

## DRAFT SQL Code - Local Food Infrastructure and Population Health - Victor Russak

```
1  /*
2  WORK IN PROGRESS DRAFT
3  Analyst: Victor Russak
4  Email: vic@vicrussak.com
5  */
6
7  /*
8  EXTRACT AND LOAD PHASE
9      Utilize ELT methodology: Import raw data as TEXT with three tables; tables
10     will be cleaned/transformed as new tables in the data warehouse.
11     Goal: Create new tables and import data from the Food Environment Atlas (my
12     alias: FEA) and NYS Community Health Indicators Report (CHIR).
13 */
14 -- Create Food Environment Atlas Access Table.
15 CREATE TABLE raw_fea_access (
16     fips TEXT,
17     "state" TEXT,
18     county TEXT,
19     laccess_pop15 TEXT,
20     laccess_pop19 TEXT,
21     pch_laccess_pop_15_19 TEXT,
22     pct_laccess_pop15 TEXT,
23     pct_laccess_pop19 TEXT,
24     laccess_lowi15 TEXT,
25     laccess_lowi19 TEXT,
26     pch_laccess_lowi_15_19 TEXT,
27     pct_laccess_lowi15 TEXT,
28     pct_laccess_lowi19 TEXT,
29     laccess_hhmv15 TEXT,
30     laccess_hhmv19 TEXT,
31     pch_laccess_hhmv_15_19 TEXT,
32     pct_laccess_hhmv15 TEXT,
33     pct_laccess_hhmv19 TEXT,
34     laccess_snap15 TEXT,
35     laccess_snap19 TEXT,
36     pch_laccess_snap_15_19 TEXT,
37     pct_laccess_snap15 TEXT,
38     pct_laccess_snap19 TEXT,
39     laccess_child15 TEXT,
40     laccess_child19 TEXT,
41     pch_laccess_child_15_19 TEXT,
42     pct_laccess_child15 TEXT,
43     pct_laccess_child19 TEXT,
44     laccess_seniors15 TEXT,
45     laccess_seniors19 TEXT,
46     pch_laccess_seniors_15_19 TEXT,
47     pct_laccess_seniors15 TEXT,
```

```
48 pct_laccess_seniors19 TEXT,
49 laccess_white15 TEXT,
50 laccess_white19 TEXT,
51 pch_laccess_white_15_19 TEXT,
52 pct_laccess_white15 TEXT,
53 pct_laccess_white19 TEXT,
54 laccess_black15 TEXT,
55 laccess_black19 TEXT,
56 pch_laccess_black_15_19 TEXT,
57 pct_laccess_black15 TEXT,
58 pct_laccess_black19 TEXT,
59 laccess_hisp15 TEXT,
60 laccess_hisp19 TEXT,
61 pch_laccess_hisp_15_19 TEXT,
62 pct_laccess_hisp15 TEXT,
63 pct_laccess_hisp19 TEXT,
64 laccess_asian15 TEXT,
65 laccess_asian19 TEXT,
66 pch_laccess_asian_15_19 TEXT,
67 pct_laccess_asian15 TEXT,
68 pct_laccess_asian19 TEXT,
69 laccess_aian15 TEXT,
70 laccess_aian19 TEXT,
71 pch_laccess_aian_15_19 TEXT,
72 pct_laccess_aian15 TEXT,
73 pct_laccess_aian19 TEXT,
74 laccess_hopi15 TEXT,
75 laccess_hopi19 TEXT,
76 pch_laccess_hopi_15_19 TEXT,
77 pct_laccess_hopi15 TEXT,
78 pct_laccess_hopi19 TEXT,
79 laccess_multir15 TEXT,
80 laccess_multir19 TEXT,
81 pch_laccess_multir_15_19 TEXT,
82 pct_laccess_multir15 TEXT,
83 pct_laccess_multir19 TEXT
84
85 );
86
87 -- Create Community Health Indicators County Trends Table.
88 CREATE TABLE raw_chir_county_trends (
89
90 topic_area TEXT,
91 indicator_title TEXT,
92 geographic_area TEXT,
93 year_type TEXT,
94 "year" TEXT,
95 measurement TEXT,
96 rate_or_percent TEXT,
97 data_source TEXT,
98 data_notes TEXT
99
100 );
101
```

```
102 -- Create the Community Health Indicators Latest Data table as a helper table. This
103 table includes the indicator_numbers which can be used as a shortened primary key (3-5
104 characters instead of 50-80 characters) later on.
105 CREATE TABLE raw_chir_latest_data (
106     county_name TEXT,
107     health_topic_number TEXT,
108     health_topic TEXT,
109     indicator_number TEXT,
110     "indicator" TEXT,
111     event_count TEXT,
112     average_number_of_denominator TEXT,
113     measure_unit TEXT,
114     percent_or_rate TEXT,
115     lower_limit_of_95pct_ci TEXT,
116     upper_limit_of_95pct_ci TEXT,
117     data_comments TEXT,
118     quartile TEXT,
119     data_years TEXT,
120     data_source TEXT,
121     mapping_distribution TEXT,
122     "location" TEXT
123 );
124
125 -- Create nysc_fips table to import a master list of FIPS codes.
126 CREATE TABLE raw_us_fips (
127     state_name TEXT,
128     county_name TEXT,
129     city_name TEXT,
130     state_code TEXT,
131     state_fips_code TEXT,
132     county_code TEXT,
133     stcnty_fips_code TEXT,
134     city_code TEXT,
135     stcntycity_fips_code TEXT
136 );
137
138 --Import raw .csv files for the four tables. The .csv file exports from the Economic
139 Research Service and the NYS Dept. of Health are encoded in WIN1252.
```

```

140 COPY raw_fea_access(fips, "state", county, laccess_pop15, laccess_pop19,
    pch_laccess_pop_15_19, pct_laccess_pop15, pct_laccess_pop19, laccess_lowi15,
    laccess_lowi19, pch_laccess_lowi_15_19, pct_laccess_lowi15, pct_laccess_lowi19,
    laccess_hhnv15, laccess_hhnv19, pch_laccess_hhnv_15_19, pct_laccess_hhnv15,
    pct_laccess_hhnv19, laccess_snap15, laccess_snap19, pch_laccess_snap_15_19,
    pct_laccess_snap15, pct_laccess_snap19, laccess_child15, laccess_child19,
    pch_laccess_child_15_19, pct_laccess_child15, pct_laccess_child19, laccess_seniors15,
    laccess_seniors19, pch_laccess_seniors_15_19, pct_laccess_seniors15,
    pct_laccess_seniors19, laccess_white15, laccess_white19, pch_laccess_white_15_19,
    pct_laccess_white15, pct_laccess_white19, laccess_black15, laccess_black19,
    pch_laccess_black_15_19, pct_laccess_black15, pct_laccess_black19, laccess_hisp15,
    laccess_hisp19, pch_laccess_hisp_15_19, pct_laccess_hisp15, pct_laccess_hisp19,
    laccess_asian15, laccess_asian19, pch_laccess_asian_15_19, pct_laccess_asian15,
    pct_laccess_asian19, laccess_aian15, laccess_aian19, pch_laccess_aian_15_19,
    pct_laccess_aian15, pct_laccess_aian19, laccess_hopi15, laccess_hopi19,
    pch_laccess_hopi_15_19, pct_laccess_hopi15, pct_laccess_hopi19, laccess_multir15,
    laccess_multir19, pch_laccess_multir_15_19, pct_laccess_multir15,
    pct_laccess_multir19)
141 FROM
142 'C:\Users\Public\fea_access.csv'
143 WITH
144 (FORMAT CSV, HEADER TRUE, DELIMITER ',', ENCODING 'WIN1252');
145
146 COPY raw_chir_county_trends(topic_area, indicator_title, geographic_area, year_type,
    "year", measurement, rate_or_percent, data_source, data_notes)
147 FROM
148 'C:\Users\Public\chir_county_trend.csv'
149 WITH
150 (FORMAT CSV, HEADER TRUE, DELIMITER ',', ENCODING 'WIN1252');
151
152 COPY raw_chir_latest_data(county_name, health_topic_number, health_topic,
    indicator_number, "indicator", event_count, average_number_of_denominator,
    measure_unit, percent_or_rate, lower_limit_of_95pct_ci, upper_limit_of_95pct_ci,
    data_comments, quartile, data_years, data_source, mapping_distribution, "location")
153 FROM
154 'C:\Users\Public\Community_Health_Indicator_Reports__CHIRS__Latest_Data_20250820.csv'
155 WITH
156 (FORMAT CSV, HEADER TRUE, DELIMITER ',', ENCODING 'WIN1252');
157
158 COPY raw_us_fips(state_name, county_name, city_name, state_code, state_fips_code,
    county_code, stcnty_fips_code, city_code, stcntycity_fips_code)
159 FROM
160 'C:\Users\Public\State_County_and_City_FIPS_Reference_Table_20260119.csv'
161 WITH
162 (FORMAT CSV, HEADER TRUE, DELIMITER ',', ENCODING 'WIN1252');
163
164 -- Runtime of script for the EXTRACT AND LOAD PHASE: 6.043 seconds
165
166 /* DATA EXPLORATION PHASE
167 Goal 1: Check that the data was loaded correctly.
168 Goal 2: Check Data Years in both tables to find common years between them.
169 */
170

```

```

171 -- Check fips codes were imported correctly (no loss of leading zeros) and there are
172 -- no entry errors.
173 SELECT fips, COUNT(*)
174 FROM raw_fea_access
175 WHERE LENGTH(fips) != 5 OR fips !~ '^[0-9]+$'
176 GROUP BY fips
177 ORDER BY fips; -- Found 318 instances of lost leading zeros. Will need to address in
178 -- transformation.
179
180 -- Data years in Food Environment Atlas are in the column names.
181 SELECT *
182 FROM raw_fea_access
183 LIMIT 0 -- Result: raw_fea_access only contains 2015 and 2019 data.
184
185 --Checking if CHIR has 2015 and 2019 data.
186 SELECT "year"
187 FROM
188     raw_chir_county_trends
189 WHERE
190     "year" = '2015' OR
191     "year" = '2019'
192 GROUP BY
193     "year"
194 ORDER BY
195     "year" ASC; -- Result: CHIR has both 2015 and 2019. This analysis will compare
196 -- both years.
197
198 /* TRANSFORMATION PHASE
199     Goal 1: Create a clean fea_access19 table (2019 data) and a clean fea_access15
200     (2015 data) table from the raw table.
201     Goal 2: Create a clean chir_county_trends table joined with chir_latest_data
202     so that the Indicator Number can be used as primary key.
203     Goal 3: Add FIPS codes to the new CHIR table for use as a foreign key.
204 */
205
206 --Create clean fea_access2019 table from the raw table. Only bringing in population
207 -- counts and percentages, not percent change, nor demographic data.
208 CREATE TABLE fea_access19 AS
209 SELECT
210     CAST(LPAD(fips, 5, '0') AS VARCHAR(5)) AS fips, -- LPAD fixes the loss of leading
211     -- zeros due to passing the data through Excel. Constrains to 5 characters.
212     CAST("state" AS VARCHAR(2)) AS "state",
213     CAST(county AS VARCHAR(50)) AS county,
214     CASE -- Checks for alpha values due to possible entry error and checks for
215     -- null indicators '-9999' and '-8888' as specified in data documentation and changes
216     -- these to null.
217         WHEN laccess_pop19 ~ '^[0-9.]+' THEN CAST(laccess_pop19 AS NUMERIC)
218         ELSE NULL
219     END AS laccess_pop19,
220     CASE
221         WHEN pct_laccess_pop19 ~ '^[0-9.]+' THEN CAST(pct_laccess_pop19 AS
222     NUMERIC)
223         ELSE NULL
224     END AS pct_laccess_pop19,

```

```

215         CASE
216             WHEN laccess_lowi19 ~ '^[0-9.]+' THEN CAST(laccess_lowi19 AS NUMERIC)
217             ELSE NULL
218         END AS laccess_lowi19,
219         CASE
220             WHEN pct_laccess_lowi19 ~ '^[0-9.]+' THEN CAST(pct_laccess_lowi19 AS
NUMERIC)
221             ELSE NULL
222         END AS pct_laccess_lowi19,
223         CASE
224             WHEN laccess_hhmv19 ~ '^[0-9.]+' THEN CAST(laccess_hhmv19 AS NUMERIC)
225             ELSE NULL
226         END AS laccess_hhmv19,
227         CASE
228             WHEN pct_laccess_hhmv19 ~ '^[0-9.]+' THEN CAST(pct_laccess_hhmv19 AS
NUMERIC)
229             ELSE NULL
230         END AS pct_laccess_hhmv19,
231         CASE
232             WHEN laccess_snap19 ~ '^[0-9.]+' THEN CAST(laccess_snap19 AS NUMERIC)
233             ELSE NULL
234         END AS laccess_snap19,
235         CASE
236             WHEN pct_laccess_snap19 ~ '^[0-9.]+' THEN CAST(pct_laccess_snap19 AS
NUMERIC)
237             ELSE NULL
238         END AS pct_laccess_snap19
239 FROM raw_fea_access;
240
241 -- Add primary key
242 ALTER TABLE fea_access19 ADD PRIMARY KEY (fips);
243
244 -- How many null indicators were removed?
245 SELECT *
246 FROM
247     fea_access19
248 WHERE
249     laccess_pop19 IS NULL; -- Result: Three rows had null indicators: fips 02063,
02066, 02261.
250
251 --Create clean fea_access2015 table from the raw table. Only bringing in population
counts and percentages, not percent change, nor demographic data.
252 CREATE TABLE fea_access15 AS
253 SELECT
254     CAST(LPAD(fips, 5, '0') AS VARCHAR(5)) AS fips, -- LPAD fixes the loss of leading
zeros due to passing the data through Excel. Constrains to 5 characters.
255     CAST("state" AS VARCHAR(2)) AS "state",
256     CAST(county AS VARCHAR(50)) AS county,
257     CASE -- Checks for alpha values due to possible entry error and checks for
null indicators '-9999' and '-8888' as specified in data documentation and changes
these to null.
258         WHEN laccess_pop15 ~ '^[0-9.]+' THEN CAST(laccess_pop15 AS NUMERIC)
259         ELSE NULL
260     END AS laccess_pop15,

```

```

261         CASE
262             WHEN pct_laccess_pop15 ~ '^[0-9.]+$' THEN CAST(pct_laccess_pop15 AS
NUMERIC)
263             ELSE NULL
264         END AS pct_laccess_pop15,
265         CASE
266             WHEN laccess_lowi15 ~ '^[0-9.]+$' THEN CAST(laccess_lowi15 AS NUMERIC)
267             ELSE NULL
268         END AS laccess_lowi15,
269         CASE
270             WHEN pct_laccess_lowi15 ~ '^[0-9.]+$' THEN CAST(pct_laccess_lowi15 AS
NUMERIC)
271             ELSE NULL
272         END AS pct_laccess_lowi15,
273         CASE
274             WHEN laccess_hhnv15 ~ '^[0-9.]+$' THEN CAST(laccess_hhnv15 AS NUMERIC)
275             ELSE NULL
276         END AS laccess_hhnv15,
277         CASE
278             WHEN pct_laccess_hhnv15 ~ '^[0-9.]+$' THEN CAST(pct_laccess_hhnv15 AS
NUMERIC)
279             ELSE NULL
280         END AS pct_laccess_hhnv15,
281         CASE
282             WHEN laccess_snap15 ~ '^[0-9.]+$' THEN CAST(laccess_snap15 AS NUMERIC)
283             ELSE NULL
284         END AS laccess_snap15,
285         CASE
286             WHEN pct_laccess_snap15 ~ '^[0-9.]+$' THEN CAST(pct_laccess_snap15 AS
NUMERIC)
287             ELSE NULL
288         END AS pct_laccess_snap15
289 FROM raw_fea_access;
290
291 -- Add primary key
292 ALTER TABLE fea_access15 ADD PRIMARY KEY (fips);
293
294 -- How many null indicators were removed?
295 SELECT *
296 FROM
297     fea_access15
298 WHERE
299     laccess_pop15 IS NULL; -- Result: Two rows had null indicators: fips 02063,
02066.
300
301 -- Create clean us_fips master table with state and county FIPS only.
302 CREATE TABLE us_fips AS
303 SELECT
304     state_name,
305     county_name,
306     CAST(state_code AS VARCHAR(2)) AS state_code,
307     CAST(LPAD( state_fips_code, 2, '0') AS VARCHAR(2)) AS state_fips_code, --
LPAD fixes the loss of leading zeros due to passing the data through Excel. Constrains
to 5 characters.

```

```

308         county_code,
309         CAST(LPAD(stcnty_fips_code, 5, '0') AS VARCHAR(5)) AS stcnty_fips_code -- LPAD
fixes the loss of leading zeros due to passing the data through Excel. Constrains to 5
characters.
310 FROM
311     raw_us_fips
312 WHERE
313     stcnty_fips_code IS NOT NULL -- Puerto Rico does not have a county code and is
not needed for this analysis.
314 GROUP BY -- removing duplicate stcnty_fips so it can act as primary key
315     state_name,
316     county_name,
317     state_code,
318     state_fips_code,
319     county_code,
320     stcnty_fips_code;
321
322 ALTER TABLE us_fips ADD PRIMARY KEY (stcnty_fips_code);
323
324 --Create master table for the Indicator Numbers in raw_chir_latest_data. Will use
later for join and primary key creation.
325 CREATE TABLE indicator_master AS
326 SELECT
327     indicator_number,
328     "indicator"
329 FROM
330     raw_chir_latest_data
331 GROUP BY
332     indicator_number,
333     "indicator";
334
335 ALTER TABLE indicator_master ADD PRIMARY KEY (indicator_number);
336
337 -- Create chir_trends_1519 table and join county FIPS codes and Indicator Numbers from
helper tables.
338 CREATE TABLE chir_trends_1519 AS
339 SELECT DISTINCT ON (us.stcnty_fips_code, im.indicator_number, raw.year) -- The PK
Columns
340     raw.topic_area AS topic_area,
341     UPPER(TRIM(REPLACE(REPLACE(raw.geographic_area, 'County', ''), '.', ''))) AS
geographic_area, -- Removes the string "County" and '.' from the county names to match
FIPS master table.
342     raw.year,
343     raw.measurement,
344     CASE -- Checks for alpha values due to possible entry error and checks for
null indicators '-9999' and '-8888' as specified in data documentation and changes
these to null.
345         WHEN raw.rate_or_percent ~ '^[0-9.]+$' THEN CAST(raw.rate_or_percent
AS NUMERIC)
346         ELSE NULL
347         END AS rate_or_percent,
348     'NY' AS state_code, -- adding state code so FIPS can be joined on state and
county names
349     us.stcnty_fips_code,

```

```
350         im.indicator_number,
351         im.indicator
352 FROM
353     raw_chir_county_trends AS raw
354 LEFT JOIN us_fips AS us
355     ON us.state_code = 'NY'
356     AND UPPER(TRIM(REPLACE(REPLACE(raw.geographic_area, 'County', ''), '.', '')))
357 = TRIM(us.county_name)
358 LEFT JOIN indicator_master AS im
359     ON TRIM(im.indicator) = TRIM(raw.indicator_title)
360 WHERE
361     "year" IN ('2015', '2019') AND
362     topic_area IN ('Cardiovascular Disease Indicators', 'Cirrhosis, Diabetes, and
363     Kidney Indicators') -- Only including relevant chronic disease indicators. Excluding
364     Obesity Topic Area since the general population obesity indicators exist as duplicates
365     in the Cardiovascular topic area.
366 ORDER BY
367     us.stcnty_fips_code,
368     im.indicator_number,
369     raw.year,
370     raw.topic_area DESC -- Drops duplicates from the Cardiovascular topic area and
371     keeps them in Cirrhosis since all three duplicates are related to kidney disease
372 ;
373 ALTER TABLE chir_trends_1519 ADD PRIMARY KEY("year", stcnty_fips_code,
374 indicator_number);
375
376 /*
377
378 Next Steps:
379 I've successfully created clean copies of chir_trends_1519 which contains 2015 and
380 2019 data
381 for the topic areas Cardiovascular and Kidney Disease. These include Obesity
382 Indicators.
383
384 I also successfully created fea_access15 and fea_access19.
385
386 I now need to create a table that combines chir_trends_1519 with fea_access15 and
387 fea_access19.
388 The fea tables are in Wide format and will need to be converted to long format before
389 continuing.
390
391 */
```