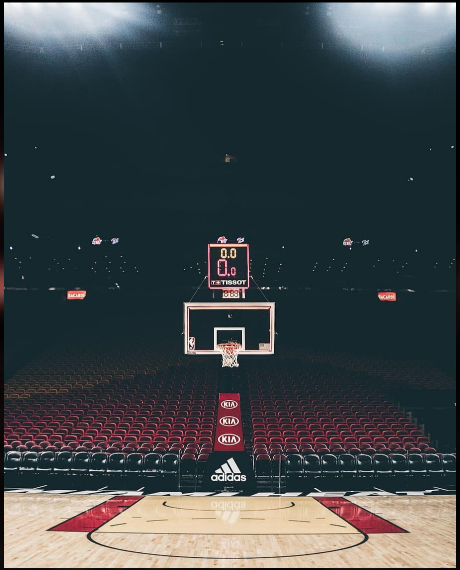
Predicting NBA Players’ Salaries Using Stats

Who We Are : Data Robots Sports Agency



Leadership Team :

- CEO : Margaret Wharton
- CFO : Jared Roussel
- CTO : Nathan Haile
- COO : Jaeik Park

Data Robots is industry leading sport agency company based in Atlanta, GA. We have 250 NBA players currently signed. We are negotiating salary on their behalf & suggesting how each player can improve his salary. We are getting a consultancy fee based on their negotiated salary, which is our main source of income.

Data Sources

NBA Player Salary Dataset (2017-2018)
NBA Player Stats 2017-2018

	Unnamed: 0	Player	Tm	season17_18
0	1	Stephen Curry	GSW	34682550.0
1	2	LeBron James	CLE	33285709.0
2	3	Paul Millsap	DEN	31269231.0
3	4	Gordon Hayward	BOS	29727900.0
4	5	Blake Griffin	DET	29512900.0

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	1	Alex Abrines/abrinal01	SG	24	OKC	75	8	1134	115	291	...	0.848	26	88	114	28	38	8	25	124	353
1	2	Quincy Acy/acyqu01	PF	27	BRK	70	8	1359	130	365	...	0.817	40	217	257	57	33	29	60	149	411
2	3	Steven Adams/adamsst01	C	24	OKC	76	76	2487	448	712	...	0.559	384	301	685	88	92	78	128	215	1056
3	4	Bam Adebayo/adebaba01	C	20	MIA	69	19	1368	174	340	...	0.721	118	263	381	101	32	41	66	138	477
4	5	Arron Afflalo/afflaar01	SG	32	ORL	53	3	682	65	162	...	0.846	4	62	66	30	4	9	21	56	179

Data Sources

NBA Player Salary Dataset (2017-2018)
NBA Player Stats 2017-2018

	Unnamed: 0	Player	Tm	season17_18
0	1	Stephen Curry	GSW	34682550.0
1	2	LeBron James	CLE	33285709.0
2	3	Paul Millsap	DEN	31269231.0
3	4	Gordon Hayward	BOS	29727900.0
4	5	Blake Griffin	DET	29512900.0

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
0	1	Alex Abrines/abrinal01	SG	24	OKC	75	8	1134	115	291	...	0.848	26	88	114	28	38	8	25	124	353
1	2	Quincy Acy/acyqu01	PF	27	BRK	70	8	1359	130	365	...	0.817	40	217	257	57	33	29	60	149	411
2	3	Steven Adams/adamsst01	C	24	OKC	76	76	2487	448	712	...	0.559	384	301	685	88	92	78	128	215	1056
3	4	Bam Adebayo/adebaba01	C	20	MIA	69	19	1368	174	340	...	0.721	118	263	381	101	32	41	66	138	477
4	5	Arron Afflalo/afflaar01	SG	32	ORL	53	3	682	65	162	...	0.846	4	62	66	30	4	9	21	56	179

Data Cleaning

In [28]:

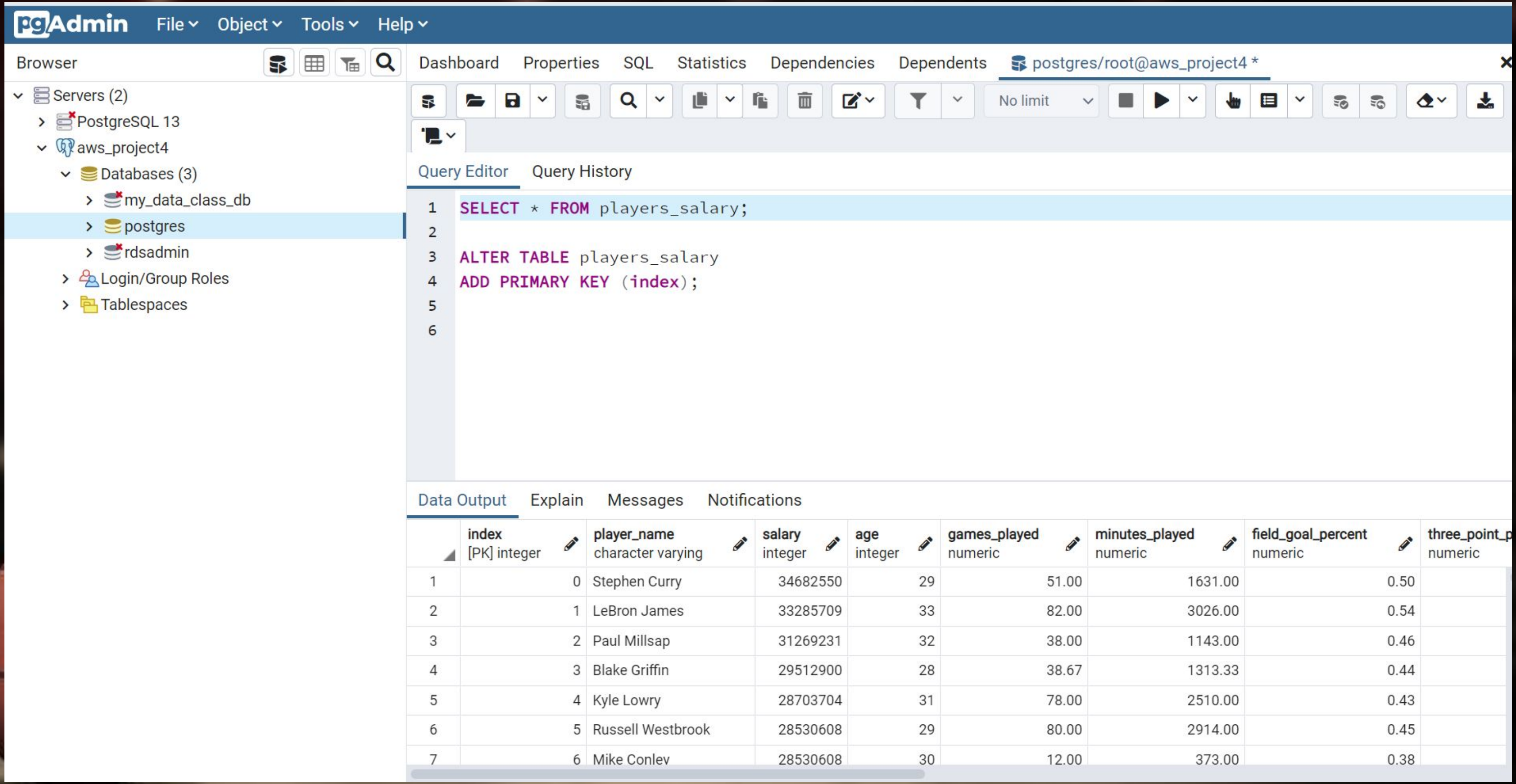
```python
# merge stats and salary on player_name
stats_salary_df=pd.merge(salary_final_df, stats_final_df, on="player_name")
stats_salary_df
```

Out[28]:

	player_name	salary	age	games_played	minutes_played	field_goal_percent	three_point_percent	two_point_percent	effective_field_goal	free_throw_pe
0	Stephen Curry	34682550.0	29.0	51.000000	1631.000000	0.495000	0.423	0.595	0.618	0.9
1	LeBron James	33285709.0	33.0	82.000000	3026.000000	0.542000	0.367	0.603	0.590	0.7
2	Paul Millsap	31269231.0	32.0	38.000000	1143.000000	0.464000	0.345	0.506	0.509	0.6
3	Blake Griffin	29512900.0	28.0	38.666667	1313.333333	0.437333	0.345	0.481	0.493	0.7
4	Kyle Lowry	28703704.0	31.0	78.000000	2510.000000	0.427000	0.399	0.474	0.553	0.8

	player_name	salary	age	games_played	minutes_played	field_goal_percent	three_point_percent	two_point_percent	effective_field_goal	free_throw_perce
0	Stephen Curry	34682550.0	29.0	51.000000	1631.000000	0.495000	0.423	0.595	0.618	0.9210
1	LeBron James	33285709.0	33.0	82.000000	3026.000000	0.542000	0.367	0.603	0.590	0.7310
2	Paul Millsap	31269231.0	32.0	38.000000	1143.000000	0.464000	0.345	0.506	0.509	0.6960
3	Blake Griffin	29512900.0	28.0	38.666667	1313.333333	0.437333	0.345	0.481	0.493	0.7846
4	Kyle Lowry	28703704.0	31.0	78.000000	2510.000000	0.427000	0.399	0.474	0.553	0.8540

AWS PostgreSQL Database



SQLAlchemy

[102]:

```
engine = create_engine(f'postgresql://root:[redacted]@mypostgresdb.cdxrpdwcb1ik.us-west-2.rds.amazonaws.com/postgres')
conn = engine.connect()
```

[103]:

```
Base=automap_base()
Base.prepare(conn,reflect=True)
```

[104]:

```
Base.classes.keys()
```

Out[104]:

```
['players_salary']
```

[105]:

```
session=Session(engine)
```

[106]:

```
players_salary_class=Base.classes.players_salary
```

[107]:

```
players_query = session.query(players_salary_class.index,players_salary_class.player_name,players_salary_class.salary,pla
df=pd.DataFrame(players_query,columns=['index','player_name','salary','age','games_played','minutes_played','field_goal_p
df
```

Out[107]:

	index	player_name	salary	age	games_played	minutes_played	field_goal_percent	three_point_percent	two_point_percent	effective_field_goal	fr
0	0	Stephen Curry	34682550	29	51.00	1631.00	0.50	0.42	0.60	0.62	
1	1	LeBron James	33285709	33	82.00	3026.00	0.54	0.37	0.60	0.59	

- Process of integrating our data from AWS PostgresSQL database.

Our NBA Salary Prediction Models

We built several supervised machine learning models to predict salary.

The 2 models below are selected in terms of accuracy:

- Random forest model
  - Determine whether a player is on top 20 salary or not
  - High accuracy
- Multi linear regression model
  - Suggest salary amount depending on their performance
  - Relatively low accuracy

Correlation Matrix

	index	salary	age	games_played	minutes_played	field_goal_percent	three_point_percent	two_point_percent	effective_field_goal	free_throw_percent	total_rebounds	assist	steal	blocking	points
index	1	-0.87	-0.39	-0.34	-0.53	-0.21	-0.1	-0.14	-0.23	-0.23	-0.46	-0.42	-0.44	-0.34	-0.55
salary	-0.87	1	0.34	0.25	0.49	0.18	0.1	0.12	0.19	0.19	0.45	0.47	0.42	0.34	0.58
age	-0.39	0.34	1	0.062	0.061	0.029	0.06	0.047	0.12	0.18	-0.00012	0.094	0.01	-0.018	0.048
games_played	-0.34	0.25	0.062	1	0.85	0.32	0.14	0.21	0.38	0.18	0.66	0.51	0.67	0.49	0.68
minutes_played	-0.53	0.49	0.061	0.85	1	0.25	0.21	0.15	0.32	0.24	0.76	0.7	0.84	0.53	0.91
field_goal_percent	-0.21	0.18	0.029	0.32	0.25	1	-0.049	0.79	0.85	-0.13	0.45	0.094	0.17	0.46	0.26
three_point_percent	-0.1	0.1	0.06	0.14	0.21	-0.049	1	-0.12	0.3	0.3	-0.023	0.16	0.15	-0.094	0.2
two_point_percent	-0.14	0.12	0.047	0.21	0.15	0.79	-0.12	1	0.7	-0.21	0.29	0.032	0.099	0.32	0.17
effective_field_goal	-0.23	0.19	0.12	0.38	0.32	0.85	0.3	0.7	1	0.059	0.35	0.11	0.21	0.32	0.29
free_throw_percent	-0.23	0.19	0.18	0.18	0.24	-0.13	0.3	-0.21	0.059	1	0.024	0.21	0.18	-0.083	0.28
total_rebounds	-0.46	0.45	-0.00012	0.66	0.76	0.45	-0.023	0.29	0.35	0.024	1	0.46	0.59	0.78	0.73
assist	-0.42	0.47	0.094	0.51	0.7	0.094	0.16	0.032	0.11	0.21	0.46	1	0.74	0.24	0.73
steal	-0.44	0.42	0.01	0.67	0.84	0.17	0.15	0.099	0.21	0.18	0.59	0.74	1	0.4	0.77
blocking	-0.34	0.34	-0.018	0.49	0.53	0.46	-0.094	0.32	0.32	-0.083	0.78	0.24	0.4	1	0.53
points	-0.55	0.58	0.048	0.68	0.91	0.26	0.2	0.17	0.29	0.28	0.73	0.73	0.77	0.53	1

Multi Linear Regression Model

- Player's salary = (\$561,941 x age) + (\$2,185 x Minutes played) + (\$423 x Total rebounds) - (\$4 x assist) + (\$5137 x Steal) +(\$ 45,334 x Blocking) +(\$9,043 x points) - (\$151,833 x game played)
- Higher salary players tend to :

  > Not too young

  > Less number of game but play longer

  > Better at steal, blocking

  > Earn higher points

Random Forest Model

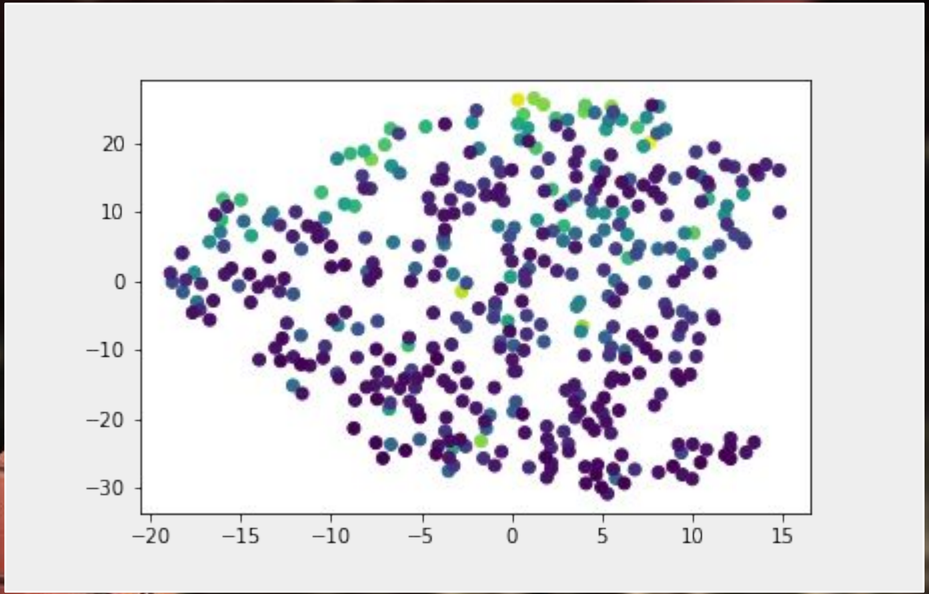
- When player's performance is given, we can see if he is on top 20 salary or not.
- Accuracy for train data : 100% / 96.2% for test data

In [259]:
```
clf_scaled = RandomForestClassifier(random_state=1).fit(X2_train_scaled, y2_train)
y_pred = clf_scaled.predict(X2_test_scaled)
print(classification_report(y2_test, y_pred))
print(f'Training Score: {clf_scaled.score(X2_train_scaled, y2_train)}')
print(f'Testing Score: {clf_scaled.score(X2_test_scaled, y2_test)}')
```

```
              precision    recall  f1-score   support

           0       0.96      1.00      0.98       100
           1       1.00      0.33      0.50         6

    accuracy                           0.96       106
   macro avg       0.98      0.67      0.74       106
weighted avg       0.96      0.96      0.95       106

Training Score: 1.0
Testing Score: 0.9622641509433962
```

Unsupervised Model

- Cluster Analysis

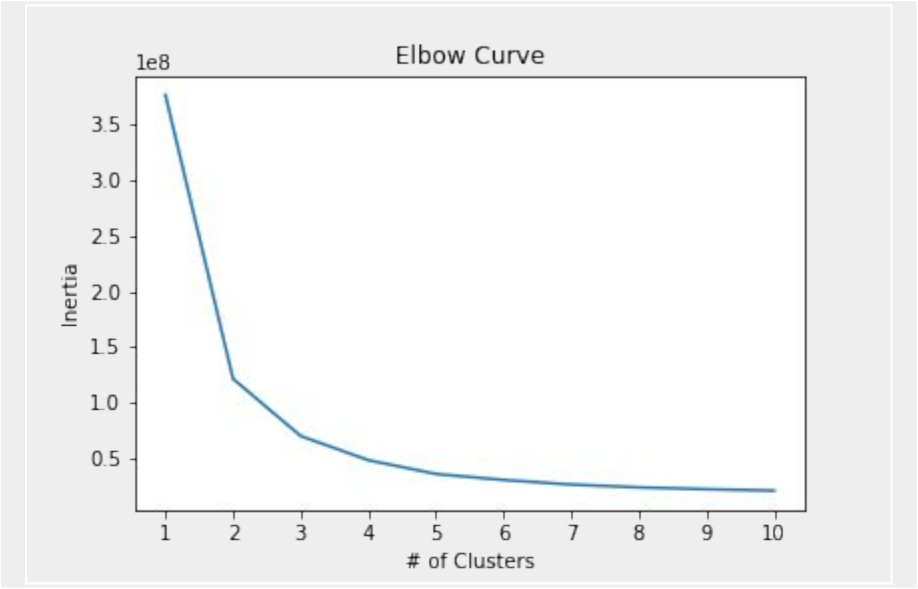


-Analyzed for any notable clusters targeting our salary label.

-Small notable cluster in green variables, which signify the highest paid players.

Unsupervised Model

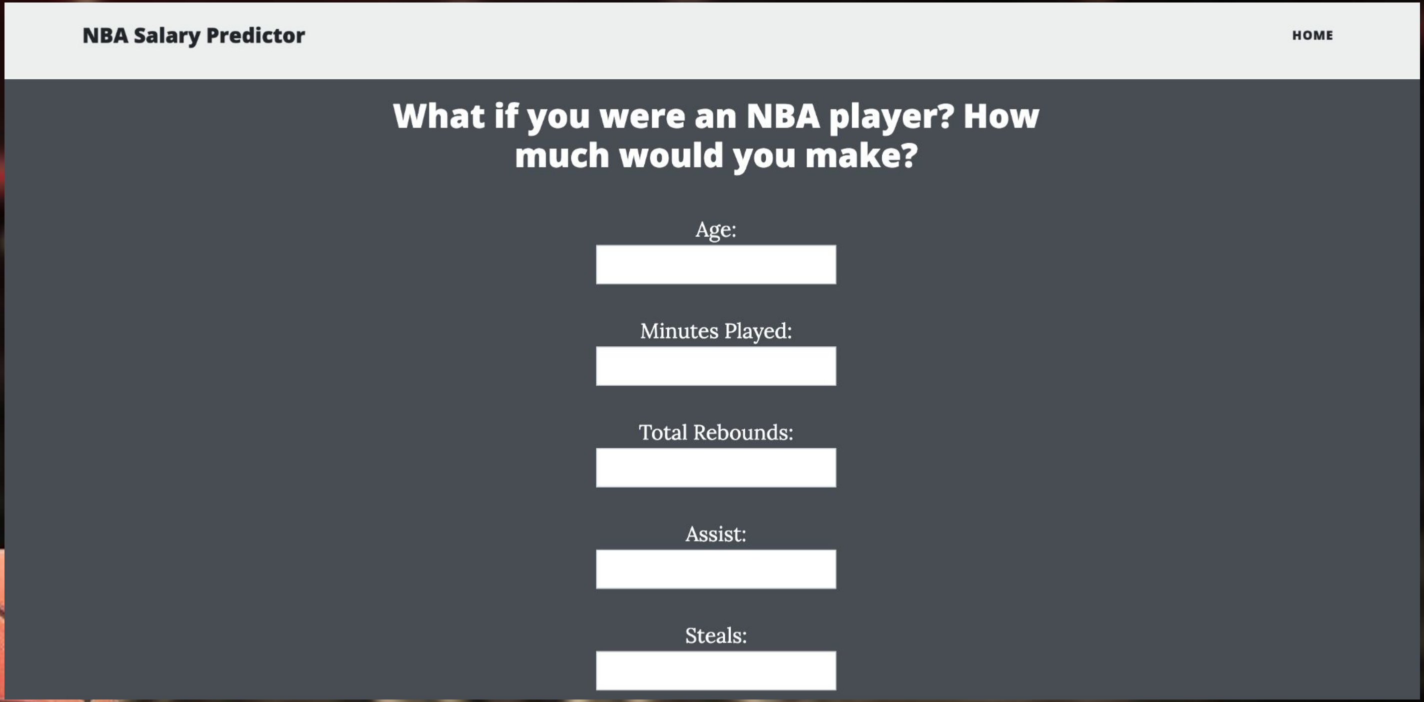
- Elbow Curve Analysis



-From our analysis we can see that number of clusters at the elbow is three clusters which the data can be grouped.

-This grouping can help organizations understand the three different level of players there are based on stats and their pay ranges.

Salary Prediction App



https://mwhar.github.io/datarobots-project4/


Thank You!
NBAStore.com
STANLEY CUP CHAMPIONS 2013-14
STANLEY CUP CHAMPIONS 2011-12
OAKLAND
22
OAKLAND
23
TAYLOR
18
GRETZKY
30
16
@ESPNLOSANGELES
Spectrum
TOSHIBA