

The Pomona-Pitzer baseball team uses a TrackMan unit to record detailed pitch-level and batted-ball data, such as velocity, spin, movement, and exit velocity. At present, this information is downloaded as large CSV files and manually processed in spreadsheets, which makes it slow, repetitive, and difficult to reproduce analyses. My capstone project will use the TrackMan API to build a cloud-based database that stores this data in a structured, queryable format and updates automatically on a regular schedule.

In addition to this database, I will develop automated tools to generate the pre- and post-game scouting reports that the team currently creates by hand. The advanced data science aspect of the project involves designing a reproducible ETL (Extract, Transform, Load) pipeline that can handle the limitations of the API and streamline reporting. This work will transition the team's analytics from one-off spreadsheets to a more sustainable infrastructure, making it easier for future analysts to query, visualize, and extend the data for new projects.