



DATA CLEANING & PRE-PROCESSING

- ▶ Loaded profiles, LatLon, addr, and sharedCensus datasets.
- ▶ Merged datasets on the 'location' column.
- ▶ Removed age outliers (e.g., 109, 110).
- ▶ Corrected data points like "drinks desperately".
- ▶ Handled missing values.
- ▶ Recategorized 'education' for clarity.
- ▶ Set boundaries for US geographic data.

AGE & ORIENTATION

- The **age distribution peaks between 25-30**, indicating most users are within this age range.
- **Most users identify as straight**, followed by gay and bisexual.

