**NFL Statistics and Arrests – Do better teams have more player arrests?**

STAT 440 Final Project – Fall 2019

Group Members: Josh Janda, Jack Massey, Ashley Arroyo and Isaias Lopez

**Introduction**

The two data files we are using are:

* <https://data.world/chrishayles/nfl-data/workspace/file?filename=receiving.csv>
* <https://www.kaggle.com/patrickmurphy/nfl-arrests>

Statement of the research interest in preparing the data:

Overall, our interest is to gain insight from both datasets alone, as well as together to see how player arrests affect teams and their statistics.

Questions we are attempting to answer:

* Do better teams have a greater number of arrests?
* What makes a `better` team?
* What teams have more arrests? Where are they located?
* What defines a `good` player?
* What do the distribution of player statistics look like?

**Methods**

**Description of the original data files**

The *receiving* dataset contains 14 variables and 67,761 observations regarding receiving player statistics from the 2000-2016 regular season. Some variables we deem important from this dataset are:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| # | | Variable | Type | Length | Format | Label | |
| 1 | | Name | Char | 28 |  | Last, First Name of Player | |
| 3 | Rec | | Num | 8 |  | | Total Number of Receives |
| 7 | Td | | Num | 8 |  | | Number of Touchdowns |
| 2 | Team | | Char | 3 |  | | Team of Player |
| 4 | Yds | | Num | 8 |  | | Total Number of Yards |

The *arrestincidents* dataset contains 8 variables and 850 observations regarding player arrests from 2000-2017. Some variables we deem important from this dataset are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Variable | Type | Length | Format | Label |
| 5 | case1 | Char | 12 |  | Incident Type |
| 6 | category | Char | 36 |  | Incident Crime Categories |
| 1 | date | Num | 8 | MMDDYY10. | Date of Incident |
| 7 | description | Char | 256 |  | Description of Crime |
| 3 | name | Char | 28 |  | Name of Player |
| 8 | outcome | Char | 1 |  | Incident outcome description |
| 4 | position | Char | 3 |  | Player's Position |
| 2 | team | Char | 3 |  | Team of Player |

**Description of the guidelines used to validate the data**

In order to validate the data, certain data requirements were used in order to classify certain observations as invalid. Some data requirements include:

* verifying that there are no missing values
* verifying that there are no extreme outliers
* pct (percentage of first downs for each receive) cannot be greater than 100
* fuml (number of fumbles lost) cannot be greater than fum (number of fumbles)
* rec (total number of receives) cannot be greater than tgt (total number of targets)
* lng (longest gain/receive for week) cannot be greater than 100

**Description of the issues needed to be cleaned and how it will be done**

Both datasets were merged using the variable *team*. However, there were several team names within the *receiving* dataset that were not the same as the team names within the *arrestincidents* dataset. Thus, the set function was used in order to set team names equal within both datasets.

After checking for extreme outliers, it was found that there was one incident where *rec* (number of total numbers of receives) was greater than *tgt* (total number of targets). Thus, the set function was also used in order to fix the invalid observation.

**Description of additional data preparation that you performed**

* Please include more
* In order check whether better teams have more arrests, both data sets were merged using the inner join function.

**Results**

Figure 1: Top Five Players with Most yards in a Game Per Season

|  |  |  |
| --- | --- | --- |
| Last, First Name of Player | Season Year | MaxYds |
| **Smith, Jimmy** | **2000** | **291** |
| **Gardner, Rod** | **2001** | **208** |
| **Burress, Plaxico** | **2002** | **253** |
| **Boldin, Anquan** | **2003** | **217** |
| **Bennett, Drew** | **2004** | **233** |

Figure 2: Top Five Players with Most Passing Yards

|  |  |
| --- | --- |
| Team  of Player | TotYds |
| **NO** | **76701** |
| **IND** | **73155** |
| **GB** | **71558** |
| **NE** | **71081** |
| **PHI** | **68355** |