# Task – Exploratory Data Analysis (EDA)

Using the dataset attached, conduct an EDA and explain the findings of your analysis. It should contain appropriate visualizations and explanations to assist readers in understanding how these elaborations are arrived at as well as their implications.

You may find the following questions to be helpful in your EDA. Do note that you are required to go beyond these questions.

1. Regions with the highest incident and lowest incident rates
2. Summary on the top 5 type of incidents
3. Top 5 common body parts injured

## Deliverable

1. You are free to use either R (RMarkdown/Shiny), Python (Jupyter Notebook), SQL, Power BI or Tableau in your data cleansing and visualization
2. If required, you can also present your findings in a PDF document with the corresponding codes attached

## Evaluation

In the submitted deliverables, you are required to:

1. Outline the steps taken in the EDA process
2. Explain the purpose of each step through appropriate comments
3. Explain the conclusions drawn from each step
4. Explain the interpretation of the various statistics generated and how they impact your analysis
5. Generate clear, meaningful and understandable visualizations that support your findings
6. Organize the submitted documents so that it is clear and easy to understand

You will be assessed on the usefulness and clarity of visualizations, accuracy and depth of your insights, presentation flow, and structure of your analysis.

Please note that your submission will be heavily penalised for any of the following conditions:

1. Code file missing from the submission
2. Code cannot be compiled/executed
3. Explanations missing or unclear in the submitted documents

## Data

### Dataset description

This dataset contains row level information for incident events retrieved from the incident reporting system found in MSD. The attributes are described in the tables below.

### List of Attributes

|  |  |
| --- | --- |
| **Table: Injury** | |
| **Attribute** | **Description** |
| Incident ID | Unique ID for each Incident |
| Injury | Type of injury sustained |
| Hazard | Hazard category |
| Sub Category | Sub category of hazard |
| Body Part | Body part injured |
| Treatment | Treatment given |
| Incident Status | Status of incident case |
| Investigation Type | Type of investigation undertaken |

|  |  |
| --- | --- |
| **Table: Demographic** | |
| **Attribute** | **Description** |
| Employee Name | Name of employee |
| Gender | Gender of Employee |
| Relationship | Relationship of Employee with Organization |
| Position Type | Job Title |
| Age Band | Age Group |
| Ethnicity | Ethnicity Group |
| Salary Range | Salary Group |

|  |  |
| --- | --- |
| **Table: Event** | |
| **Attribute** | **Description** |
| Incident ID | Unique ID for each Incident |
| Incident Date | Date of incident logged into system |
| Site 1 Name | Type of location |
| Site 2 Name | Location where incident occurred |
| Site 3 Name | Sub location where incident occurred |
| Incident Work Area Name | Area where incident occurred |
| Event Type Name | Category of system |
| Person Employee Name | Name of employee affected by incident |
| Incident Categories Full Name | Type of incident that occurred |
| Potential Consequence | Severity of incident |
| Action Taken | Aftermath actions taken |