FLIP00 Final Assessment

Cong Ma

2020-10-08

Overview

Problem Definition

Kaggle Project Introduce

Overall Research Ideas

Overall Research Ideas

Data visualization

Check the data ,Data cleaning,Data process

Data visualization

Build the model

Build the model

Build and train the model and produce the submission data

Conclusion

Problem Definition

Kaggle Project Introduce

Defn

Kobe Bryant marked his retirement from the NBA by scoring 60 points in his final game as a Los Angeles Laker on Wednesday, April 12, 2016. Drafted into the NBA at the age of 17, Kobe earned the sport's highest accolades throughout his long career. Using 20 years of data on Kobe's swishes and misses, can you predict which shots will find the bottom of the net? This competition is well suited for practicing classification basics, feature engineering, and time series analysis. Practice got Kobe an eight-figure contract and 5 championship rings. What will you get from it?

Overall Research Ideas

Overall Research Ideas

- Step One - Check the data ,Data cleaning,Data process
- Step Two - Data visualization
- Step Three - Build the model and select the optimal parameters
- Step Three - Visualization parameters and process the test data

Data visualization

Check the data ,Data cleaning,Data process

- First,let's check the data.

(30697, 25)

Out[4]:

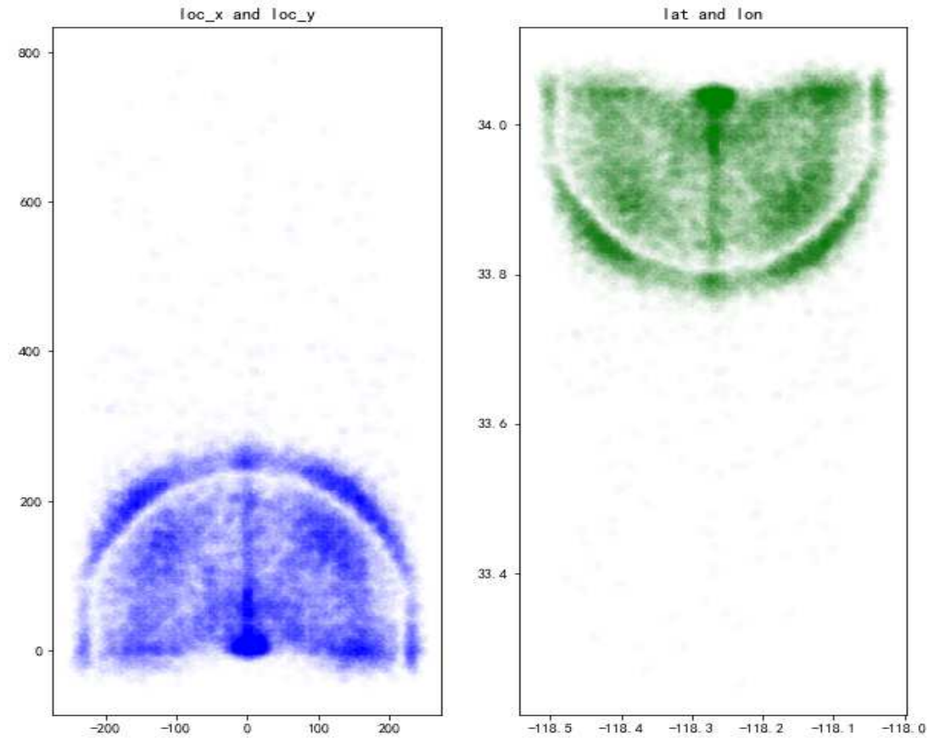
	action_type	combined_shot_type	game_event_id	game_id	lat	loc_x	loc_y	lon	minutes_remaining	period	...	shot_type	shot_zone_area	sho
0	Jump Shot	Jump Shot	10	20000012	33.9723	167	72	-118.1028	10	1	...	2PT Field Goal	Right Side(R)	
1	Jump Shot	Jump Shot	12	20000012	34.0443	-157	0	-118.4268	10	1	...	2PT Field Goal	Left Side(L)	
2	Jump Shot	Jump Shot	35	20000012	33.9093	-101	135	-118.3708	7	1	...	2PT Field Goal	Left Side Center (LC)	
3	Jump Shot	Jump Shot	43	20000012	33.8693	138	175	-118.1318	6	1	...	2PT Field Goal	Right Side Center(RC)	
4	Driving Dunk Shot	Dunk	155	20000012	34.0443	0	0	-118.2698	6	2	...	2PT Field Goal	Center(C)	R

5 rows × 25 columns

- There must be variables in so much data that have nothing to do with our model, and we will remove them in some way.

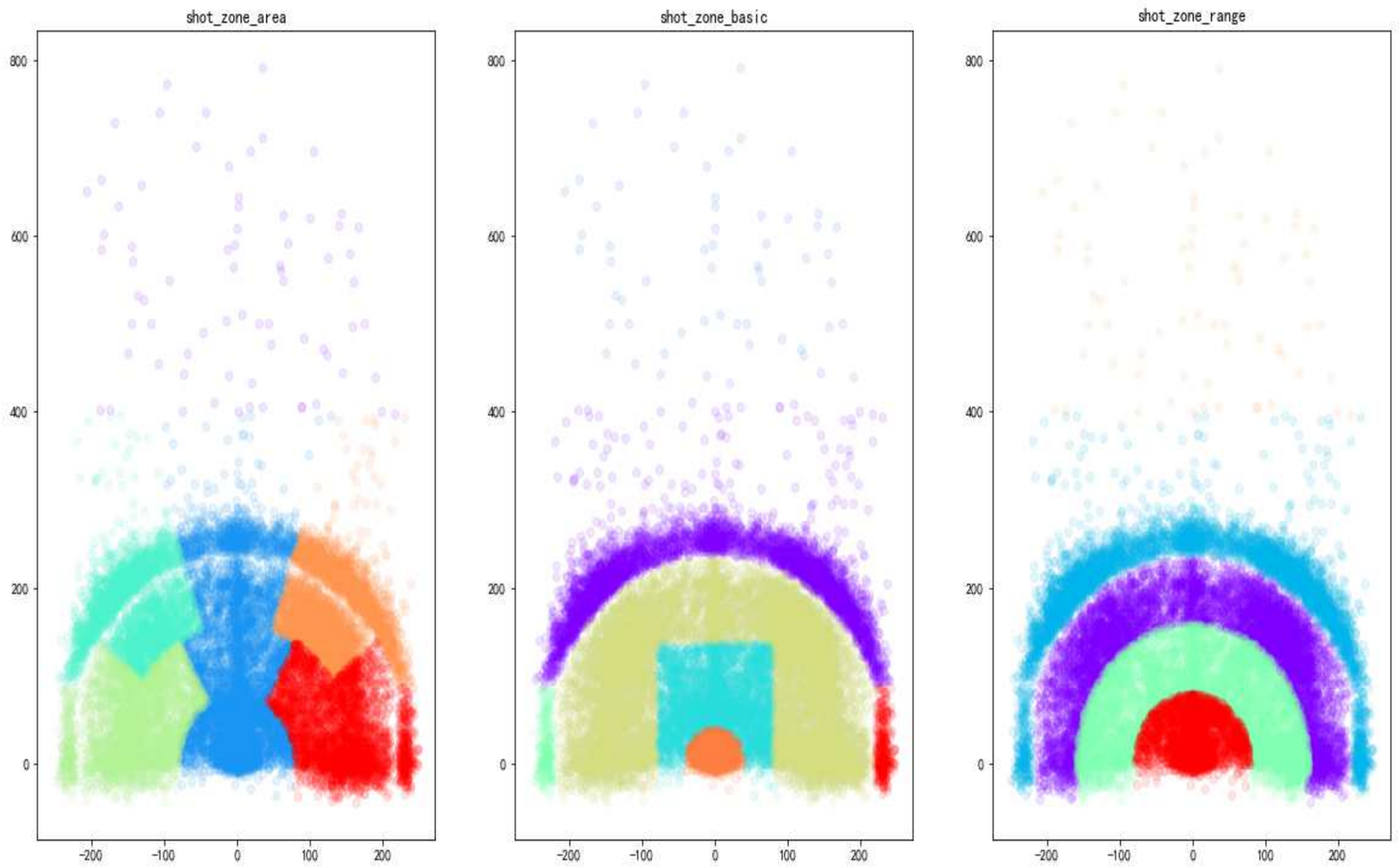
Data visualization

- After we delete blank data,we create charts to check the relationship between variable.



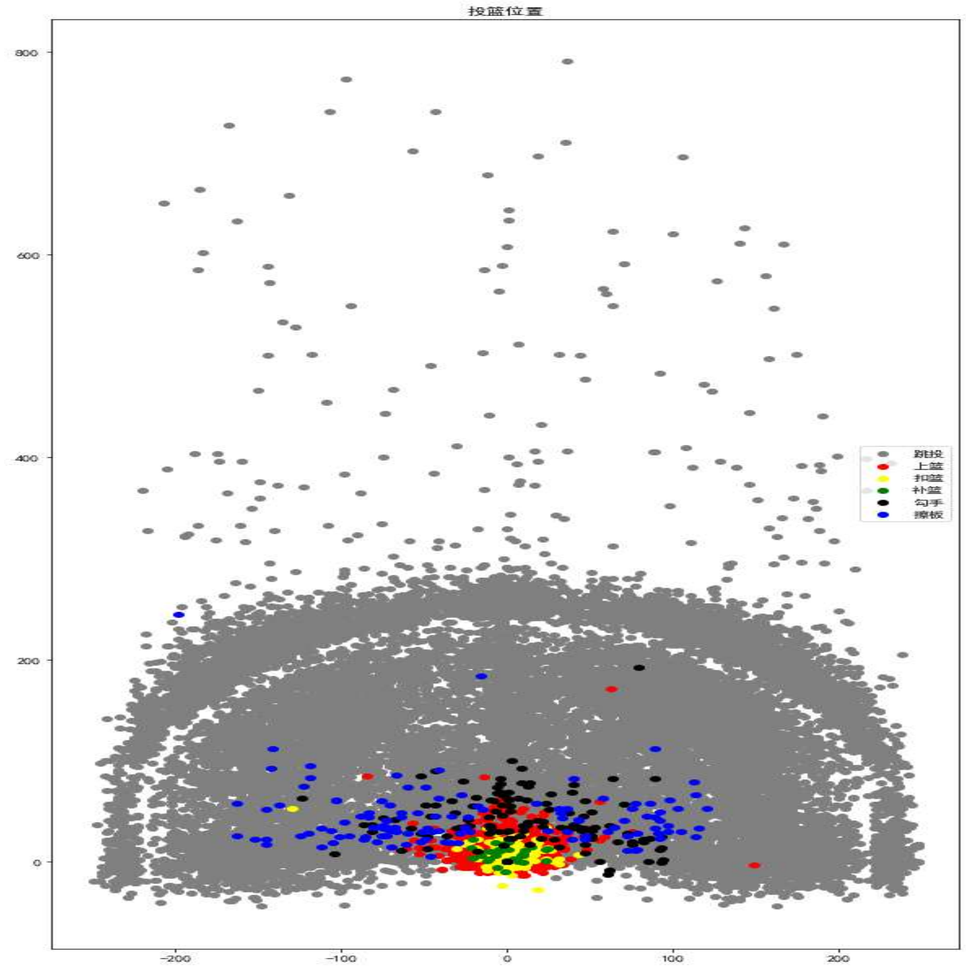
- We can make sure that the shooting position and latitude and longitude are basically consistent with the court.

Data visualization



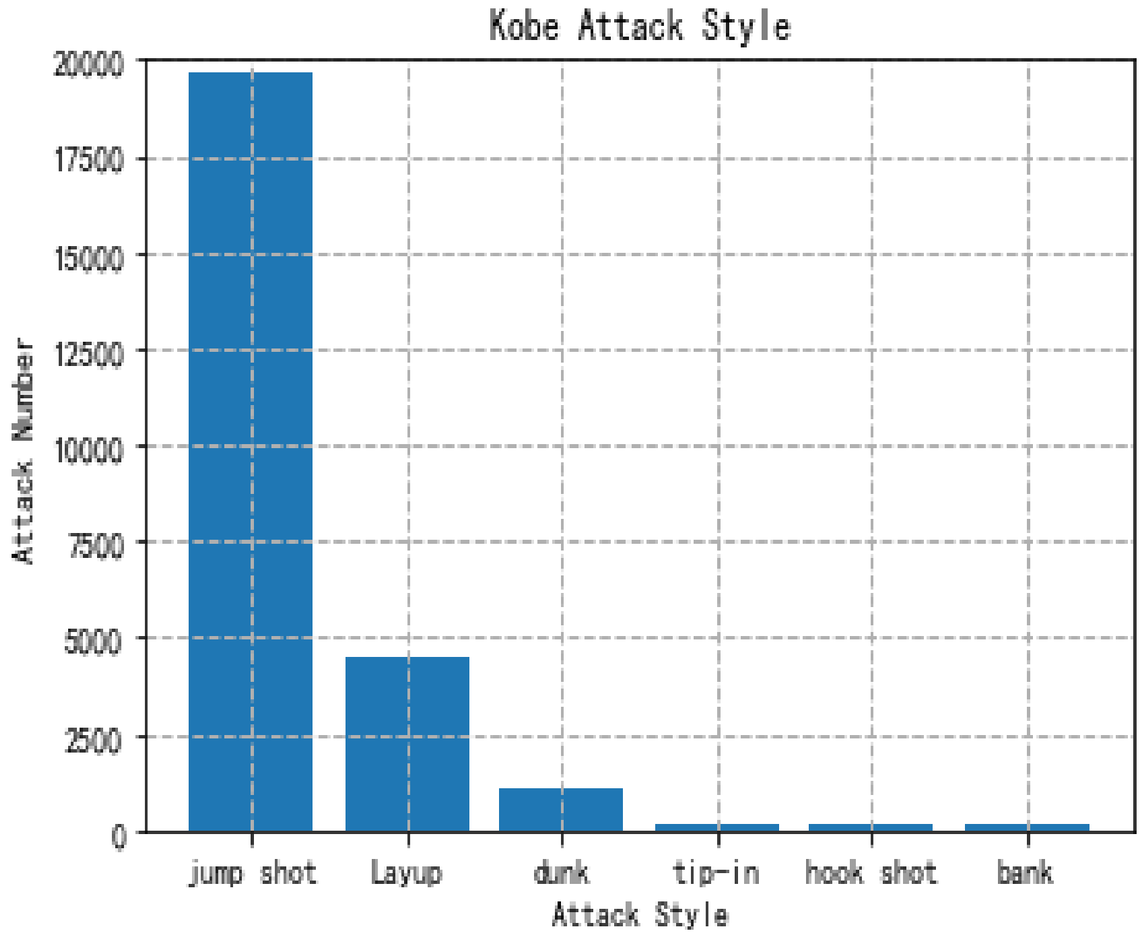
- The shotzone also has the same characteristic whth the lat and lon.

Data visualization



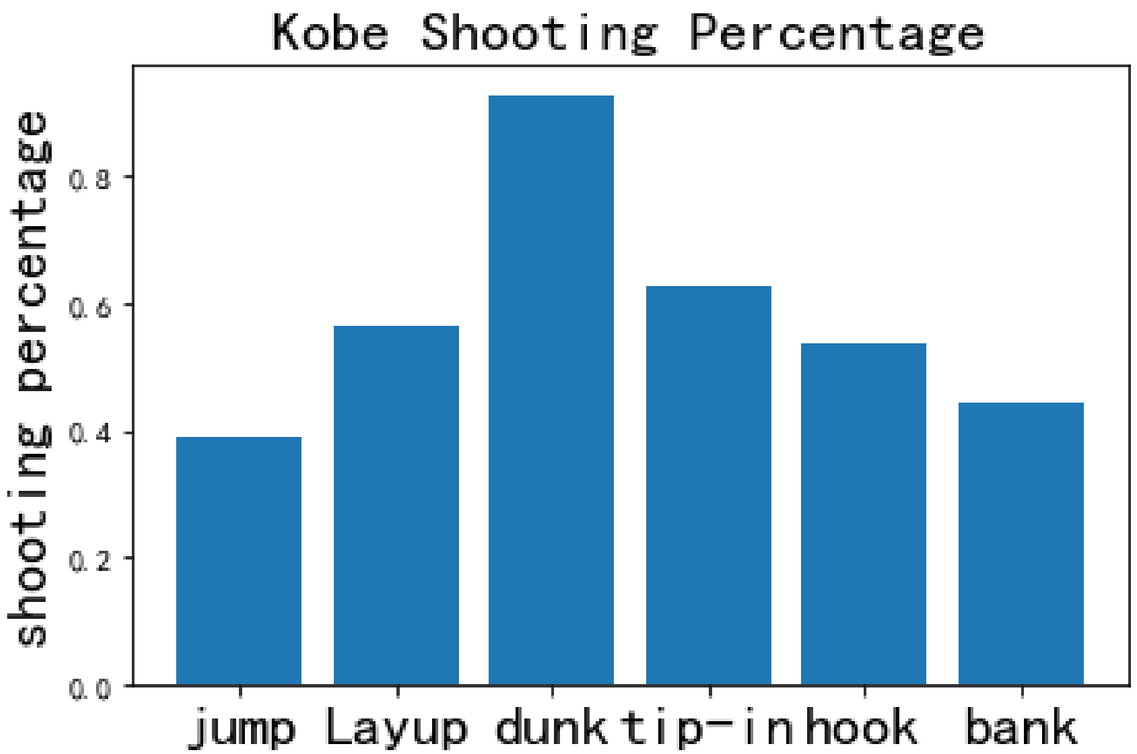
- The shottype also has the same characteristic whth the lat and lon.

Data visualization



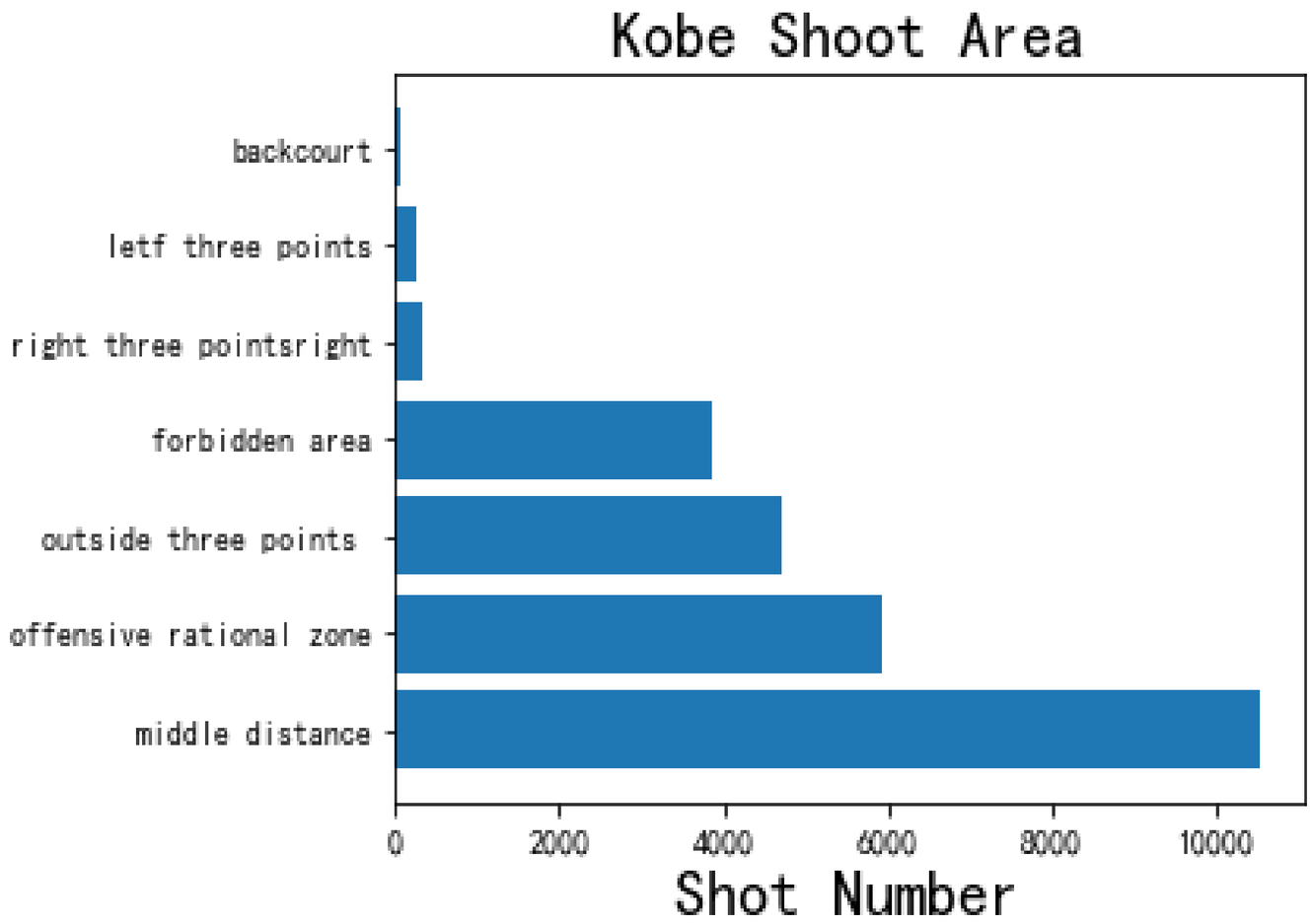
- Kobe's way of shooting, it is clear that the most jump shot.

Data visualization



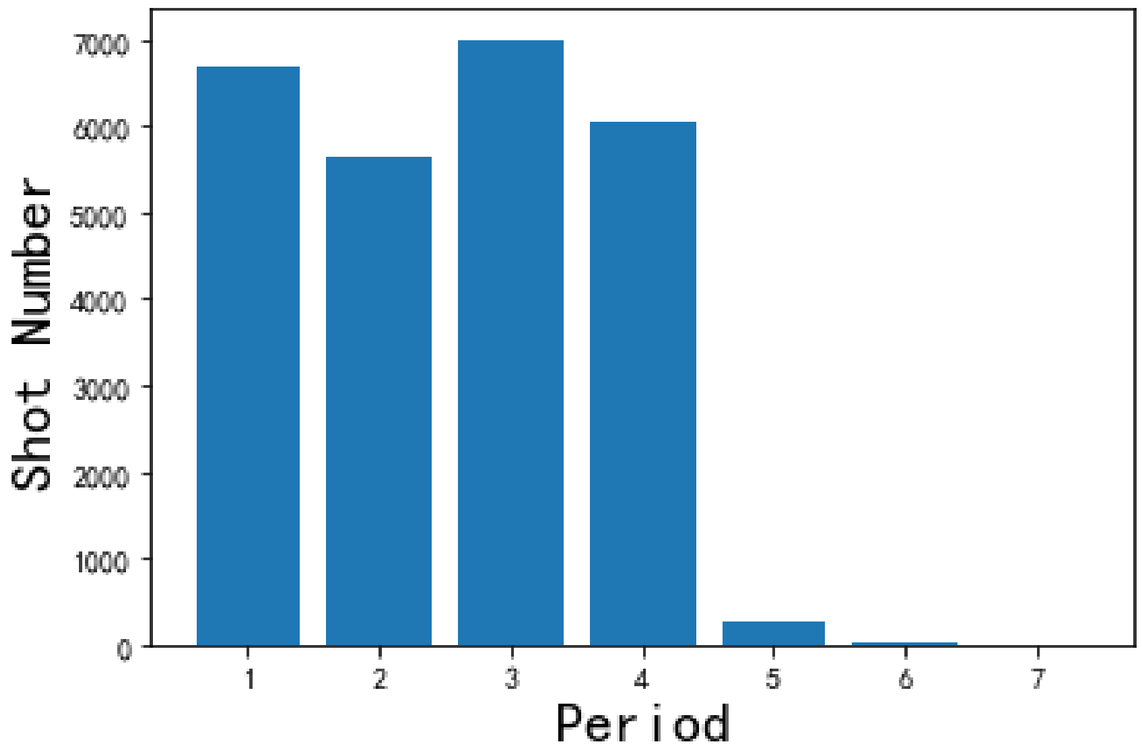
- On the shooting percentage, the dunk is undoubtedly the highest.

Data visualization



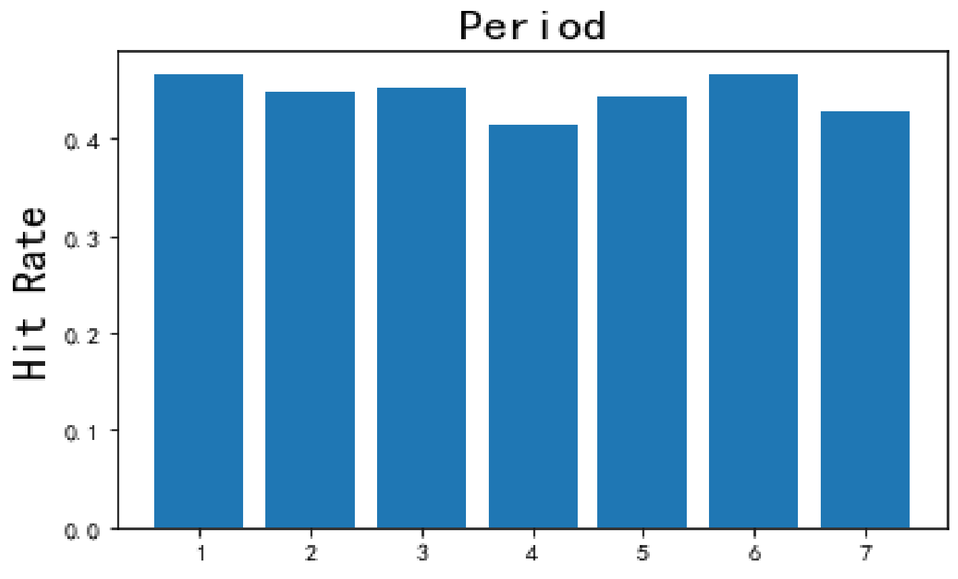
- From the shooting distance, most of Kobe is in the middle distance.

Data visualization



- Kobe's shooting times in each quarter are the most in the third quarter and the second in the first quarter.
- Because the 5, 6 and 7 quarters are overtime games, naturally much less.

Data visualization



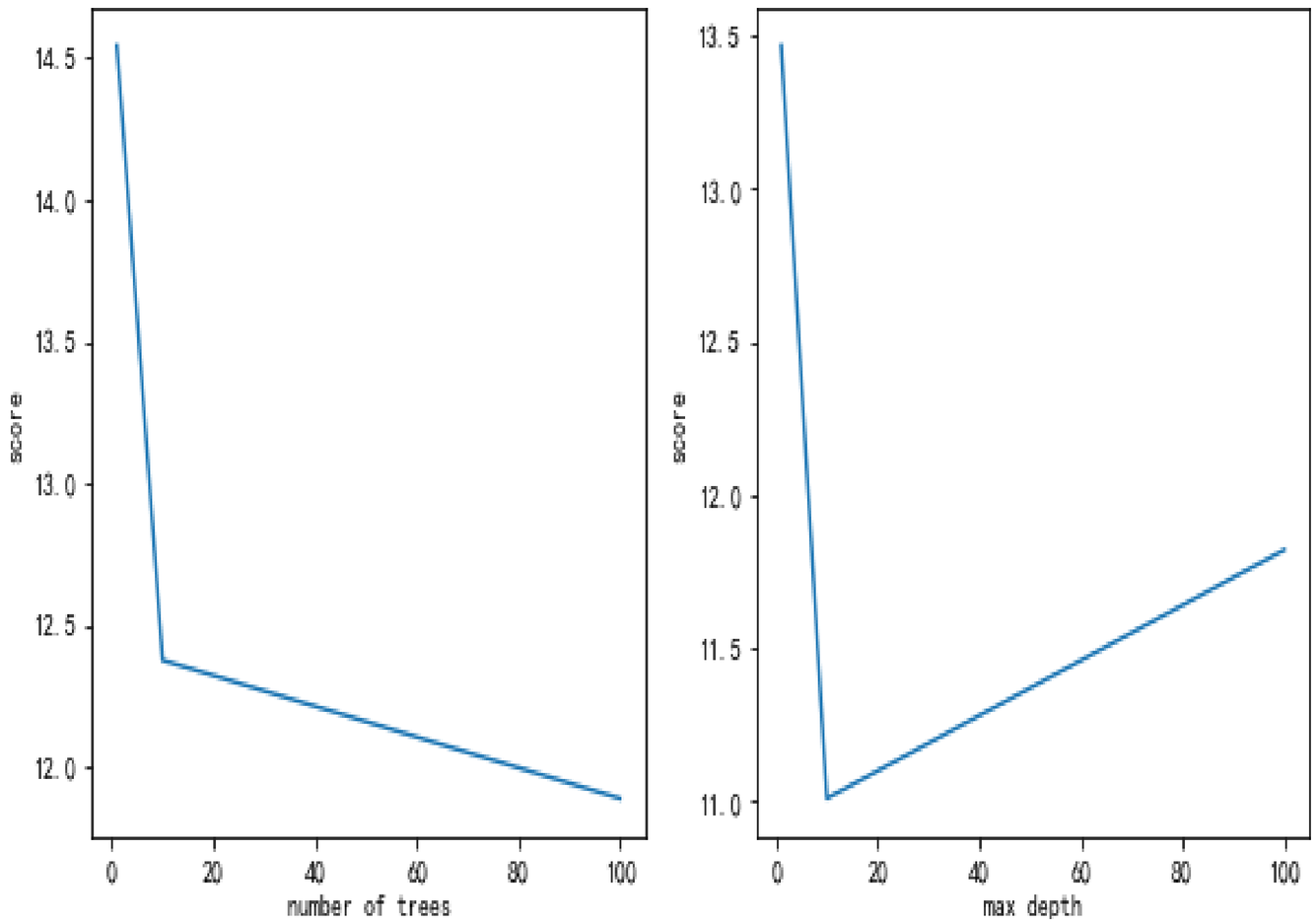
- The shooting percentage of each section can be seen that Kobe's shooting percentage is the lowest in the fourth quarter.
- It can be seen that physical strength has some influence on Kobe's shooting percentage.

Build the model

Build the model

- We need to eliminate duplicate data before we build models.
- Encoded the data by one-hot.
- Choose the minimum value of loss and the number of optimal trees, the optimal value of maximum depth.

Build and train the model and produce the submission data



Conclusion

Conclusion

Thanks For Your Listening

Made by MC

QingDao technological university