Memorandum

|  |  |
| --- | --- |
| to: | Dr. Landowski |
| from: | Joyce Woznica |
| subject: | Project Proposal for IST 652 |
| date: | May 8, 2020 |
|  |  |

I am pleased to submit my project proposal for IST 652 at Syracuse University for the Spring 2020 semester at 7:00 PM on Tuesday evenings.

# Topic

Horse Race Track Infractions and Incidents on New York State Tracks from 2009 - 2019

# Movitation

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 and to protect the riders, drivers and of course, the horses. 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) will be studied. For this project, I will be using at least two (2) of the following sources if not all of three sources.

## New York State Race Track Rulings from 1985 to Present

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 at racing events. Rulings are captured and detailed in this data set.

| **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 |

Source: <https://data.ny.gov/Government-Finance/Horse-Racing-Rulings-Beginning-1985/igam-2tkj>

## Incident Data

Within the state of New York, any incident, death or breakdown of any horse at any facility under the Gaming Commissions’ jurisdiction is reported. Once the data is verified, it is entered into a database that contains thirteen components that detail over four thousand incidents at New York State equine racetracks starting in 2009 to present day. For this discussion, data from 2009 to 2019 was used.

| **Data Label** | **Data Type** | **Data Description** |
| --- | --- | --- |
| Year | Numeric | Year of Incident |
| Incident Date | Date | Date of Incident |
| Incident Type | Text | Type of Incident: Death, Breakdown, Injury, etc. |
| Track | Text | Location of Incident |
| Inv Location | Text | Location on Track |
| Division | Text | Racing division: Harness or Thoroughbred |
| Weather Condition | Text | Weather conditions at time of incident |
| Horse Name | Text | Horse name |
| Trainer | Text | Trainer’s Name |
| Jockey Driver | Text | Jockey or Driver’s Name |
| Incident Description | Text | Text of incident |
| Death or Injury | Text | Death or Injury |

Source: <https://data.ny.gov/Government-Finance/Equine-Death-and-Breakdown/q6ts-kwhk>

## Licensing Data

Participants in New York State racing are required to obtain an occupational license. The character and integrity of all participants are important to the continuity of the sport. Every application is reviewed, updated and carefully processed by the Board’s Licensing unit before a license is issued. Also, by statute, fingerprinting is required of each applicant. Where it is believed that an applicant's background and associates may warrant a finding that his/her participation in racing would be inconsistent with the public interest or the best interest of racing generally, a thorough investigation is conducted before action is taken in the form of a denial, suspension or revocation of his/her license.

| **Data Label** | **Data Type** | **Data Description** |
| --- | --- | --- |
| Person ID | Numeric | Unique Identifier of the Licensee |
| Name | Text | Name of the Licensee |
| Occupation | Text | Occupation licensed |
| Eligibility | Text | Eligible or ineligible to race |
| Division | Text | Racing division: Harness or Thoroughbred |
| License | Text | License Number |
| Expires | Date | Expiration Date |

Source: <https://data.ny.gov/Government-Finance/Horse-Racing-Licensing/cz9u-yj7m>

# Analysis

As part of reviewing this data set, the following information is to be analyzed.

## Infractions

* Which track have the most infractions?
* Are there more infractions in Harness or Thoroughbred Racing?
* Are certain individuals receiving higher numbers of infractions that others?
* What occupation has the most infractions?
* What are the number infractions over the years by track?
* Are their common threads or words in the descriptions of the infraction?
* What is the average suspension length?

## Fines

* What is the highest fines paid by Individual or Occupation?
* What is the average fine amount by Occupation?
* Are certain tracks likely to see more or higher fines?

## Relationship to Incidents and Infractions

* Is there any correlation between incidents and recent infractions?
* Do individuals that have a high number of infractions have a high number of incidents?
* Are their common threads or words in the descriptions of the infraction as related to a possible incident on the same day?

# data Preparation Plan

A subset of the data will used for this assignment to only contain data for the years of 2009 – 2019. This will be done so that the rulings dataset can be used in conjunction with the other data sets which only have data from the year 2009. Data may be removed or manipulated to obtain the proper results and/or graphs for analysis purposes.

All data will be read into a *pandas* dataframe for manipulation. Outputs with summarized or gathered information from the source data was also placed into a dataframe before plotting with *seaborn*.

The code will be well documented and will be separated into the following sections:

* Package Import  
  In this section, the packages that will be used as part of the program are imported.
* Data Load, Structure and Preparation  
  In this section, the data is loaded by reading the *csv* file into a *pandas* dataframe for manipulation. A full path name was provided to the datafile for reading. As mentioned, earlier years were removed from the data so the data only reflects the last decade. Missing data was investigated, but there was no need to remove missing data from the data set.
* Analysis and Graphing  
  The next sections in the program consists of manipulating the data into different subsets to answer specific questions and for graphing purposes. The *seaborn* package was used for graphing.