Exploratory Data Analysis (EDA) is an approach to analysing data sets to summarize their main characteristics, often with visual methods. Following are the different steps involved in EDA :

* Data Collection
* Data Cleaning
* Data Preprocessing
* Data Visualisation

### Data Collection

Data collection is the process of gathering information in an established systematic way that enables one to test hypothesis and evaluate outcomes easily.

After getting data we need to check the data-type of features.

There are following types of features :

* numeric
* categorical
* ordinal
* datetime
* coordinates

In order to know the data types/features of data, we need to run following command:

train\_data.dtypes

or

train\_data.info()

Let’s have a look to the statistical summary about our dataset.

train\_data.describe()

### Data Cleaning

Data cleaning is the process of ensuring that your data is correct and useable by identifying any errors in the data, or missing data by correcting or deleting them.

Once the data is clean we can go further for data preprocessing.

### Data Preprocessing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. It includes normalisation and standardisation, transformation, feature extraction and selection, etc. The product of data preprocessing is the final training dataset.

### Data Visualisation

Data visualisation is the graphical representation of information and data. It uses statistical graphics, plots, information graphics and other tools to communicate information clearly and efficiently.