The General Pipeline to be approached for a Data Science Problem is as follows :

Data Ingestion

Revisit the Pipeline for Fresh Data

**Data Interpretation**

**Modelling The Data**

**Visualizing The Data**

**Data Cleaning**

**Raw Data**

**Gather Required Data**

1.Gather Required Data :

**Data Acquisition**

**Database**

**Strict Checks/Filters for the Reliability of the data**

**Un-Structured Data Sources**

**Structured Data Sources**

Without Data we can not proceed with our Data Science Problem. Emphasis must be laid on obtaining the data which is Authentic and Reliable. Strict Checks must be Done while obtaining the Data and then Extract the data into a usable format(.csv,JSON,XML,etc).

2.Preparing and Cleaning Data:

**Storage**

**Augmentation**

**Transformation**

**Anomaly Removal**

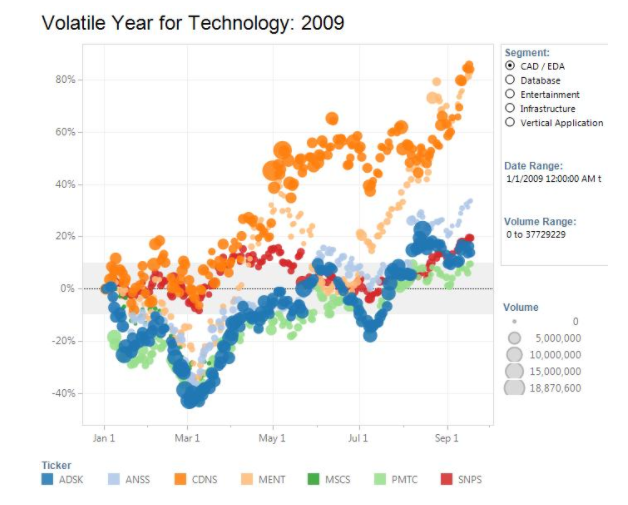
**Data Cleansing**

(Duplicates, Nulls Removal)

Here The Prime Emphasis is on Removal of the Duplicate Values, Null Values and other Anomalies because the results of our Machine Learning model are good only when the Data is Relevant and clean.

In this stage we have to thoroughly examine the data and identify the errors and eliminate them.

3.Exploration and Visualisation Of Data:



Here we try to find the Patterns and extract key values from our Data. We have to Use Different types of Visualization Techniques so that secrets in the data could be revealed.

The Primary objective is to find out the patterns through Visualization and charts will lead us to the important feature Extractions using Statistics to identify and test Significant Variables in our Data.

4.Modelling the Data(Machine Learning):

**Training Datasets**

**Prepared Data**

**Visualization Datasets**

**Validate Results**

**Training Results**

**Model Training**

**Hyper Parameter Tuning**

**Models**

The main objective here is to do the in-depth analytics, mainly the creation of relevant machine learning models, like predictive model/algorithm to answer the problems related to predictions.

The second important objective here is to evaluate and refine our own model. Any machine learning model can’t be superlative on the first attempt. We will have to increase its accuracy by training it with fresh ingestion of data, minimizing losses, etc.

Some of the Commonly Used Methods are :

1. Supervised/Unsupervised Algorithms,

2. Classiffication Model

5.Interpreting the Data:

Interpreting the data is more like communicating your findings to the non-technical audience.

The objective of this step is to first identify the business insight and then correlate it to your data findings.

6.Revisiting the Model:

As our model is in production, it becomes important to revisit and update our model periodically, depending on how often we receive new data or as per the changes in the nature of the business.

If we don’t, our model will degrade over time and won’t perform as well, leaving our business to degrade too

Summary:

* Identify the Business Problem and start with reasonable objective.
* Obtain the relevant data from authentic and reliable sources.
* After obtaining the data, Clean the data and remove Anomalies
* Explore the data with Visualizations and find the key Patterns.
* Model the Data with Different Machine Learning Algorithms.
* Interpret the Data with Evaluation.
* Revisit the model and update the Model Periodically