Data Science Project Proposal

CS-4981

Draven Schilling

10-12-19

The First Robotics competition is a high school level robotics competition where teams across the world are introduced to a new engineering challenge and are given 10 weeks to design, build, and test their robots then in the following weeks are eligible to compete at local, regional and potentially championship tournaments to put their designs to the test. As a previous team member and head strategist I have always found it interesting looking at the associations between teams and their performance year to year. Since the game changes every year and students are only eligible to stay on the team until they graduate It is interesting to notice the consistency of some teams and volatility of others.

In this data science proposal, I would like to analyze team performance over the years to identify correlations to what makes a successful team. The dataset I have acquired was created by scraping "The Blue Alliance API" which is the largest resource for storing FIRST competition data. From the API I was able to obtain regional data for the team and which competitions they go to; I was also able to get competition regional data and team statistics per competition (wins : losses : ties). Finally, I was able to obtain team historical data such as total number of competition wins, championship wins, and awards in general.

Some initial hypothesizes I have about the data is that there may be regional "hotspots" of good teams such as the Silicon Valley area or the east coast. I also hypothesize that teams with more total awards go to more competitions and generally have higher opr's (offensive power rating) and ccwm's (calculated contribution to winning margin). I also predict teams that have any championship wins are consistently better than average teams or that regional winners usually have higher opr's then dpr's (defensive power rating).

These are just a few of many predictions I have about the data which could ultimately contribute to answering the big question "what should I do to be a consistently good team".

```
In [4]:   ▶  realestate_readings = pd.read_json("tba-data.txt")
             realestate_readings.head()

Out[4]:
```

	address	ccwm	district	dpr	event_size	event_type_string	lat	lng	losses	opr	...	rank	state_prov	team_key	tie
0	7555 Falconridge Blvd NE #10, Calgary, AB T3J ...	24.430202	None	10.577721	3	Regional	51.120263	-113.948629	0	35.007923	...	1	AB	frc359	
1	7555 Falconridge Blvd NE #10, Calgary, AB T3J ...	15.876843	None	9.988855	3	Regional	51.120263	-113.948629	1	25.865698	...	2	AB	frc6485	
2	7555 Falconridge Blvd NE #10, Calgary, AB T3J ...	23.483059	None	9.607736	3	Regional	51.120263	-113.948629	2	33.090795	...	3	AB	frc2122	
3	7555 Falconridge Blvd NE #10, Calgary, AB T3J ...	29.129244	None	8.370246	3	Regional	51.120263	-113.948629	1	37.499490	...	4	AB	frc4627	
4	7555 Falconridge Blvd NE #10, Calgary, AB T3J ...	19.534209	None	9.936691	3	Regional	51.120263	-113.948629	3	29.470900	...	5	AB	frc5015	

5 rows × 21 columns