Adam Sneath // April 17, 2019 // Pitt Analytics Team // CMS Stars Project // Verification Process

*All code files used for this processing, plotting, and dashboard design can be found at the public Github repo:* [*https://github.com/adsneath/Highmark-2018-CMS-Stars-Shiny-Dashboard*](https://github.com/adsneath/Highmark-2018-CMS-Stars-Shiny-Dashboard)

**Received Two Data Sets**

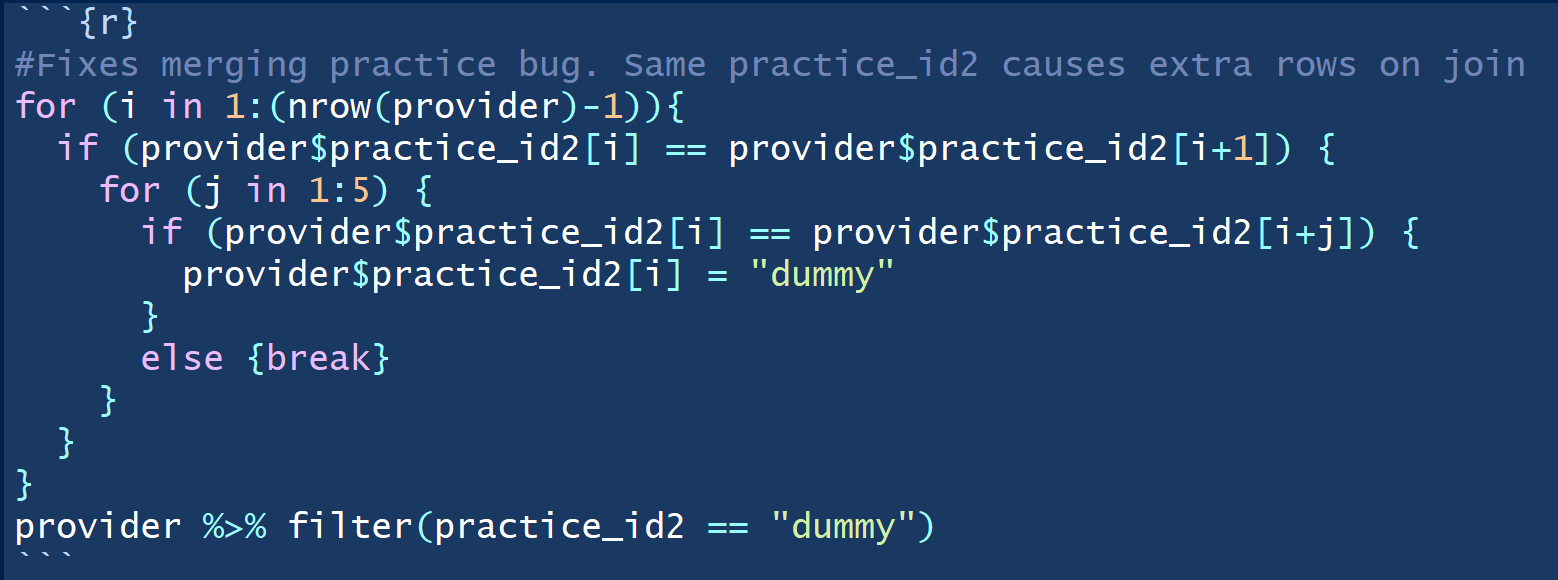
All data (and data dictionaries) received from Highmark Data Engineer Matthew Glover

* provider\_20181229.sas7bdat – “Provider”
* tab1\_practice\_level.sas7bdat – “Practice

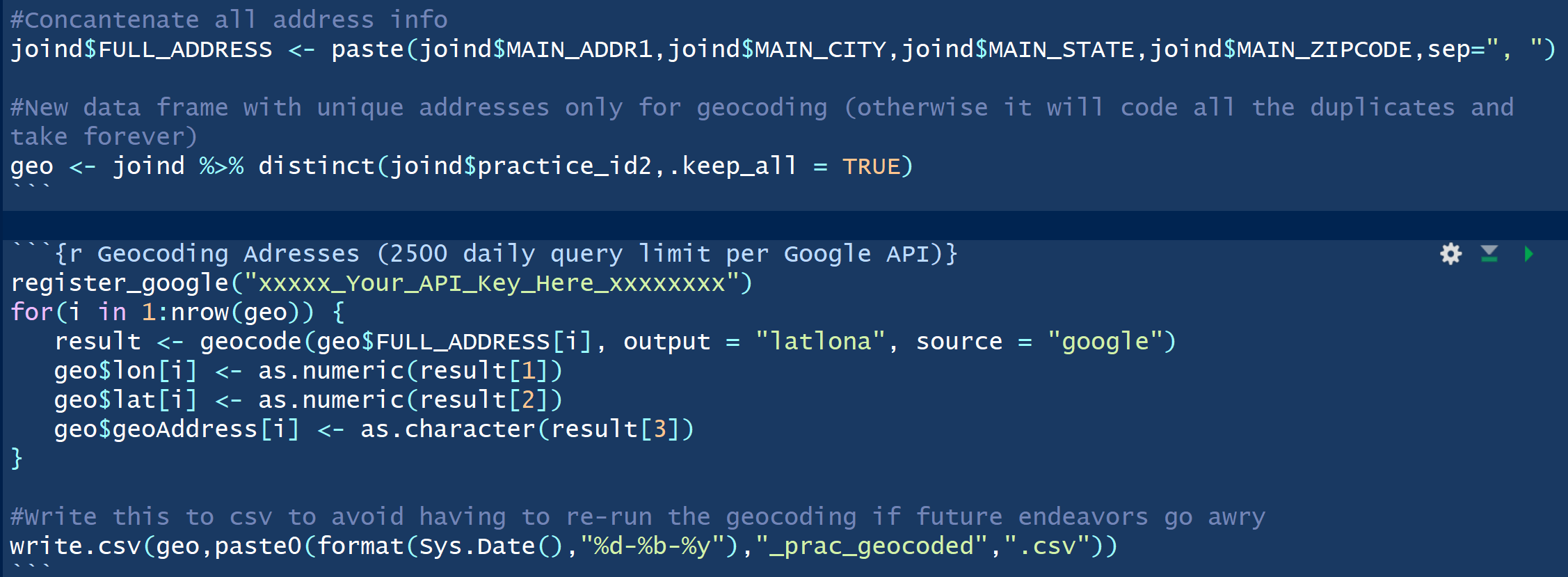
**Data Manipulation**

Let [1] and [2] refer to Appendices 1 and 2 (data dictionaries provided by Matt)

1. Recognized that practice\_id2 [1] is unique identifier
   1. PRACTICE\_ID uniquely identifies some practices that have since merged; for example, two (or three) practices might have different PRACTICE\_ID but just one practice\_id2 for the merged location; therefore practice\_id2 uniquely identifies all practices.
2. Wrote for loop to assign “dummy” to practice\_id2, of the old practices (within the merge), to avoid the creation of extra rows on joining by practice\_id2. For instance, if the join notices.

****

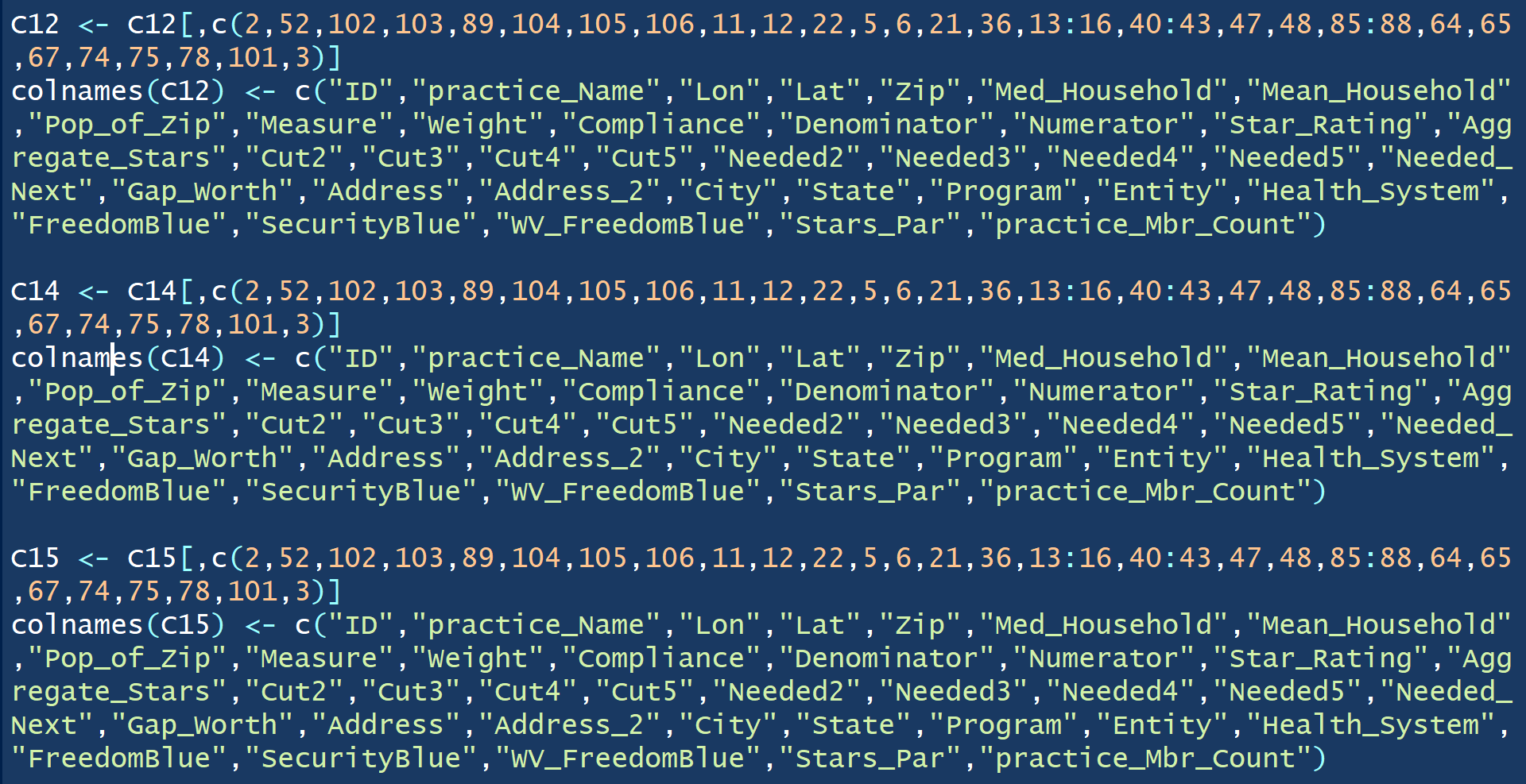
1. Join datasets using joined <- left\_join(practice,provider,by="practice\_id2") so all practice data records have address attributes
2. Concatenate address information to combine street, city, state and zip in preparation for Google API geocoding
3. Geocode (convert address to lat & lon) all addresses using free Google API & for loop

****

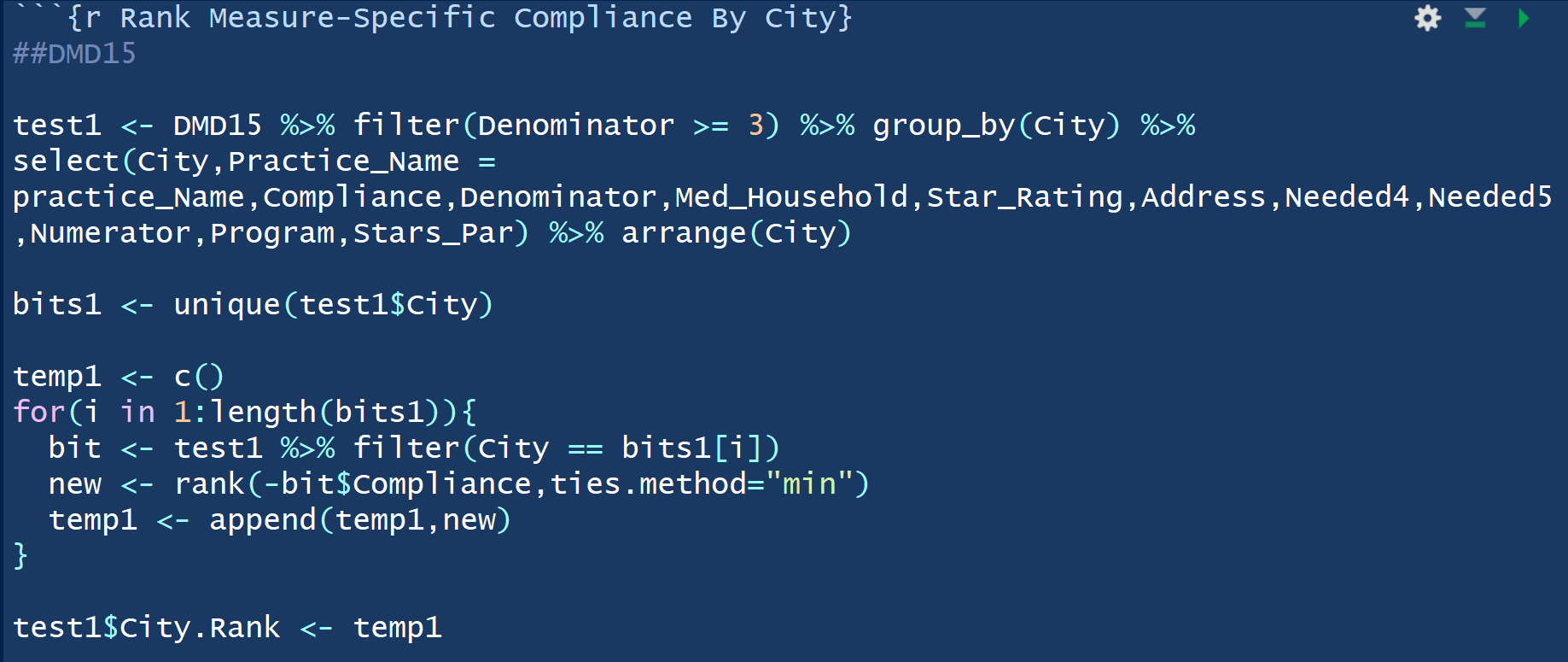
1. Join lat & lon back to practice data

**Data Filtering and Cleaning (see “Process Code Files” 🡪 “4-17 Clean Master” on Github)**

1. Filter data to individual data frames (in R) for each of the following 16 Star Measures: C01, C02, C07, C12, C13, C14, C15, C17, C20, C21, D11, D12, D13, D14, DMD15, DMC17
2. Remove denominator = NA or denominator = 0 so they don’t interfere with plotting (they also have NA or 0 values of compliance / star\_rating and we don’t want these on plots)
3. Cut data down to most relevant columns and reorder them most logically



1. Write each data frame to its own .csv file for later use in the Dashboard
2. Implement algorithm to rank each practice (with at least 3 in denominator) in terms of compliance within their city, and join this City.Rank back to the practice data

****

1. Write these “ranked” data frames to their own .csv files for later use in the dashboard.

*All .csv files needed for loading the dashboard are on Github.*

**Verification**

Verification was performed by common sense and low-level analyses. For example:

* After joining provider data to practice, it is ensured that the resulting joined data has the same number of rows as the original practice data.
* After geocoding, about 10 lat/lon pairs are chosen at random and checked on google maps to make sure they’re at the correct address.
* The denominators of all records in each .csv file are summed and checked against the original “all contracts” spreadsheet (containing total eligible member population for each measure) and they agree.
* The City.Rank algorithm is easily confirmed using the interactive data table (on the Dashboard) and searching a city then filtering compliance in descending order. Notice that the highest compliance has the highest City.Rank.

**Appendix 1 [Provider Data Dictionary]**

|  |  |  |  |
| --- | --- | --- | --- |
| **Alphabetic List of Variables and Attributes** | | | |
| **#** | **Variable** | **Type** | **Label** |
| **1** | STATUS | Char | Active/Terminated [Terminated IDs may have practice\_id2/practice\_name2 because they combined/turned into other practices] |
| **2** | group\_shld | Char | [Use practice ID] |
| **3** | PRACTICE\_ID | Char | Practice ID |
| **4** | practice\_name | Char | The practice name |
| **5** | PRACTICE\_NPI | Char | Federal NPI associated with practice |
| **6** | PRACTICE\_EFF\_DT | Num | Practice Effective Date |
| **7** | tax\_id | Char | Federal Tax ID for Practice\_ID |
| **8** | REGION | Char | Highmark Region |
| **9** | crosswalk\_id | Char | use practice\_id2 |
| **10** | crosswalk\_name | Char | use practice\_name2 |
| **11** | crosswalk\_tax\_id | Char | Use tax\_id2 |
| **12** | XREF\_EFF\_DT | Num | crossref effective date |
| **13** | XREF\_EFF\_AFF\_DT | Num | crossref affiliation effective date |
| **14** | practice\_id2 | Char | The current practice ID |
| **15** | practice\_name2 | Char | The current practice name |
| **16** | tax\_id2 | Char | Federal Tax ID for Current Practice ID |
| **17** | PROGRAM\_NAME | Char | Incentive Program Name [True Performance] |
| **18** | entity\_name | Char | True Performance Entity Name |
| **19** | ENTITY\_ID | Char | ENTITY\_ID |
| **20** | health\_system\_name | Char | Name of affiliated health system |
| **21** | HEALTH\_SYSTEM\_ID | Char | HEALTH\_SYSTEM\_ID |
| **22** | hospital\_system\_id | Char | ID of affiliated hospital system |
| **23** | hospital\_system\_name | Char | Name of affiliated hospital system |
| **24** | INDEMNITY\_PAR | Char | [Use Stars Par] |
| **25** | KHPW | Char | KHPW |
| **26** | PBS | Char |  |
| **27** | FREEDOMBLUE | Char | [Use Stars Par] |
| **28** | SECURITYBLUE | Char | [Use Stars Par] |
| **29** | WV\_INDEMNITY\_PAR | Char | [Use Stars Par] |
| **30** | WV\_PPO | Char | [Use Stars Par] |
| **31** | WV\_FREEDOMBLUE | Char | [Use Stars Par] |
| **32** | DE\_INDEMNITY\_PAR | Char | [Use Stars Par] |
| **33** | DE\_FEP | Char | [Use Stars Par] |
| **34** | DE\_PPO | Char | [Use Stars Par] |
| **35** | COMMBLUE\_MEDADV | Char | [Use Stars Par] |
| **36** | non\_par | Char | Does not participate |
| **37** | BS\_MVEI\_ID | Char | BS\_MVEI\_ID |
| **38** | MAIN\_ADDR1 | Char | Physical Practice Address |
| **39** | MAIN\_ADDR2 | Char | Physical Practice Address |
| **40** | MAIN\_CITY | Char | Physical Practice Address |
| **41** | MAIN\_STATE | Char | Physical Practice Address |
| **42** | MAIN\_ZIPCODE | Char | Physical Practice Address |
| **43** | MAIN\_COUNTY | Char | Physical Practice Address |
| **44** | MVC\_SEQ\_N | Num |  |
| **45** | MV\_ID | Num |  |
| **46** | fax | Char | Fax number |
| **47** | MAIL\_ADDR1 | Char | MAIL\_ADDR1 |
| **48** | MAIL\_ADDR2 | Char | MAIL\_ADDR2 |
| **49** | MAIL\_CITY | Char | MAIL\_CITY |
| **50** | MAIL\_STATE | Char | MAIL\_STATE |
| **51** | MAIL\_ZIPCODE | Char | MAIL\_ZIPCODE |
| **52** | MAIL\_COUNTY | Char | MAIL\_COUNTY |
| **53** | entity\_number | Char | True Performance Entity ID |
| **54** | stars\_par | Char | Eligible for Participation in Stars Incentive Program |

**Appendix 2 [Practice Data Dictionary]**

|  |  |  |  |
| --- | --- | --- | --- |
| **Alphabetic List of Variables and Attributes** | | | |
| **#** | **Variable** | **Type** | **Label** |
| **1** | f | Char | Practice Blue Shield ID |
| **2** | mbr\_count | Num | Number of Attributed Members for Entity/Practice/Health System |
| **3** | measure\_cd | Char | HEDIS Measure Code |
| **4** | denominator | Num | Eligible Population (Denominator) |
| **5** | numerator | Num | Number of Gaps Closed |
| **6** | br | Num | # of Eligible Population Beyond Remediation |
| **7** | open | Num | Open Gaps (that can still be changed) |
| **8** | pred\_readm | Num | [Use Observed] |
| **9** | measure\_cat | Char | Class |
| **10** | measure\_ds | Char | Measure Name |
| **11** | measure\_wt | Num | Measure Weight |
| **12** | star2 | Num | 2 Star Compliance |
| **13** | star3 | Num | 3 Star Compliance |
| **14** | star4 | Num | 4 Star Compliance |
| **15** | star5 | Num | 5 Star Compliance |
| **16** | row | Num | row |
| **18** | gaps\_closed | Num | # of Gaps Closed (Numerator) |
| **19** | ytd\_comp | Num | Actual YTD Compliance Rate |
| **20** | star\_rating | Num | Current/Trend Measure Star Rating |
| **21** | compliance | Num | Compliance Rate |
| **22** | display\_denominator | Num | Calculation Denominator |
| **23** | display\_numerator | Num | Calculation Numerator |
| **24** | pot\_star\_rating | Num | Maximum Potential Star Rating (after beyond remediation taken into account) |
| **25** | gaps\_addressed | Num | # of Gaps Addressed and On Track (Trend) |
| **26** | trend | Num | Trend Compliance Rate |
| **27** | max\_pot | Num | Maximum Potential Compliance Rate (after beyond remediation taken into account) |
| **28** | closed | Num | [Use Gaps Closed] |
| **29** | observed | Num | Observed Readmissions |
| **30** | expected | Num | Expected Readmissions [All-Cause Readmissions Only] |
| **31** | star\_rating\_ind | Num | Whether the Star Rating counts for this measure |
| **32** | trend\_wt | Num | Points after weight is multiplied by star rating |
| **33** | max\_trend\_wt | Num | The weight times 5 |
| **34** | pot\_trend\_wt | Num |  |
| **35** | aggr\_star\_rating | Num | Total Star Rating of Entity/Practice/Health System/etc |
| **36** | points\_earned | Num | [Not for Stars Incentive Program] |
| **37** | sum\_weights | Num | Weights used for all cause readmissions risk calculation |
| **39** | needed2 | Num | Numerator Change for 2 Stars |
| **40** | needed3 | Num | Numerator Change for 3 Stars |
| **41** | needed4 | Num | Numerator Change for 4 Stars |
| **42** | needed5 | Num | Numerator Change for 5 Stars |
| **43** | missing\_points | Num | max\_trend\_wt - trend\_wt |
| **44** | adj\_star\_rating | Num | [Not for Stars Incentive Program] |
| **45** | adj\_aggr\_difference | Num | [Not for Stars Incentive Program] |
| **46** | needed | Num | Numerator Change to get to next Star Rating |
| **47** | gap\_worth | Num | The worth of a gap towards changing the Star Rating |