Homework 1

Joyce LJ Woznica

# Introduction

Each race conducted at a New York Thoroughbred racetrack is observed by three stewards: one employed by the Racing and Wagering Board, one employed by the racing association and one employed by the Jockey Club. At the harness tracks, each race is observed by three judges who all serve as employees of the Racing and Wagering Board. The stewards’ and judges’ viewing stand is located near the finish line of each racetrack and is equipped with several television monitors to permit the viewing of multiple angles of each race. The stewards and judges observe the races and the race grounds to ensure that all conduct is in accordance with rules and regulations.

The hopes are that with citing individuals breaking the rules will help mitigate risk as well as encourage compliance with rules in the future. Since the horse racing industry is plagued with accidents resulting in equine injury, rules exist to attempt to minimize these incidents. Further investigation can shed light on contributing factors and provide insight into what leads to the high number of infractions. Armed with this information, tracks can put the proper constraints in place in the future to minimize infractions that could also minimize incidents.

# Data

To begin such an investigation, a subset of both harness and thoroughbred racetracks within New York State (NYS) has been studied. Although three datasets are available and will be used for my final project in the course, for this homework assignment, only the rulings data set was used for analysis.

## Rulings (Infractions) Data

There are few racing jurisdictions that have equaled New York’s strong stand on integrity in horse racing. Over the past decade, the Board has enabled more capable and qualified people to become stewards and presiding judges and now have a cadre of qualified individuals who have passed newly instituted coursework are monitoring horse racing events.

|  |  |  |
| --- | --- | --- |
| **Data Label** | **Data Type** | **Data Description** |
| Fine Year | Numeric | Year of ruling |
| Notice | Text | Notice ID number |
| Full Name | Text | Licensee’s full name |
| Race Track | Text | Issuing location |
| Race Type | Text | Racing division: Harness or Thoroughbred; “?” indicates unknown |
| Type | Text | Type of ruling: Fine, Suspension, Denial |
| Notice Date | Date | Date of Notice for ruling |
| Rules | Text | Rules violated; public reference list of rules available at http://www.gaming.ny.gov/statutes.php |
| Ruling Text | Text | Text description of the ruling |
| Occupation | Text | Licensee’s occupation |
| Fine Amount | Numeric | Fine amount in dollars |
| Suspension Start Date | Date | Suspension starting date |
| Suspension End Date | Date | Suspension ending date |
| Days Suspended | Text | Number of days suspended or description of period of suspension |

This data was obtained from the New York State government website: <https://data.ny.gov/Government-Finance/Horse-Racing-Rulings-Beginning-1985/igam-2tkj>

This data contains 36,790 rows and 14 columns.

## Overall Data Approach

A subset of the data was used for this assignment to only contain data from the years of 2009 – 2019 so that this dataset could be used in conjunction with the other data sets which only has data from the year starting in 2009. This reduced the overall dataset to 11,516 rows and 14 columns.

No data was removed because when evaluated, the missing data was not required and therefore would not affect the data results.

All data was read into a *pandas* dataframe for manipulation. Outputs with summarized or gathered information from the source data was also placed into a dataframe before output into file formats.

# Analysis

As part of reviewing this data set, the following questions could be explored and analyzed:

### Overall Infractions

#### Infractions by Race Track

Question 1: Which tracks have the most infractions?

In order to review which tracks have had the most infractions, the rulings data set is grouped by “Race Track” and then the number of matching observations for each track is counted to provide a new smaller data frame which I called *rulingsDFbyTrack* which I then reflected in a bar graph of number of infractions by race track shown below.

GRAPH HERE

#### Infractions by Race Type

Question 2: Are there more infractions in Harness or Thoroughbred Racing?

#### Infractions by Individual

Question 3: Are certain individuals receiving higher numbers of infractions that others?

#### Infractions by Occupations

Question 4: What occupation has the most infractions?

#### Infractions over Years by Track

Question 5: What are the number infractions over the years by track?

### Overall Fines

#### Average Fine by Occupation

Question 1: What is the average fine amount by occupation?

#### Average Fine by Year

Question 2: What is the average fine by year (2009 – 2019)?

#### Fines by Year by Occupation

Question 3: What are the total fines per year by Occupation?

#### Fines by Year by Track

Question 4: Are there tracks that have significantly higher overall fines that others by year?

### Common Infraction

Question 1: Are their common threads or words in the descriptions of the infraction?

# Conclusion