**College Football Statistics**

College football Saturdays are some of the bests of the year. I choose this topic because I thought it would be nice to have a database of all college football stats for the last few years. Derived from sources is incorporated data of team stats from 2018 through 2020. I felt that having a non-relational database was the best for this dataset.

**Sources (Extract):**

The data comes from the NCAA. The source csv files came from Kaggle.com.

* <https://www.kaggle.com/jeffgallini/college-football-team-stats-2019>
* <https://www.kaggle.com/mhixon/college-football-statistics>

**Transform:**

1. Added years for each file.

2. Added unique IDs.

3. Split Conferences from their respective Teams.

4. Update column names.

5. Append each file to each other.

**Load:**

The tables loaded into pgAdmin from the original CSVs were:

- 'offense\_passing'

- 'defense\_passing'

- 'penalties'

- 'offense\_punts'

- 'defense\_punts'

- 'offense\_redzone'

- 'defense\_redzone'

- 'offense\_rushing'

- 'defense\_rushing'

- 'sacks'

- 'turnovers'

- 'win\_loss'

- 'offense'

- 'defense'

- 'offense\_firstdowns'

- 'defense\_firstdowns'

- 'offense\_fourthdowns'

- 'defense\_fourthdowns'

- 'offense\_kickoffs'

- 'defense\_kickoffs'

Example Tables:

offense\_firstdowns

Table

Description automatically generated

penalties

Table

Description automatically generated

The above tables were chosen as I believed these were the best way to group the available data. I broke them down this way because I thought that they were the right size without overloading each table. I believe these give the user querying the database the best results when comparing teams and their respective stats.