Flip00 Project Final Presentation

Jin Chen
HuNan University

November 21, 2019

Overview

Problem Statement

- Problem Definition
- Data Set

Exploratory Data Analysis

- Data Visualization

Data Preparation

- Data Cleaning
- Data Transformation

Methods

Forecast Results

- Forecast Results

Conclusion

Thanks for attending and welcome for questions

Problem Statement

Problem Definition

Defn The data contains the location and circumstances of every field goal attempted by Kobe Bryant took during his 20-year career. The task is to predict whether the basket went in (shot_made_flag).

Train Data and Test Data

There are 30697 lines of data in the training set.I will split the dataset as training sets and testing sets. They have removed 5000 of the shot_made_flags (represented as missing values in the csv file). These are the test set shots for which we need submit a prediction. We are provided a sample submission file with the correct shot_ids needed for a valid prediction.

Data Set

Defn The action_type,shot_made_flag, shot_type and shot_zone_area are part of the attributes of each sample, the fllowings are the meaning of some attributes.

Data List

Attribute	Note
action_type	Jumpshot,Layup,Dunk,Tipshot,Hookshot,Bankshot
loc_x ,loc_y	shots point
shot_made_flag	1=Yes,0=No
shot_type	2PT Field Goal,2PT Field Goal
shot_zone_area	shots area by area
shot_zone_basic	shots area by NBA rules
shot_zone_range	shots area by radius

Exploratory Data Analysis

Data Visualization

Exp

Use EDA to plot the distribution of the data, can observate the data intuitively and find the relation between the attribute values.

- Figures
  - Histogrm
  - Scatter Plot
  - Line Chart

Data Visualization

Exp It can be seen that dunk is the highest hit rate, followed by bank shot is about 80%, while jump shot and Tip shot are relatively difficult.

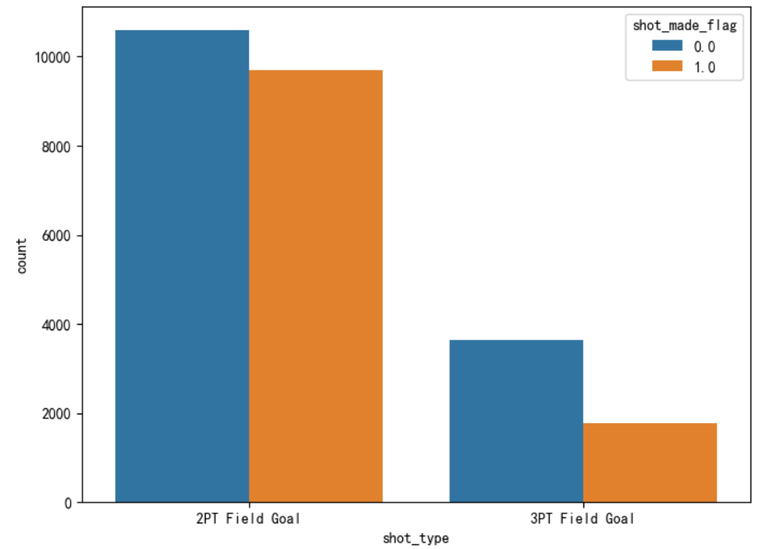

Figure 1: The hit distribution histogram of two shot types

Data Visualization

Exp

It can be seen that dunk is the highest hit rate, followed by bank shot is about 80%, while jump shot and Tip shot are relatively difficult.

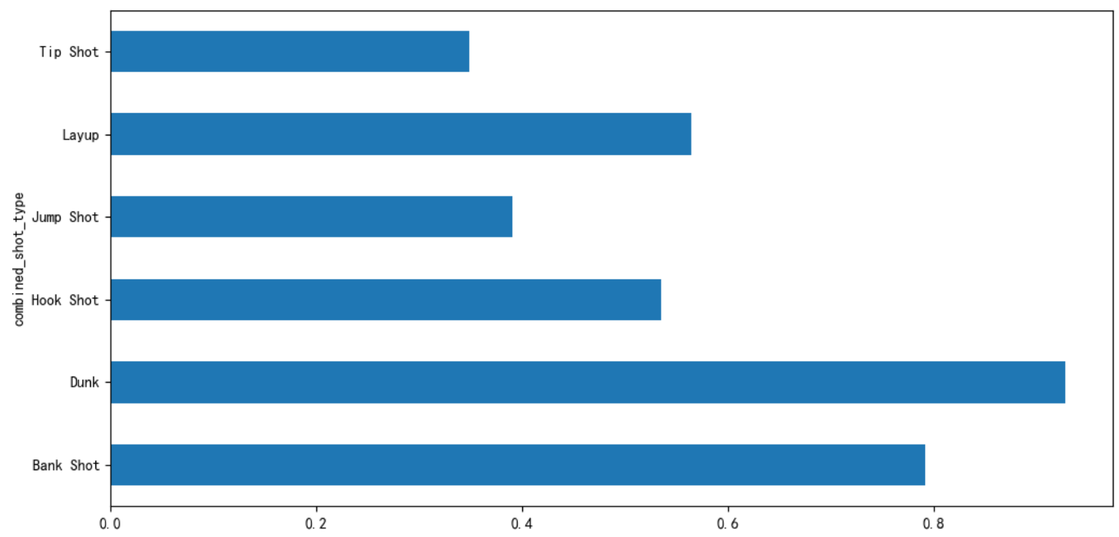

Figure 2: the shot accuracy of various action type

Data Visualization

Exp Using the scatter plot we can combine multiple categorical value series on to the same chart distinguishing them using color or variation in symbol. Lets get some understanding about the different zones and the shots made from zones.

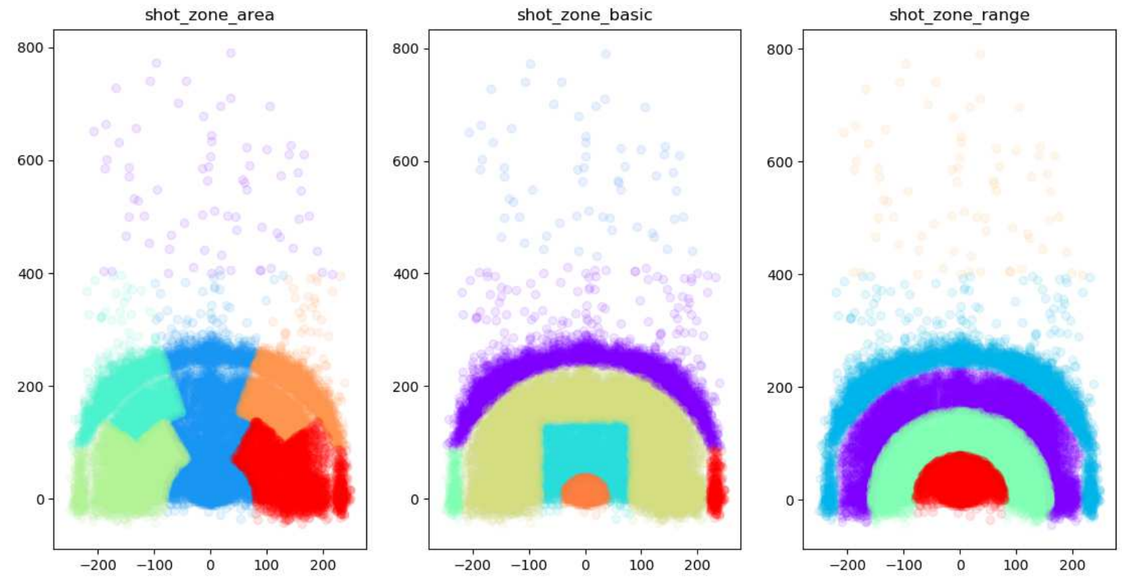

Figure 3: Division of shooting area

Data Visualization

Exp The line chart can not only show the quantity, but also clearly see the increase and decrease of data. Lets now see the Kobe's shots positioning with the time and distance.

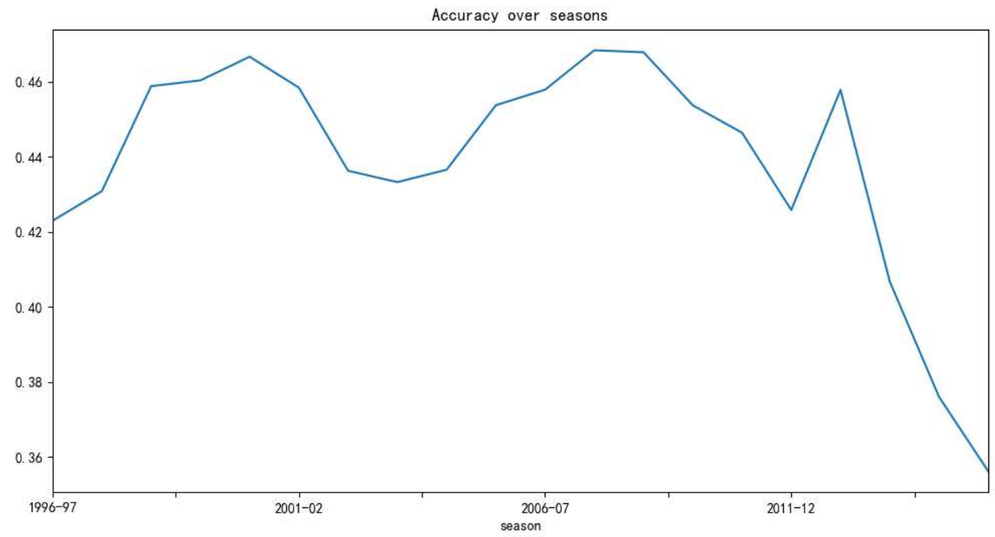

Figure 4: shot accuracy of each seasons

Data Visualization

Exp The line chart can not only show the quantity, but also clearly see the increase and decrease of data. Lets now see the Kobe's shots positioning with the time and distance.

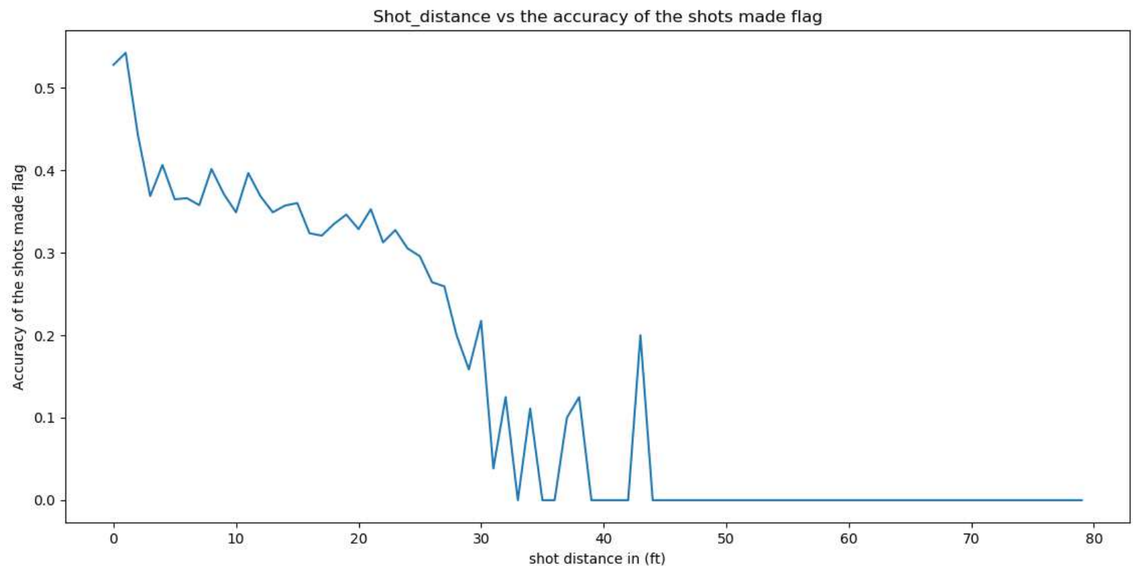

Figure 5: Shot_distance vs the accuracy of the shots made flag

Data Preparation