**Final Report: ETL Project**

Project Members: Max Patterson, Anastasios Tomazos, Erin Wills

Abstract

The objective of this project is to first *extract* National Football League (“NFL”) data from at least two different sources, then *transform* it into a comprehensible and usable format, followed by *loading* the data into a central database. Although there seems to be numerous sources for NFL data, the dearth of free and publicly accessible data sources is a major stumbling block for the interested fanbase, like us, in performing significant football analytics. Be that as it may, we found NFL data from two different sources with which we created useful databases comprised of player and gameday data for the 2009 – 2013 seasons.

Breakdown of Tasks

The overall ETL process is comprised of Extract, Transform, and Load subprocesses.

*Extract*

Our data sources are extracted from the **nflscrapR-data** repository and **NFLsavant.com**. **nflscrapR-data** was created by a group of [Carnegie Mellon University statistical researchers](http://www.stat.cmu.edu/) who publicly released this data**,** which is an R package that “uses an API maintained by the NFL to scrape, clean, parse, and output clean datasets at the individual play, player, game, and season levels.”[[1]](#footnote-1) NFLsavant.com is a web site that obtains publicly-available NFL play-by-play data on the internet.

Our plan is to create the following two databases, which includes:

* Player data from the nflscrapR-data and append to this dataset the NFLsavant player information to include college information
* Take game data from nflscrapR-data and append NFLsavant weather data

**nflscrapR-data:**

CSV files: Gameday information from 2009-2013

Player rosters from 2009-2013

**NFLsavant.com**:

CSV files: [CSV of All NFL Weather From 1960 to 2013](http://nflsavant.com/dump/weather_20131231.csv)

CSV files: player info from 2009-2013

*Transform*

*Load*

1. See, https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016 [↑](#footnote-ref-1)