

This exercise follows the five steps of the process of exploratory data analysis (EDA):

Ask, Prepare, Process, Analyze, Share.

### **Ask:**

As we set out to explore the data, here are some questions we must start with:

- What is the problem you are trying to solve?
- What metrics will you use to measure your data to achieve your objective?
- Who are the stakeholders?
- Who is your audience?

The dataset I've chosen explores a personal area of interest. For a business, questions about how insights from the analysis would help the client make decisions must also be included.

With this EDA, we will try to answer the following questions:

1. Which countries have had the most journalist deaths since 1992?
2. Are certain mediums targeted more than the others? If yes, which are the top two mediums that record the highest deaths?
3. Which beats/coverage areas record the highest journalist deaths?

### **Prepare:**

As we clean the data in this stage, we must also ensure that data-integrity best practices are followed and take into account concerns such as, where the data will be stored, who will have access to the dataset, and more.

We clean the data in this stage.

Some of the tasks I had to perform to clean the dataset used in this exercise included: removing duplicates, removing extra spaces, deleting variables not relevant to my analysis, treating missing values, separating day, month, and year, and a few more.

The original dataset can be viewed [here](#).

The clean dataset can be viewed [here](#).

### **Process:**

In this stage we must ask ourselves questions such as which tools we would like to use, if we are sure the data is clean, and if we have documented the data cleaning process.

I used a combination of spreadsheets and R to clean the dataset for this exercise.

Documenting the cleaning process is a good practice if you are working with a team of analysts or if you would need to refer to this exercise again in the future.

You can access the R file for data-cleaning [here](#).

**Analyze:**

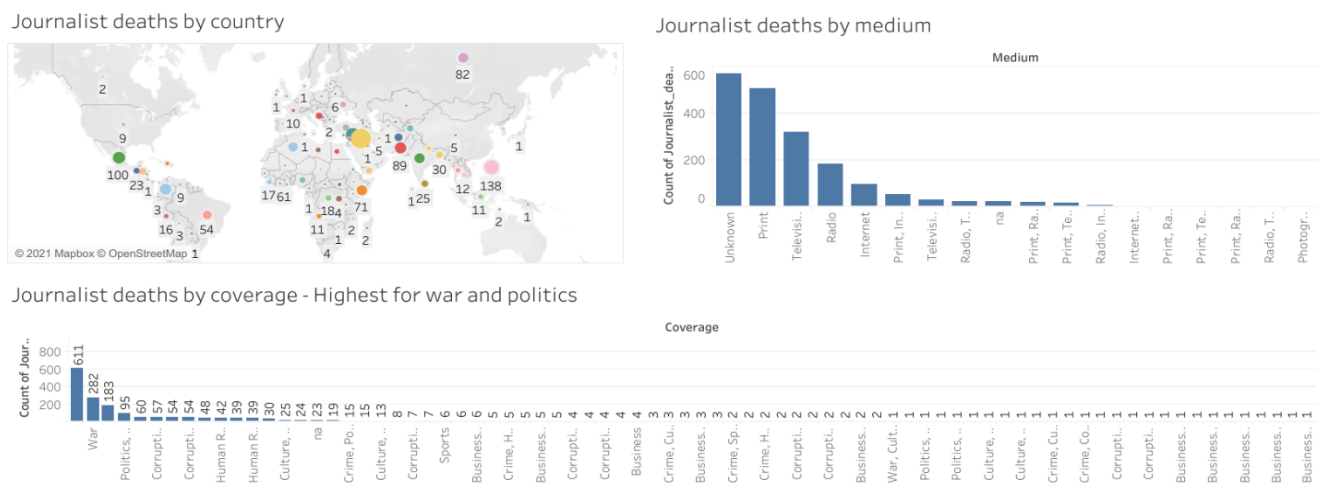
This is the stage we have been preparing the dataset for.

In this stage we aggregate the data, confirm the data is clean, and decipher trends and insights.

**Share:**

This is the last stage before we act on our findings.

I have put together a Tableau dashboard to share the insights from this exercise.



The dashboard can be accessed [here](#).