Premier League Fantasy Soccer

By: Luis Bejaran, Megan Cogguillo, Julissa Guzman, and Max Izotov

Project Motivation

In fantasy soccer:

- Participants use the information available to them about player statistics to build a roster of players to compete against fellow gamers and the rosters they have put together.
- The goal is to build a team of players whose stats will be accumulated and measured against the other competitors in the league.

Our team decided to use machine learning models and interactive visualizations to help users assess how to spend their limited fantasy budget to easily analyze key player metrics from past and current seasons to build their fantasy roster.

About the Data

The English Premier League is the top level English football league system.

It includes 20 clubs, and the seasons run from August to May with each team playing 38 matches (playing all 19 other teams both home and away).

We obtained historical data of player statistics from 2017 through the current 2020 season which include overall minutes played, overall goals scored, assists, and penalties to name a few.

Fantasy Rules: Quick Overview

Squad Size

To join the game select a fantasy squad of 15 players, consisting of:

- 2 Goalkeepers
- 5 Defenders
- 5 Midfielders
- 3 Forwards

Budget

The total value of your initial squad must not exceed £100 million.

Players Per Team

You can select up to 3 players from a single Premier League team.

Scoring

During the season, the fantasy players will be allocated points based on their performance in the games.

Data Clean Up

The Dataset

Action	Points
For playing up to 60 minutes	1
For playing 60 minutes or more (excluding stoppage time)	2
For each goal scored by a goalkeeper or defender	6
For each goal scored by a midfielder	5
For each goal scored by a forward	4
For each goal assist	3
For a clean sheet by a goalkeeper or defender	4
For a clean sheet by a midfielder	1
For every 3 shot saves by a goalkeeper	1
For each penalty save	5
For each penalty miss	-2
Bonus points for the best players in a match	1-3
For every 2 goals conceded by a goalkeeper or defender	-1
For each yellow card	-1
For each red card	-3
For each own goal	-2

The dataset contains information ranging from the minutes the player plays in a single game to the how much it will cost to acquire the player for your fantasy team.

Some of the player specific metrics captured are the number of assists, penalties and free-licks, a players ICT Index, which is a soccer players Influence, the degree which a player has made an impact on a match, Creativity, assess a players performance in terms of producing scoring opportunities, and Threat, examines a player’s threat on goal.

The initial data contained over 67 metrics for 645 players.

The original dataset included a lot of columns that we decided to drop as they were not needed for scoring purposes of the fantasy soccer game.

```python
#create dataframe
players_df = pd.DataFrame(json['elements'])
players_df
```

	chance_of_playing_next_round	chance_of_playing_this_round	code	cost_change_event	cost_change_event_fall	cost_change_start	cost_change_sta
0	0.0	0.0	37605	0	0	-3	3
1	0.0	0.0	39476	0	0	-2	2
2	100.0	100.0	41270	0	0	-1	1
3	50.0	50.0	54694	0	0	-7	7
4	100.0	100.0	58822	0	0	-4	4
...	...	...	...	...	...	...	...
640	NaN	NaN	481626	0	0		
641	NaN	NaN	448487	0	0		
642	NaN	NaN	209353	0	0		
643	NaN	NaN	465551	0	0		
644	100.0	100.0	73314	0	0		

645 rows × 67 columns

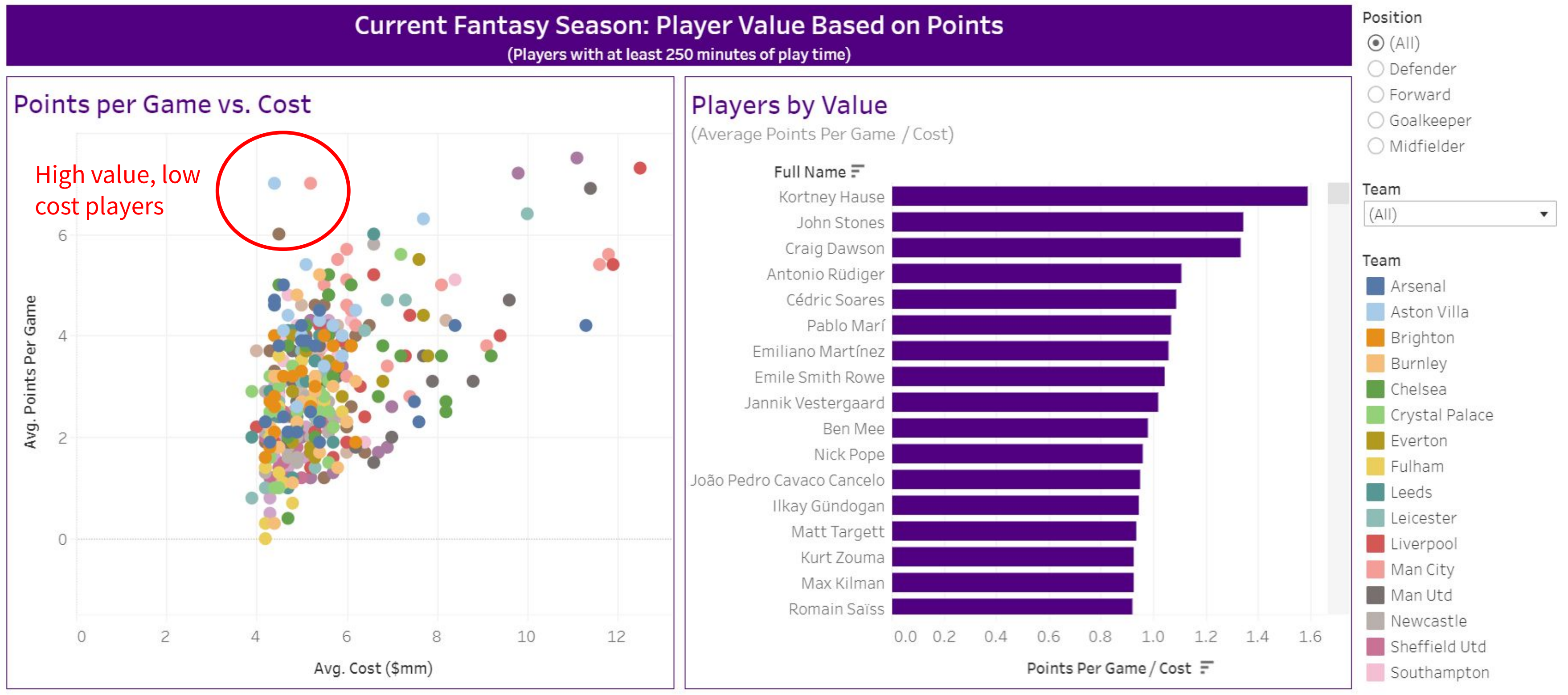
```python
#sort data by points
cut_players_df_sorted = cut_players_df.sort_values('total_points', ascending=False)
cut_players_df_sorted
```

	id	second_name	first_name	value_season	value_form	form	total_points	points_per_game	team	team_code	...	yellow_cards	red_cards	save
519	388	Kane	Harry	12.8	0.5	6.0	143	7.5	Spurs	6	...	1	0	0
395	302	Borges Fernandes	Bruno Miguel	12.5	0.3	3.7	142	7.1	Man Utd	1	...	4	0	0
521	390	Son	Heung-Min	14.4	0.4	4.4	141	7.4	Spurs	6	...	0	0	0
331	254	Salah	Mohamed	10.5	0.1	1.6	131	6.9	Liverpool	14	...	0	0	0
265	224	Vardy	Jamie	11.6	0.2	1.8	116	6.4	Leicester	13	...	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
289	196	Douglas	Barry	0.0	0.0	0.0	0	0.0	Leeds	2	...	0	0	0
277	237	Benkovic	Filip	0.0	0.0	0.0	0	0.0	Leicester	13	...	0	0	0
264	223	Ward	Danny	0.0	0.0	0.0	0	0.0	Leicester	13	...	0	0	0
0	1	Özil	Mesut	0.0	0.0	0.0	0	0.0	Arsenal	3	...	0	0	0
39	31	Taylor	Neil	-0.2	0.0	-0.2	-1	-1.0	Aston Villa	7	...	1	0	0

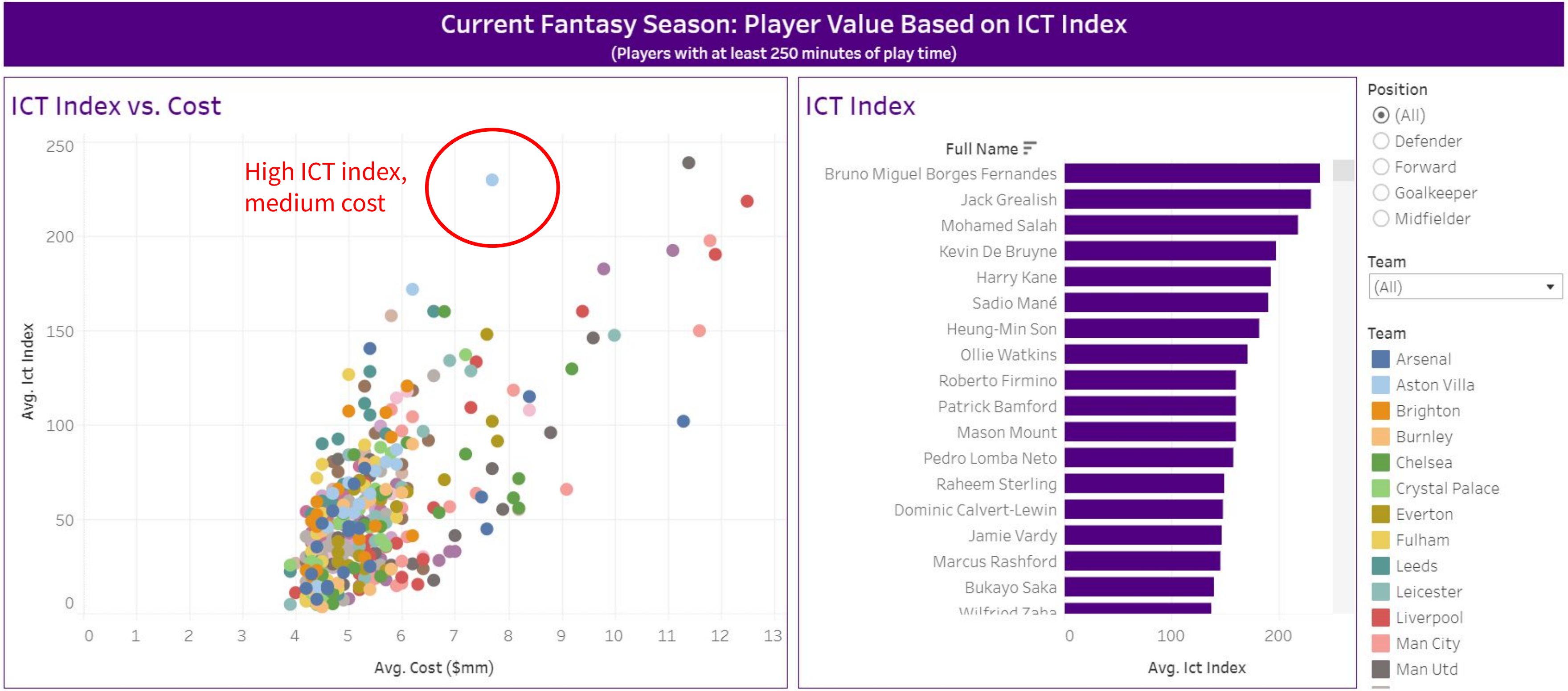
645 rows × 34 columns

Visualizations

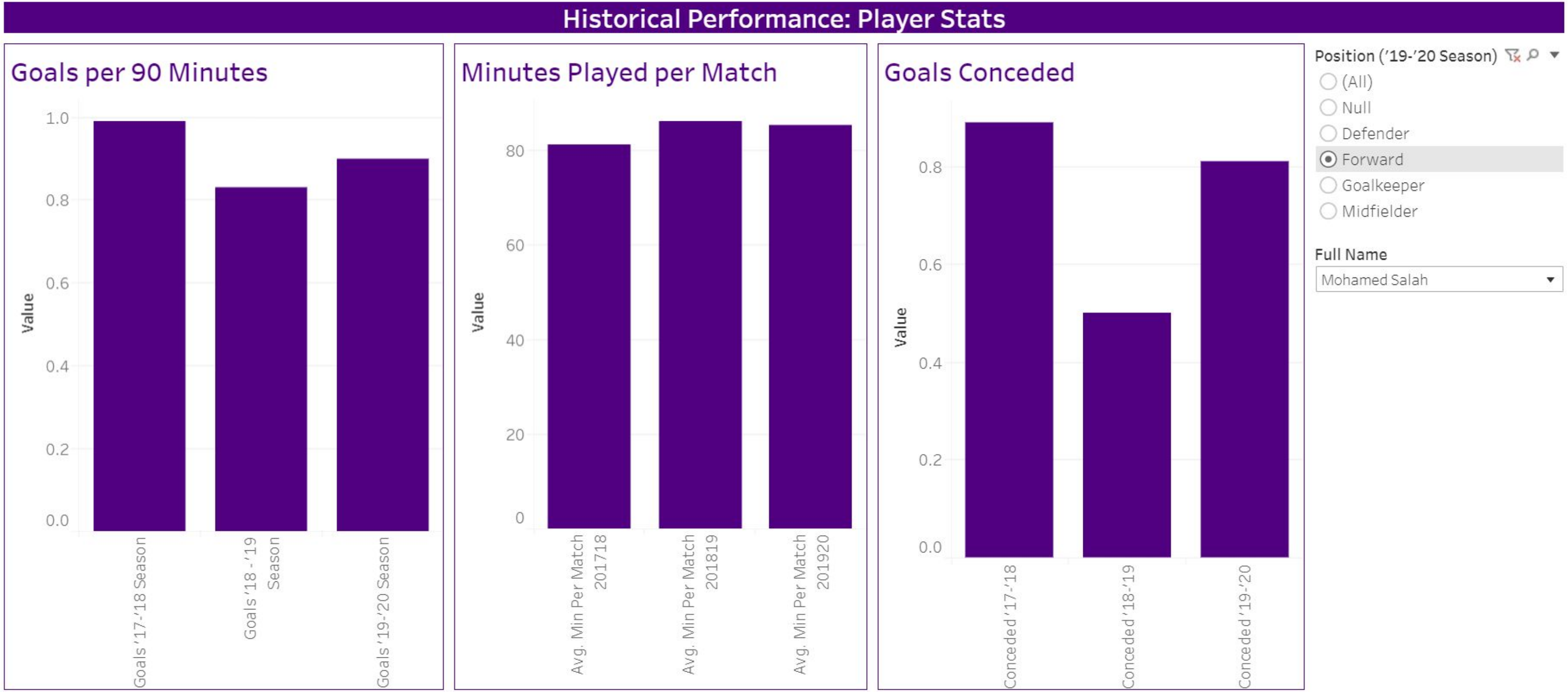
Assessing Value: High Points, Low Cost



Assessing Value: High ICT Index, Low Cost



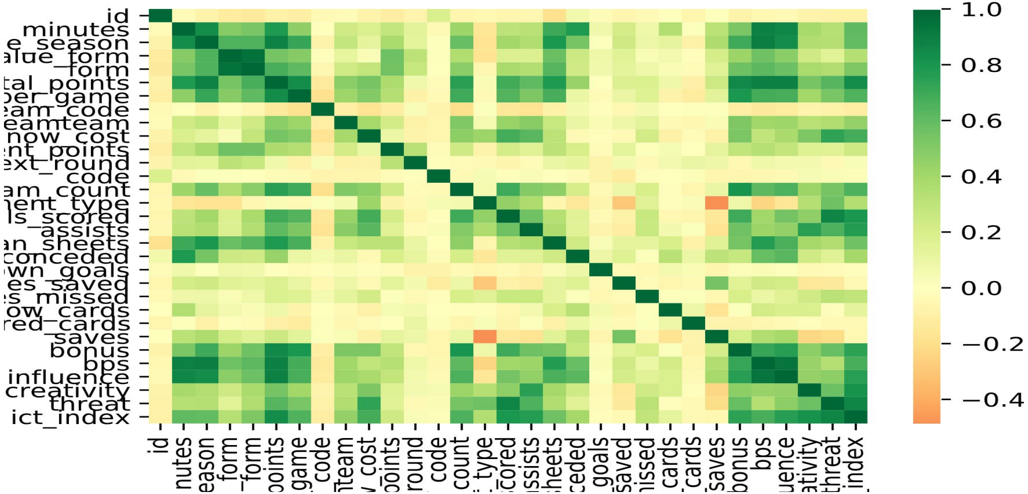
Easily Access Historical Player Stats

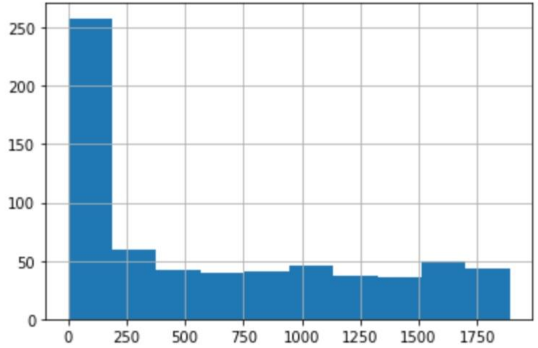


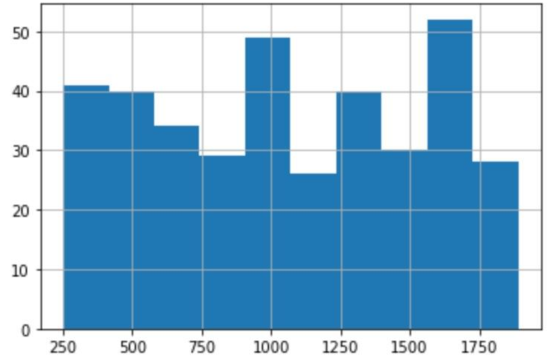
Machine Learning Model

Initial Identification of Important Variables

Valuing Relationships







Model 1: Using Neural Networks to Learn

We used *tensorflow* and *sklearn* to transform, test, and split our existing data from previous seasons to train our model. We then executed the model on an updated data set with this season's statistics. We put more emphasis on this season's statistics return a more accurate forecast for points generation in the near future.

Model 2: Elastic Net Regression to Predict Variables

Using an Elastic Net Regression model and a simplified data set with the most effective points-generating events, users are able to input expected variables themselves to predict variables such as points. A Mean Absolute Error of .090 suggest that our model will likely be more accurate than a biased decision based on fandom.

```
In [ ]:  #Make a prediction
         data = df.values
         X, y = data[:, :-1], data[:, -1]
         model = ElasticNet(alpha=1.0, l1_ratio=0.5)
         model.fit(X, y)

         row = [###'INSERT ROW VALUES HERE'###]

         yhat = model.predict([row])
         print('Predicted: %.3f' % yhat)
```

maximizo Merge pull request #53 from maximizo/maxbranch ... faa09f3 3 minutes ago 114 commits

.ipynb_checkpoints	adding cut data resource	24 minutes ago
__pycache__	update html files	17 minutes ago
api	adding cut data resource	24 minutes ago
static	update html files	17 minutes ago
templates/html	update html files	17 minutes ago
Points Learning.ipynb	adding elastic net regression	43 minutes ago
README.md	first commit	11 days ago
app.py	update flask structure	yesterday
currentpoints_to_csv.ipynb	final points learning file + clean up	1 hour ago
cut_data.ipynb	worked on elastic_net_regression	10 minutes ago
elastic_net_regression.ipynb	final elastic regression push	4 minutes ago
json_points.ipynb	file + clean up	1 hour ago
points_learning.ipynb	final points learning file + clean up	1 hour ago
points_learning.sav	points_learning.sav add	2 hours ago
sns heatmap.png	adding png lr visuals	2 days ago

Final Insights

In Fantasy sports, the winning teams are determined by the real-life statistics of the athletes.

- The goal of our project was to provide an interactive web application that gamers could use to assess potential picks for their teams.
  - Essentially compare and contrast player statistic.

Ideally, future versions of this project would include a model that could assess the players’ monetary value and their predicted metrics.

Questions?