

GradCafe is anonymous and self-reported, which introduces bias. Users selectively share data, often higher scores or unusual outcomes, so GPAs and GREs appear inflated and acceptance rates don't reflect actual stats. Field entries like "status," "term," and "degree" are inconsistent, frequently incomplete, or contain typos.

The platform changes over time and posts aren't validated. Even after cleaning and LLM normalization, some records remain ambiguous or incorrect. With no verification, inaccurate or misleading entries persist. This makes the dataset useful for practicing scraping, cleaning, and SQL analysis but unreliable for official admissions trends.