Dataset Details

We have provided a real-world dataset containing case data as captured by eCourts for 739,670 hearings across 134,699 unique civil cases. This data has been drawn from over 20 years and covers only disposed cases across 8 case types in the Karnataka High Court (restricted to Bangalore Bench) – ***dataset explanation in Appendix.***

Civil cases at the high court level typically go through roughly the following stages:

- Office objections / pre-admission
- Admission & Framing of charges
- Evidence
- Arguments
- Interlocutory applications (at any stage) or additional submissions
- Settlement or revoked (at any stage)
- Orders/Judgments
- Final Disposal

Hearings we have been unable to reasonably classify have been tagged as Other.

This dataset gives an indication of:

- how long cases of different types have taken to be disposed, how many hearings they have gone through, and how frequently they have been listed;
- how purposes and stages of each hearing have been captured and how much time cases spent in each stage;
- combinations of judges who presided over specific hearings.

However, this is historical data and any algorithm being designed for the future should be able to think afresh about the scheduling challenge while still:

- taking into account a range of complex factors (intelligently discerning between signal and noise),
- navigating shortcomings or limitations of data capture and quality in the available metadata, and
- allowing judges (the main users) to establish their personal preferences and customise the ultimate designed approach.

The case table contains:

- Two unique identifiers: the **CNR number**, which is a 16-digit number unique across India. The first 4 digits indicate the High Court (in this case, KAHC for Karnataka HC) and the last 4 digits indicate the year the case was filed (i.e. 2000) as well as the **Combined Case Number**, which contains an abbreviation for the case type, an assigned case number (by the court registry), as well as the year of filing.
- There are 8 civil **case types** chosen for this dataset:

RSA	Regular Second Appeal
CRP	Civil Revision Petition
RFA	Regular First Appeal
CA	Company Appeal
CCC	Civil Contempt Petition
CP	Civil Petition
MISC.CVL	Miscellaneous for Civil

CMP	Civil Miscellaneous Petition

- **Name of High Court** (Karnataka), **Court name** (Principal Bench at Bengaluru) are uniform across the dataset while **Court Number** varies based on the real allocated courtroom
- **Current Status** is Disposed for all cases provided in the dataset
- The dataset contains the following dates:
    - **Date filed**: The date when the lawyer filed a case.
    - **Registration Date**: The date when the case was registered by the court, which may differ slightly from the filing date, depending on how long the scrutiny process took.
    - **Last sync date**: The date on which the data was extracted from the eCourts database.
    - **Decision date**: The date on which a final judgment is given in a case, and it is considered disposed by the court.
- **Nature of Disposal, Outcome** and **Binary**: The nature of disposal provides some clarity into how the case was decided by the court, but this field is often not filled.
    - It contains two components, the first is merely 'contested' or 'uncontested', which we have separated as the **binary** column while the second is more detailed and is separated under the **outcome** column.
- **NJDG Judge name** is usually a marker of whether the case has been allocated to a particular courtroom or is yet to be allocated. This determines the 'queue' into which a case enters for the purpose of scheduling.
- The **Petitioner** and the **Respondent** columns contain the names of the parties who have filed the case. Sometimes demographic characteristics, such as gender, age or protected class may be taken to list a case on an urgent basis.
    - These columns have been ***removed from the dataset*** to protect personal information. Participants should feel free to generate synthetic data for party names if relevant for their intended outputs.
- **Disposal time adjusted** is the number of days that each case took to be disposed and is calculated as (Decision Date – Date filed +1). This is because some cases were filed and disposed on the same day, so we wanted to ensure they were calculated as 1 day for disposal instead of 0.

Some columns such as **Police Station, Acts** and **Section** may not be specifically relevant for building a scheduling algorithm but provide further information on the nature and range of cases being filed.

The hearing table contains:

- **Hearing_ID** is an unique identifier that combines the Case number and the date of hearing,
- **Petitioner** and **Respondent** information along with their listed **Advocates** (for the purpose of generating the daily cause list)
- **Before Honorable Judges** – indicates the combination of judges that have heard a specific hearing. If there are multiple judges, there are separated in successive columns – cases may be heard by a single judge or a division bench (where a combination of judges may sit together, for civil cases, this is typically in cases with more complicated questions of law or a greater commercial value).
- **Business on Date** is the specific date of the hearing
- **Purpose of Hearing** is the way the court captures what was due to happen in each hearing – we have consolidated these in a **Re-mapped stage** column found later in the dataset

We've retained a number of fields that are un-populated to give an indication of data that may be synthetically produced to build the algorithm (for example allocation to a specific court 'code' or day's board). These may be ignored by participants if they do not wish to utilise them.

Although, we want participants to be informed by this data but develop a fresh model based on clear assumptions. They may set some logical bounds (example: the number of cases to be capped in a day's schedule) or allow for addition or deletion from the 'optimised' list due to practical considerations such as last minute changes or cases demanding urgent attention.

Guiding Principles for Judicial Scheduling Exercise

These guiding principles will help participants design practical, realistic, and dynamic scheduling models while considering operational constraints, judicial time, and case priorities. Participants are expected to use the provided data as a foundation but develop a fresh scheduling model based on clear assumptions. Logical bounds may be set (e.g., capping the number of cases per day) or adjusted to account for practical considerations such as last-minute changes or urgent cases.

Principle 1 – Court Working Days & Time Horizon:
High Courts operate 5 days a week with pre-set vacation days.[1] The algorithm should simulate case scheduling over a 2-year period to allow cases to appear multiple times, progress through stages, and move toward disposal.

Principle 2 – Courtroom Allocation:
Cases are allocated to a specific courtroom or 'queue' for first hearings, determined by case type. The scheduler should handle multiple courtrooms (minimum 5), each presided over by a single judge.

Principle 3 – Limited Judicial Time:
Judges have finite hearing time each day. Using dataset insights, participants can estimate a daily maximum number of cases per courtroom. Each hearing's exact duration is unknown.[2] Unheard cases are adjourned with no strict rescheduling policy; moreover, ideal frequency of listing is not fixed.[3] Cases may be categorised as "ripe" or "unripe", e.g., cases listed for "non-compliance with office orders" could be handled asynchronously to optimise judicial time.

Principle 4 – Case Tracking & Alerts:
No case should be overlooked. Cases may go unlisted if parties or lawyers do not actively follow up. The algorithm should include mechanisms to flag long-pending cases or suggest reasons for non-listing. Participants may introduce synthetic fields or front-end options to capture such information.

Principle 5 – Real-World Dynamism:
The scheduler should account for case inflow and outflow, approximating monthly filing and disposal rates using the NJDG data or as observed from the provided dataset.

---

[1] These can be found on high court websites (https://judiciary.karnataka.gov.in/calendar.php) - 2025 is estimated to have 192 working days

[2] Delhi High Court Zero Pendency Project: https://www.dakshindia.org/delhi-pilot-project/ and DAKSH Time and Motion Study: https://www.dakshindia.org/time-and-motion-study/

[3] The dataset does not capture typical or average time taken for various kinds of hearings – in other work, this has been captured through time-and-motion studies.

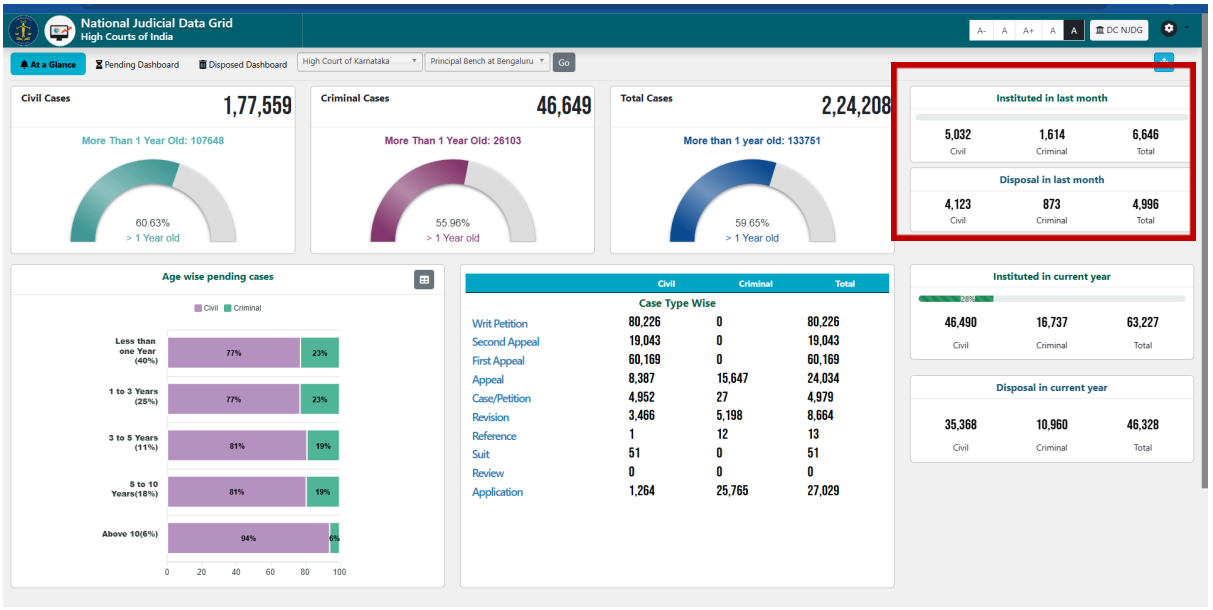

Figure 1: National Judicial Grid as of September 22nd 2025

Note: Participants may come up with additional assumptions or model conditions as they progress – we would just ask that these are made explicit. Similarly, they may choose not to take all the information into account for simplicity's sake or generate additional fields of synthetic data or utilise AI/LLM tools to facilitate the algorithm, if desired.