



Pythagorean Expectation and English Soccer

In soccer, teams score goals, and we can calculate Pythagorean Expectations based on goals scored and goals conceded.

The structure of competition in soccer in most countries around the world is different from the sports we have looked at so far. Rather than leagues operating as independent entities, they are connected through a hierarchical system, sometimes called "the pyramid". In England, the English Premier League is at the top of the pyramid (it used to be called the First Division) and contains 20 teams.

Beneath the Premier League is The Football League Championship (it used to be called Division Two) and it contains 24 teams. The Premier League and the Championship are linked via the system of promotion and relegation. At the end of each season, the three worst performing teams (measured by points won in competition) are relegated to play Championship soccer in the following season, to be replaced by the three best performing teams in the Championship. Beneath the Championship are two more leagues - League One (formerly Third Division) and League Two (formerly Fourth Division). These leagues are also linked, hierarchically, through promotion and relegation. Thus it makes sense to think of these four divisions as part of a common system.

In any one season, there are 92 teams in the system. Even though teams compete in different divisions, we can define both win percentage and Pythagorean Expectation for each team, in order to see how well the data fits.

In each of the four divisions, every team plays every other team twice in a season, once at home and once away. There is no playoff, so the champion is the team at the end of the season with the largest number of points (3 points for a win, one for a draw (tie)). Unlike the leagues we have looked at so far, draws are not only possible but are quite common. We need to adjust our definition of win percentage for this case. We could create a statistic such as the percentage of maximum possible points, but instead, we do something simpler- we give a value of 1 for a win, 0 for a loss, and 1/2 for a draw.

We now follow the same procedure we have used to date.

```
In [1]: # Load the packages
import pandas as pd
import numpy as np
import statsmodels.formula.api as smf
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: # Load the data.
# Our data covers the 2017/18 season
Eng18 = pd.read_excel('..../Data/Week 1/Engsoccer2017-18.xlsx')
print(Eng18.columns.tolist())
['Div', 'Date', 'HomeTeam', 'AwayTeam', 'FTHG', 'FTAG', 'FTR']

In [3]: # We can see what our dataframe looks like simply by typing its name
Eng18
```

Out[3]:

	Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR
0	EPL	2017-11-08 00:00:00	Arsenal	Leicester	4	3	H
1	EPL	2017-12-08 00:00:00	Brighton	Man City	0	2	A
2	EPL	2017-12-08 00:00:00	Chelsea	Burnley	2	3	A
3	EPL	2017-12-08 00:00:00	Crystal Palace	Huddersfield	0	3	A
4	EPL	2017-12-08 00:00:00	Everton	Stoke	1	0	H
5	EPL	2017-12-08 00:00:00	Southampton	Swansea	0	0	D
6	EPL	2017-12-08 00:00:00	Watford	Liverpool	3	3	D
7	EPL	2017-12-08 00:00:00	West Brom	Bournemouth	1	0	H
8	EPL	13/08/2017	Man United	West Ham	4	0	H
9	EPL	13/08/2017	Newcastle	Tottenham	0	2	A
10	EPL	19/08/2017	Bournemouth	Watford	0	2	A
11	EPL	19/08/2017	Burnley	West Brom	0	1	A
12	EPL	19/08/2017	Leicester	Brighton	2	0	H
13	EPL	19/08/2017	Liverpool	Crystal Palace	1	0	H
14	EPL	19/08/2017	Southampton	West Ham	3	2	H
15	EPL	19/08/2017	Stoke	Arsenal	1	0	H
16	EPL	19/08/2017	Swansea	Man United	0	4	A
17	EPL	20/08/2017	Huddersfield	Newcastle	1	0	H
18	EPL	20/08/2017	Tottenham	Chelsea	1	2	A
19	EPL	21/08/2017	Man City	Everton	1	1	D
20	EPL	26/08/2017	Bournemouth	Man City	1	2	A
21	EPL	26/08/2017	Crystal Palace	Swansea	0	2	A
22	EPL	26/08/2017	Huddersfield	Southampton	0	0	D
23	EPL	26/08/2017	Man United	Leicester	2	0	H
24	EPL	26/08/2017	Newcastle	West Ham	3	0	H
25	EPL	26/08/2017	Watford	Brighton	0	0	D
26	EPL	27/08/2017	Chelsea	Everton	2	0	H
27	EPL	27/08/2017	Liverpool	Arsenal	4	0	H
28	EPL	27/08/2017	Tottenham	Burnley	1	1	D
29	EPL	27/08/2017	West Brom	Stoke	1	1	D
...
2006	FL2	21/04/18	Wycombe	Accrington	0	4	A
2007	FL2	24/04/18	Coventry	Lincoln	2	4	A
2008	FL2	24/04/18	Morecambe	Cambridge	0	0	D
2009	FL2	24/04/18	Newport County	Accrington	2	1	H
2010	FL2	24/04/18	Yeovil	Forest Green	0	0	D
2011	FL2	28/04/18	Accrington	Lincoln	1	0	H
2012	FL2	28/04/18	Cheltenham	Coventry	1	6	A

2013	FL2	28/04/18	Chesterfield	Wycombe	1	2	A
2014	FL2	28/04/18	Colchester	Swindon	0	0	D
2015	FL2	28/04/18	Crawley Town	Crewe	1	2	A
2016	FL2	28/04/18	Grimsby	Notts County	2	1	H
2017	FL2	28/04/18	Luton	Forest Green	3	1	H
2018	FL2	28/04/18	Morecambe	Barnet	0	1	A
2019	FL2	28/04/18	Newport County	Cambridge	2	1	H
2020	FL2	28/04/18	Port Vale	Carlisle	1	2	A
2021	FL2	28/04/18	Stevenage	Exeter	3	1	H
2022	FL2	28/04/18	Yeovil	Mansfield	2	3	A
2023	FL2	2018-05-05 00:00:00	Chesterfield	Newport County	1	0	H
2024	FL2	2018-05-05 00:00:00	Barnet	Chesterfield	3	0	H
2025	FL2	2018-05-05 00:00:00	Cambridge	Port Vale	5	0	H
2026	FL2	2018-05-05 00:00:00	Carlisle	Newport County	1	1	D
2027	FL2	2018-05-05 00:00:00	Coventry	Morecambe	0	0	D
2028	FL2	2018-05-05 00:00:00	Crewe	Cheltenham	2	1	H
2029	FL2	2018-05-05 00:00:00	Exeter	Colchester	1	0	H
2030	FL2	2018-05-05 00:00:00	Forest Green	Grimsby	0	3	A
2031	FL2	2018-05-05 00:00:00	Lincoln	Yeovil	1	1	D
2032	FL2	2018-05-05 00:00:00	Mansfield	Crawley Town	1	1	D
2033	FL2	2018-05-05 00:00:00	Notts County	Luton	0	0	D
2034	FL2	2018-05-05 00:00:00	Swindon	Accrington	3	0	H
2035	FL2	2018-05-05 00:00:00	Wycombe	Stevenage	1	0	H

2036 rows × 7 columns

In [4]: # Once again our data is in the form of game results. We first identify whether the result was a win for the home team (H), # the away team (A) or a draw (D). We also create the counting variable.

```
Eng18['hwinvalue']=np.where(Eng18['FTR']=="H",1,np.where(Eng18['FTR']=="D",.5,0))
Eng18['awinvalue']=np.where(Eng18['FTR']=="A",1,np.where(Eng18['FTR']=="D",.5,0))
Eng18['count']=1
```

In [5]: # Once again we have to create separate dfs to calculate home team and away team performance.
Here is the home team df, including only the variables we need.

```
Enghome = Eng18.groupby(['HomeTeam','Div'])[['count','hwinvalue','FTHG','FTAG']].sum().reset_index()
Enghome = Enghome.rename(columns={'HomeTeam':'team','count':'Ph','FTHG':'FTHGh','FTAG':'FTAGh'})
Enghome
```

Out[5]:

	team	Div	Ph	hwinvalue	FTHGh	FTAGh
0	AFC Wimbledon	FL1	23	11.0	25	30
1	Accrington	FL2	23	18.5	42	19
2	Arsenal	EPL	19	16.0	54	20
3	Aston Villa	FLCH	23	17.5	42	19
4	Barnet	FL2	23	11.0	24	25
5	Barnsley	FLCH	23	9.5	25	32
6	Birmingham	FLCH	23	11.5	21	24
7	Blackburn	FL1	23	18.0	46	20
8	Blackpool	FL1	23	13.0	37	29
9	Bolton	FLCH	23	11.0	25	33
10	Bournemouth	EPL	19	9.5	26	30
11	Bradford	FL1	23	11.0	28	32
12	Brentford	FLCH	23	14.5	37	24
13	Brighton	EPL	19	11.0	24	25
14	Bristol City	FLCH	23	14.0	41	28
15	Bristol Rvs	FL1	23	14.0	38	30
16	Burnley	EPL	19	9.5	16	17
17	Burton	FLCH	23	6.5	19	43
18	Bury	FL1	23	9.0	20	30
19	Cambridge	FL2	23	15.5	38	23
20	Cardiff	FLCH	23	18.0	40	16
21	Carlisle	FL2	23	12.0	31	23
22	Charlton	FL1	23	14.0	31	24
23	Chelsea	EPL	19	13.0	30	16
24	Cheltenham	FL2	23	11.0	31	31
25	Chesterfield	FL2	23	9.5	27	33
26	Colchester	FL2	23	12.5	30	23
27	Coventry	FL2	23	15.0	36	24
28	Crawley Town	FL2	23	10.0	30	30
29	Crewe	FL2	23	12.0	32	32
...
62	Oxford	FL1	23	12.0	34	32
63	Peterboro	FL1	23	14.0	37	26
64	Plymouth	FL1	23	14.5	37	30
65	Port Vale	FL2	23	10.0	26	29
66	Portsmouth	FL1	23	13.5	33	21
67	Preston	FLCH	23	13.0	27	22
68	QPR	FLCH	23	14.5	38	31
69	Reading	FLCH	23	9.0	25	35
70	Rochdale	FL1	23	12.0	24	24
71	Rotherham	FL1	23	16.5	45	23
72	Scunthorpe	FL1	23	13.0	28	23
73	Sheffield United	FLCH	23	14.5	33	20
74	Sheffield Weds	FLCH	23	11.5	37	31

75	Shrewsbury	FL1	23	16.0	32	17
76	Southampton	EPL	19	7.5	20	26
77	Southend	FL1	23	15.5	38	21
78	Stevenage	FL2	23	13.5	42	27
79	Stoke	EPL	19	7.5	20	30
80	Sunderland	FLCH	23	6.5	23	39
81	Swansea	EPL	19	7.5	17	24
82	Swindon	FL2	23	11.5	29	36
83	Tottenham	EPL	19	15.0	40	16
84	Walsall	FL1	23	12.0	30	31
85	Watford	EPL	19	10.0	27	31
86	West Brom	EPL	19	7.5	21	29
87	West Ham	EPL	19	10.0	24	26
88	Wigan	FL1	23	17.0	37	11
89	Wolves	FLCH	23	18.5	47	18
90	Wycombe	FL2	23	14.5	43	35
91	Yeovil	FL2	23	10.5	29	26

92 rows × 6 columns

In [6]: # Now we create the mirror image df for the away team results.

```
Engaway = Eng18.groupby('AwayTeam')[['count','awinvalue', 'FTHG','FTAG'].sum().reset_index()
Engaway = Engaway.rename(columns={'AwayTeam':'team','count':'Pa','FTHG':'FTHGa','FTAG':'FTAGa'})
```

Out[6]:

	team	Pa	awinvalue	FTHGa	FTAGa
0	AFC Wimbledon	23	9.0	28	22
1	Accrington	23	13.5	27	34
2	Arsenal	19	6.0	31	20
3	Aston Villa	23	12.0	23	30
4	Barnet	23	6.0	40	22
5	Barnsley	23	6.5	40	23
6	Birmingham	23	5.0	44	17
7	Blackburn	23	16.0	20	36
8	Blackpool	23	9.5	26	23
9	Bolton	23	5.5	41	14
10	Bournemouth	19	7.0	31	19
11	Bradford	23	11.5	35	29
12	Brentford	23	11.0	28	25
13	Brighton	19	4.5	29	10
14	Bristol City	23	11.0	30	26
15	Bristol Rvs	23	7.5	36	22
16	Burnley	19	10.5	22	20
17	Burton	23	9.0	38	19
18	Bury	23	5.0	41	21
19	Cambridge	23	8.0	37	18
20	Cardiff	23	13.5	23	29
21	Carlisle	23	13.0	31	31
22	Charlton	23	11.5	27	27
23	Chelsea	19	11.5	22	32
24	Cheltenham	23	8.0	42	36
25	Chesterfield	23	4.5	50	20
26	Colchester	23	10.5	29	23
27	Coventry	23	11.5	23	28
28	Crawley Town	23	11.5	36	28
29	Crewe	23	7.5	43	30
...
62	Oxford	23	8.5	34	27
63	Peterboro	23	9.5	34	31
64	Plymouth	23	10.0	29	21
65	Port Vale	23	8.0	38	23
66	Portsmouth	23	9.5	35	24
67	Preston	23	14.0	24	30
68	QPR	23	6.0	39	20
69	Reading	23	8.0	35	23
70	Rochdale	23	8.0	33	25
71	Rotherham	23	11.0	30	28
72	Scunthorpe	23	14.5	27	37
73	Sheffield United	23	10.0	35	29
74	Sheffield Weds	23	10.0	29	22
75	Shrewsbury	23	15.0	22	28
76	Southampton	19	7.0	30	17
77	Southend	23	7.5	41	20
78	Stevenage	23	7.0	38	18
79	Stoke	19	5.5	38	15
80	Sunderland	23	8.5	41	29
81	Swansea	19	5.0	32	11
82	Swindon	23	12.5	29	38
83	Tottenham	19	12.0	20	34
84	Walsall	23	7.5	35	23
85	Watford	19	5.0	33	17

86	West Brom	19	5.0	27	10
87	West Ham	19	6.0	42	24
88	Wigan	23	17.5	18	52
89	Wolves	23	16.0	21	35
90	Wycombe	23	15.5	25	36
91	Yeovil	23	7.5	49	30

92 rows × 5 columns

In [7]: # Merge the home team and away team results

```
Eng18 = pd.merge(Enghome, Engaway, on = ['team'])
Eng18
```

Out[7]:

	team	Div	Ph	hwinvalue	FTHGh	FTAGh	Pa	awinvalue	FTHG _a	FTAG _a
0	AFC Wimbledon	FL1	23	11.0	25	30	23	9.0	28	22
1	Accrington	FL2	23	18.5	42	19	23	13.5	27	34
2	Arsenal	EPL	19	16.0	54	20	19	6.0	31	20
3	Aston Villa	FLCH	23	17.5	42	19	23	12.0	23	30
4	Barnet	FL2	23	11.0	24	25	23	6.0	40	22
5	Barnsley	FLCH	23	9.5	25	32	23	6.5	40	23
6	Birmingham	FLCH	23	11.5	21	24	23	5.0	44	17
7	Blackburn	FL1	23	18.0	46	20	23	16.0	20	36
8	Blackpool	FL1	23	13.0	37	29	23	9.5	26	23
9	Bolton	FLCH	23	11.0	25	33	23	5.5	41	14
10	Bournemouth	EPL	19	9.5	26	30	19	7.0	31	19

In [8]: # Sum the results by home and away measures to get the team overall performance for the season

```
Eng18['W'] = Eng18['hwinvalue']+Eng18['awinvalue']
Eng18['G'] = Eng18['Ph']+Eng18['Pa']
Eng18['GF'] = Eng18['FTHGh']+Eng18['FTAGa']
Eng18['GA'] = Eng18['FTAGh']+Eng18['FTHGa']
Eng18
```

Out[8]:

	team	Div	Ph	hwinvalue	FTHGh	FTAGh	Pa	awinvalue	FTHG _a	FTAG _a	W	G	GF	GA
0	AFC Wimbledon	FL1	23	11.0	25	30	23	9.0	28	22	20.0	46	47	58
1	Accrington	FL2	23	18.5	42	19	23	13.5	27	34	32.0	46	76	46
2	Arsenal	EPL	19	16.0	54	20	19	6.0	31	20	22.0	38	74	51
3	Aston Villa	FLCH	23	17.5	42	19	23	12.0	23	30	29.5	46	72	42
4	Barnet	FL2	23	11.0	24	25	23	6.0	40	22	17.0	46	46	65
5	Barnsley	FLCH	23	9.5	25	32	23	6.5	40	23	16.0	46	48	72
6	Birmingham	FLCH	23	11.5	21	24	23	5.0	44	17	16.5	46	38	68
7	Blackburn	FL1	23	18.0	46	20	23	16.0	20	36	34.0	46	82	40
8	Blackpool	FL1	23	13.0	37	29	23	9.5	26	23	22.5	46	60	55
9	Bolton	FLCH	23	11.0	25	33	23	5.5	41	14	16.5	46	39	74
10	Bournemouth	EPL	19	9.5	26	30	19	7.0	31	19	16.5	38	45	61
11	Bradford	FL1	23	11.0	28	32	23	11.5	35	29	22.5	46	57	67
12	Brentford	FLCH	23	14.5	37	24	23	11.0	28	25	25.5	46	62	52
13	Brighton	EPL	19	11.0	24	25	19	4.5	29	10	15.5	38	34	54
14	Bristol City	FLCH	23	14.0	41	28	23	11.0	30	26	25.0	46	67	58
15	Bristol Rvs	FL1	23	14.0	38	30	23	7.5	36	22	21.5	46	60	66
16	Burnley	EPL	19	9.5	16	17	19	10.5	22	20	20.0	38	36	39
17	Burton	FLCH	23	6.5	19	43	23	9.0	38	19	15.5	46	38	81
18	Bury	FL1	23	9.0	20	30	23	5.0	41	21	14.0	46	41	71
19	Cambridge	FL2	23	15.5	38	23	23	8.0	37	18	23.5	46	56	60
20	Cardiff	FLCH	23	18.0	40	16	23	13.5	23	29	31.5	46	69	39
21	Carlisle	FL2	23	12.0	31	23	23	13.0	31	31	25.0	46	62	54
22	Charlton	FL1	23	14.0	31	24	23	11.5	27	27	25.5	46	58	51
23	Chelsea	EPL	19	13.0	30	16	19	11.5	22	32	24.5	38	62	38
24	Cheltenham	FL2	23	11.0	31	31	23	8.0	42	36	19.0	46	67	73
25	Chesterfield	FL2	23	9.5	27	33	23	4.5	50	20	14.0	46	47	83
26	Colchester	FL2	23	12.5	30	23	23	10.5	29	23	23.0	46	53	52
27	Coventry	FL2	23	15.0	36	24	23	11.5	23	28	26.5	46	64	47
28	Crawley Town	FL2	23	10.0	30	30	23	11.5	36	28	21.5	46	58	66
29	Crewe	FL2	23	12.0	32	32	23	7.5	43	30	19.5	46	62	75
...
62	Oxford	FL1	23	12.0	34	32	23	8.5	34	27	20.5	46	61	66
63	Peterboro	FL1	23	14.0	37	26	23	9.5	34	31	23.5	46	68	60
64	Plymouth	FL1	23	14.5	37	30	23	10.0	29	21	24.5	46	58	59
65	Port Vale	FL2	23	10.0	26	29	23	8.0	38	23	18.0	46	49	67
66	Portsmouth	FL1	23	13.5	33	21	23	9.5	35	24	23.0	46	57	56
67	Preston	FLCH	23	13.0	27	22	23	14.0	24	30	27.0	46	57	46
68	QPR	FLCH	23	14.5	38	31	23	6.0	39	20	20.5	46	58	70
69	Reading	FLCH	23	9.0	25	35	23	8.0	35	23	17.0	46	48	70
70	Rochdale	FL1	23	12.0	24	24	23	8.0	33	25	20.0	46	49	57
71	Rotherham	FL1	23	16.5	45	23	23	11.0	30	28	27.5	46	73	53
72	Scunthorpe	FL1	23	13.0	28	23	23	14.5	27	37	27.5	46	65	50
73	Sheffield United	FLCH	23	14.5	33	20	23	10.0	35	29	24.5	46	62	55
74	Sheffield Weds	FLCH	23	11.5	37	31	23	10.0	29	22	21.5	46	59	60
75	Shrewsbury	FL1	23	16.0	32	17	23	15.0	22	28	31.0	46	60	39
76	Southampton	EPL	19	7.5	20	26	19	7.0	30	17	14.5	38	37	56
77	Southend	FL1	23	15.5	38	21	23	7.5	41	20	23.0	46	58	62
78	Stevenage	FL2	23	13.5	42	27	23	7.0	38	18	20.5	46	60	65

79	Stoke	EPL	19	7.5	20	30	19	5.5	38	15	13.0	38	35	68
80	Sunderland	FLCH	23	6.5	23	39	23	8.5	41	29	15.0	46	52	80
81	Swansea	EPL	19	7.5	17	24	19	5.0	32	11	12.5	38	28	56
82	Swindon	FL2	23	11.5	29	36	23	12.5	29	38	24.0	46	67	65
83	Tottenham	EPL	19	15.0	40	16	19	12.0	20	34	27.0	38	74	36
84	Walsall	FL1	23	12.0	30	31	23	7.5	35	23	19.5	46	53	66
85	Watford	EPL	19	10.0	27	31	19	5.0	33	17	15.0	38	44	64
86	West Brom	EPL	19	7.5	21	29	19	5.0	27	10	12.5	38	31	56
87	West Ham	EPL	19	10.0	24	26	19	6.0	42	24	16.0	38	48	68
88	Wigan	FL1	23	17.0	37	11	23	17.5	18	52	34.5	46	89	29
89	Wolves	FLCH	23	18.5	47	18	23	16.0	21	35	34.5	46	82	39
90	Wycombe	FL2	23	14.5	43	35	23	15.5	25	36	30.0	46	79	60
91	Yeovil	FL2	23	10.5	29	26	23	7.5	49	30	18.0	46	59	75

92 rows × 14 columns

In [9]: # Create the win percentage and Pythagorean Expectation

```
Eng18['wpc'] = Eng18['W']/Eng18['G']
Eng18['pyth'] = Eng18['GF']**2/(Eng18['GF']**2 + Eng18['GA']**2)
Eng18
```

Out[9]:

	team	Div	Ph	hwinvvalue	FTHGh	FTAGh	Pa	awinvvalue	FTHGa	FTAGa	W	G	GF	GA	wpc	pyth
0	AFC Wimbledon	FL1	23	11.0	25	30	23	9.0	28	22	20.0	46	47	58	0.434783	0.396375
1	Accrington	FL2	23	18.5	42	19	23	13.5	27	34	32.0	46	76	46	0.695652	0.731880
2	Arsenal	EPL	19	16.0	54	20	19	6.0	31	20	22.0	38	74	51	0.578947	0.677974
3	Aston Villa	FLCH	23	17.5	42	19	23	12.0	23	30	29.5	46	72	42	0.641304	0.746114
4	Barnet	FL2	23	11.0	24	25	23	6.0	40	22	17.0	46	46	65	0.369565	0.333701
5	Barnsley	FLCH	23	9.5	25	32	23	6.5	40	23	16.0	46	48	72	0.347826	0.307692
6	Birmingham	FLCH	23	11.5	21	24	23	5.0	44	17	16.5	46	38	68	0.358696	0.237970
7	Blackburn	FL1	23	18.0	46	20	23	16.0	20	36	34.0	46	82	40	0.739130	0.807785
8	Blackpool	FL1	23	13.0	37	29	23	9.5	26	23	22.5	46	60	55	0.489130	0.543396
9	Bolton	FLCH	23	11.0	25	33	23	5.5	41	14	16.5	46	39	74	0.358696	0.217379
10	Bournemouth	EPL	19	9.5	26	30	19	7.0	31	19	16.5	38	45	61	0.434211	0.352419
11	Bradford	FL1	23	11.0	28	32	23	11.5	35	29	22.5	46	57	67	0.489130	0.419876
12	Brentford	FLCH	23	14.5	37	24	23	11.0	28	25	25.5	46	62	52	0.554348	0.587049
13	Brighton	EPL	19	11.0	24	25	19	4.5	29	10	15.5	38	34	54	0.407895	0.283890
14	Bristol City	FLCH	23	14.0	41	28	23	11.0	30	26	25.0	46	67	58	0.543478	0.571629
15	Bristol Rvs	FL1	23	14.0	38	30	23	7.5	36	22	21.5	46	60	66	0.467391	0.452489
16	Burnley	EPL	19	9.5	16	17	19	10.5	22	20	20.0	38	36	39	0.526316	0.460064
17	Burton	FLCH	23	6.5	19	43	23	9.0	38	19	15.5	46	38	81	0.336957	0.180387
18	Bury	FL1	23	9.0	20	30	23	5.0	41	21	14.0	46	41	71	0.304348	0.250074
19	Cambridge	FL2	23	15.5	38	23	23	8.0	37	18	23.5	46	56	60	0.510870	0.465558
20	Cardiff	FLCH	23	18.0	40	16	23	13.5	23	29	31.5	46	69	39	0.684783	0.757880
21	Carlisle	FL2	23	12.0	31	23	23	13.0	31	31	25.0	46	62	54	0.543478	0.568639
22	Charlton	FL1	23	14.0	31	24	23	11.5	27	27	25.5	46	58	51	0.554348	0.563956
23	Chelsea	EPL	19	13.0	30	16	19	11.5	22	32	24.5	38	62	38	0.644737	0.726929
24	Cheltenham	FL2	23	11.0	31	31	23	8.0	42	36	19.0	46	67	73	0.413043	0.457221
25	Chesterfield	FL2	23	9.5	27	33	23	4.5	50	20	14.0	46	47	83	0.304348	0.242801
26	Colchester	FL2	23	12.5	30	23	23	10.5	29	23	23.0	46	53	52	0.500000	0.509523
27	Coventry	FL2	23	15.0	36	24	23	11.5	23	28	26.5	46	64	47	0.576087	0.649643
28	Crawley Town	FL2	23	10.0	30	30	23	11.5	36	28	21.5	46	58	66	0.467391	0.435751
29	Crewe	FL2	23	12.0	32	32	23	7.5	43	30	19.5	46	62	75	0.423913	0.405956
...	
62	Oxford	FL1	23	12.0	34	32	23	8.5	34	27	20.5	46	61	66	0.445652	0.460691
63	Peterboro	FL1	23	14.0	37	26	23	9.5	34	31	23.5	46	68	60	0.510870	0.562257
64	Plymouth	FL1	23	14.5	37	30	23	10.0	29	21	24.5	46	58	59	0.532609	0.491454
65	Port Vale	FL2	23	10.0	26	29	23	8.0	38	23	18.0	46	49	67	0.391304	0.348476
66	Portsmouth	FL1	23	13.5	33	21	23	9.5	35	24	23.0	46	57	56	0.500000	0.508849
67	Preston	FLCH	23	13.0	27	22	23	14.0	24	30	27.0	46	57	46	0.586957	0.605592
68	QPR	FLCH	23	14.5	38	31	23	6.0	39	20	20.5	46	58	70	0.445652	0.407067
69	Reading	FLCH	23	9.0	25	35	23	8.0	35	23	17.0	46	48	70	0.369565	0.319822
70	Rochdale	FL1	23	12.0	24	24	23	8.0	33	25	20.0	46	49	57	0.434783	0.424956
71	Rotherham	FL1	23	16.5	45	23	23	11.0	30	28	27.5	46	73	53	0.597826	0.654829
72	Scunthorpe	FL1	23	13.0	28	23	23	14.5	27	37	27.5	46	65	50	0.597826	0.628253
73	Sheffield United	FLCH	23	14.5	33	20	23	10.0	35	29	24.5	46	62	55	0.532609	0.559616
74	Sheffield Weds	FLCH	23	11.5	37	31	23	10.0	29	22	21.5	46	59	60	0.467391	0.491597
75	Shrewsbury	FL1	23	16.0	32	17	23	15.0	22	28	31.0	46	60	39	0.673913	0.702988
76	Southampton	EPL	19	7.5	20	26	19	7.0	30	17	14.5	38	37	56	0.381579	0.303885
77	Southend	FL1	23	15.5	38	21	23	7.5	41	20	23.0	46	58	62	0.500000	0.466704
78	Stevenage	FL2	23	13.5	42	27	23	7.0	38	18	20.5	46	60	65	0.445652	0.460064
79	Stoke	EPL	19	7.5	20	30	19	5.5	38	15	13.0	38	35	68	0.342105	0.209438
80	Sunderland	FLCH	23	6.5	23	39	23	8.5	41	29	15.0	46	52	80	0.326087	0.297012
81	Swansea	EPL	19	7.5	17	24	19	5.0	32	11	12.5	38	28	56	0.328947	0.200000
82	Swindon	FL2	23	11.5	29	36	23	12.5	29	38	24.0	46	67	65	0.521739	0.515148
83	Tottenham	EPL	19	15.0	40	16	19	12.0	20	34	27.0	38	74	36	0.710526	0.808624
84	Walsall	FL1	23	12.0	30	31	23	7.5	35	23	19.5	46	53	66	0.423913	0.392045
85	Watford	EPL	19	10.0	27	31	19	5.0	33	17	15.0	38	44	64	0.394737	0.320955
86	West Brom	EPL	19	7.5	21	29	19	5.0	27	10	12.5	38	31	56	0.328947	0.234562
87	West Ham	EPL	19	10.0	24	26	19	6.0	42	24	16.0	38	48	68	0.421053	0.332564
88	Wigan	FL1	23	17.0	37	11	23	17.5	18	52	34.5	46	89	29	0.750000	0.904017
89	Wolves	FLCH	23	18.5	47	18	23	16.0	21	35	34.5	46	82	39	0.750000	0.815525

```

90    Wycombe   FL2 23     14.5   43   35 23     15.5   25   36 30.0 46 79 60 0.652174 0.634184
91    Yeovil    FL2 23     10.5   29   26 23     7.5   49   30 18.0 46 59 75 0.391304 0.382275

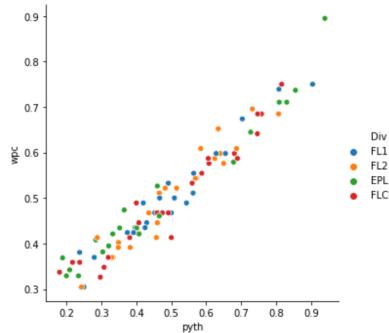
```

92 rows × 16 columns

```
In [10]: # Plot the data
# Seaborn allow us to color code teams based on division

sns.relplot(x="pyth", y="wpc", data=Eng18,hue='Div')

Out[10]: <seaborn.axisgrid.FacetGrid at 0x7f65b9545780>
```



Self test

run sns.relplot again, but this time write y="W" instead of y="wpc". What do you find? Does it make a difference?

```
In [11]: # Run the regression

pyth_lm = smf.ols(formula = 'wpc ~ pyth', data=Eng18).fit()
pyth_lm.summary()
```

```
Out[11]: OLS Regression Results
Dep. Variable: wpc R-squared: 0.934
Model: OLS Adj. R-squared: 0.933
Method: Least Squares F-statistic: 1265.
Date: Tue, 13 Jul 2021 Prob (F-statistic): 8.87e-55
Time: 04:12:28 Log-Likelihood: 187.39
No. Observations: 92 AIC: -370.8
Df Residuals: 90 BIC: -365.7
Df Model: 1
Covariance Type: nonrobust

coef std err t P>|t| [0.025 0.975]
Intercept 0.1803 0.010 18.804 0.000 0.161 0.199
pyth 0.6502 0.018 35.561 0.000 0.614 0.687

Omnibus: 3.561 Durbin-Watson: 2.070
Prob(Omnibus): 0.169 Jarque-Bera (JB): 2.898
Skew: 0.327 Prob(JB): 0.235
Kurtosis: 3.574 Cond. No. 6.86
```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Self test

Run the regression above but instead write 'wpc ~ result' instead of 'wpc ~ result' in the line starting pyth_lm. What difference does this make?

Conclusion

Notwithstanding the different organizational structures of soccer, we have found the Pythagorean Expectation model fits the data well.

We have now looked at league results from four different sports and found that the Pythagorean model fits the data well in three of the four.

But we now want to consider a different question: does the Pythagorean model work as a forecasting model? We address this question in the next notebook.

```
In [ ]: 
```