Project Proposal

Group Members: Edmond Mui, Vanshika Bansal

Kaggle.com Data: `https://www.kaggle.com/lantanacamara/hong-kong-horse-racing`

1 Project Goal

Since the late 1800s, horse racing has intrigued gamblers across the world, generating a worldwide market of over $100 billion USD. Gambling on horses in Hong Kong is particularly interesting because of the small horse pool and limited number of jockeys and trainers. This results in many of the same horses, jockeys, and trainers racing against one another more frequently than they normally would. However, this has not made picking horses any easier for gamblers as the Hong Kong Jockey Club recorded a record breaking $216 billion HKD turnover for the 2017 financial year. Thus, our project's goal is to first determine which horse will win under a specific set of racing conditions and then to potentially come up with a long-term betting strategy for Hong Kong horse racing.

2 Applications of Dataset

The dataset we will use for this project is the actual result of several races in Hong Kong through 2014 to 2017. This dataset contains several attributes of horses that raced, including their weights, the lanes they ran in, their jockeys and trainers, and their positions at different times throughout the race. The data set also contains information specific to races, such as the race course, track condition, length of the track, and the date of the race. This information can also be used to stipulate other conditions, such as the weather on the day of the race using historical weather data in Hong Kong available online. Using several forms of analysis and cross validation, we can use the data available to determine which horses are most likely to win given a specific set of race conditions. In addition, we can also use the results of these races to determine whether or not any generalized betting strategies are profitable.

3 Problem Significance

The significance of our project, if successful, is through its value as patented research. If we can accurately determine which horses will win races, horse breeders (a multi-billion dollar industry) would be willing to purchase our findings to more accurately breed horses to win specific races. In addition, if we can develop profitable betting strategies, there will be heavy demand for our research from participants of horse race betting. Although finding profitable research will be difficult (and likely to have been tried by many previously), we believe that the specific characteristics of Hong Kong horse racing are favorable when using data analytics to generate insightful results and findings.