master-dataset.ipynb

Scope: Construct the working dataset based on three files: FanGraphs-Pitching-Standard.csv, FanGraphs-Pitching-Advanced.csv, and FanGraphcs-Pitching-Value.csv. Export the created dataset to a .csv file for us in WAR-Predictor.ipynb.

In [1]:
```python
import pandas as pd
import os
```

In [3]:
```python
#Reading FanGraphs data into Pandas dataframes
pitching_standard = pd.read_csv('FanGraphs-Pitching-Standard.csv')
pitching_advanced = pd.read_csv('FanGraphs-Pitching-Advanced.csv')
pitching_value = pd.read_csv('FanGraphs-Pitching-Value.csv')
```

In [4]:
```python
#Removing irrelevant statistics from pitching_value
pitching_value = pitching_value.drop(columns=["RA9-WAR", "BIP-Wins", "LOB-Wins", "FDP-Wins", "RAR", "Dollars"])

#Joining datasets to create master dataset
pitching_mds = pd.merge(pitching_standard, pitching_advanced, on=["playerid", "Season", "Name", "Team"])
pitching_mds = pd.merge(pitching_mds, pitching_value, on=["playerid", "Season", "Name", "Team"])

#Removing duplicate columns
pitching_mds = pitching_mds.drop(columns=["ERA_y"])
pitching_mds = pitching_mds.rename(columns={"ERA_x" : "ERA"})

#Export master dataset to csv file
pitching_mds.to_csv('pitching-masterdataset.csv', index=False)
```

In [ ]: