

# Project1\_DataCleaning

February 8, 2021

```
[1]: import pandas as pd
```

```
[2]: df = pd.read_csv("/Users/thomasoakley/Project_1_2012_Datasets/  
↳Project1_DataSheet.csv")
```

```
[3]: df.head()
```

```
[3]:
```

	Team	Name	Position
0	Orioles	Dylan Bundy	RHP
1	Orioles	Kevin Gausman	RHP
2	Orioles	Jonathan Schoop	2B
3	Orioles	Nick Delmonico	3B
4	Orioles	Xavier Avery	OF

```
[4]: df.count()
```

```
[4]: Team          600  
     Name          600  
     Position      600  
     dtype: int64
```

```
[5]: print(df.columns)
```

```
Index(['Team', 'Name', 'Position'], dtype='object')
```

We now can create sub dataframes from our original data set. One is going to be strictly pitchers, while the other will consist of all other positional players.

```
[6]: pitchers = df.loc[(df['Position'] == 'RHP') | (df['Position'] == 'LHP')]
```

```
[7]: fielders = df.loc[(df['Position'] != 'RHP') & (df['Position'] != 'LHP')]
```

Just going to do a quick double check to make sure that both dataframes add up to our total of 600 players

```
[8]: pitchers.count()
```