The Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and features of the data that might be unexpected.

- **Remember: All data has stories to tell.**

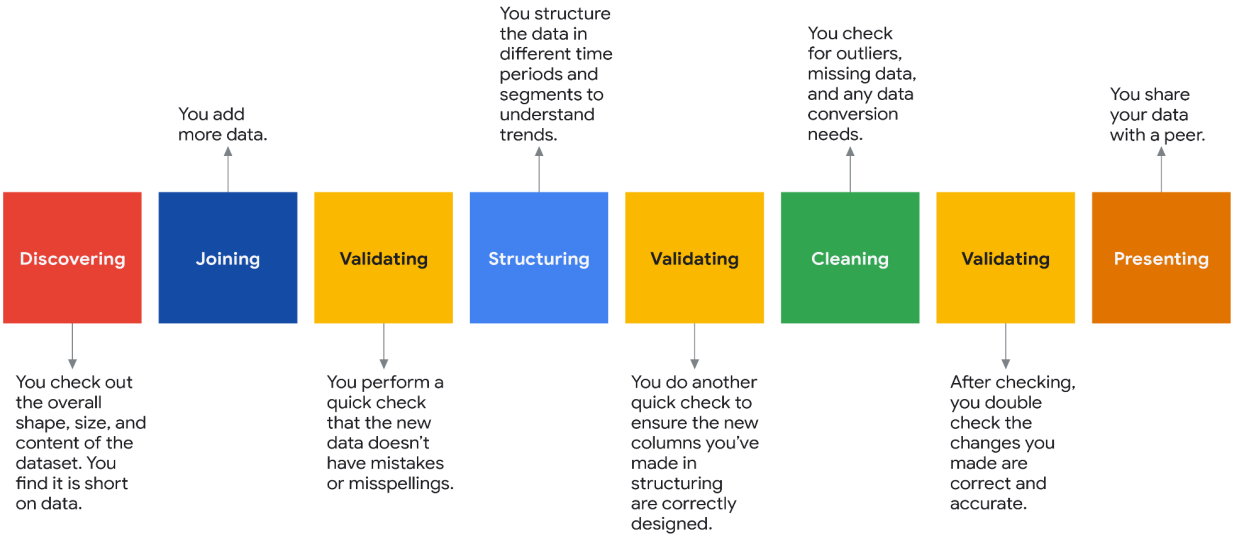
The six practices of EDA are iterative and non-sequential

Exploratory data analysis (EDA) is not like a cake recipe. It is ***not*** a step-by-step process you follow. Instead, the six practices of EDA are iterative and non-sequential.

- **Iterative**: Relating to or involving repetition of a process
- **Non-sequential**: Not arranged in or following an order or sequence.

Because of the varying nature of datasets, the approach to exploring that data will be different each time. That means that you will need to use your logic and experience throughout the EDA process to determine which of the six practices to utilize, how many times to apply them, and when in the process you should apply them.

Visual



1. **Discovering**: You check out the overall shape, size, and content of the dataset. You find it is short on data.
2. **Joining**: You add more data.
3. **Validating**: You perform a quick check that the new data doesn't have mistakes or misspellings.
4. **Structuring**: You structure the data in different time periods and segments to understand trends.
5. **Validating:** You do another quick check to ensure the new columns you've made in structuring are correctly designed.
6. **Cleaning**: You check for outliers, missing data, and needs for conversions or transformations,
7. **Validating**: After cleaning, you double check the changes you made are correct and accurate,
8. **Presenting**: You share your dataset with a peer.