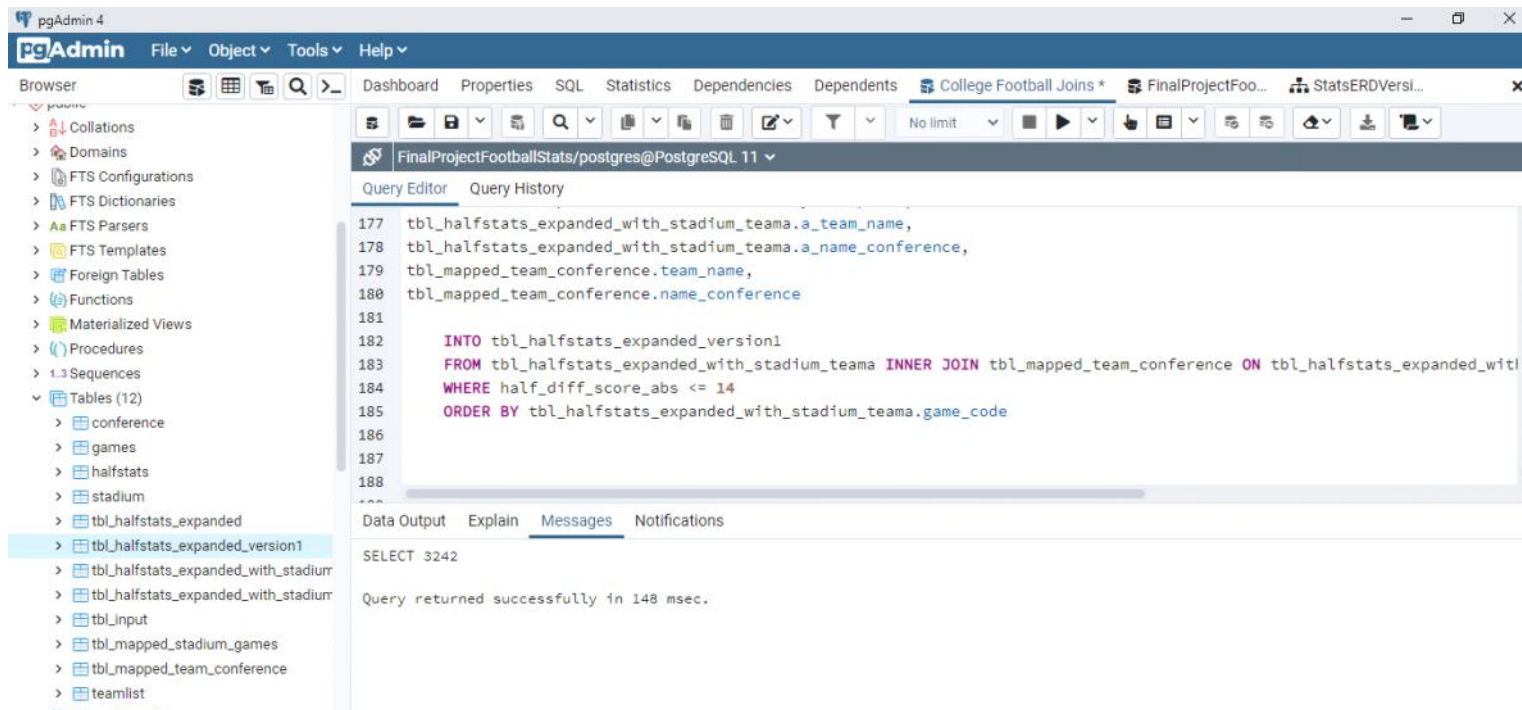


## SQL/Database work

Tuesday, February 01, 2022 5:26 PM

Database was created in PGAdmin  
There are multiple tables in the database

- 1) Conference (records:27), games distinct (Records: 4918), stadium (records:166) and teamlist (records:247) are direct from Kaggle 2008-2013 (with small adjustments to conferences that may have changed between those five years)
- 2) Halfstats (records: 9590) was imported from game statistics to align with what is needed for machine learning
- 3) Halfstats expanded (records 4795) was imported from csv, it was created from halfstats (in excel) to place Team A and Team B within a distinct primary key = game\_code.--and was matched to tbl\_mapped\_stadium\_games (records: 4918).
- 4) Mapped stadium games and team conference was required to turn each game code into meaningful team names, stadium names and conference names.
- 5) Halfstats expanded version 1 is the gold copy expanded form of data to be used for machine learning and visuals. Where clauses can be created to trim anomalies (or not) as the team proceeds.
  - a. Gold Copy Half Stats trimmed to halftime score <=14 points (records 3248)
- 6) Table structure below (see ERD in folder)



Steps in PGADMIN for joining data

--Step 1 Bring in normalized and distinct conference name into table = conference  
--Step 2 Create a table that maps conference against teamlist conference code (tbl\_mapped\_team\_conference)

```
SELECT
teamlist.team_code,
teamlist.team_name,
teamlist.conference_code,
conference.name_conference,
conference.subdivision
INTO mapped_team_conference
FROM teamlist INNER JOIN conference ON teamlist.conference_code=conference.conference_code;
```

--Step 3 Bring in all games from 2008-2013  
--Step 4 Bring in stadium codes that dropped since 2008, 123 North Tex and 26-Ft Laud  
--Step 5 Right Join Stadium Name from Stadium.stadium\_code to games.stadium code

```
SELECT
games.game_code,
games.visit_date,
games.stadium_code,
stadium.stadium_name,
games.home_team_code,
games.visit_team_code,
games.site,
games.year,
stadium.city,
stadium.state,
stadium.capacity,
stadium.surface,
stadium.year_opened,
```

```

stadium.surfaceint
INTO tbl_mapped_stadium_games
FROM stadium RIGHT JOIN games ON stadium.stadium_code=games.stadium_code
ORDER BY games.game_code,home_team_code;

```

--Step 6 input halfstats\_expanded where Team a and Team b are on distinct game\_code  
--Step 7 link halfstats\_expanded to mapped stadium games in order to get stadium details

```

SELECT
tbl_halfstats_expanded.year,
tbl_halfstats_expanded.game_code,
tbl_halfstats_expanded.a_team_code,
tbl_halfstats_expanded.b_team_code,
tbl_halfstats_expanded.a_result,
tbl_halfstats_expanded.b_result,
tbl_halfstats_expanded.a_points,
tbl_halfstats_expanded.b_points,
tbl_halfstats_expanded.a_rush,
tbl_halfstats_expanded.b_rush,
tbl_halfstats_expanded.a_offense_rush_yards,
tbl_halfstats_expanded.b_defense_rush_yards,
tbl_halfstats_expanded.a_defense_rush_yards,
tbl_halfstats_expanded.b_offense_rush_yards,
tbl_halfstats_expanded.a_offense_pass_yards,
tbl_halfstats_expanded.b_defense_pass_yards,
tbl_halfstats_expanded.a_defense_pass_yards,
tbl_halfstats_expanded.b_offense_pass_yards,
tbl_halfstats_expanded.a_pass,
tbl_halfstats_expanded.b_pass,
tbl_halfstats_expanded.a_incomplete_pass,
tbl_halfstats_expanded.b_incomplete_pass,
tbl_halfstats_expanded.a_penalty,
tbl_halfstats_expanded.b_penalty,
tbl_halfstats_expanded.a_penalty_yards,
tbl_halfstats_expanded.b_penalty_yards,
tbl_halfstats_expanded.half_diff_score,
tbl_halfstats_expanded.half_diff_score_abs,
tbl_halfstats_expanded.half_leader,
tbl_halfstats_expanded.hometeam,
tbl_halfstats_expanded.b_home,
tbl_mapped_stadium_games.visit_date,
tbl_mapped_stadium_games.stadium_code,
tbl_mapped_stadium_games.stadium_name,
tbl_mapped_stadium_games.site,
tbl_mapped_stadium_games.city,
tbl_mapped_stadium_games.state,
tbl_mapped_stadium_games.capacity,
tbl_mapped_stadium_games.surface,
tbl_mapped_stadium_games.year_opened,
tbl_mapped_stadium_games.surfaceint
INTO tbl_halfstats_expanded_with_stadium
FROM tbl_halfstats_expanded RIGHT JOIN tbl_mapped_stadium_games ON
tbl_halfstats_expanded.game_code=tbl_mapped_stadium_games.game_code
ORDER BY tbl_halfstats_expanded.game_code

```

--Step 8 Link halfstats with stadium for team and conferences for team a and team b

```

SELECT
tbl_halfstats_expanded_with_stadium.year,
tbl_halfstats_expanded_with_stadium.game_code,
tbl_halfstats_expanded_with_stadium.a_team_code,
tbl_halfstats_expanded_with_stadium.b_team_code,
tbl_halfstats_expanded_with_stadium.a_result,
tbl_halfstats_expanded_with_stadium.b_result,
tbl_halfstats_expanded_with_stadium.a_points,
tbl_halfstats_expanded_with_stadium.b_points,
tbl_halfstats_expanded_with_stadium.a_rush,
tbl_halfstats_expanded_with_stadium.b_rush,
tbl_halfstats_expanded_with_stadium.a_offense_rush_yards,
tbl_halfstats_expanded_with_stadium.b_defense_rush_yards,
tbl_halfstats_expanded_with_stadium.a_defense_rush_yards,
tbl_halfstats_expanded_with_stadium.b_offense_rush_yards,
tbl_halfstats_expanded_with_stadium.a_offense_pass_yards,
tbl_halfstats_expanded_with_stadium.b_defense_pass_yards,
tbl_halfstats_expanded_with_stadium.a_defense_pass_yards,
tbl_halfstats_expanded_with_stadium.b_offense_pass_yards,
tbl_halfstats_expanded_with_stadium.a_pass,
tbl_halfstats_expanded_with_stadium.b_pass,
tbl_halfstats_expanded_with_stadium.a_incomplete_pass,
tbl_halfstats_expanded_with_stadium.b_incomplete_pass,
tbl_halfstats_expanded_with_stadium.a_penalty,
tbl_halfstats_expanded_with_stadium.b_penalty,
tbl_halfstats_expanded_with_stadium.a_penalty_yards,
tbl_halfstats_expanded_with_stadium.b_penalty_yards,
tbl_halfstats_expanded_with_stadium.half_diff_score,
tbl_halfstats_expanded_with_stadium.half_diff_score_abs,
tbl_halfstats_expanded_with_stadium.half_leader,
tbl_halfstats_expanded_with_stadium.hometeam,
tbl_halfstats_expanded_with_stadium.b_home,
tbl_halfstats_expanded_with_stadium.visit_date,
tbl_halfstats_expanded_with_stadium.stadium_code,
tbl_halfstats_expanded_with_stadium.stadium_name,
tbl_halfstats_expanded_with_stadium.site,
tbl_halfstats_expanded_with_stadium.city,
tbl_halfstats_expanded_with_stadium.state,
tbl_halfstats_expanded_with_stadium.capacity,
tbl_halfstats_expanded_with_stadium.surface,

```

```
tbl_halfstats_expanded_with_stadium.surfaceint,
tbl_halfstats_expanded_with_stadium.year_opened,
tbl_mapped_team_conference.team_name,
tbl_mapped_team_conference.name_conference

    INTO tbl_halfstats_expanded_with_stadium_teama
FROM tbl_halfstats_expanded_with_stadium INNER JOIN tbl_mapped_team_conference ON
tbl_halfstats_expanded_with_stadium.a_team_code=tbl_mapped_team_conference.team_code
ORDER BY tbl_halfstats_expanded_with_stadium.game_code

SELECT
tbl_halfstats_expanded_with_stadium_teama.year,
tbl_halfstats_expanded_with_stadium_teama.game_code,
tbl_halfstats_expanded_with_stadium_teama.a_team_code,
tbl_halfstats_expanded_with_stadium_teama.b_team_code,
tbl_halfstats_expanded_with_stadium_teama.a_result,
tbl_halfstats_expanded_with_stadium_teama.b_result,
tbl_halfstats_expanded_with_stadium_teama.a_points,
tbl_halfstats_expanded_with_stadium_teama.b_points,
tbl_halfstats_expanded_with_stadium_teama.a_rush,
tbl_halfstats_expanded_with_stadium_teama.b_rush,
tbl_halfstats_expanded_with_stadium_teama.a_offense_rush_yards,
tbl_halfstats_expanded_with_stadium_teama.b_defense_rush_yards,
tbl_halfstats_expanded_with_stadium_teama.a_defense_rush_yards,
tbl_halfstats_expanded_with_stadium_teama.b_offense_rush_yards,
tbl_halfstats_expanded_with_stadium_teama.a_offense_pass_yards,
tbl_halfstats_expanded_with_stadium_teama.b_defense_pass_yards,
tbl_halfstats_expanded_with_stadium_teama.a_defense_pass_yards,
tbl_halfstats_expanded_with_stadium_teama.b_offense_pass_yards,
tbl_halfstats_expanded_with_stadium_teama.a_pass,
tbl_halfstats_expanded_with_stadium_teama.b_pass,
tbl_halfstats_expanded_with_stadium_teama.a_incomplete_pass,
tbl_halfstats_expanded_with_stadium_teama.b_incomplete_pass,
tbl_halfstats_expanded_with_stadium_teama.a_penalty,
tbl_halfstats_expanded_with_stadium_teama.b_penalty,
tbl_halfstats_expanded_with_stadium_teama.a_penalty_yards,
tbl_halfstats_expanded_with_stadium_teama.b_penalty_yards,
tbl_halfstats_expanded_with_stadium_teama.half_diff_score,
tbl_halfstats_expanded_with_stadium_teama.half_diff_score_abs,
tbl_halfstats_expanded_with_stadium_teama.half_leader,
tbl_halfstats_expanded_with_stadium_teama.hometeam,
tbl_halfstats_expanded_with_stadium_teama.b_home,
tbl_halfstats_expanded_with_stadium_teama.visit_date,
tbl_halfstats_expanded_with_stadium_teama.stadium_code,
tbl_halfstats_expanded_with_stadium_teama.stadium_name,
tbl_halfstats_expanded_with_stadium_teama.site,
tbl_halfstats_expanded_with_stadium_teama.city,
tbl_halfstats_expanded_with_stadium_teama.state,
tbl_halfstats_expanded_with_stadium_teama.capacity,
tbl_halfstats_expanded_with_stadium_teama.surface,
tbl_halfstats_expanded_with_stadium_teama.surfaceint,
tbl_halfstats_expanded_with_stadium_teama.year_opened,
tbl_halfstats_expanded_with_stadium_teama.a_team_name,
tbl_halfstats_expanded_with_stadium_teama.a_name_conference,
tbl_mapped_team_conference.team_name,
tbl_mapped_team_conference.name_conference

    INTO tbl_halfstats_expanded_version1 (aka goldcopyhalfstats.csv)
FROM tbl_halfstats_expanded_with_stadium_teama INNER JOIN tbl_mapped_team_conference ON
tbl_halfstats_expanded_with_stadium_teama.b_team_code=tbl_mapped_team_conference.team_code
WHERE half_diff_score_abs <= 14
ORDER BY tbl_halfstats_expanded_with_stadium_teama.game_code
```

SQLite/Python/VScode		
Created database initial_model_input.sqlite	Table 1: initial_model_input	Can be used for baseline predictive model inputs (from goldcopy/kaggle 2008/2013)
	Table 2: user_input_java_winner	Can be used to store predicted winners for the model