



IPE-205 | Python in Industrial Engineering | Machine Learning to FIFA 18

Summary of the Poster:

We created a predictive model that is able to forecast how good a soccer player is based on their game statistics (using Python in kaggle.com). We visualize the data with histogram, pie chart, boxplot. From the scatterplot we can see that increase in value of a player increases the overall rating of the player. The Linear Regression model helps us to predict the future overall rating of the player.

Steps for the analysis:

- Preparing the data
- Understanding the data
- Histogram
- Machine learning algorithms for building our predictive model
- Linear regression
- Boxplot
- Scatterplot
- Pie chart

Code

```
plt.hist(df.Overall, bins=16, alpha=0.6, color='b')
plt.title("#Players per Overall")
plt.xlabel("Overall")
plt.ylabel("Count")
```

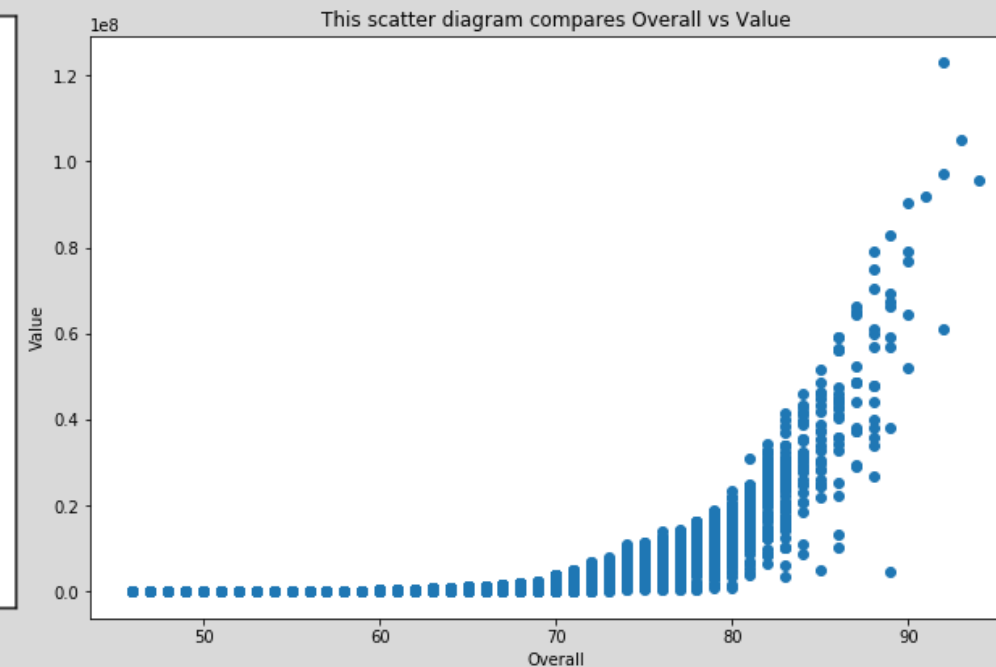
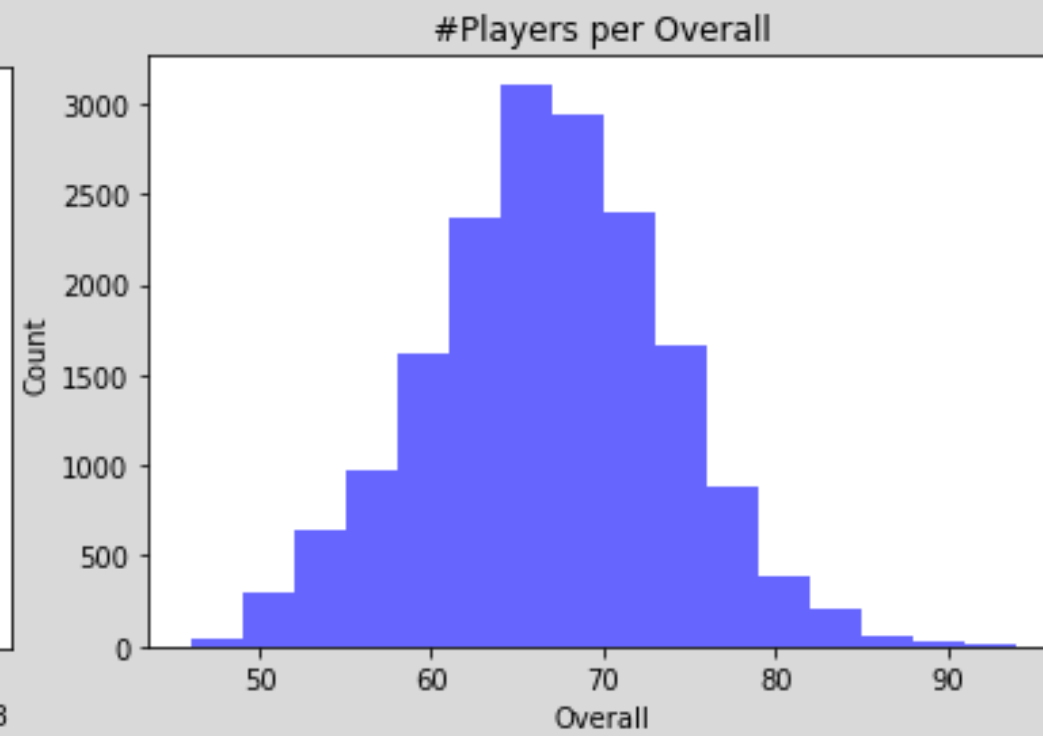
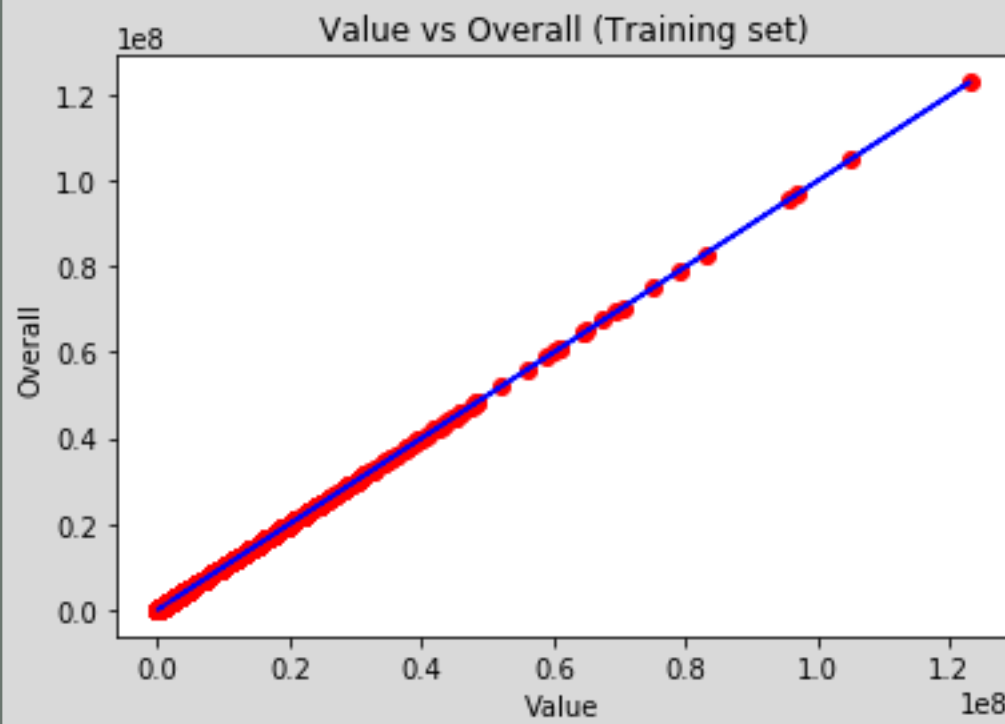
```
from sklearn import linear_model
regr = linear_model.LinearRegression(fit_intercept=False)
regr.fit(xtrain, ytrain)
y_pred = regr.predict(xtest)
plt.scatter(xtest, ytest, color='black')
plt.plot(xtest, y_pred, color='blue', linewidth=3)
plt.xlabel("Value") plt.ylabel("Overall")
plt.show()
```

```
import seaborn as sns
fig = plt.figure(1, figsize=(300, 6))
ax = fig.add_subplot(111)
sns.boxplot(x=df['Nationality'], y=df['Overall'], data=df)
plt.scatter(x=df['Overall'], y=df['Value'])
```

```
players = df[['Name', 'Age', 'Nationality']].dropna()
players.groupby("Nationality").Name.count().sort_values(ascending=False).head(9).plot(kind="pie")
```

Submitted to:
Tanmoy Das

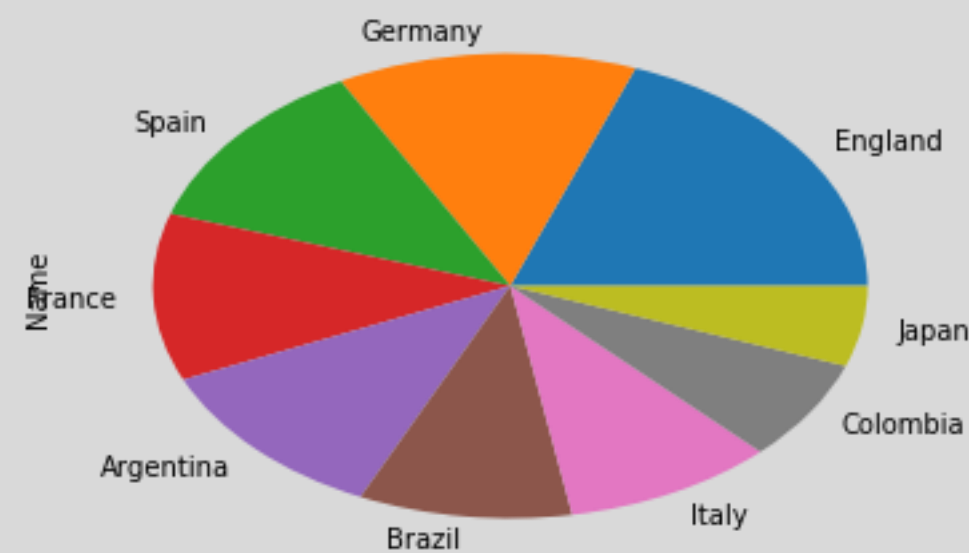
Submitted by:
201736012
201736044
201736045
201636045



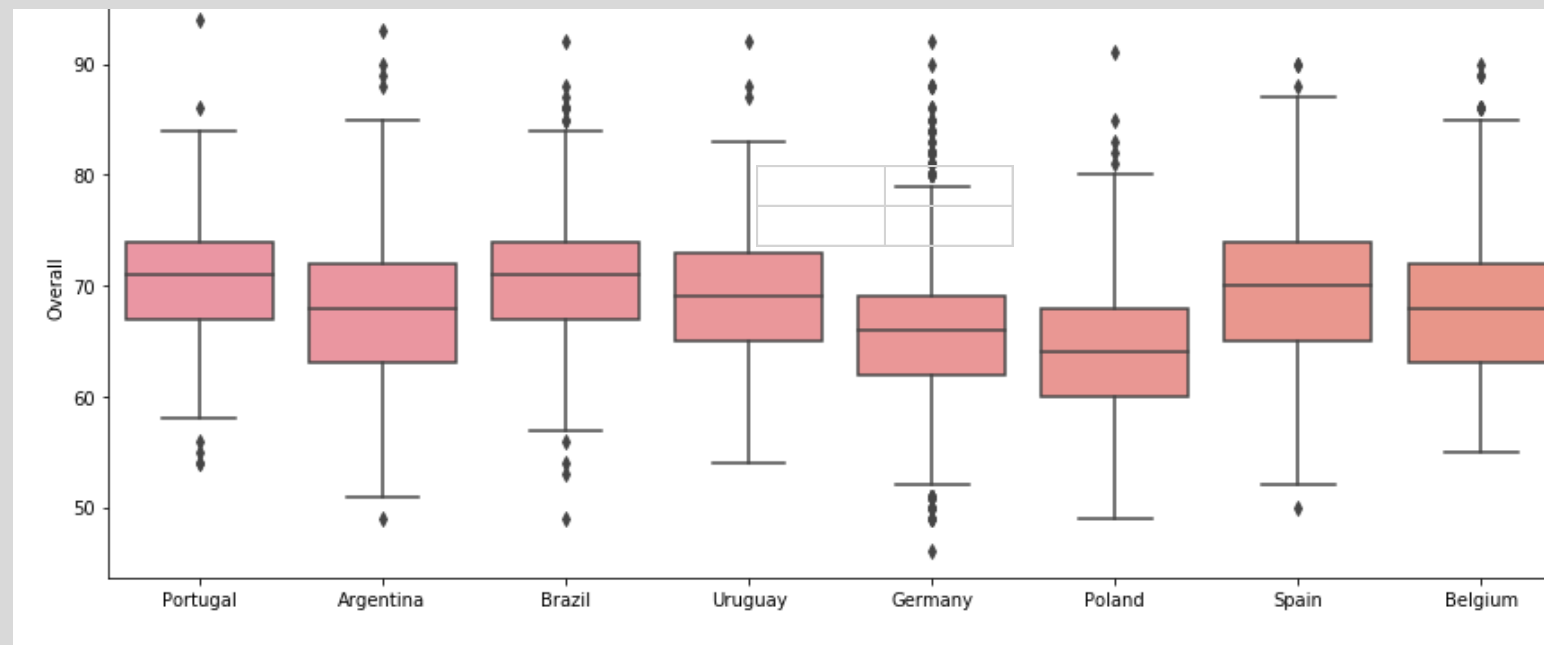
Linear Regression

Histogram

Scatterplot



Pie chart



Boxplot

	Name	Age	Nationality	Overall	Value	Agility	Finishing
0	Cristiano Ronaldo	32	Portugal	94	95500000	89	94
1	L. Messi	30	Argentina	93	105000000	90	95
2	Neymar	25	Brazil	92	123000000	96	89
3	L. Suárez	30	Uruguay	92	97000000	86	94
4	M. Neuer	31	Germany	92	61000000	52	13
5	R. Lewandowski	28	Poland	91	92000000	78	91
6	De Gea	26	Spain	90	64500000	60	13
7	E. Hazard	26	Belgium	90	90500000	93	83
8	T. Kroos	27	Germany	90	79000000	71	76
9	G. Higuain	29	Argentina	90	77000000	75	91

Dataset

Kaggle Link:

<https://www.kaggle.com/sdsourav80/poster-presentation>

Github Link:

<https://github.com/SdSourav/Poster-Presentation>