Data understanding

1. Corr
2. univariate
   1. Artist
      1. distinct\_values
      2. value\_counts
   2. Auth
   3. Gender -- Keep
   4. Level -- Keep
      1. A screenshot of a cell phone

         Description automatically generated
   5. location: string (nullable = true) -- too long, only keep the state
   6. A screenshot of a cell phone

      Description automatically generated
   7. method: string (nullable = true)
   8. page: string (nullable = true) --2 values define churn
   9. registration: long (nullable = true) -- User's registration timestamp
   10. sessionId: long (nullable = true)
   11. song: string (nullable = true)
   12. status: long (nullable = true) -- should change type
   13. ts: long (nullable = true)
   14. userAgent: string (nullable = true)
   15. userId: string (nullable = true)
3. features
   1. Distinct values
   2. Missing values
   3. Cate:
   4. Num
   5. Id:
4. Bivariate
   1. Gender vs Churn
   2. Level vs Churn
   3. Registration cs Churn
5. average number of songs per session
6. Total number of thumbs down
7. Total number of thumbs up
8. Total number of errors encountered
9. Number of Add to Playlist events
10. Number of Add Friends events
11. Total number of songs played