesday at midnight CT.

Requirements

1. You may spend a **maximum of 8 hours** working on this exercise. If you do not complete the exercise that is OK, this is intended to benchmark capacity.
2. Pull the New York Motor Vehicle Collisions dataset from Kaggle, load this data into a database technology of your choice:
   a. At HAAS Alert we use PostgreSQL,
3. Write SQL queries to address the following questions:
   a. Where do collisions occur and how many collisions were there?
   b. Which vehicles are involved in the most collisions?
   c. Describe the kind of collisions that occur:
      i. Which combination of vehicles are involved?
      ii. What causes collisions?
   d. Summarize the causality profile of collisions.
4. Using Python, visualize your key finds and results.
5. Produce an 8-10 minute presentation of your findings and your method:
   a. 4-5 Minutes: Presentation of finds suitable for a business stakeholder
   b. 4-5 Minutes: Discussion of methods, technical challenges, and code

Outputs

To submit your Take Home Exercise send an email to connor@haasalert.com with:
1. A pdf of your presentation,
2. A github link to your code,
3. A link to your video presentation - preferably a Loom.

References

Dataset URL: https://www.kaggle.com/datasets/ishmaelkiptoo/motor-vehicle-collisions