**Event Log for Incident Management Process**

**Dataset**

Number of instances: 141,712

Number of Features: 36

Link: <https://drive.google.com/file/d/1Qzn5vJgFQpYIqC-5o4f2x2wtGh30uywO/view?usp=sharing>

Description: This is an event log for an incident management process extracted from an audit system in Brazil. The event log is enriched with data loaded from a relational database underlying a corresponding process-aware information system. The additional information of the features is attached at the end of the document.

This project is divided into 3 deliverables, Preprocessing & EDA, Prediction & Clustering and Final Report. For all deliverables you are required to make a report that you will compile and submit in Deliverable 3 i.e., Final Report. The weightage for each deliverable is also given alongside each of them.

**D1: Preprocessing & EDA**

For this deliverable, focus on initial data exploration, missing data handling and graphs. This could include but not limited to:

* Removing extra information from all features in the file such as remove “Caller “, from each instance of the column, “caller\_id”. Hint, the similar update should not be in more than 13 columns in the dataset.
* Finding missing values and replacing them by an appropriate strategy (where required).
* Calculating and visualizing summary statistics.
* Finding correlation between attributes.
* Finding out dependence between categorical and numerical attributes.
* Analyzing the “incident\_state” i.e., the current state of the incident and represent these frequencies using a histogram plot.
* Finding the Average time required to handle an incident report.
* Analyzing the average number of instances reported for one incident.
* Finding the month with highest number of incidents **reported.**