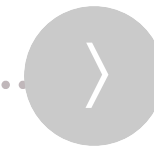


PE 프로그램 진행과정 (7.16)

01



02



03

데이터 전처리 함수 작성

RAW DATA 확보 후
함수를 통해 데이터 전처리

추 후 타 시즌 Raw Data를
빠르게 처리하기 위한 절차

데이터 병합 과정에서의 오류

EDA과정에서 발견하지 못했던
한 날짜의 두 경기가 있는 경우

merge시 데이터의 길이가 바뀌는
오류 발생

데이터 EDA

사용할 feature 선정
경기 결과와 feature와의 상관성
scatter plot

PE 프로그램 진행과정 (7.16)

04

새로운 Raw Data 수집

전처리 과정에서 데이터 손실이
있어 2019, 2020년의 데이터
추가 수집

To do list

2019 / 2020년 Raw data 수집
함수 코드를 사용하여 전처리

구장별 성적, 상대 투수 유형 별 성적
을 구하는 함수 만들기

데이터를 사용함에 있어 시즌 초반의
경기를 어느 정도 제외할 것인지에
대한 고민과 함께

MLP model code