Limitations of analytics over anonymously submitted data.
This analysis relies on crowd-submitted, largely unverified records that can contain omissions, inconsistencies, and selective reporting. Self-selection bias is strong: applicants with unusually positive or negative outcomes are more likely to post, which skews rates (e.g., acceptance share) and averages. Fields such as GPA and GRE may be reported in different scales or formats; despite light "reasonability" caps (GPA ≤ 5, GRE ≤ 400) and standardized parsing, missing values, rounding to coarse bins, and occasional field misuse (e.g., total vs. section scores) remain. Duplicates, partial edits, and posts without stable IDs are filtered, but that can remove data unevenly across programs or time. LLM-based normalization of university/program names improves grouping but can introduce mapping errors; these are mitigated by conservative defaults and audit logs.

Why some metrics differ from official standards.
Differences such as higher observed GRE metrics compared with official ETS distributions are expected for several reasons. First, the sample is not representative: posters often come from highly competitive CS programs and top-tier institutions, where applicant pools have stronger test profiles. Second, survivorship and brag bias drive over-reporting of strong outcomes and under-reporting of weaker ones. Third, program filters (e.g., focusing on CS graduate admissions for specific terms) select a subpopulation with systematically higher scores. Finally, data entry quirks: rounded scores, mixed scales, and incomplete fields can inflate or deflate means in small subgroups. Consequently, these analytics are informative for directional patterns in this community, but they should not be interpreted as population-level statistics.