**BAIS:3200**

**Final Project Report**

Group 12: Hikaru Carolin, Alexa Martin, Ozan Dogan, Reed Foster, Jenna Grant

**Introduction**

The NFL Draft took place last week, and previous datasets can help predict how a player will perform based on when they were drafted. The NFL Draft is important for NFL teams because it allows the teams to bring new talent and opportunities into their roster. For this project, our group gathered data from NFL Drafts between 2000 and 2015 in order to understand past patterns and trends of the NFL Draft. Our project will also analyze the most successful members of the drafts, and if the team they were drafted to has an impact on their success. This analysis can be helpful for NFL players, NFL teams, and football enthusiasts looking for more information about the draft.

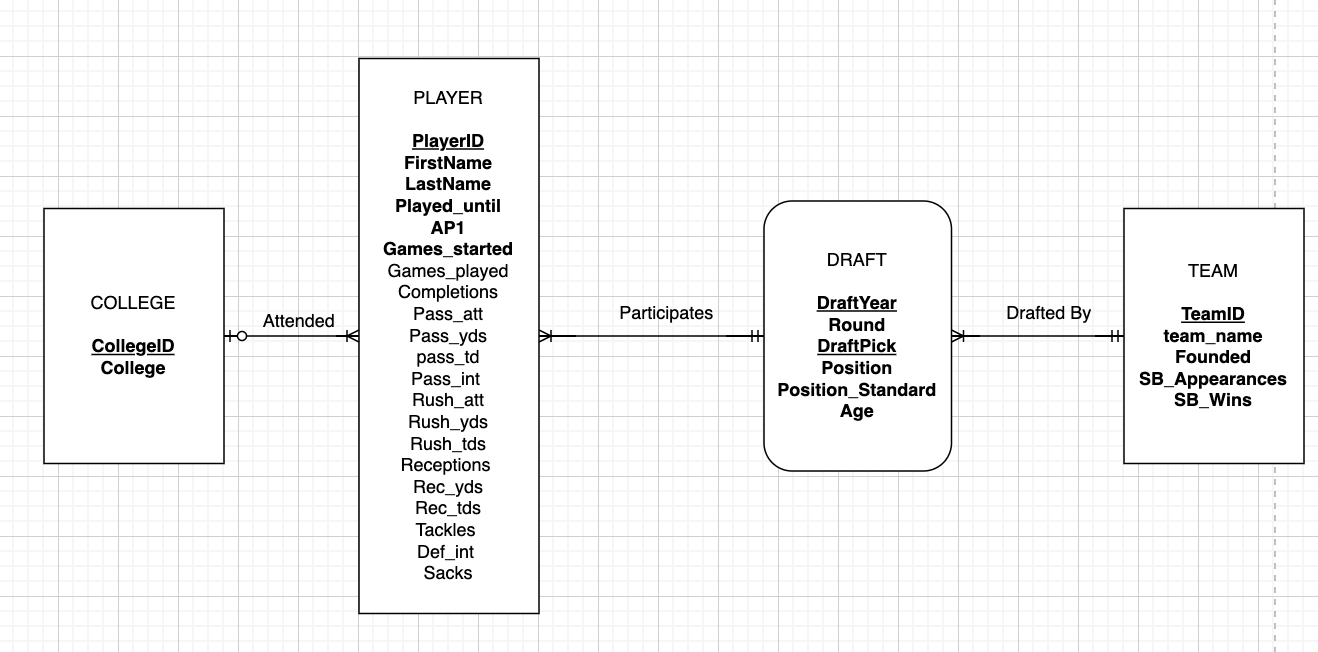
**Data**

This project uses data gathered from a Kaggle dataset about NFL players draft information and their overall performance in the NFL (<https://www.kaggle.com/datasets/ulrikthygepedersen/nfl-draft-1985-2015>). The dataset includes NFL drafts from the years of 1985 to 2015, but our group only used data from 2000 – 2015, due to how much the game of football has changed since the mid-1980s. This dataset originally included 4,079 responses and 34 categories, but our group narrowed the data to 35 categories to analyze the most relevant fields of information and have included fields for team information and created a primary identifier for colleges. The most relevant information is player description, the draft pick, and statistics while in the NFL. The player description will give us unique player identifications, while the data of the draft pick and statistics of the player in the NFL will help our group analyze information about the NFL Draft and the success of players being picked. Table 1 displays a description of the data.

*Table 1 Data dictionary*

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| PlayerID | Text | Unique ID for each player |
| Draft\_Year | Numeric | The year the player was drafted |
| Round | Numeric | The round the player was drafted in |
| Draft\_Pick | Numeric | The pick the player was selected with |
| TeamID | Text | Unique ID for each team |
| Team | Text | The team the player was drafted to |
| Year\_Founded | Number | The year the team was founded |
| SB\_Appearances | Number | The number of times the team has made the Super Bowl |
| SB\_Wins | Number | The number of times the team has won the Super Bowl |
| First\_Name | Text | The player's first name |
| Last\_Name | Text | The player’s last name |
| Position | Text | The player's specific position |
| Position Standard | Text | The player's standardized position |
| Age | Numeric | The player's age when drafted |
| Played Until | Text | The year the player played until (up to 2016), players who never played in an NFL game are listed as “DNP” (Did not play) |
| First Team All-Pro | Numeric | The number of times the player was named First Team All- Pro |
| Pro Bowl | Numeric | The number of times the player was selected for the Pro Bowl |
| Games\_Started | Numeric | Number of games the player was his team's starter |
| Games Played | Numeric | The total number of games the player played in |
| Completions | Numeric | Number of total pass completions |
| Pass\_att | Numeric | Number of passing attempts |
| Pass\_Yds | Numeric | Number of passing yards |
| Pass\_Td | Numeric | Number of passing touchdowns |
| Pass\_Int | Numeric | Number of interceptions thrown |
| Rush\_Att | Numeric | Number of rushing attempts |
| Rush\_Yds | Numeric | Number of rushing yards |
| Rush\_Td | Numeric | Number of rushing touchdowns |
| Receptions | Numeric | Number of receptions |
| Rec\_Yds | Numeric | Number of receiving yards |
| Rec\_Tds | Numeric | Number of receiving touchdowns |
| Tackles | Numeric | Number of tackles |
| Def\_Int | Numeric | Number of defensive interceptions |
| Sacks | Numeric | Number of defensive sacks |
| CollegeID | Text | Unique ID for each college a player was drafted out of |
| College | Text | The college the player was drafted out of |

The PLAYER entity is uniquely identified by PlayerID. The PLAYER entity also displays the player’s first and last names, games started and played, as well as various other stats related to performance. These stats are not all applicable for every player, as they vary by each player’s position. The DRAFT entity holds the draft details of each player including the PlayerID DraftYear and DraftPick which act as the composite primary key. Other information includes the round, position, position\_standard and age. The COLLEGE Tables unique identifier is CollegeID and the name of the college. The fourth entity, TEAM is uniquely identified by TeamID. In the database, each team has its name, when they were founded, and their Superbowl appearances and wins. Figure 1 displays the Entity Relationship Diagram for this data.



*Figure 1 – Entity Relationship Diagram (ERD)*

Based on this ERD, we normalized the data and created a relational schema with 4 tables. Figure 2 displays the graphical relational schema of our NFL Draft database. The COLLEGE table acts as a parent table to PLAYER, where CollegeID is a foreign key in the PLAYER table. The PLAYER table and the TEAM table act as parent tables for the DRAFT table where PlayerID from the PLAYER table is part of the composite primary key for the DRAFT table, and TeamID from the TEAM table is a foreign key in the DRAFT table.



*Fig. 2 Graphical relational schema*

**Database Implementation**

To implement the database into Oracle APEX, we used CREATE table commands to input our 4 tables from the relational schema.

COLLEGE

As a parent table for PLAYER, COLLEGE was created and populated first:

Create Table COLLEGE (

CollegeID VarChar2(5) not null,

College varchar2(50) not null,

Constraint College\_PK Primary Key (CollegeID));

INSERT INTO COLLEGE VALUES (‘ORE’, ‘Oregon’);

TEAM

As one of the parent tables for DRAFT, TEAM was created and populated second :

Create Table TEAM (

TeamID Char(3) not null,

Team\_Name VarChar2(11) not null,

Founded Number (4,0) not null,

SB\_Appearances Number (\*,0) not null,

SB\_Wins Number (\*,0) not null,

Constraint TEAM\_PK Primary Key (TeamID));

INSERT INTO TEAM VALUES (‘TAM’, ‘Buccaneers’, 1976, 1, 1);

\*Note: The order in which the COLLEGE and TEAM table were created and populated are interchangeable

PLAYER

Being a parent table for DRAFT, the PLAYER table must be created and populated before the DRAFT table but after the COLLEGE table:

CREATE TABLE PLAYER (

PlayerID VARCHAR2(20) NOT NULL,

Firstname VARCHAR2(25) NOT NULL,

Lastname VARCHAR2(25) NOT NULL,

CollegeID VARCHAR2(5) NOT NULL,

Played\_Until VARCHAR2(4) DEFAULT 'DNP',

ap1 NUMBER(2,0) NOT NULL,

pro\_bowl NUMBER(2,0) NOT NULL,

Games\_Started NUMBER(\*,0) NOT NULL,

Games\_Played NUMBER(\*,0) DEFAULT 0,

Completions NUMBER(\*,0) DEFAULT 0,

Pass\_att NUMBER(\*,0) DEFAULT 0,

pass\_yds NUMBER(\*,0) DEFAULT 0,

pass\_td NUMBER(\*,0) DEFAULT 0,

pass\_int NUMBER(\*,0) DEFAULT 0,

rush\_att NUMBER(\*,0) DEFAULT 0,

rush\_yds NUMBER(\*,0) DEFAULT 0,

rush\_tds NUMBER(\*,0) DEFAULT 0,

receptions NUMBER(\*,0) DEFAULT 0,

rec\_yds NUMBER(\*,0) DEFAULT 0,

rec\_tds NUMBER(\*,0) DEFAULT 0,

tackles NUMBER(\*,0) DEFAULT 0,

def\_int NUMBER(\*,0) DEFAULT 0,

sacks NUMBER(\*,0) DEFAULT 0,

CONSTRAINT PLAYER\_PK PRIMARY KEY (playerid),

CONSTRAINT PLAYER\_FK FOREIGN KEY (CollegeID) REFERENCES COLLEGE (CollegeID));

INSERT INTO PLAYER (PlayerID, Firstname, Lastname, CollegeID, Played\_Until, ap1, pro\_bowl, Games\_Started, Games\_Played, Completions, Pass\_att, pass\_yds, pass\_int, rush\_att, rush\_yds, rush\_tds) VALUES (‘WinsJa00’, ‘Jameis’, ‘Winston’, ‘FSU’, ‘2016’, 0, 1, 2, 26, 540, 913, 6722, 42, 25, 86, 311, 7);

DRAFT

CREATE table DRAFT (

PlayerID VARCHAR2(20) not null,

Draft\_Year Number(4,0) not null check (Draft\_Year > 1999 AND Draft\_Year < 2016),

Draft\_Round Number (1,0) not null,

Draft\_Pick Number (3,0) not null,

TeamID Char(3) not null,

Position VarChar2(3) not null Check (position in ('QB','OLB','WR','T','DE','RB','CB','NT','DT','C','G','FS','ILB','SS','TE','FB','S','P','LS','LB','OL','DB','K','DL')),

Position\_Standard Varchar2(2) not null check (position\_standard in ('QB','LB','WR','T','DE','RB','DB','DT','C','G','TE','FB','P','LS','K')),

Age Number (2,0) not null,

Constraint DRAFT\_PK Primary Key (PlayerID, Draft\_Year, Draft\_Pick),

Constraint DRAFT\_Player\_FK Foreign Key (PlayerID) References PLAYER (PlayerID),

Constraint DRAFT\_Team\_FK Foreign Key (TeamID) References TEAM (TeamID));

INSERT INTO DRAFT VALUES (‘WinsJa00’, 2015, 1, 1,‘TAM’, ‘QB’, ‘QB’, 21);

**Analysis**

This analysis is intended to identify specific relationships and information relating to NFL Draft data by identifying key statistics related to draft picks per team, the average games played by first round draft picks, the most picked position each year, the player with the most rushing yards, andquarterback performance.

**Question 1**: Draft Picks Per Team

What is the total number of draft picks made by each team between the years 2000 and 2015? To answer this research question, our group created a JOIN Query that counts the number of draft picks per team each year. From there the query creates a total the number of picks, between 2000 and 2015, and lists from highest to lowest amount.

SELECT TEAM.Team\_Name, COUNT(DRAFT.Draft\_Pick) AS Total\_Draft\_Picks

FROM TEAM

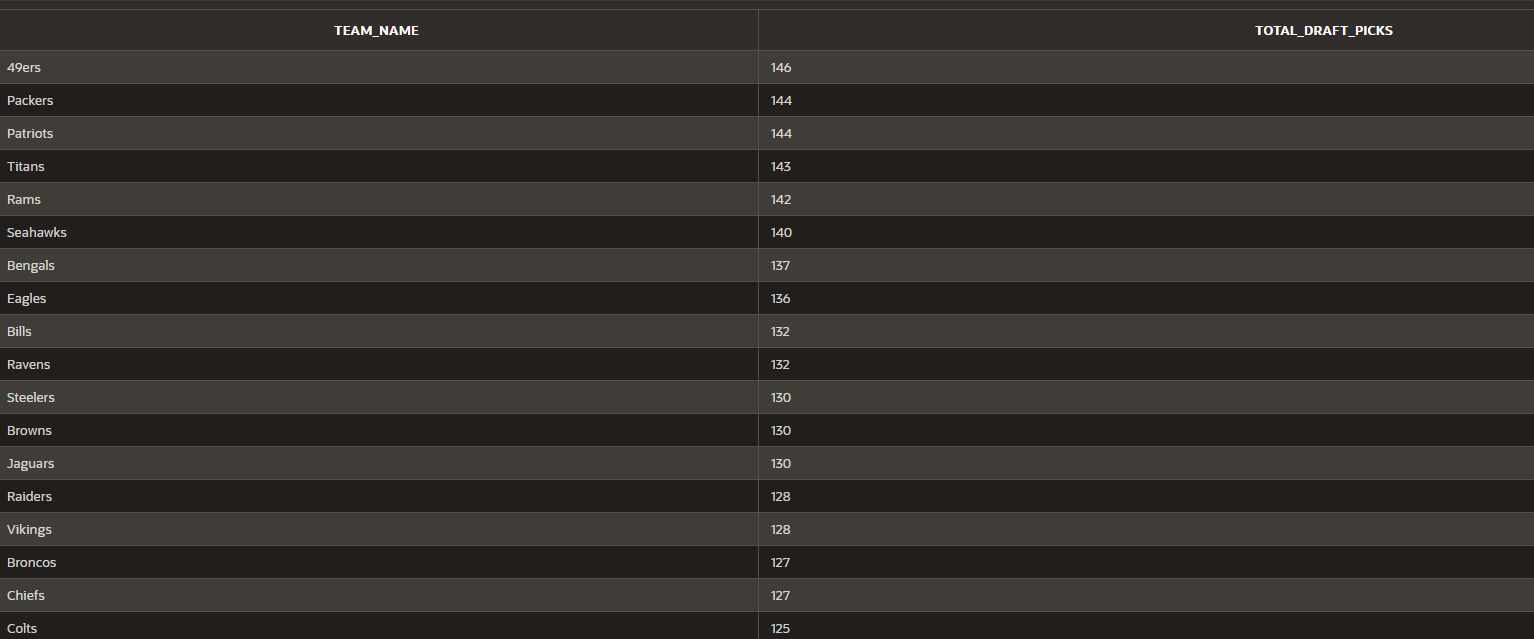
JOIN DRAFT ON TEAM.TeamID = DRAFT.TeamID

WHERE DRAFT.Draft\_Year BETWEEN 2000 AND 2015

GROUP BY TEAM.Team\_Name

ORDER BY Total\_Draft\_Picks DESC;

The table produced as a result of the query is shown below. From the table, we were able to conclude that the 49ers were the team with the greatest number of picks, 146. The Saints were the lowest with only 107 picks between 2000 and 2015. While this query returns 32 rows, the first 18 are shown below (Fig. 3):



*Fig. 3: Most Draft Picks*

**Question 2**: Average Games Played by First Round Draft Picks

What is the number of games played by first round draft picks from 2000-2015? This question can be answered with a GROUP BY query that returns each first round draft picks name, the NFL team, and the number of games played. This can help analyze whether there is a pattern between being a first-round draft pick and averaging a higher number of games.

SELECT Team\_Name, Firstname, Lastname, AVG(Games\_Played) as Avg\_Games\_Played

FROM PLAYER

JOIN DRAFT ON PLAYER.PlayerID = DRAFT.PlayerID

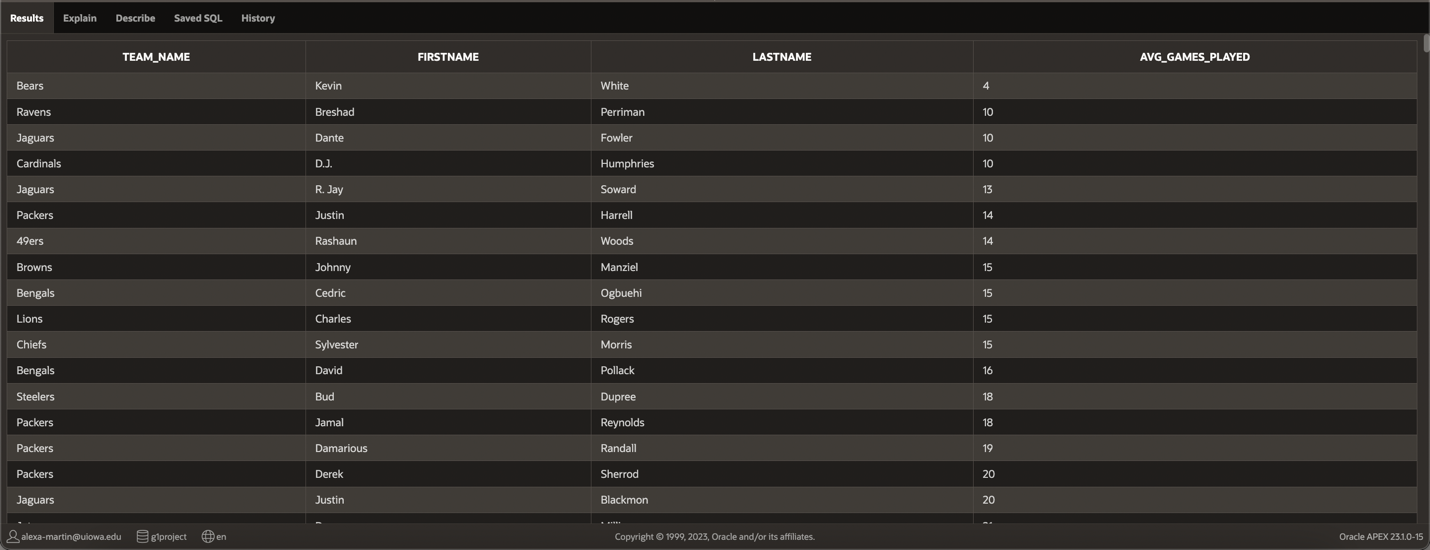
JOIN TEAM ON DRAFT.TeamID = TEAM.TeamID

WHERE Draft\_Round = 1 AND Draft\_Year BETWEEN 2000 AND 2015

GROUP BY Team\_Name, Firstname, Lastname

ORDER BY Avg\_Games\_Played;

The table’s results are shown below, and it can be concluded that the Bears first-round pick Kevin White played in the least number of games (4 games), while the Raiders first-round pick Sebastian Janikowski played in the most amount of games (262). This data can also help find a pattern if a team has multiple players with high or low numbers of games which can help the team strategize for future drafts. (Fig 4.)



*Fig. 4: Average Games played*

**Question #3**: Most Picked Position each Year

What are the most common positions picked in the NFL draft and how has this changed over time? To answer this question, our group constructed a SQL query that retrieves data from the DRAFT table. This query uses a GROUP BY clause to group the data by "draft\_year" and "position," and a subquery to find the maximum count of players drafted for each year. The HAVING clause filters the results to only show the most picked position(s) for each year, and the ORDER BY clause sorts the results in ascending order by draft year. Overall, this query allows us to analyze the data and identify the most commonly picked positions in the NFL draft over time, providing insights into draft trends and strategies.

SELECT draft\_year as Year, position, COUNT(\*) AS Drafted

FROM DRAFT

GROUP BY draft\_year, position

HAVING COUNT(\*) = (

SELECT MAX(cnt)

FROM (

SELECT draft\_year, position, COUNT(\*) AS cnt

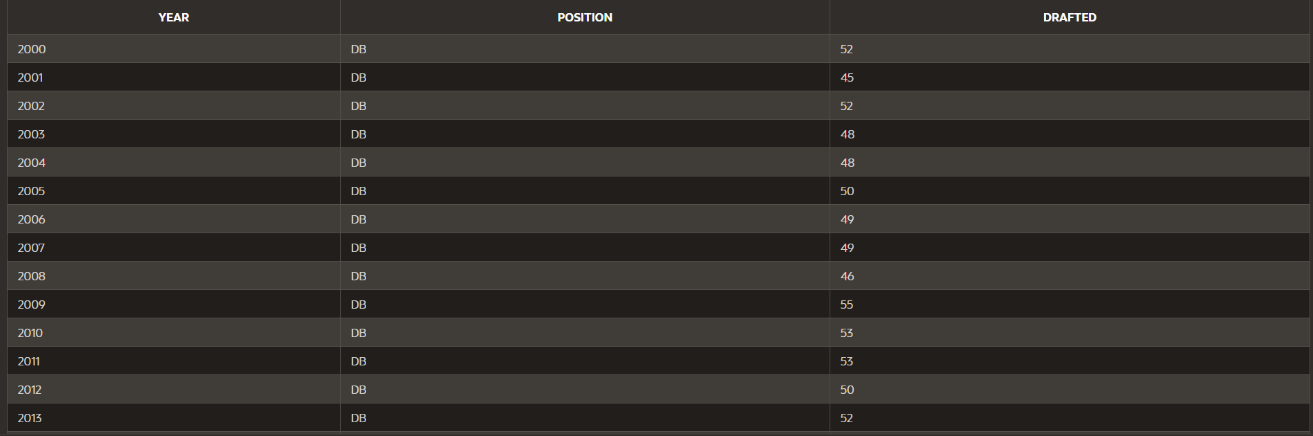
FROM DRAFT

GROUP BY draft\_year, position) t

WHERE t.draft\_year = DRAFT.draft\_year)

ORDER BY draft\_year ASC;

The query’s results are shown in the table below. From the table, it can be concluded that the most picked position in the NFL draft between the years 2000 to 2015 was Defensive Back (DB). This position was most picked in every year except 2015, when Wide Receiver (WR) became the most picked position. The fact that WR was the most picked position in 2015 suggests a shift in the league's offensive strategies and the increasing emphasis on passing the ball. Although the query returns 16 rows, only the first 14 are shown below (Fig. 5).



*Fig. 5: Most Picked Positions*

**Question #4:** Player with most rushing yards

Which running back drafted between the years 2000 and 2015 had the most rushing yards? To answer this question, our group created a SELECT query that retrieves data from the PLAYER and DRAFT tables by joining them under the common column, PlayerID. This specific query is a complex SELECT statement that includes both JOIN and subquery attributes as well as a MAX function to return the result.

SELECT FirstName, LastName, rush\_yds as "Rushing Yards"

FROM PLAYER

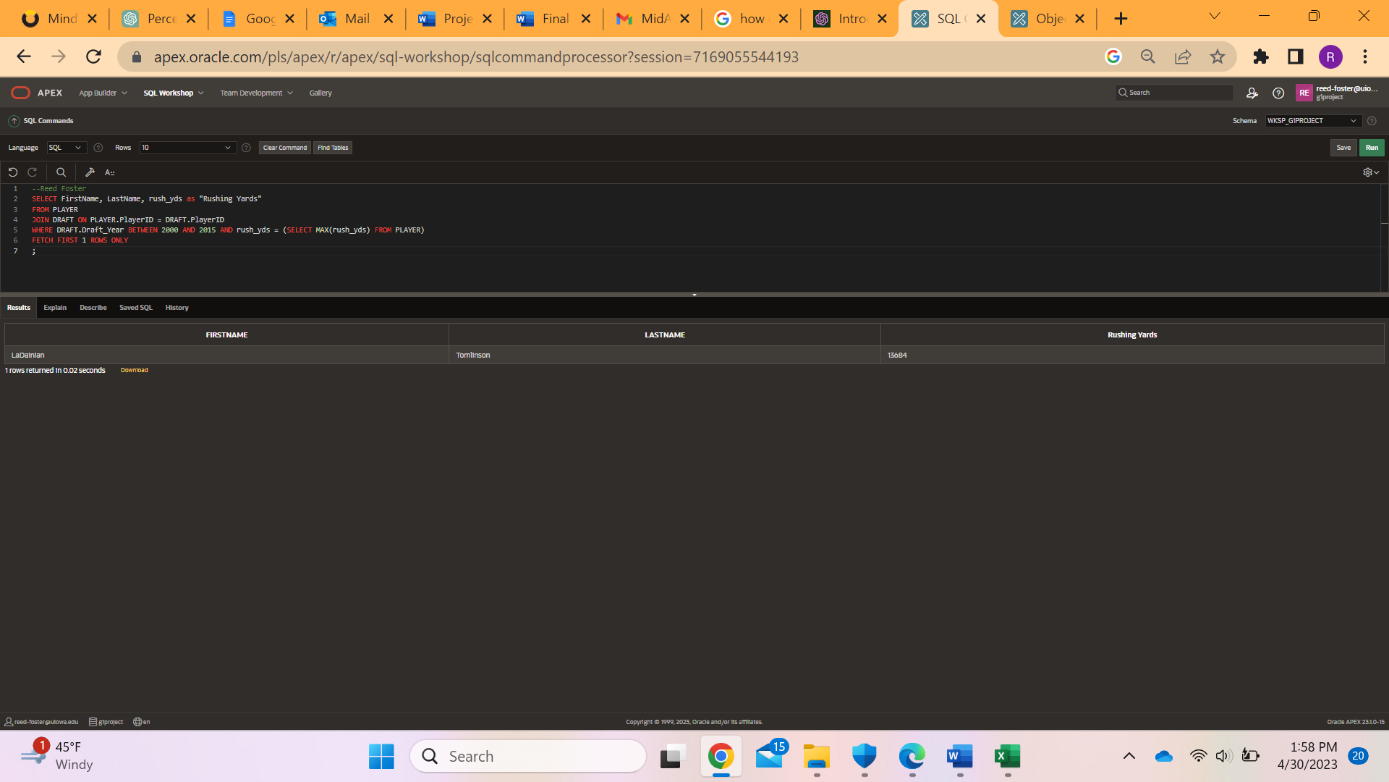
JOIN DRAFT ON PLAYER.PlayerID = DRAFT.PlayerID

WHERE DRAFT.Draft\_Year BETWEEN 2000 AND 2015 AND rush\_yds = (SELECT MAX(rush\_yds) FROM PLAYER)

FETCH FIRST 1 ROWS ONLY

;

The table’s results show that LaDainian Tomlinson has the most rushing yards of running backs from the NFL draft between the years of 2000 and 2015. (Fig. 6)



*Fig. 6: Most rushing yards*

**Question #5:** Quarterback Performance

Which quarterbacks drafted between 2000 - 2014 have had some of the best careers? Can we use passing yards per season and touchdown/interception ratios to determine which QBs with at least 100 pass attempts have been the most successful? With it being difficult to determine a quarterback's skill after just one year, QBs who played only one season are not included. To exclude quarterbacks that have only played for one or zero seasons, we cast played\_until as an integer, and excluded quarterbacks who were in the league for one year or did not play. We included the quarterbacks who had a TD/Int ratio above the average for the dataset. (Fig. 7)

SELECT firstname, lastname, draft\_year, ROUND((pass\_yds / (CAST(played\_until AS INTEGER) - draft\_year)), 1) AS yards\_per\_season, (pass\_td ||'/'|| pass\_int) AS td\_int\_ratio

FROM player

JOIN draft ON player.playerid = draft.playerid

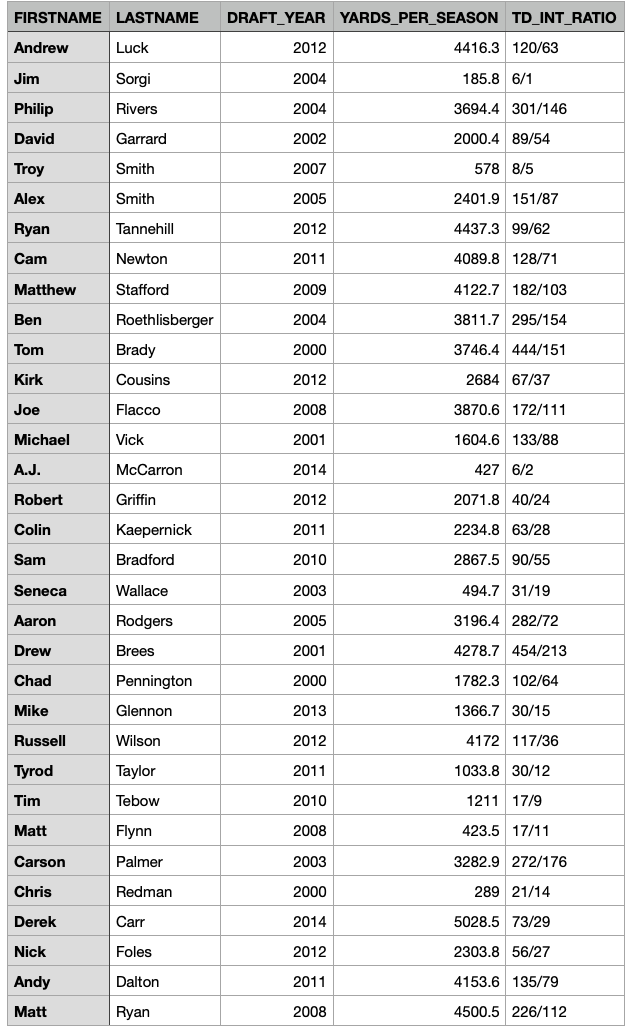
WHERE played\_until <> 'DNP'

HAVING (CAST(played\_until AS INTEGER) - draft\_year) > 1 and pass\_att > 100

AND pass\_td / pass\_int >

(SELECT AVG(pass\_td) / AVG(pass\_int) FROM player)

GROUP BY firstname, lastname, draft\_year, pass\_yds, pass\_att, played\_until, pass\_td, pass\_int;



*Fig 7: Top QB Careers*

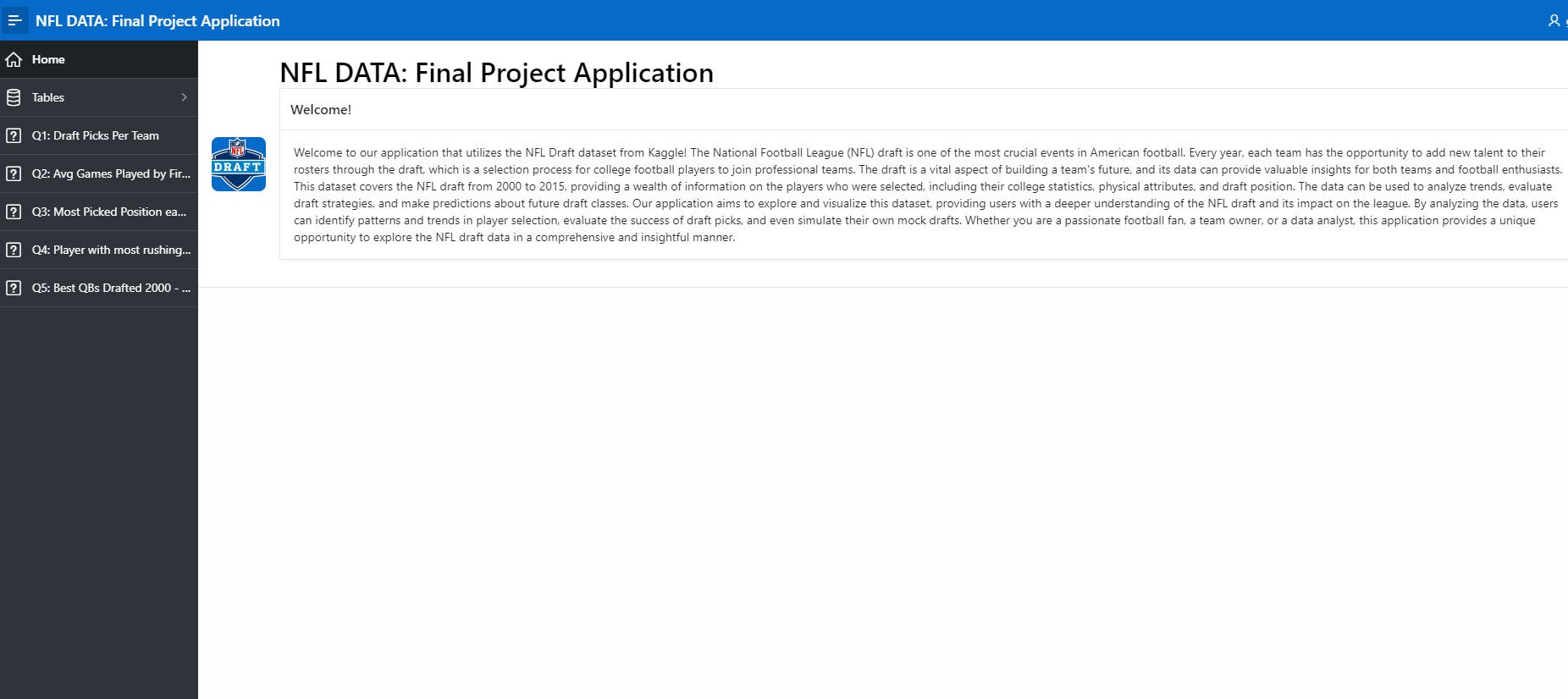
The results of the query return many household names that have been successful in the league. Despite there being 198 QBs in this dataset, this query only returned 27 players, which displays how hard it is for a quarterback to succeed in this league. It also shows that many QBs do not even get to throw 100 passes, which is reassuring to the players who begin to lose confidence in themselves as a starter, showing them that only a few are even above the average of touchdown/interception ratios.

**Web Design:**

Oracle Application Public Link: <https://apex.oracle.com/pls/apex/r/g1project/nfl-data-final-project-application/home?session=114924929818165>

Home Page

The home page of our application includes a brief introduction as to what our application is about and the dataset that we used from Kaggle and a Creative Commons-licensed image of the NFL draft logo (<https://en.wikipedia.org/wiki/National_Football_League_Draft)>. It highlights the importance of the data we collected, and the different ways real NFL analysts can collect data about the draft. We wanted to give users a look inside the relevance of our data while also catching their attention! Figure 8 shows a screenshot of the home page.



*Fig. 8 Home Page*

Tables

Within the database we made tables to explain the data that factors into our analysis and reports. Using these tables, it is easy to find and sort through the data to fully comprehend our analysis of the NFL draft data which can help to find patterns in past and future drafts. Using these tables, we can create other pages to answer the questions that we have asked about the NFL draft. Our table pages also have a description for easy understandability for users looking at the webpage.

A screenshot of a computer

Description automatically generated

*Fig. 9 COLLEGE*

A screenshot of a computer

Description automatically generated

*Fig. 10 TEAM*

*A screenshot of a computer

Description automatically generated with medium confidence*

*Fig.11 PLAYER*

*A screenshot of a computer

Description automatically generated*

*Fig. 12 DRAFT*

Queries

Question 1: The first question in our application shows the total number of picks each team in our database accumulated from 2000 to 2015. I displayed the results of the query in a classic report as well as an accompanied bar chart. On each of our questions we added a text box to restate the research question and the general results of each query. In my results section, I included the teams with the highest number as well as the least number of picks during the time period. This page is shown below in Figure 13.

A screenshot of a computer

Description automatically generated with medium confidence

*Fig. 13 TOTALPICKS*

Question 2:

What is the number of games played by first round draft picks from 2000-2015? Which teams will have the most first-round draft picks playing on their team? Will this help them strategize for the draft in upcoming seasons? The table’s results are shown in figure label (Fig 14), and it can be concluded that the Bears first-round pick, Kevin White, played in the least number of games (4 games), while the Raiders first-round pick, Sebastian Janikowski, played in the most amount of games (262). This data can also help find a pattern if a team has multiple players on certain teams with high or low numbers of games which can help the team strategize for future drafts.

Graphical user interface

Description automatically generated

*Fig. 14: Average Games Played by First-Round Draft Picks*

Question 3:

What are the most common positions picked in the NFL draft and how has this changed over time? The query’s results are shown in the table below in figure 15 (Fig 15). From the table, it can be concluded that the most picked position in the NFL draft between the years 2000 to 2015 was Defensive Back (DB). This position was most picked in every year except 2015, when Wide Receiver (WR) became the most picked position. The fact that WR was the most picked position in 2015 suggests a shift in the league's offensive strategies and the increasing emphasis on passing the ball.

Table

Description automatically generated

*Fig. 15: Most Picked Position Each Year*

Question 4:

Which running back drafted between the years 2000 and 2015 had the most rushing yards? To answer this question, our group created a SELECT query that retrieves data from the PLAYER and DRAFT tables by joining them under the common column, PlayerID. This specific query is a complex SELECT statement that includes both JOIN and subquery attributes as well as a MAX function to return the result. The table’s results show that LaDainian Tomlinson has the most rushing yards of running backs from the NFL draft between the years of 2000 and 2015 in figure 16 (Fig 16).

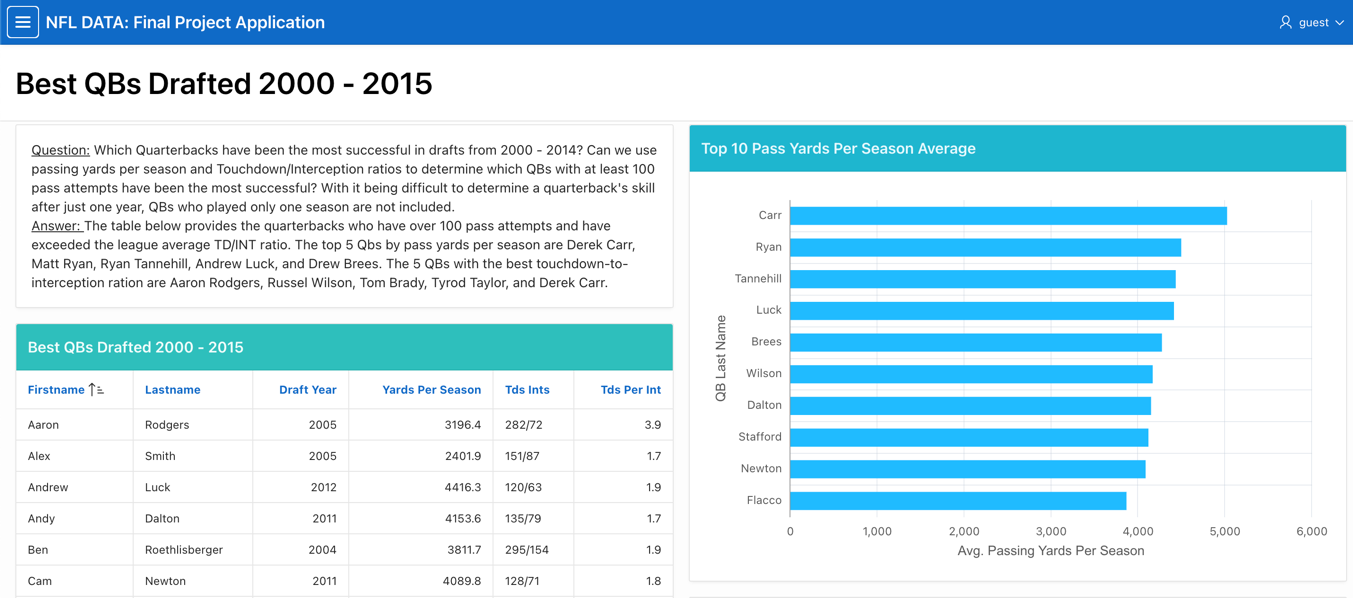
Graphical user interface, text, application

Description automatically generated

*Fig. 16: Player with the Highest Number of Rushing Yards*

Question 5:

The final research question concerns which quarterbacks that were drafted between 2000 – 2015 have had the best performances in their careers. The charts are split between the top five quarterbacks with the highest average passing yards per season highest TD/INT ratio. We first chose quarterbacks who have played more than one season, have thrown at least 500 passes, and who have touchdown-to-interception ratios that are higher than the average for all quarterbacks. Results were presented as a classic table and as two interactive charts (Figure 17a and 17b).



*Fig. 17a Best QBs: Pass Yards per Season*



*Fig. 17a Best QBs: Touchdown-to-Interception Ratio*