**.Report.**

Analysis of SPI Matches Dataset

ESPN’s Soccer Power Index (SPI) is an international and club rating system designed to be the best possible representation of team’s current overall skill level. As opposes to other systems like the FIFA rankings, SPI is forward-looking and predictive.

**Dataset**

The dataset used for this project was found FiveThirtyEight. The basic idea of analysing the Soccer Power Index (SPI) dataset is to get fair idea about the factors affecting the establishment of different types of leagues at different places in all over the world.

This kind of analysis can be done using the data, by studying the factors such as:

|  |  |  |  |
| --- | --- | --- | --- |

* Season
* Leagues
* SPI
* Teams,etc

**Tools and Libraries**

* Python
* Jupyter Notebook
* Pandas
* Numpy
* Seaborn
* Matplotlib
* Plotly & Cufflinks

**Data Description**

This dataset contains 54,411 rows and the following 23 columns:

1. season
2. date
3. league\_id
4. league
5. team1
6. team2
7. spi1
8. spi2
9. prob1
10. prob2
11. probtie
12. proj\_score1
13. proj\_score2
14. importance1
15. importance2
16. score1
17. score2
18. xg1
19. xg2
20. nsxg1
21. nsxg2
22. adj\_score1 and adj\_score2.

**Data cleaning**

I made the following changes

* Deleted the columns like importance1, importance2, xg1, xg2, nsxg1, nsxg2, adj\_scoe1, adj\_score2, score1 and score2 as they have more null values and were not important for my analysis

**EDA**

I looked at the different-different trends of the data and below are few highlights of the analysis.

* Analysed season by reference of leagues.
* Removed unwanted columns
* Plotted graphs and pie charts to specific columns.
* Sorted Values of specific columns.
* Calculated mean values for specific columns
* Plotted mean values for specific columns,etc…