**BACKGROUND**

Indian Railways owing to its cost-efficiency remains the most convenient and sought-after mode of transportation. One of the major problems encountered by the users is to make a decision about the train they would opt for, based on the delay, seat availability, connectivity and other factors. We have observed that the data that is present accounts for only a past few days of record and reaching to the conclusion would often require intensive search thus consuming a lot of time.

Our aim is to collect relevant information, analyse the data and create visual based report which are easier to comprehend and would expedite the decision making process. The interactive visual reports would be of great assistance to the user, but at times the user might not be equipped with sufficient ability to apprehend the visual reports to arrive at conclusion. For such scenarios we aim to simplify the task further by making use of the data at our disposal to predict or suggest the train to the user measured on various metrics like delay, type of train etc.

One of the major challenge that we face is the collection of data, the historical data is not available directly at one source, from where it could be obtained.

Following mentioned links justify the same:

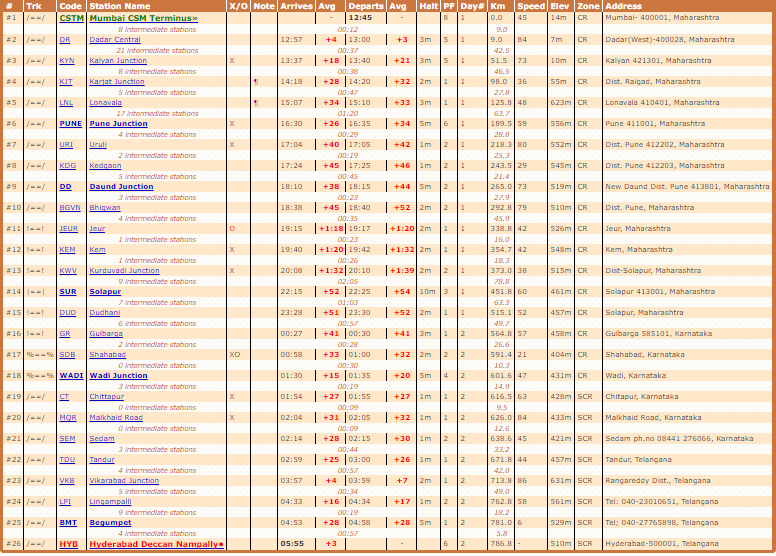
<https://www.quora.com/How-can-I-get-a-past-years-data-for-arrival-departure-delays-for-Indian-Railways>

<http://www.indiamike.com/india/indian-railways-f10/indian-railway-ontime-performance-train-delay-t139389/>

<https://www.quora.com/How-can-I-get-statistical-data-related-to-Indian-railways-of-last-5-or-10-year>s

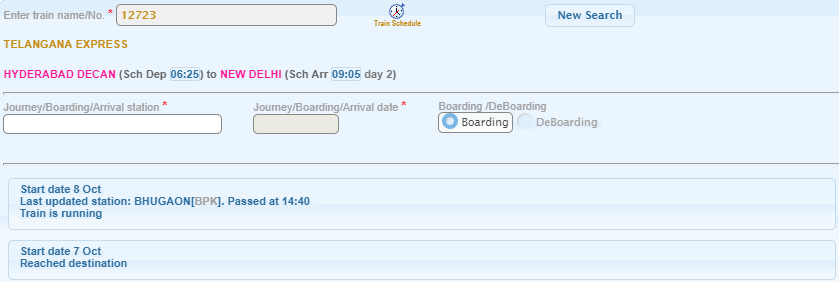
There are few sites on internet which would provide user past information about a couple of days. Some of them has been listed below:

* IndianRailInfo: <https://indiarailinfo.com/>



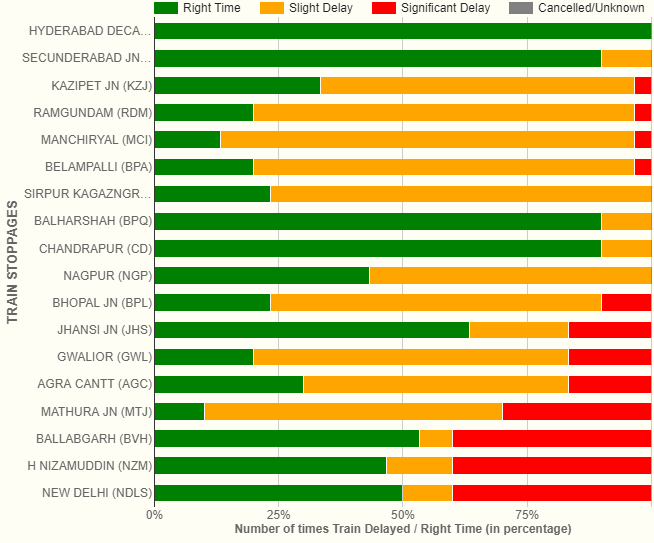
Data displayed on this site might not be precise. Moreover, it shows the data corresponding to average delay only.

* NTES: <https://enquiry.indianrail.gov.in/ntes/>



Provides the running instances corresponding to past two to three runs only.

* Etrain:<https://etrain.info/in?PAGE=runningHistory--12349--1m#!PAGE=runningHistory--12723--1m>



Provides number of times the train has been delayed but with no precise details.

**PROBLEM FORMULATION**

Usually for a given route, there are many trains that travel via the same route as desired by the person, but this necessarily doesn’t mean that all the trains travelling on same route are going to take same time, some of them might get delayed frequently while for some other delay might be seldom. The main goal of the project is to develop application which would be helpful to the targeted audience in selecting better alternative from the available options. We intend to develop interactive visual based reports which can be customized as per the need using Power BI along with prediction of the best trains for a given route using machine learning and artificial intelligence.

**PROGRESS**

* Analysed the work flow of the site NTES (National Train Enquiry System) website.

This is the data source site from where we will collect data.

* Created a prototype for database which is in its initial stage and would be modified as per the requirement.
* Developed two different selenium script with java as a wrapper language. These selenium scripts would automate the process of dumping data into database after capturing the data from the website.
* We have used multi-threading to reduce the time taken to dump the data. When the data was being pulled sequentially the approximate time taken for dumping data for 10 trains ranged between 10-12 minutes, while with the help of parallel programming this figure has come down to 3-4 minutes.
* We are presently working on the database design.
* Currently working on data collection considering 30-40 trains.