National Parks trail data dictionary:

Trail\_id: Trail ID

Name: Trail Name

Area\_name: National Park containing trail

City\_name: City trail is closest to

State\_name: State containing trail

Country\_name: Country containing trail (all United States)

\_geoloc: Lat/Long of trail starting point

Popularity: internal alltrails rank of popularity, maybe found by # of views on site?

Don’t know how this id different from visitor\_usage

Length: length of trail in meters

Elevation\_gain: elevation gain of trail in meters

Difficulty\_rating: 1:”easy”, 3:”moderate”, 5:”hard”, and 7:”hard”

Route\_type: factor w/three levels: “out and back”, “loop”, ”point to point”

Visitor\_usage: 1:”light”, 2:”moderate”, 3:”heavy”, 4:”heavy” (contains 253 NAs –

recoded to 0)

Avg\_rating: rating given by users on site; rounds to half-point mark (so, 0, .5, 1, 1.5…)

Num\_reviews: number of reviews given on website

Features: features of trail (dog-friendly, forest, lake, kids, views, etc)

Activities: things to do on trail (hiking, birding, walking, snowshoeing, etc)

Units: type of units to use? I think this is worthless, all units in doc are metric)

Tasks:

* Recode \_geoloc to lat & long columns
* One-hot encode route\_type
* One-hot encode difficulty/visitor\_usage?
* Remove country\_name and units, they’re irrelevant
* Recode features and activities