DATA OPS CHALLENGE By Will Burke

APPROACH AND **TOOLS**







DATA CLEANING STEPS

01 - GAINING INSIGHTS

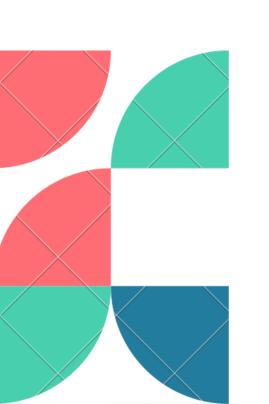
Getting summaries and descriptive information to reveal inconsistencies

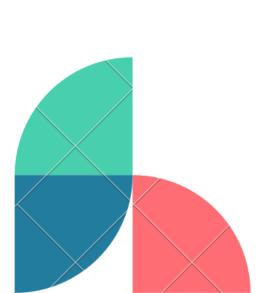
02 - FIXING INCONSISTENCIES

Fixing inconsistencies across tables before merging.

03 - SHAPING DATA

Shaping data to fit requested deliverable.





PRODUCING MERGED TABLE

```
# Creating table called "std_member_info" (takes a second)
cursor.execute("DROP TABLE IF EXISTS std_member_info")

# Ordering columns for EACH table because UNION considers column order, NOT column name
colNames = "Person_Id, First_Name, Last_Name, Dob, Age, Gender, Street_Address, State, C
r1_str = "SELECT " + colNames + " FROM roster_1"
r2_str = "SELECT " + colNames + " FROM roster_2_fixed"
r3_str = "SELECT " + colNames + " FROM roster_3"
r4_str = "SELECT " + colNames + " FROM roster_4"
r5_str = "SELECT " + colNames + " FROM roster_5"

# Executing query
query = "CREATE TABLE std_member_info AS " + r1_str + " UNION " + r2_str + " UNION " + r
result = cursor.execute(query)
```

QUERY RESULTS

- How many distinct members are eligible in April 2022?
 There are 99807 distinct members eligible in April 2022.
- How many members were included more than once?
 There were 10845 members included more than once.
- What is the breakdown of members by payer?
 There were two payers: Madv and Mdcd. Madv has 37352 members, with an average age of about 50 years. Mdcd has 62455 members, also with an average age of about 50 years.
- How many members live in a zip code with a food_access_score lower than 2?
 There are 7685 members as described above
- What is the average social isolation score for the members?

 The average social isolation score for the members described above is 2.98, compared to an average of 3.07 across the whole dataset.
- Which members live in the zip code with the highest algorex_sdoh_composite_score?

 45 members in zip code 95950

SCALING + ONWARDS

Imagine receiving a roster every week from this client. How would your script need to change?

- Completely generalized code to reduce coding time.
- Pre-defined script for searching for inconsistencies.

