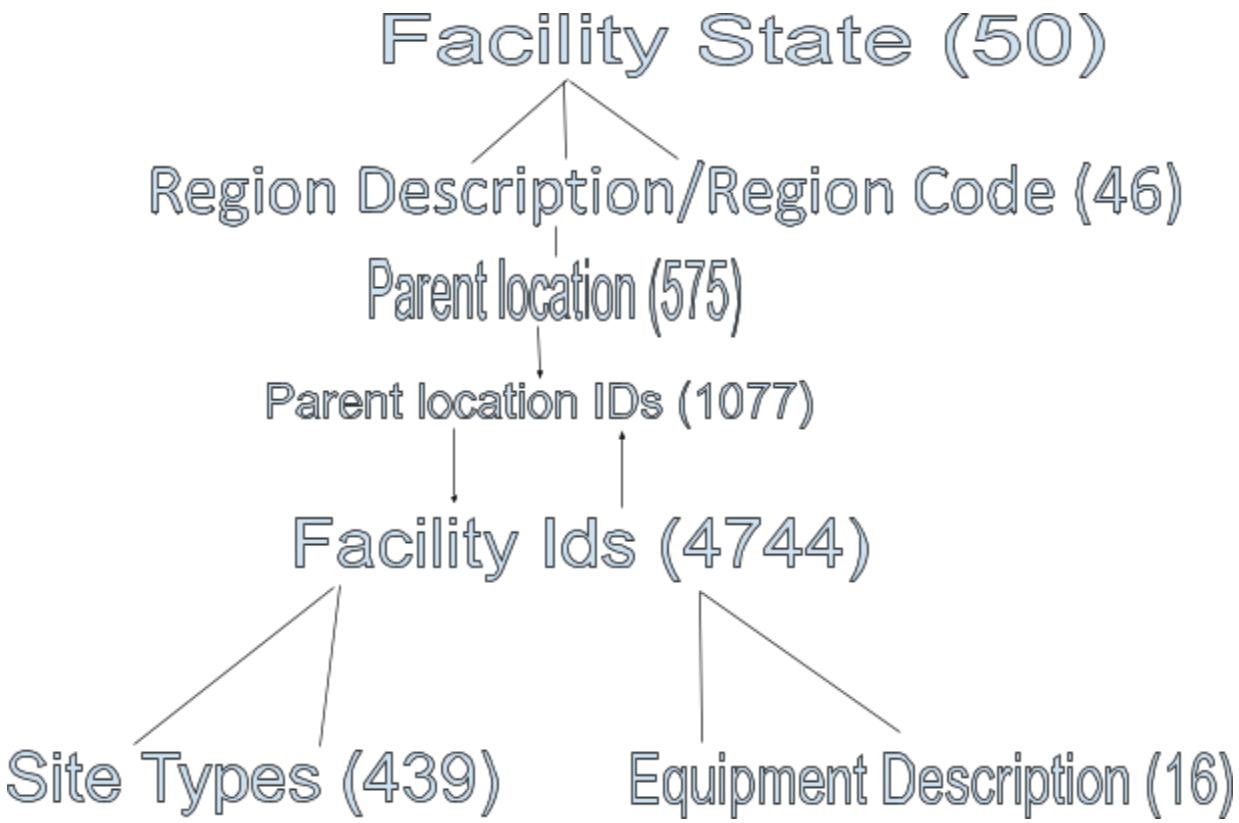
Summary Analysis from the DW sheet

1. Total Number of columns - 35
2. Total Number of rows - 8984097
3. Total unique Region codes- 47
4. Total unique Region Description- 46
5. Total unique Parent location Ids- 1077
6. Total unique Parent location - 575
7. Total unique Site Types - 439
   Eg array(['STANDARD NONELECTRIC', 'GROUP TENT ONLY AREA NONELECTRIC', nan,
   'GROUP SHELTER NONELECTRIC', 'TENT ONLY NONELECTRIC',
   'Nature Tour', 'Hike', 'Historic Tour', 'CABIN NONELECTRIC',
   'GROUP STANDARD NONELECTRIC', 'Day Use Fee',
   'Buckskin Gulch Day Use – Activity Scan and Pay',
   'Wire Pass Day Use - Activity Scan and Pay',
   'White House Day Use - Activity Scan and Pay', 'Entrance',
8. Total unique User Type - 5
   Eg array([nan, 'Overnight', 'Day', 'Activity', 'Multi'], dtype=object)
9. Total unique Inventory type -14
   Eg (['CAMPING', 'LOTTERY_CAMPING', 'TICKET','ACTIVITY PASS', 'DONATION', 'PERMIT', 'LOTTERY_PERMIT', 'VEHICLE_PERMIT', 'TIMED_ENTRY','VENUE_RESERVATIONS', 'LOTTERY_TICKET', 'QUEUE_LOTTERY', 'RENTAL','TREE_PERMIT'],
10. Total unique Facility IDs - 4744
11. Total unique facility states - 50
12. Total unique Equipment Descriptions - 16
    Eg - array(['Tent', nan, 'Large Tent Over 9X12`', 'Vehicle', 'Pop up', 'RV', 'Car', 'Trailer', 'Fifth Wheel', 'Caravan/Camper Van','Pickup Camper', 'Small Tent', 'Boat', 'Hammock', 'RV/Motorhome', 'Horse'], dtype=object)
13. There are 22 categorical and 13 non-categorical columns



Find the Null values and percentage

1. There are 15111 missing values in region description and region code column which is about 0.168% -

   Below are the columns having the most missing values with % in dataset

```
tranfee                 100.000000
usefee                  100.000000
attrfee                  99.999955
equipmentdescription     57.044275
equipmentlength          54.079681
usetype                  51.327807
legacyfacilityid         39.793159
sitetype                 36.533455
numberofpeople           35.360749
enddate                  31.648445
nights                   31.628766
facilityzip              22.977167
facilitystate            22.822372
customerzip              19.655876
```

```
facilitylongitude           18.694901
facilitylatitude            18.694901
```

Questions answered from Analysis

1. Calculate how many sites are associated with each region.
2. Calculate how many equipment descriptions are associated with each region.
3. Calculate how many regions are associated with each facility state.
4. Created df by Calculating the missing values per region for each inventory type
5. Created df by Calculating the missing values per region for each site type
6. what columns are categorical what not categorical
7. Created dataframes to calculate the (null,min,max,stdev) for numerical columns (eg number of people) per region per inventory type (eg camping)
8. Created dataframes and csv files for categorical columns to count the number of eg. site types per region and per inventory type
9. Created dataframes and csv files for categorical columns to count the number of eg. facility state per region and per inventory type.