Hi Asrar! Abhimanyu from the grading team here.

You have scored a 3/10 in the Data Science Capstone Proposal and an 8/10 is required to pass this successfully.

1. Data (1/2.5)
   * Please mention the link from where the data will be accessed from. (0/1)

* <https://data.sfgov.org/Public-Safety/Fire-Department-Calls-for-Service/nuek-vuh3>
  + Please provide more insight into the variables that you are going to use in the project. (0/0.5)
* The dataset consists of 34 features composed of mainly timestamps and categorical features. Timestamps relate to day/time of when 911 calls were made/dispatched/arrived. Categorical variables consists of the reason for the call/type of call and how the calls were treated. Numerical features mainly consist of coordinates. The dataset also includes numerical features that are to be treated as categorical features (Zip Code / Fire Station Number and etc).
  + The dataset is readily available so web scraping is not required. (1/1)
* Use of Specialization (0/2.5)
  + Please provide more detail on the techniques that you would be using in the capstone project. This can range from data preprocessing to any form of statistical analysis. (0/1)
* Since most of the features are categorical with high cardinality, preprocessing will include reducing unique values to reduce dimensions and also assessing relationships between categorical features using chi-sq values and contingency tables.
  + Please provide more detail on the specialization that you intend to use in the capstone project. (0/1.5)
* My specialization was on Big Data. Since PySpark has an API with python, I will surely be implementing PySpark method and attributes through Pandas.

1. Product/Business Impact (2/5)
   * The student has clearly outlined the problem that he/she is attempting to solve. (1/1)
   * The student discusses the biggest challenges that he/she might face in the project. (1/1)
   * Please provide more detail on how the proposed solution will be valuable to the end-user. (0/1)

* Given the location, time and day of 911 calls, the end user will be able to predict if it is woth dispatching all resources to the incident. Instead, it can perhaps dispatch less resources to assess whether it warrants a bigger response. This would result in resource savings.
  + Please define project goals/milestones that you aim to achieve in this project. (0/1)
* The goal is to exercise working with categorical features and treating high dimensional datasets. What models to employ if the features/target are not linearly related. By working on this dataset, I can explore one-hot-encoding and target encoding to treat categorical features. I can then use feature importance attribute to extract only most important features.
  + Please provide more detail on how the end-user will use your product. What is the kind of input will you take from the user. (0/1)
* The enduser will able to process new data and assess whether it is cost effective to dispatch all resources to the incident. If model shows a high concentration of fake 911 calls from a specific geometrical area, then perhaps the enduser can use caution and restraint to dispatch resources there.

Overall try to structure your project proposal into sections so that it easier for you and the person reviewing it to navigate through the project. You can use the rubric points above to structure it better.