Submission

Ready to submit your project? Collect the following files:

1. A pdf document containing your answers to the rubric questions. This file should document your data wrangling process.
2. Your Python code for Lesson 6 quizzes, as well as any additional code you used in auditing and cleaning your dataset for the final project.
3. A text file containing a link to the map position you wrangled in your project, a short description of the area and a reason for your choice.
4. An .osm file containing a sample part of the map region you used (around 1 - 10 MB in size). See instructor notes.
5. A text file containing a list of Web sites, books, forums, blog posts, github repositories etc that you referred to or used in this submission (Add N/A if you did not use such resources).

Project 3 specs: <https://docs.google.com/document/d/1TpfNxDzUjhibq9Qb8cOQHtlvZUelft-W0fb7pCTTyYE/pub?embedded=True>

Example: <https://docs.google.com/document/d/1F0Vs14oNEs2idFJR3C_OPxwS6L0HPliOii-QpbmrMo4/pub?embedded=True>

* Problems encountered in your map (1)
* Overview of the Data (2)
* Other ideas about the datasets (3)

1. Data issues:
   1. Street names differ - addr:street
   2. Zip codes differ - [addr:postcode](http://wiki.openstreetmap.org/wiki/Key:addr:postcode)
   3. addr:state
2. After downloading an Open Street Map provided area of South Lake Tahoe I’ve run some data correction checks against the data, below are some of the data quality issues that I’ve uncovered.
   1. Street names have different endings. For example, ‘Market St’ instead of ‘Market Street’ or ‘Heybourne Rd’ instead of ‘Heybourne Road’.
      1. 'St': set(['Market St'])
      2. 'Rd': set(['Hawkins Ranch Rd', 'Heybourne Rd', 'River Ranch Rd'])
      3. 'Ln': set(['Library Ln'])
      4. 'Blvd': set(['7081 N. Lake Blvd', 'Lake Tahoe Blvd'])
      5. '27.4': set(['Hwy 80 PM 27.4'])

All of the above are updated to match the majority, which in this case are full names.

* 1. Some of the Zip codes have state abbreviations attached. The majority of the Zip codes don’t, therefore I assume that the below are incorrect and remove the state abbreviations
     1. CA 96143: 2
     2. CA 96161: 4
     3. NV 89511: 2
     4. NV 89413: 2
     5. NV 89449: 2
     6. NV 89511: 2
     7. NV 89701: 2
     8. NV 89703: 2
     9. NV 89706: 2
     10. RG6 1LT: 2 <= does not look like a US Zip code at all, after short Googling found that this is a UK zip code, therefore I will update this value to NA
  2. Lastly, some of the state names are abbreviated and some are written in full
     1. CA: 354
     2. California: 2
     3. Nevada: 2
     4. NV: 248

Since the majority of the names are abbreviated I will update the full names to be abbreviated as well

1. File description:
   1. Size:
      1. south-lake-tahoe\_california.osm circa 116 mb
      2. south\_taho.json circa xxx mb
   2. Number of nodes:
   3. Number of Ways
   4. Number of tags:
   5. Number of unique users:
   6. Number of some nodes (café, street, etc.)
2. Other stats from the data set:
   1. Top contributing user:
   2. Second Top contributing user:
   3. Top node:
   4. Top tag attribute:
   5. Most active day (timestamp based):
   6. Most active month (timestamp based):