ETL Technical Report
Ray Hearne & Howard Bartee, Jr.

April 27th, 2020

Introduction

In this project, due to social distancing and the virtual upcoming NFL Draft, we have been tasked to evaluate the First Round Draft Picks over the last 10 years and determine what positions the players played, what colleges they attended, and the individual conference of those individuals selected. A ReadMe File was created to discussing this information as well regarding the process we used in this project on the NFL Draft for 2020. Data has been obtained from various sites, including www.nfl.com and www.profootballreference.com to help develop the data set for this process. The data was migrated to a production data base called PGMADMIN.

The data on this database makes it possible for teams to evaluate and make a determination on draft picks for the first-ever virtual draft in 2020. Ultimately, we are looking at the trends on how team are drafting from the Power 5 Conferences as well as the Non Power 5 Conferences.

In the Data Cleanup & Analysis, we did the following as we performed the ETL Process.

Extract

The sources of data was extracted from www.nfl.com and www.profootballreference.com.

Transform

After the data was extracted into 11 Excel Files from 2009 to 2019, they were joined by merging them into one Master CSV File called NFLDraft20092010_ETL.csv. Dataframes were created, columns were cleaned, and then data was filtered by 5 compoonents including Draft Year, Round, Player, Position, and Conference.

Load

Next, we decided to use Jupyter Notebook by creating new dataframes and then used PGAdmin Database to ultimately create tables.

The final tables that we used in the production database include the total number of 1st Round Draft Picks in the SEC Conference, which had the highest number of draft picks selected in this round from 2009-2019. Additional tables can be ascertained from this the database upon request.

Conclusion

As a result of this ETL project, we will be able to provide this information for the teams to use in preparation for the Virtual Draft 2020. As part of the task assigned, this concluded our final technical report additionally with the above information and steps required to reproduce our ETL process. It is saved in Github as hbarteejr/ETL-Project-2020.