EDA – train.csv

* Data Summary

A picture containing table

Description automatically generated

* + No duplicate rows
  + No null latitude/longitude values
* Cleaning
  + Text cleaning function (Lowercase, remove all punctuation)
    - name, address, city, state, zip, country, url, categories
  + Remove spaces
    - zip
  + Round to 6 decimal places
    - lat, lon
  + Some processing of **phone** field needed
* Feature ideas
  + Length of field
  + Extract a language from field