

Considerations -

Game Table - Schedule\_week has playoff, extended season, superbowl, numbers, etc… we left the values in so the column is int and strings, but we can adjust as you needed. We decided to go this way as we assumed you will likely be doing analysis based on schedule\_week where that will be your category.

NFL stadium surface – The stadium surface contains values such as the following

“Grass, Turf (1969-1970)”

“FieldTurf”

“Grass”

“Hellas Matrix Turf”

On top of these there are values that are empty. We decided to split “Grass, Turf …” at the comma and leave just Grass behind. All other values are unchanged.

Another issue while importing the game information was the fact the team acronym was listed, we can’t tie this back to an individual team very reliably as there are the same abbreviation with a different team name. We took the team\_name that was used most often over the last year.

NFL Teams – Raiders had the wrong acronym and so did Jacksonville. We updated them both.

Staging table – We created a staging table to allow for the ingestion of information more easily and quickly. This is a temporary table that will only be created at the beginning of our replication and removed once we are done.

Customer Table – Just so you are aware there are customers with the same first name and last names in your database but we implimented ways to prevent that from affect our manipulation.

Bet Amounts -

A screenshot of a computer program

Description automatically generated

The bet amounts have been moved over as they were specified, but it seems we likely have a substantial data quality issue with them. The amount many customers are betting is substantially more than their income. We have moved all the data over as requested but just so you are aware this issue is present.

Logging –

Instead of using print statements we have implemented logging within our pipeline, this will allow us to see errors arise and investigate them more easily. We also have the ability to adjust the level of logging.

Credentials –

The database credentials are in a .env file for added security while we are accessing, we have provided this.

Schema –

We have created the ability to have a second set of the data in the database in a different location if you would like for us to experiment or make any changes.