

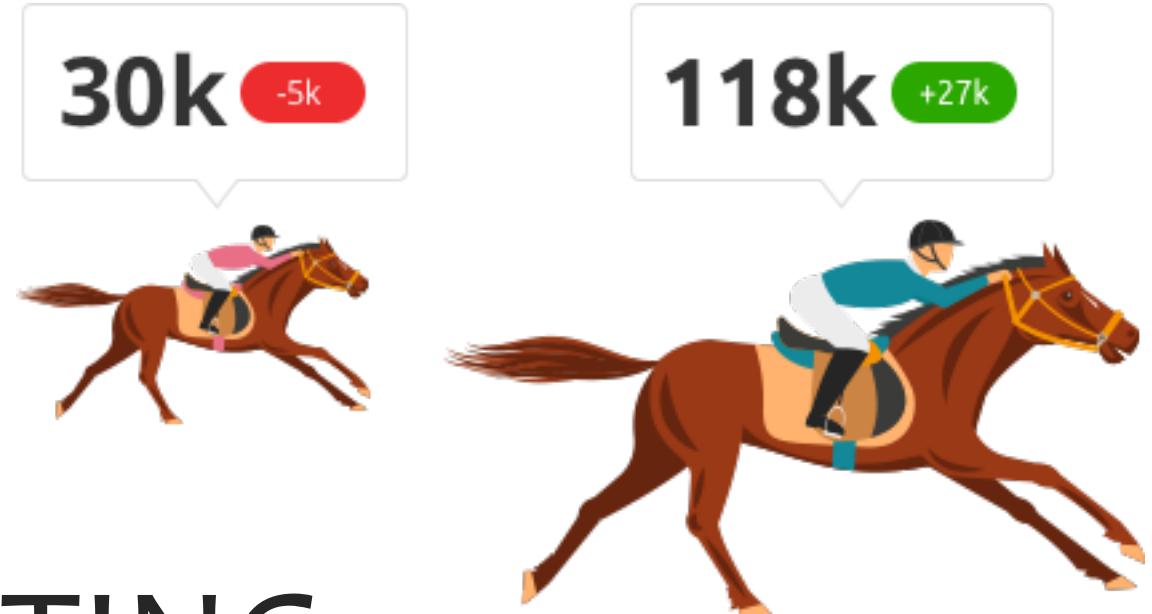


Python Project

Team Members: KARISHMA **VELAGALETI** | JIANG **SHANG** |
SATYARTHA PRAKASH **POPPASANI** | ADITYA **TRIPATHI**



SIMON
BUSINESS
SCHOOL



BUSINESS SETTING

Business Context and Problem Statement



Horse Racing & Our Problem Statement

HORSE RACING

- ❖ Most sought after gambling sport
 - Most Lucrative gambling activity



PROBLEM STATEMENT

- ❖ Build a predictive model for the following:
 1. To identify the finishing position of all horses in a race
 2. To identify the winning horses in a race



OUR DATA

Sources & challenges



Our DataSet

❖ Data:

- Horse Racing (n= 19997)
 - ✓ **Features:** Races_distance, trainer, actual_weight, jockey, win_odds
- Races (n=1562)
 - ✓ **Features:** track, class, race_distance

❖ Source: Kaggle

	finishing_position	horse_number	horse_name	horse_id	jockey	trainer	actual_weight	declared_horse_weight
1								
2	7	6	A BEAUTIFUL	T421	C Y Ho	W Y So	120	1082
3	8	6	A BEAUTIFUL	T421	B Prebble	W Y So	124	1080
4	10	7	A BEAUTIFUL	T421	H W Lai	W Y So	120	1079
5	6	8	A BEAUTIFUL	T421	H W Lai	W Y So	119	1060
6	7	11	A BEAUTIFUL	T421	C Y Ho	W Y So	115	1062
7	10	14	A BEAUTIFUL	T421	H W Lai	W Y So	114	1066
8	9	2	A BEAUTIFUL	T421	C K Tong	W Y So	130	1064
9	7	2	A BEAUTIFUL	T421	H W Lai	W Y So	132	1041
10	11	3	A BEAUTIFUL	T421	H W Lai	W Y So	128	1054
11	1	7	A BEAUTIFUL	T421	H W Lai	W Y So	126	1057
12	5	1	A BEAUTIFUL	T421	H W Lai	W Y So	133	1060
13	12	5	A FAST ONE	S424	M Demuro	D E Ferraris	126	1093
14	12	5	A FAST ONE	S424	C Reith	D E Ferraris	125	1077
15	6	7	A FAST ONE	S424	R Hughes	D E Ferraris	123	1077
16	8	11	A FAST ONE	S424	D Whyte	D E Ferraris	121	1078
17	9	12	A FAST ONE	S424	U Rispoli	D E Ferraris	120	1072
18	9	10	A FAST ONE	S424	C Y Ho	D E Ferraris	116	1058
19	3	14	A FAST ONE	S424	C Y Ho	D E Ferraris	114	1059



Challenges



Data Cleaning

Converting data into dummy values

More features required



RULES : ‘before’ Get Set Go...



SIMON
BUSINESS
SCHOOL



Rules of the game!

❖ Rule #1:

- number of horses participating in a single race can not exceed 14

❖ Rule #2:

- Horses are ranked pre-race by horse number. A handicap system is installed to balance out each horse's natural ability by assigning heaviest weight to the best horse, and it is assigned to No.1

❖ Rule #3:

- Gate number for each horse is drawn two days before the event, the lower the draw number, the closest to the inside rail. Therefore better performance is expected

❖ Rule #4:

- Horses are also categorized by class (class 1 being the best cohort). Horses generally compete within their own class with a few exceptions



ANALYSIS & RESULTS

- Exploratory Analysis
 - Predictive Analysis – Part 1
 - Predictive Analysis – Part 2

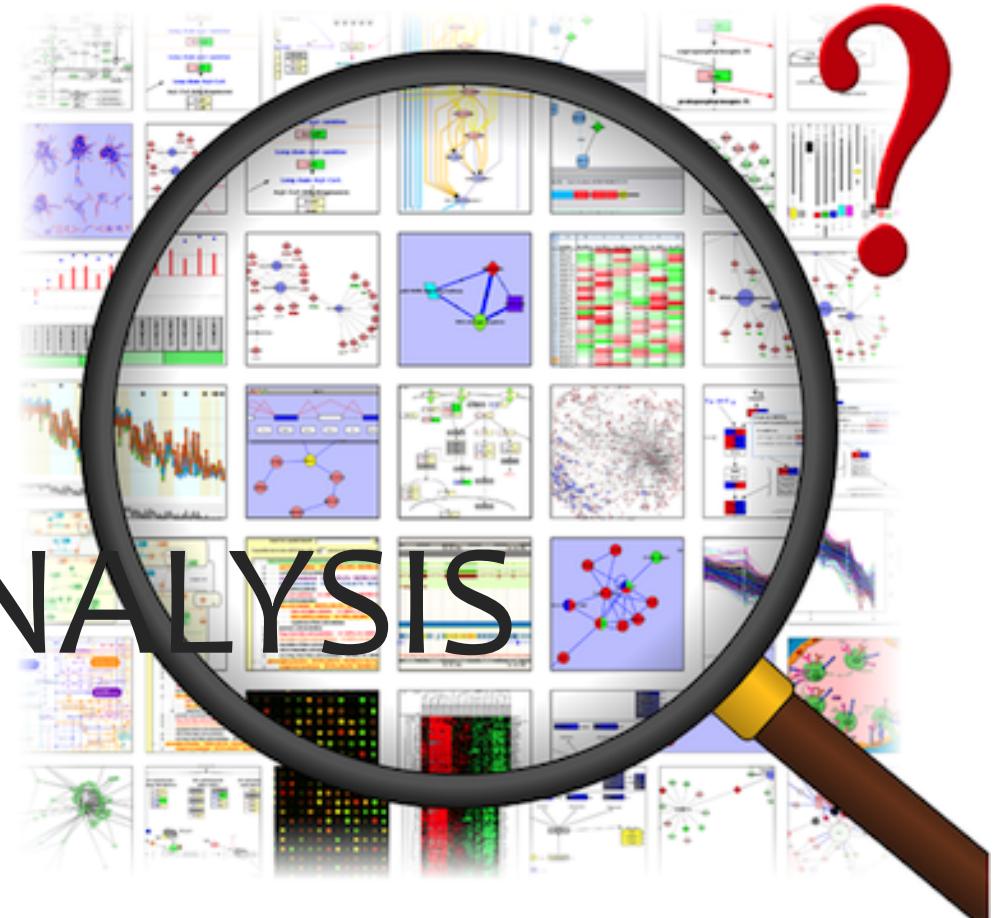


SIMON
BUSINESS
SCHOOL



EXPLORATORY ANALYSIS

Let's find out what our data speaks?

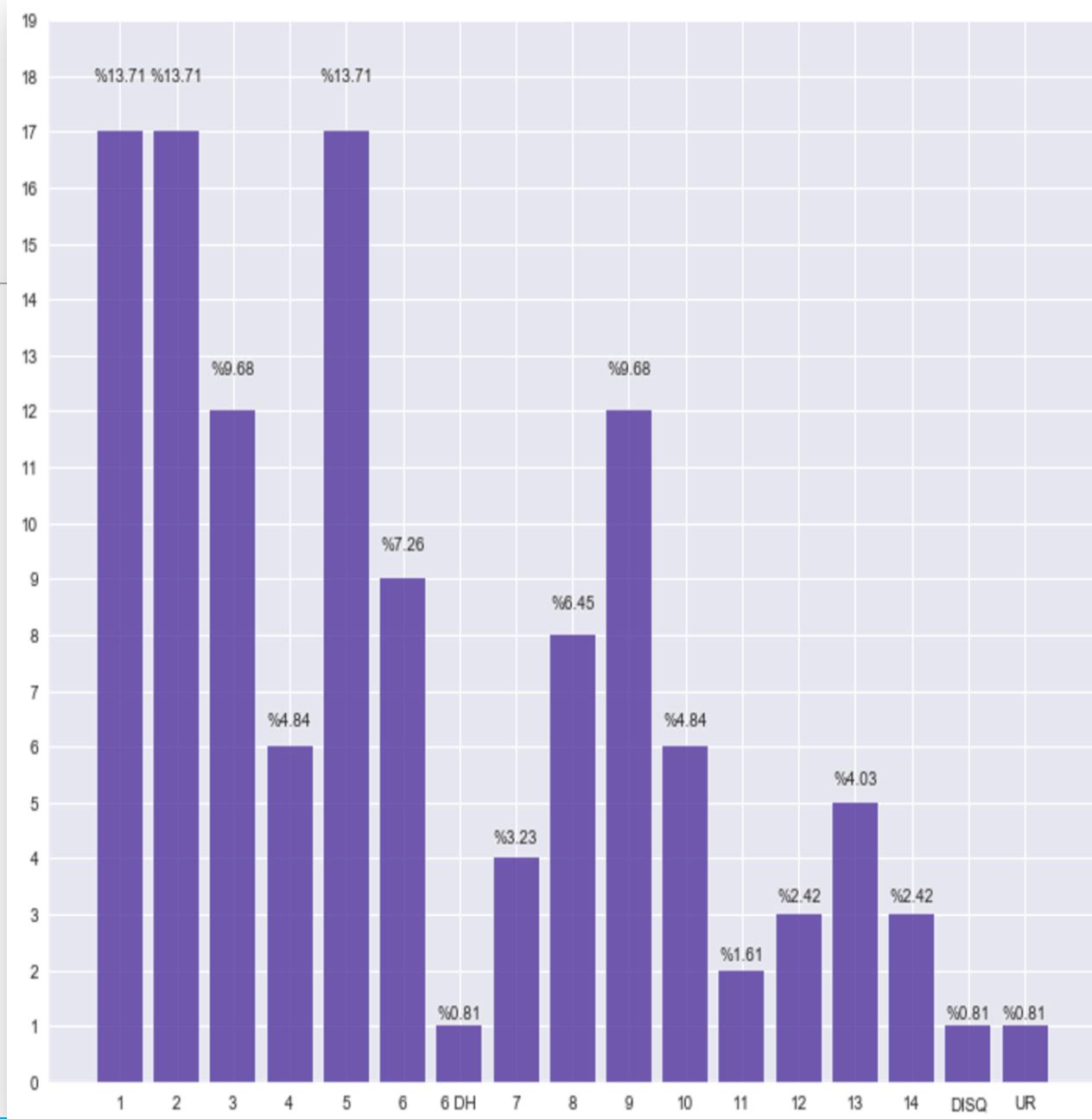


SIMON
BUSINESS
SCHOOL
UNIVERSITY OF
ROCHESTER



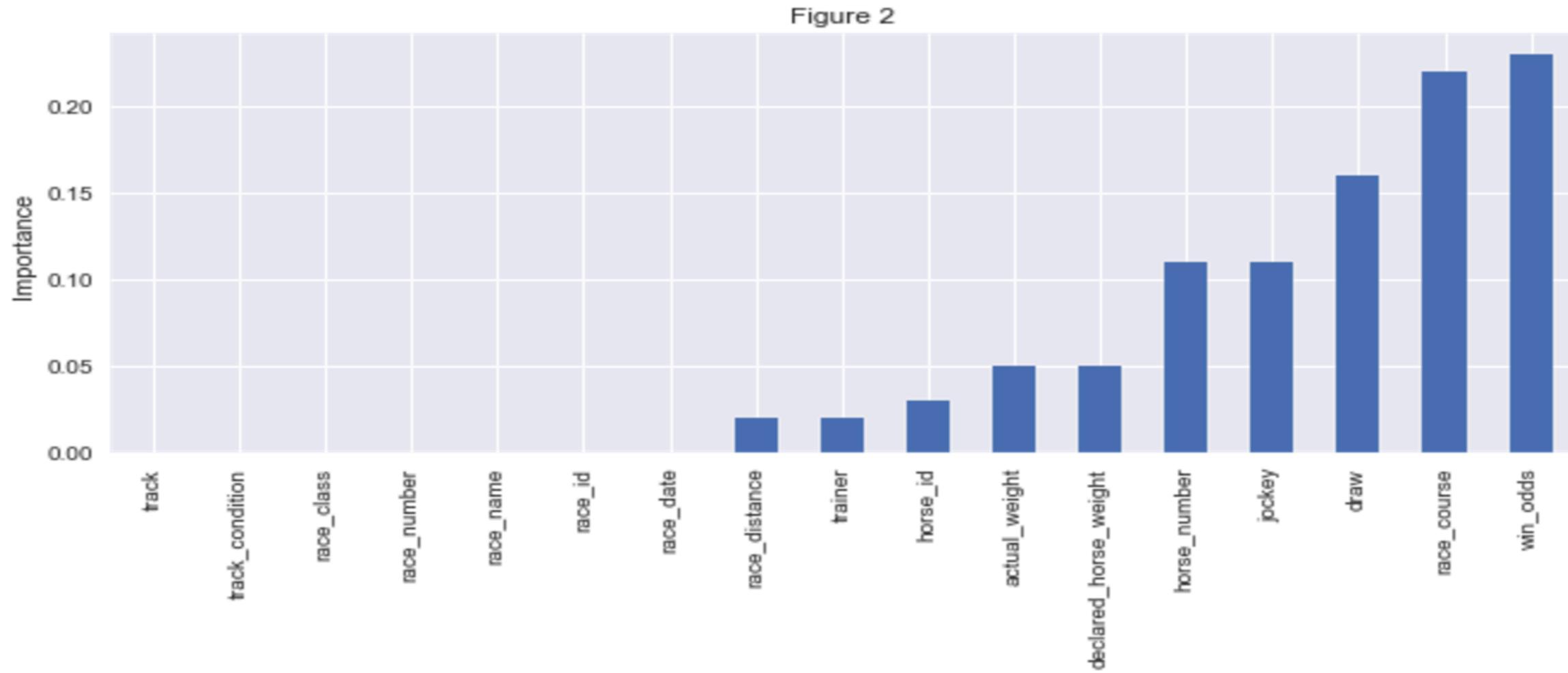
Two Questions

- ❖ Out of 1561 races, what is the possibility that horse_number==1, and draw is also ==1?
- ❖ In these 124 cases where there is the apparent advantage, can those horses always win?





Which attribute is the greatest determinant?



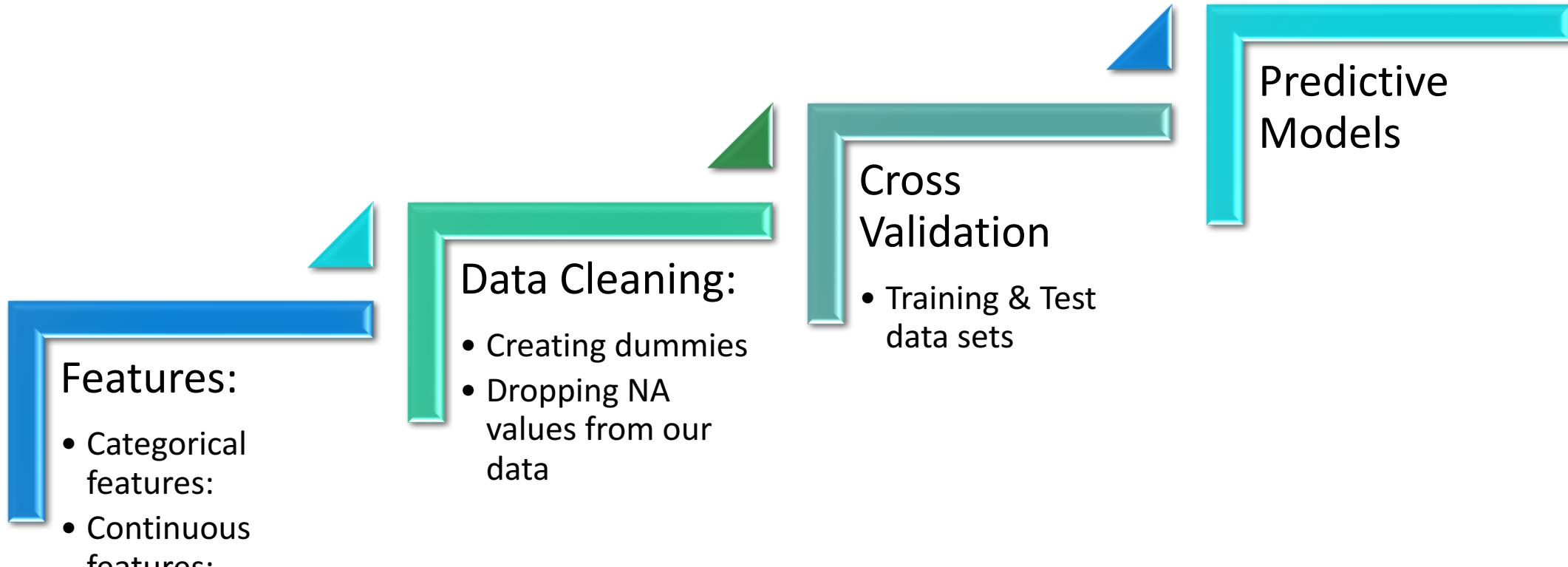


DATA ANALYSIS: PART 1

Predict the finishing position of horses for a given race



Methodology





Analysis

Model Used	Training	Test
Random Forest	12.95%	13%
Neural Network	27.60%	10.20%
K-Nearest Neighbour	13.64%	10.01%
Decision Tree Classifier	99.97%	8.78%
Boosting(DT as base)	9.18%	8.65%
Bagging	9.12%	8.26%
Gaussian Mixture	-63.18%	-64.35%



DATA ANALYSIS: PART 2

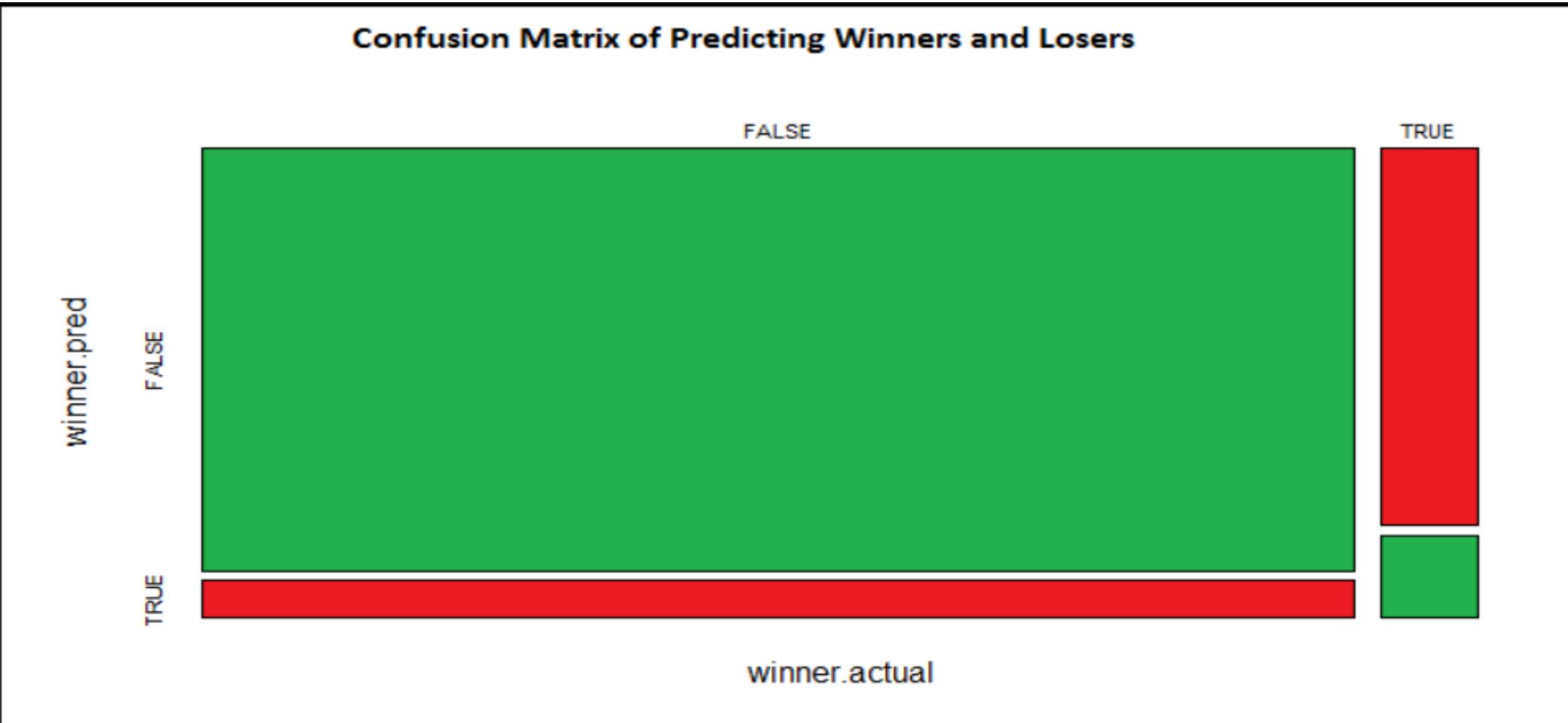
Predict if a horse will win the race or not



SIMON
BUSINESS
SCHOOL



Analysis – Identify the winning horse





Conclusion



Analysis 1: Finishing position of the horses

- Accuracy of the Model
- Best Model

Analysis 2: Winning horse



- Good accuracy but speaks very little about real world

