NCAA Basketball Tournament Database
Final Report

Brian Collins, Jack Coyne, Jake Burger, Lucas Regina, Tommy Jaeger

Web Application:
https://apex.oracle.com/pls/apex/r/database202323/basketball-data/home?session=7266016698525

Introduction

The NCAA basketball tournament, informally known as "march madness", captivates sports fans annually with its thrilling competitive spirit. Recognizing the diverse spectrum of audience engagement, from casual viewers to fervent basketball aficionados, our team is dedicated to developing a sophisticated, easily accessible database. This resource will cater to those seeking an in-depth exploration of team performances and statistics from the 2014 NCAA Division 1 season. We will harness the power of data-driven insights to deliver a comprehensive and insightful database, encapsulating the nuances and achievements of all participating teams, with a special focus on the top contenders. This initiative aims to elevate the experience of NCAA Tournament enthusiasts by providing a rich, detailed statistical landscape.

Data

Our project utilizes a dataset initially sourced from Kaggle, which was originally complied for predicting future NCAA tournament outcomes. This data set encompasses a comprehensive record of games from the beginning of the 2003 Regular Season through the finale of the 2014 Tournament. Initially, this expansive dataset featured over 60,000 rows of unique game entries, along with their respective statistical details. To tailor the dataset for our specific use in Oracle APEX, we refined the dataset to focus exclusively on the 2014 season, resulting in a more manageable 11,495 rows. Furthermore, the original dataset contained 150 data columns, each representing different aspects of game statistics. To align with our project's objective of presenting data effectively through a Web-Based Application, we conducted an evaluation of these columns. This process led to the exclusion of columns that were either redundant, not unique, or pertinent to our goal. We ultimately narrowed the dataset down to 40 columns, each containing relevant insights that contribute to a clear and insightful representation of the 2014 NCAA season's statistics. The column names and description of columns we used are the following:

Table 1 Data Dictionary

Field	Type	Description
TeamID	Text	A unique ID for each team
TeamName	Text	The name of team
Season	Text	Year of the season
RegionCode	Text	A unique ID for each region
RegionName	Text	The name of the region
SeedNum	Numeric	The seed for teams in the tournament
GameID	Text	A unique ID for each game
DayNum	Numeric	Number of days since the start of the season
WinningTeamID	Text	ID of the team that won the game
WinnerScore	Numeric	Score of the winning team
LosingTeamID	Numeric	ID of the team that lost the game
LoserScore	Numeric	Score of the losing team
WinnerLocation	Text	Location of where the game was won

NumOTPeriods	Numeric	Indicates if there was an overtime period
WinnerFGM	Numeric	Number of field goals the winner made
WinnerFGA	Numeric	Number of field goals the winner attempted
Winner3PTM	Numeric	Number of three pointers the winner made
Winner3PTA	Numeric	Number of three pointers the winner attempted
WinnerFreeThrowMade	Numeric	Number of free throws the winner made
WinnerFreeThrowAttempted	Numeric	Number of three pointers the winner attempted
WinnerOffensiveRebounds	Numeric	Number of offensive rebounds the winner had
WinnerDefensiveRebounds	Numeric	Number of defensive rebounds the winner had
WinnerAssists		
WinnerTurnovers	Numeric	Number of assists the winner had
WinnerSteals	Numeric	Number of turnovers the winner had
WinnerBlocks	Numeric	Number of steals the winner had
WinnerPersonalFouls	Numeric	Number of blocks the winner had
LoserFGM	Numeric	Number of personal fouls the winner had
LoserFGA	Numeric	Number of field goals the loser made
Loser3PTM	Numeric	Number of field goals the loser attempted
Loser3PTA	Numeric	Number of three pointers the loser made
LoserFreeThrowMade	Numeric	Number of three pointers the loser attempted
LoserFreeThrowAttempted	Numeric	Number of free throws the loser made
LoserOffensiveRebounds	Numeric	Number of three pointers the loser attempted
LoserDefensiveRebounds	Numeric	Number of offensive rebounds the loser had
LoserAssists	Numeric	Number of defensive rebounds the loser had
LoserTurnovers	Numeric	Number of assists the loser had
LoserSteals	Numeric	Number of turnovers the loser had
LoserBlocks	Numeric	Number of steals the winner had
LoserPersonalFouls	Numeric	Number of blocks the winner had
	Numeric	Number of personal fouls the winner had

The TEAM table is the primary entity in the database, identified by TeamID, and serves as the parent table to the GAME table, which is one of the main focuses of the database as it captures comprehensive game-related data. Each team can be involved in multiple games, creating a one-to-many relationship from TEAM to GAME. REGION and TOURNEYSEED are additional tables; the former identifies the location of games and the latter associates' teams with their tournament seeding. The GAME table stores detailed statistics for each match, including team performance metrics and outcomes, with foreign keys linking back to the TEAM and REGION tables. This setup allows for a robust collection and analysis of game data within the context of teams and their regions. Figure 1 displays the ERD for this data.

Entity Relationship Diagram

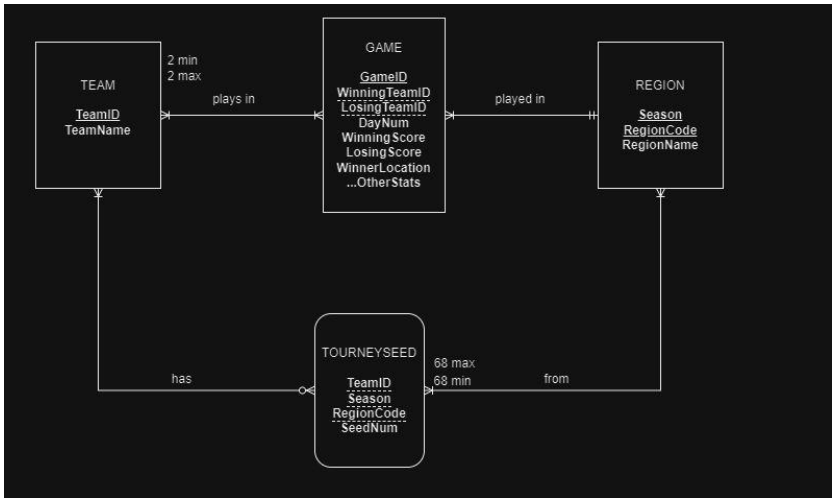


Fig. 1 Entity Relationship Diagram

Based on our ERD, we normalized the data and created a relational schema with 4 tables. Figure 2 displays the graphical relational schema of the database. Each table is interconnected, with TEAM being the parent table from which GameID in the GAME table references as a foreign key. The GAME table captures the exhaustive details of each match, while REGION and SEED further contextualize the data with location and tournament seeding specifics. Composite primary keys are utilized in tables where a single field is not sufficient to ensure uniqueness.

Relational Schema:

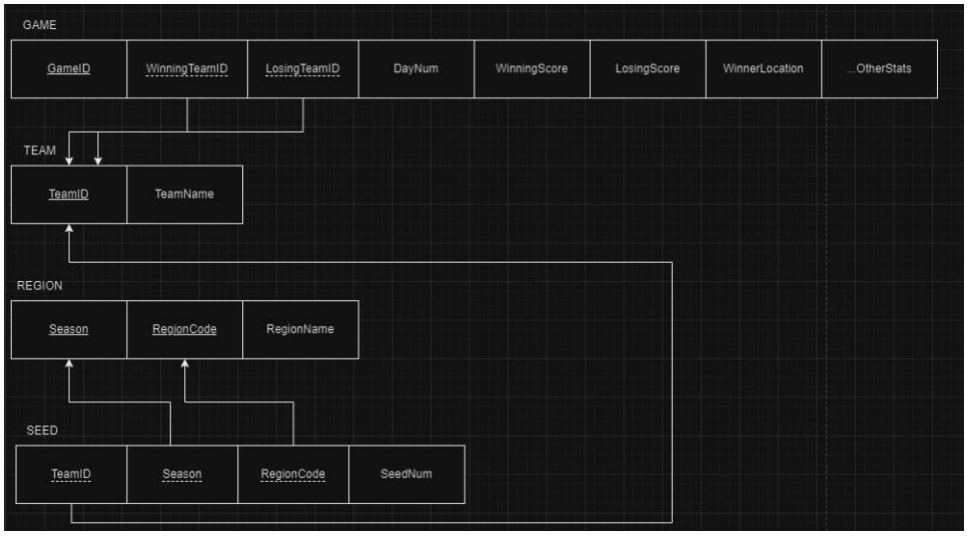


Fig. 2 Graphical Relational Schema

Database Implementation

To implement the database in APEX, we wrote CREATE TABLE commands for each table in the relational schema.

TEAM

As the parent table, TEAM was created and populated first:

```
CREATE TABLE Team (
   TeamID CHAR(4) NOT NULL,
   TeamName CHAR(225) NOT NULL,
   CONSTRAINT TEAM_PK PRIMARY KEY (TeamID)
```

REGION

```
CREATE TABLE Region (
   Season char(4) not null,
   RegionCode char(1) not null,
   RegionName varchar(25) not null,
   constraint Region_pk primary key (SeasonID, RegionCode)
```

SEED

```
CREATE TABLE SEED (
   Season NUMBER(*,0) NOT NULL,
   RegionCode CHAR(1) NOT NULL,
   SeedNum VARCHAR2(3) NOT NULL,
   TeamID CHAR(4) NOT NULL,
   CONSTRAINT SEED_PK PRIMARY KEY (Season, RegionCode, SeedNum),
   CONSTRAINT SEED_FK FOREIGN KEY (TeamID) REFERENCES TEAM (TeamID)
```

GAMES

```
CREATE TABLE Games (
   GameID CHAR(6) NOT NULL,
   Type VARCHAR(7) NOT NULL,
   Season NUMBER(4,0) NOT NULL,
   DayNum NUMBER(4,0) NOT NULL,
   WinnerScore NUMBER(3,0) NOT NULL,
   LoserScore NUMBER(3,0) NOT NULL,
   WinnerLocation VARCHAR(1) NOT NULL,
   NumOTPeriods NUMBER(1) NOT NULL,
   WinnerFGM NUMBER(2,0) NOT NULL,
   WinnerFGA NUMBER(2,0) NOT NULL,
   Winner3PTM NUMBER(2,0) NOT NULL,
   Winner3PTA NUMBER(2,0) NOT NULL,
   WinnerFreeThrowMade NUMBER(2,0) NOT NULL,
   WinnerFreeThrowAttempted NUMBER(2,0) NOT NULL,
   WinnerOffensiveRebounds NUMBER(2,0) NOT NULL,
   WinnerDefensiveRebounds NUMBER(2,0) NOT NULL,
   WinnerAssits NUMBER(2,0) NOT NULL,
   WinnerTurnovers NUMBER(2,0) NOT NULL,
```

```
    WinnerSteals NUMBER(2,0) NOT NULL,
    WinnerBlocks NUMBER(2,0) NOT NULL,
    WinnerPersonalFouls NUMBER(2,0) NOT NULL,
    LoserFGM NUMBER(2,0) NOT NULL,
    LoserFGA NUMBER(2,0) NOT NULL,
    Loser3PTM NUMBER(2,0) NOT NULL,
    Loser3PTA NUMBER(2,0) NOT NULL,
    LoserFreeThrowMade NUMBER(2,0) NOT NULL,
    LoserFreeThrowAttempted NUMBER(2,0) NOT NULL,
    LoserOffensiveRebounds NUMBER(2,0) NOT NULL,
    LoserDefensiveRebounds NUMBER(2,0) NOT NULL,
    LoserAssits NUMBER(2,0) NOT NULL,
    LoserTurnovers NUMBER(2,0) NOT NULL,
    LoserSteals NUMBER(2,0) NOT NULL,
    LoserBlocks NUMBER(2,0) NOT NULL,
    LoserPersonalFouls NUMBER(2,0) NOT NULL,
    WinningTeamID CHAR(4) not null,
    LosingTeamID CHAR(4) not null,
    CONSTRAINT GAME_PK PRIMARY KEY (GameID),
    CONSTRAINT WinningTeam_FK FOREIGN KEY (WinningTeamID) REFERENCES Team(TeamID),
    CONSTRAINT LosingTeam_FK FOREIGN KEY (LosingTeamID) REFERENCES Team(TeamID)
```

Analysis:

Question 1:

How well do teams perform in front of neutral audiences? Additionally, how common is it for teams to play in front of an unbiased audience?

```
SELECT
    Team.TeamID,
    Team.TeamName,
    SUM(CASE WHEN Team.TeamID = Games.WinningTeamID THEN 1 ELSE 0 END) AS Wins,
    SUM(CASE WHEN Team.TeamID = Games.LosingTeamID THEN 1 ELSE 0 END) AS Losses,
    ROUND(AVG(Games.WinnerScore),1) AS PointsPerGame,
    CASE
        WHEN COUNT(*) >= 10 THEN 'High'
        WHEN COUNT(*) >= 5 THEN 'Medium'
        ELSE 'Low'
    END AS NeutralGameFrequency
FROM
    Team
JOIN
    Games ON Team.TeamID IN (Games.WinningTeamID, Games.LosingTeamID)
WHERE
    Games.WinnerLocation = 'N'
GROUP BY
    Team.TeamID, Team.TeamName
ORDER BY
    Wins DESC;
```

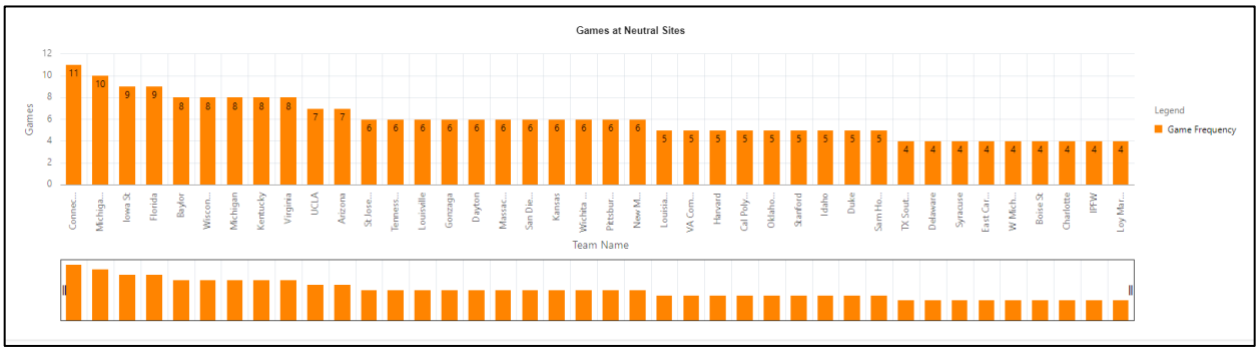


Fig. 3 Wins and Losses at Neutral Sites

Teamid	Teamname	Neutralgamefrequency	Wins	Losses	Pointspergame
1163	Connecticut	High	11	1	70.8
1277	Michigan St	High	10	2	75.4
1235	Iowa St	High	9	1	84.1
1196	Florida	High	9	1	66.4
1124	Baylor	High	8	3	74.7
1458	Wisconsin	High	8	2	75.2
1276	Michigan	High	8	3	71.5
1246	Kentucky	High	8	4	71.4
1438	Virginia	Medium	8	1	69.4
1417	UCLA	Medium	7	2	83.7
1112	Arizona	Medium	7	2	70.3
1386	St Joseph's PA	Medium	6	2	76
1397	Tennessee	Medium	6	3	73.2
1257	Louisville	Medium	6	2	79
1211	Gonzaga	Medium	6	3	78.4

Fig. 4 Frequency of Neutral Games

Description:
The graphs the teams with the most wins at neutral site games. The table provides various other statistics for neutral site games including frequency, losses, and points per game. In the top 15 teams, the teams who have most frequently played at neutral games appear to have played 12 games, while the least frequent was 8 games.

Results:
Connecticut wins most often at neutral games, with a record of 11 wins and 1 loss. However, Iowa State appears to score the most points at neutral games, with an average of 84.1 points per game.

Question 2:

How does the performance of top-seeded teams compare to that of bottom-seeded teams in tournaments? What are the statistics of each seed?

```
SELECT
  TO_CHAR(seednum, '9') AS SeedNumber,
  COUNT(*) AS Amount_Wins_Per_Seed,
  TO_CHAR(COUNT(*) * 100 / (
    SELECT COUNT(*)
    FROM games
    JOIN team ON winningteamid = team.teamid
    JOIN seed ON team.teamid = seed.teamid
    WHERE seednum IN ('03', '02', '01')
  ), '99') || '%' AS Percentage_Wins
FROM games
JOIN team ON winningteamid = team.teamid
JOIN seed ON team.teamid = seed.teamid
WHERE seednum IN ('05', '04', '03', '02', '01')
  AND seed.season='2014'
GROUP BY seednum
ORDER BY seednum;
```

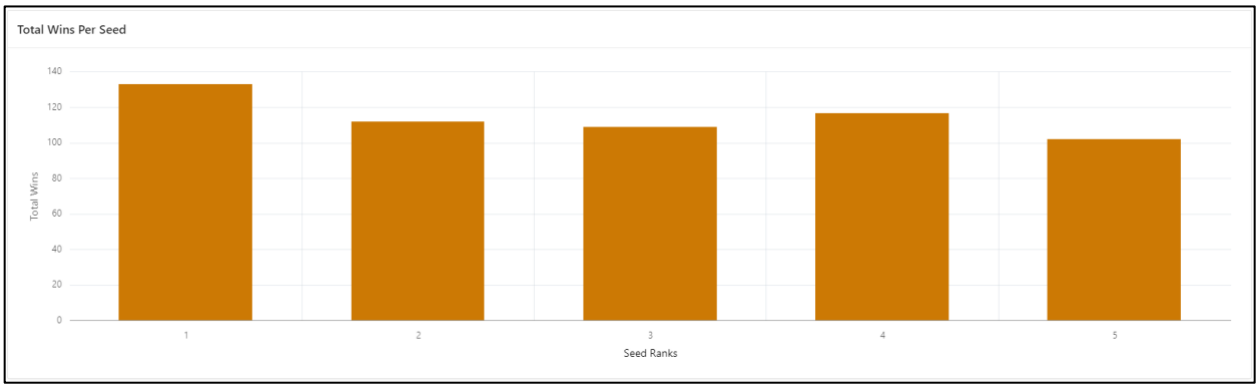


Fig. 5 Wins Per Seed

Team Seed Number	Total Wins Per Seed
1	133
2	112
3	109
4	117
5	102

Fig. 6 Top Seeded Performers

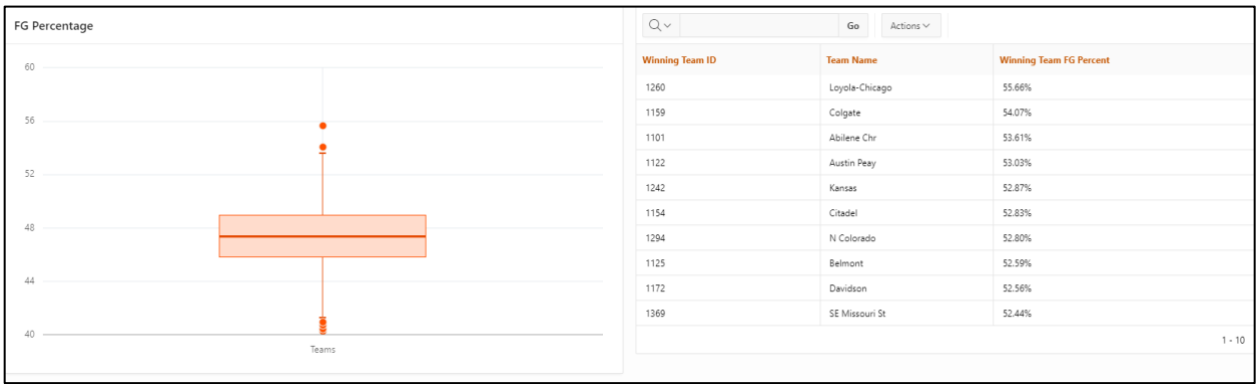
Description:

Returns the results of the top 5 seeds. This includes total wins across all seeds 1-5.

Results:
With the exception of seed 1, it appears as though the top seeds generally won the same number of games throughout the season. This suggests that, at least at the top, there isn't a difference in the number of wins.

Question 3:
What are the top ten teams with the highest field goal percentage in their winning games?

```
SELECT WinningTeamID, TeamName, TO_CHAR((SUM(WinnerFGM) / NULLIF(SUM(WinnerFGA), 0)) *
100, '99.99') || '%' AS WinningTeamFGPercentage
FROM Team
JOIN Games ON TeamID IN (WinningTeamID)
GROUP BY WinningTeamID, TeamName
ORDER BY WinningTeamFGPercentage DESC
FETCH FIRST 10 ROWS ONLY;
```



Winning Team ID	Team Name	Winning Team FG Percent
1260	Loyola-Chicago	55.66%
1159	Colgate	54.07%
1101	Abilene Chr	53.61%
1122	Austin Peay	53.03%
1242	Kansas	52.87%
1154	Citadel	52.83%
1294	N Colorado	52.80%
1125	Belmont	52.59%
1172	Davidson	52.56%
1369	SE Missouri St	52.44%

1 - 10

Fig. 7 Winning Field Goal Percentage

Description:
The query yields the top 10 basketball teams with the highest field goal percentages in their wins, including each team's ID, name, and their winning field goal percentage, ordered from highest to lowest. The boxplot provides insight into the distribution of the winning field goal percentage among all teams.

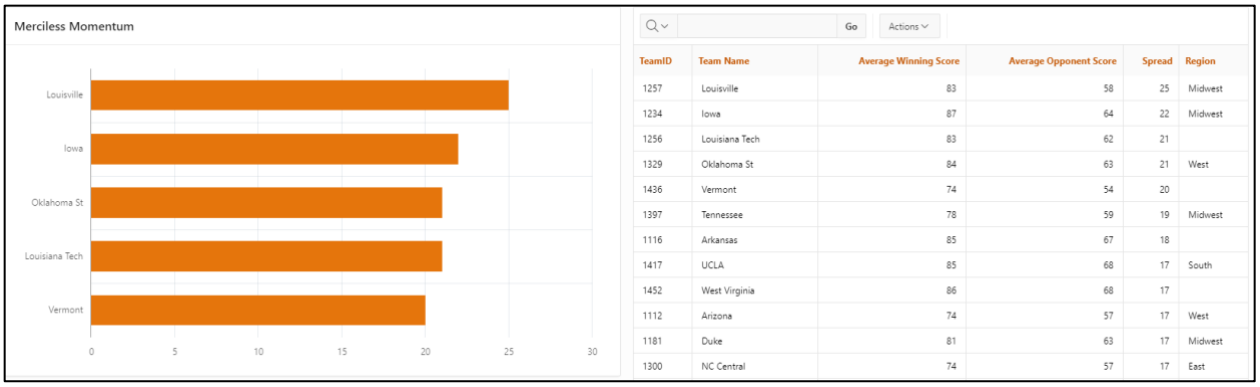
Results:
The query results showcase Loyola-Chicago leading with the highest winning field goal percentage of 55.66%, with Colgate and Abilene Christian University closely following at 54.07% and 53.61%, respectively. The boxplot reveals the median winning field goal percentage is roughly 47%, with a tight distribution typically ranging between 46% to 49%.

Question 4:

How do teams perform when they're winning or losing? Which teams are the most merciless and win by the most points? Which teams are landslide losers and concede the most points when they lose?

Merciless Momentum:

```
SELECT TEAM.TeamID, TEAM.TeamName, ROUND(AVG(WinnerScore),0) AS AvgWinnerScore,
ROUND(AVG(LoserScore),0) AS AvgOpponentScore, ROUND(AVG(WinnerScore) - AVG(LoserScore),0) AS
Spread, R.RegionName
FROM TEAM JOIN GAMES
ON TEAM.TeamID = Games.WinningTeamID
LEFT JOIN (SELECT * FROM SEED WHERE Season = '2014') S
ON TEAM.TeamID = S.TeamID
LEFT JOIN (SELECT * FROM REGION WHERE Season = '2014') R
ON S.RegionCode = R.RegionCode
GROUP BY TEAM.TeamID, Team.TeamName, R.RegionName
ORDER BY ROUND(AVG(WinnerScore) - AVG(LoserScore),0) DESC;
```



TeamID	Team Name	Average Winning Score	Average Opponent Score	Spread	Region
1257	Louisville	83	58	25	Midwest
1234	Iowa	87	64	22	Midwest
1256	Louisiana Tech	83	62	21	
1329	Oklahoma St	84	63	21	West
1436	Vermont	74	54	20	
1397	Tennessee	78	59	19	Midwest
1116	Arkansas	85	67	18	
1417	UCLA	85	68	17	South
1452	West Virginia	86	68	17	
1112	Arizona	74	57	17	West
1181	Duke	81	63	17	Midwest
1300	NC Central	74	57	17	East

Fig. 8 Merciless Momentum

Landslide Losers:

```
SELECT TEAM.TeamID, TEAM.TeamName, ROUND(AVG(LoserScore),0) AS AvgLoserScore,
ROUND(AVG(WinnerScore),0) AS AvgOpponentScore, ROUND(AVG(WinnerScore) - AVG(LoserScore),0)
AS Spread, R.RegionName
FROM TEAM JOIN GAMES
ON TEAM.TeamID = Games.LosingTeamID
LEFT JOIN (SELECT * FROM SEED WHERE Season = '2014') S
ON TEAM.TeamID = S.TeamID
LEFT JOIN (SELECT * FROM REGION WHERE Season = '2014') R
ON S.RegionCode = R.RegionCode
GROUP BY TEAM.TeamID, Team.TeamName, R.RegionName
ORDER BY ROUND(AVG(WinnerScore) - AVG(LoserScore),0) DESC;
```

TeamID	Team Name	Average Losing Score	Average Opponent Score	Spread	Region
1309	New Orleans	65	86	21	
1408	Tulane	58	78	20	
1212	Grambling	62	82	19	
1255	Longwood	65	83	19	
1115	Ark Pine Bluff	58	76	18	
1290	MS Valley St	67	84	18	
1262	MA Lowell	57	75	18	
1381	Southern Utah	57	75	18	
1224	Howard	59	75	17	
1202	Furman	59	76	17	
1354	S Carolina St	58	76	17	
1342	Presbyterian	62	79	17	

Fig. 9 Landslide Losers

Description:
The first table is titled “Merciless Momentum” and displays the teams who, when they win, typically win by the largest margins. The second table is titled “Landslide Losers” and displays the teams who, when they lose, typically lose by the largest margins. Each query returns the team’s unique ID, team name, average winner/loser Score, average opponent score, spread (the difference between the winner/loser’s score and the opponent’s score), and the team’s region if they make it to the NCAA March Madness tournament.

Results:
The five most “merciless” teams are Louisville, Iowa, Oklahoma State, Louisiana Tech, and Vermont. These teams, on average, win by over 20 points whenever they win their games. Two of these teams (Louisville and Iowa) are from the Midwest region, while Oklahoma State is from the West region; all three of these teams made it to the NCAA March Madness tournament in 2014. Despite their dominant performances, Louisiana Tech and Vermont did not make it to the tournament. The teams who lost by the biggest landslides were New Orleans, Tulane, Longwood, Grambling, and Ark Pine Bluff; all of these teams generally lost by over 18 points. None of these teams made it to the tournament.

Question 5:
How do basketball teams perform in terms of 3-pointers when they win or lose? How many three pointers do teams make when they win games? Which losing teams score the least?

```
SELECT G.GameID, G.WinningTeamID, WT.TeamName AS WinningTeamName, G.Winner3PTM,
G.LosingTeamID, LT.TeamName AS LosingTeamName, G.Loser3PTM
FROM GAMES G
JOIN TEAM WT ON G.WinningTeamID = WT.TeamID
JOIN TEAM LT ON G.LosingTeamID = LT.TeamID
```

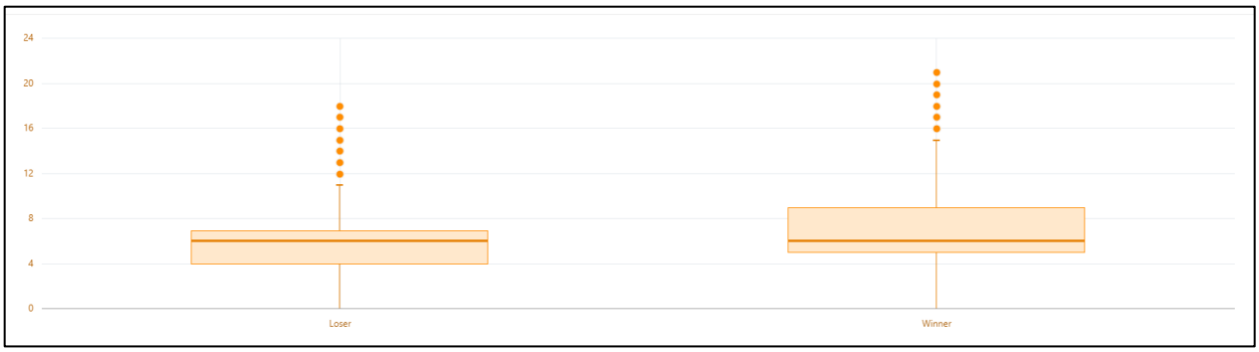


Fig. 10 Three Point Comparison

Fig. 11 Three Pointers in Winning and Losing Teams

Description:
The query returns a list of games, each with the game ID, the ID of the winning team, the ID of the losing team, the name of the winning team, the name of the losing team, and the number of three-pointers made by team for all games.

Results:
Surprisingly, it appears as though winners and losers score just about as many three pointers in their games. This suggests that two-point shots are more likely to decide the winner of basketball games. Both distributions for the number of three pointers in a game also have very low variance.

Web Design
https://apex.oracle.com/pls/apex/r/database202323/basketball-data/home?session=7266016698525

Home Page
The home page of the web application showcases the NCAA Men's Basketball Tournament Bracket, featuring an interactive design for user engagement. We also incorporated a sidebar for easy access and navigation to other sections of the application, ensuring a seamless user experience. Figure 12 shows a screenshot of the homepage.

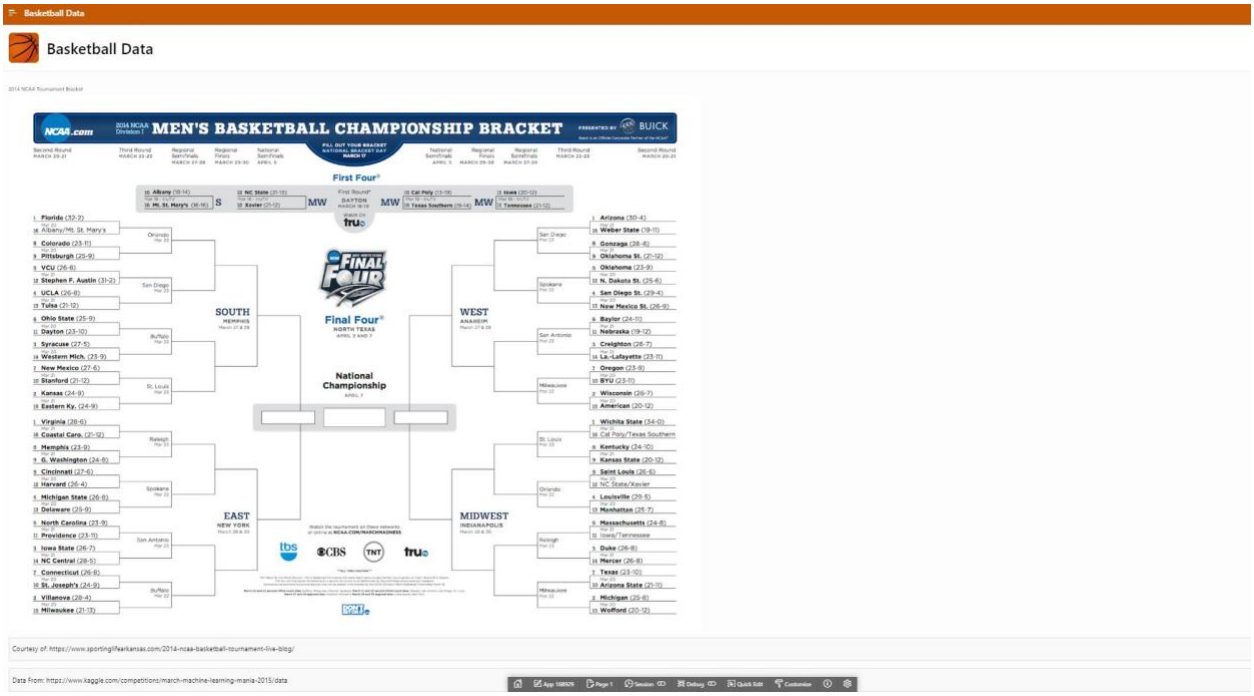


Fig. 12 Home Page

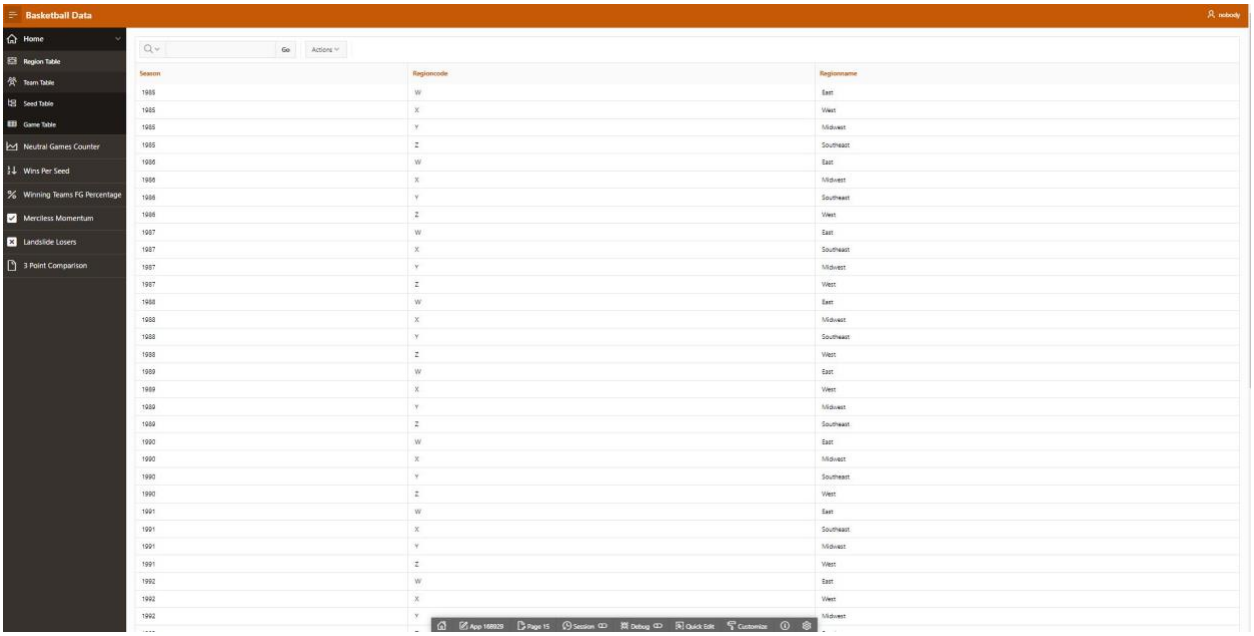


Fig. 13 Region Table Page

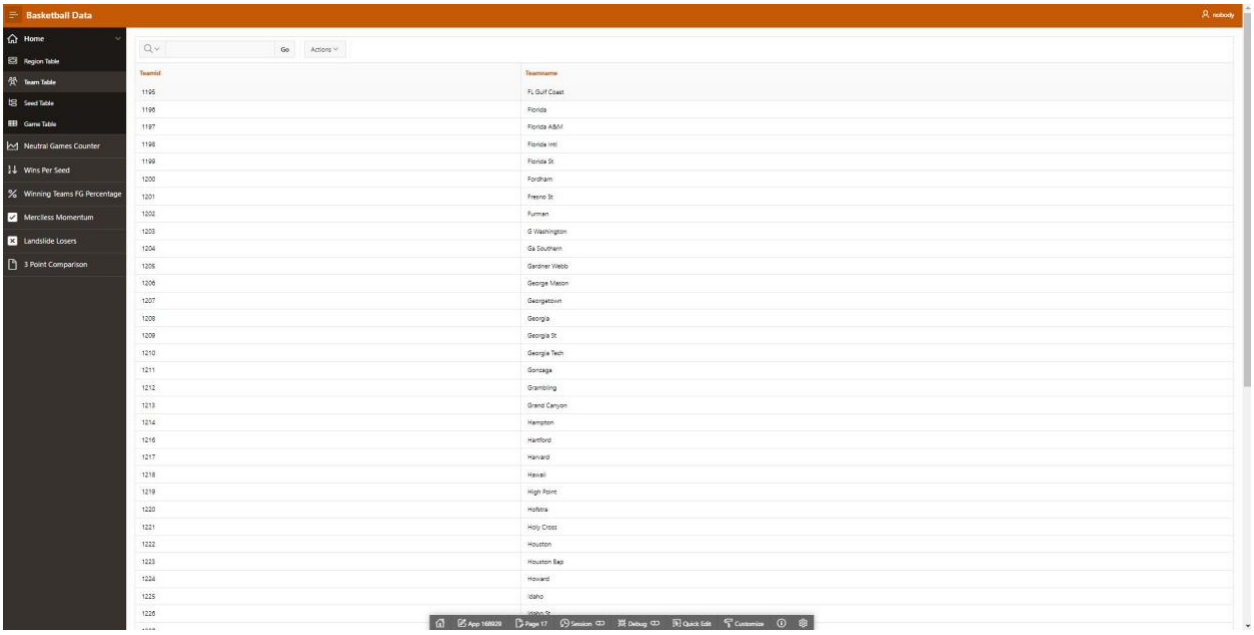


Fig. 14 Team Table Page

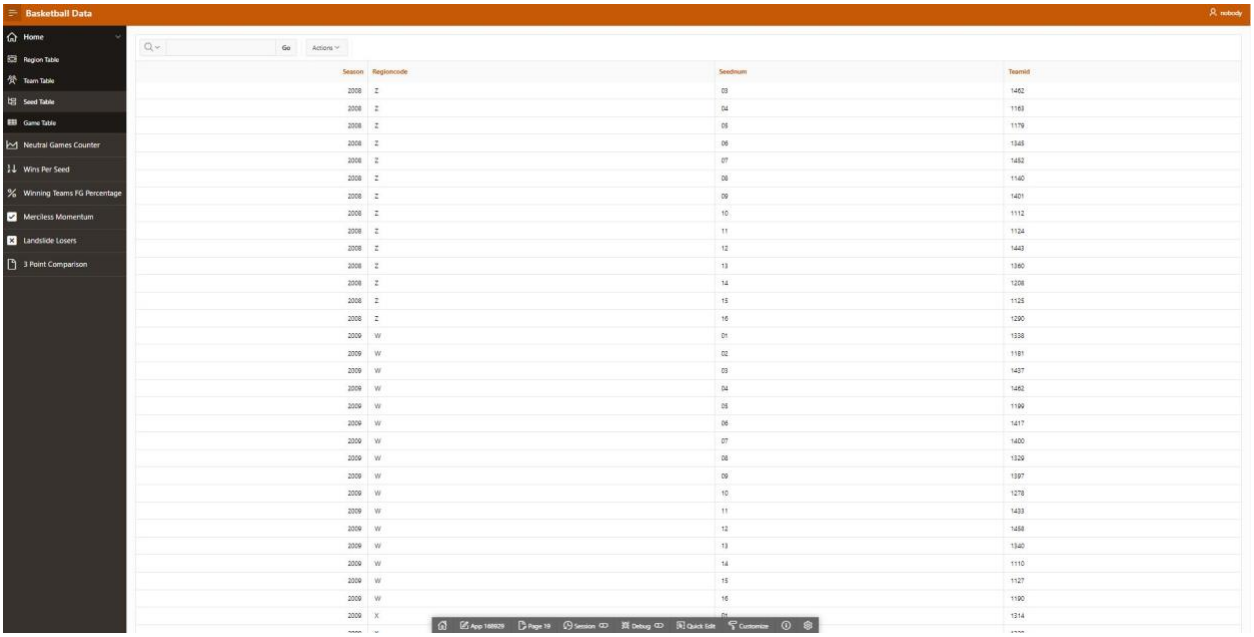


Fig. 15 Seed Table Page

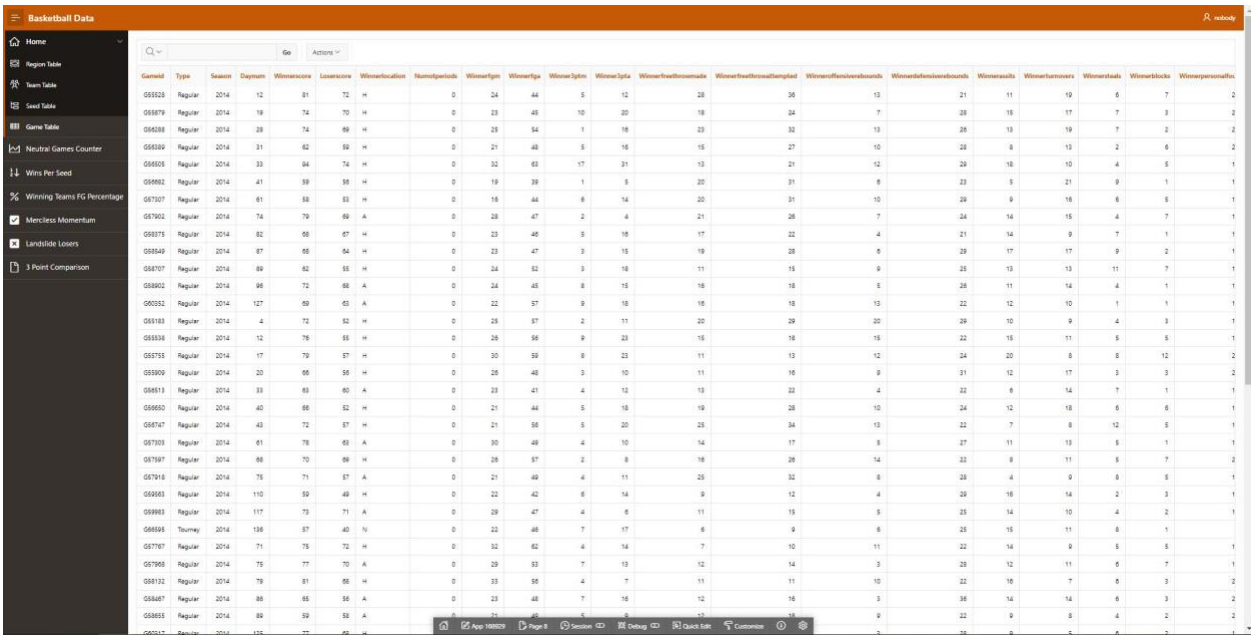


Fig. 16 Game Table Page

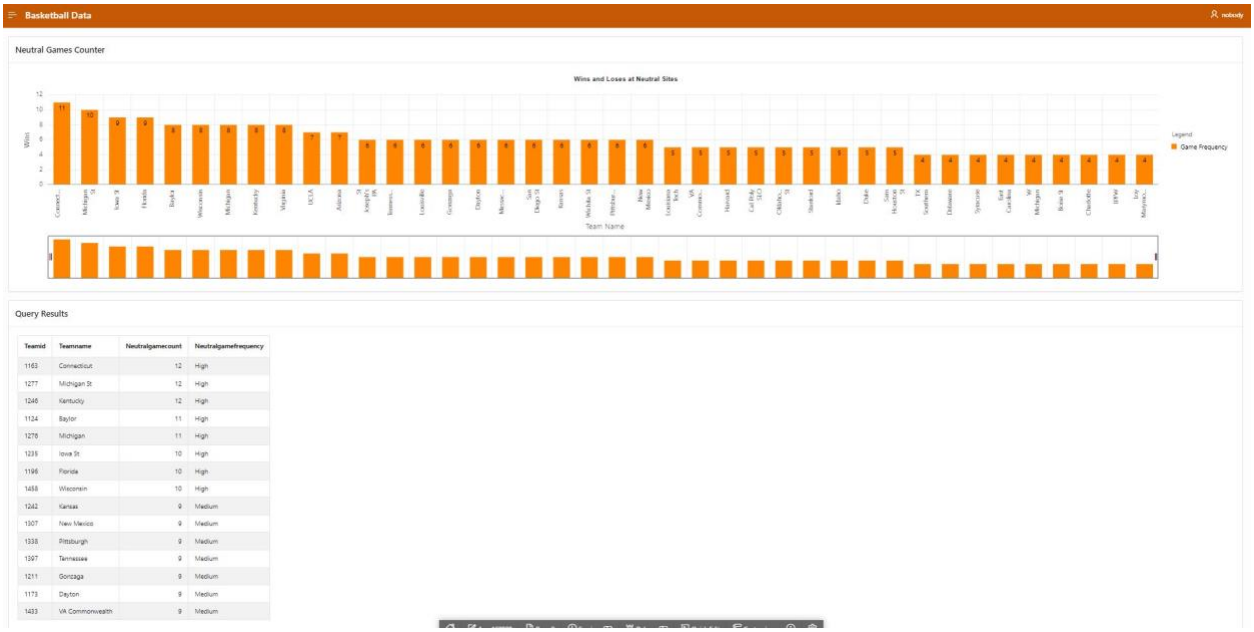


Fig. 17 Neutral Site Games Page

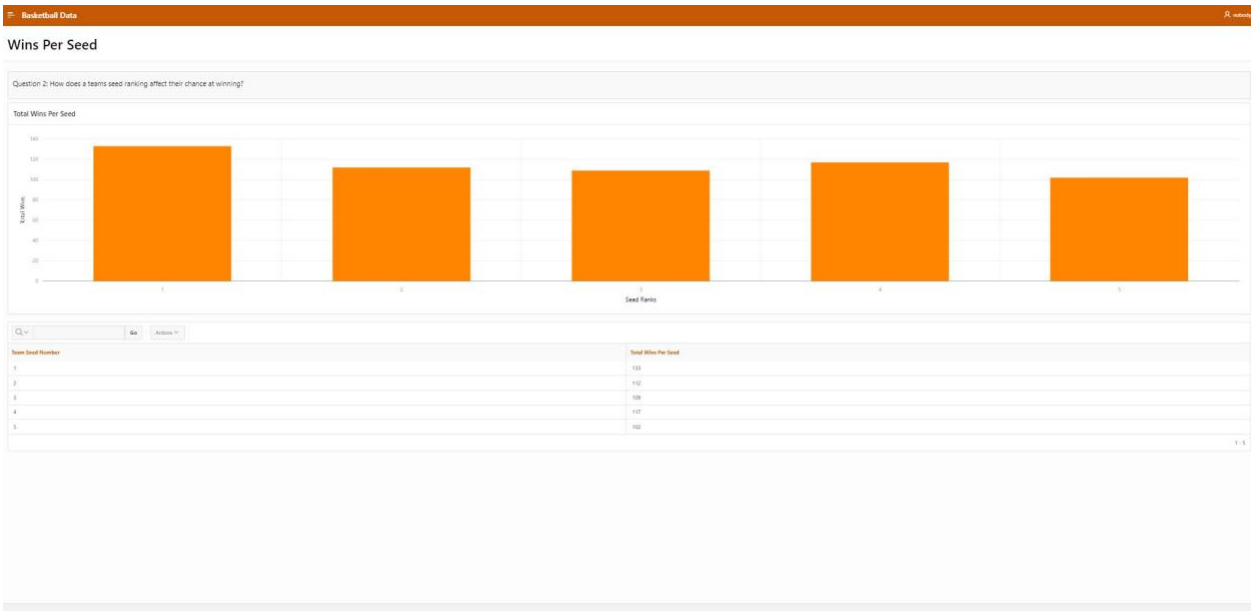


Fig. 18 Wins Per Seed Page

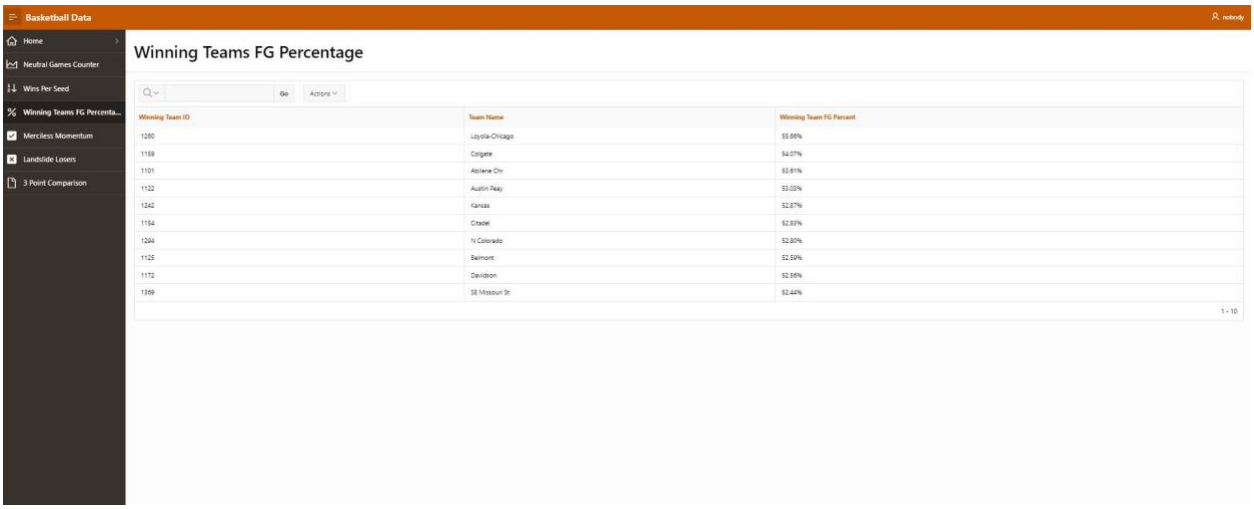


Fig. 19 Winning Team Field Goal Percentage Page

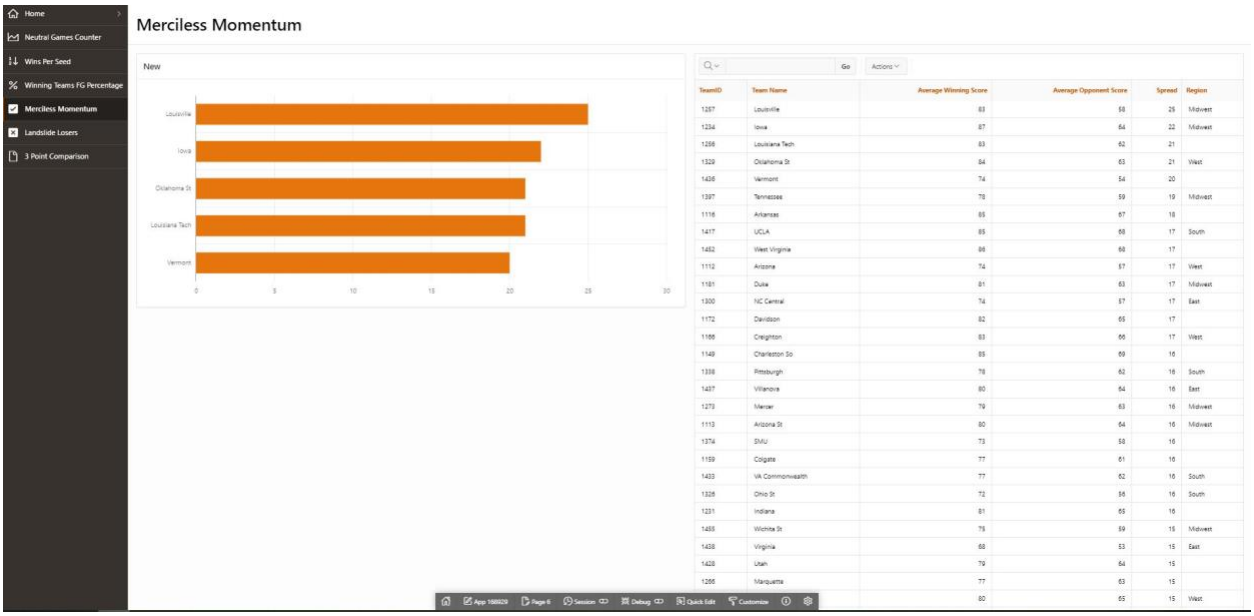


Fig. 20 Merciless Momentum Page

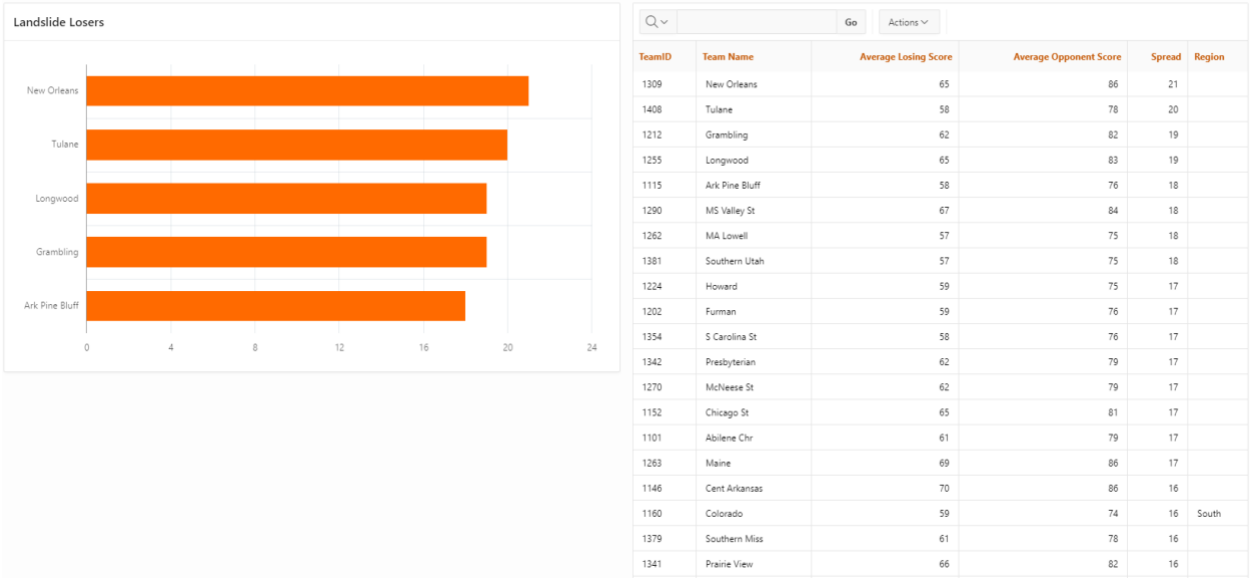


Fig. 21 Landslide Losers Page

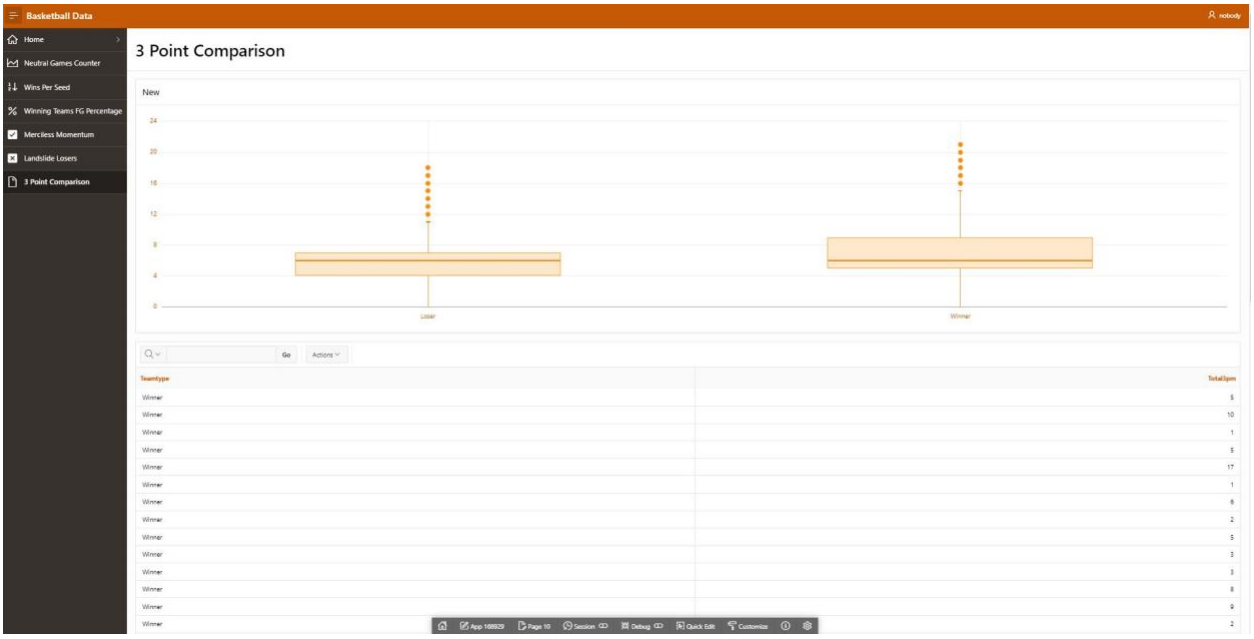


Fig. 22 Winning Team 3 Point Comparison