STAT 596 : Regression and Time Series Analysis
FINAL PROJECT PRESENTATION

NBA Salary Insights

Vishwa Patel vup4@scarletmail.rutgers.edu
Devyani Mardia dm1633@scarletmail.rutgers.edu
Rutvik Deshpande rvd39@scarletmail.rutgers.edu
Priyal Shaha ps1274@scarletmail.rutgers.edu

Background and Goals

- Identify **factors affecting** the player Salaries.
- **Leveraging player stats** and demographics across various teams, their positions, shooting styles, etc over the period of time.
- Performed Forward Selection and Backward Elimination using AIC and BIC values for parameter selection.
- Linear **Hypothesis testing** for our Linear Models.
- **Predicting Salaries** for NBA players by fitting a Linear Model.

Data Set

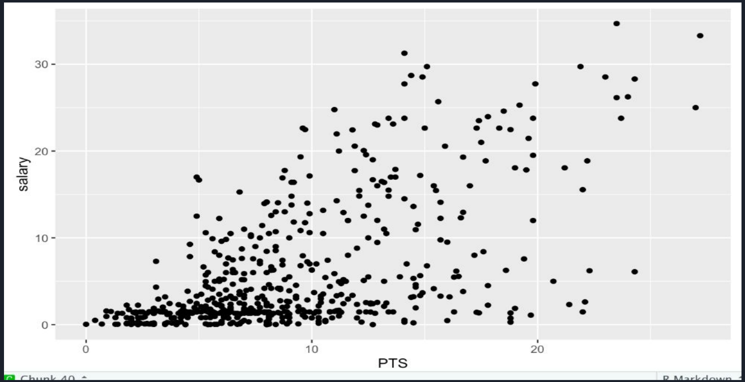
- NBA **Player Statistics** data.
- NBA **Player salaries modelling** for season 2014-15
- Data sets sources: Players.csv and salaries.csv
- Players.csv : Player **demographics and Statistics**
- Salaries.csv: **Salaries** over the period of time (1985-2018)

Data Manipulation

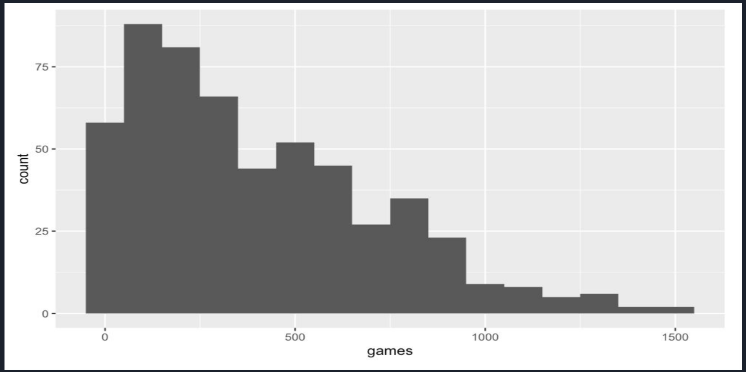
- **Unified the dataset** using player_id as key **merging** players.csv and salaries.csv.
- **Filled up missing values** in the data set.
- **Handled categorical variables** in the data set to fit our linear model.
- **Normalized** Target variable “salary” in the range of millions to user friendly variable.
- **Created new columns** for age at the end of season, height in cm, weight, etc.

Exploratory Data Analysis

Univariate Analysis



Scatter Plot : Points scored vs Salary



Histogram : No of Games played

Exploratory Data Analysis

Multicollinearity

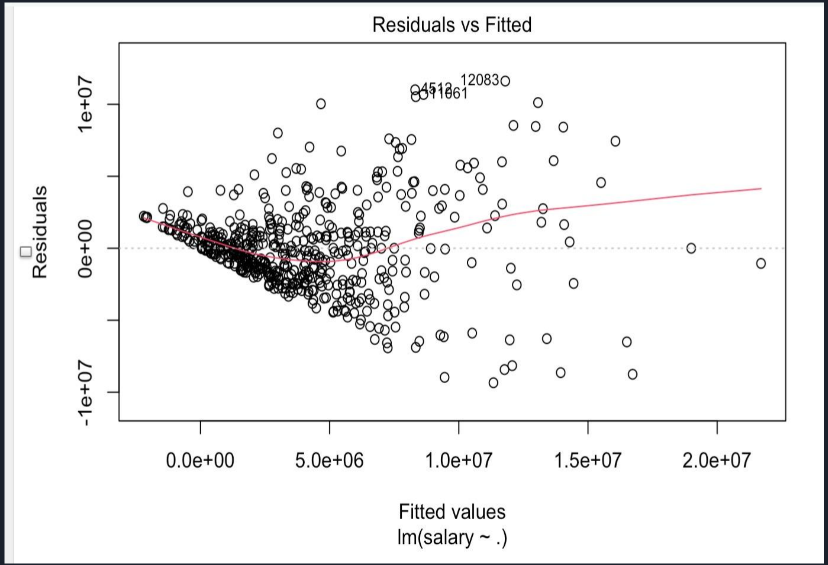
	age	AST	FG	FG3	FT	games
age	1.0000000000	0.23176645	0.09090559	0.098598622	0.172397637	0.81638598
AST	0.2317664524	1.00000000	-0.05612518	0.244184105	0.228739757	0.39810739
FG	0.0909055860	-0.05612518	1.00000000	-0.226472591	-0.178475353	0.21365929
FG3	0.0985986223	0.24418411	-0.22647259	1.000000000	0.330818851	0.13144822
FT	0.1723976365	0.22873976	-0.17847535	0.330818851	1.000000000	0.20396056
games	0.8163859761	0.39810739	0.21365929	0.131448220	0.203960563	1.00000000
PER	0.2597386765	0.39308625	$\boxed{0.69047200}$	-0.010543621	0.118698026	0.47895319
PTS	0.2842778540	0.63666352	0.19945118	0.253809447	0.317624182	0.57417772
TRB	0.1660983406	0.12714491	0.55301395	-0.181325971	-0.135426180	0.40120882
WS	$\boxed{0.6467375623}$	0.46202683	0.25998987	0.062398226	0.155952532	0.83968444
eFG	0.0882649151	-0.04759632	$\boxed{0.86069735}$	0.117402814	-0.008211145	0.19210296
height	-0.1055221519	-0.52748851	0.43816505	-0.371729580	-0.272720738	-0.03860283
position	0.1033509758	-0.10783072	-0.33197041	0.347958886	0.134588731	0.10047067
shoots	-0.0380478168	-0.09425524	-0.08581791	0.028789656	0.027476516	-0.07971788
weight	0.0165369670	-0.41335617	0.43822076	-0.375919026	-0.257100978	0.08365706
team	0.0006707831	0.01839429	0.04681294	-0.007403481	-0.053002423	0.03056276

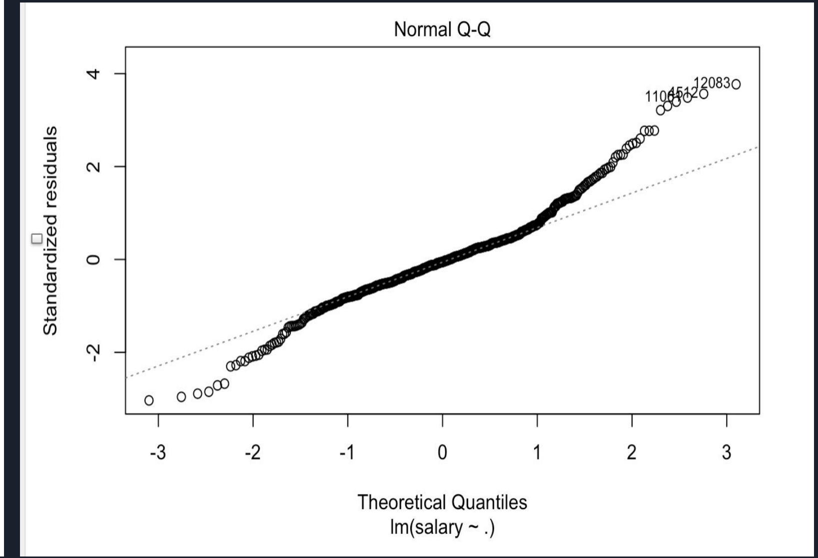
	PER	PTS	TRB	WS	eFG	height
age	0.25973868	0.2842778540	0.16609834	0.646737562	0.088264915	-0.105522152
AST	0.39308625	0.6366635188	0.12714491	0.462026828	-0.047596320	-0.527488506
FG	0.69047200	0.1994511838	0.55301395	0.259989869	0.860697355	0.438165050
FG3	-0.01054362	0.2538094465	-0.18132597	0.062398226	0.117402814	-0.371729580
FT	0.11869803	0.3176241823	-0.13542618	0.155952532	-0.008211145	-0.272720738
games	0.47895319	0.5741777233	0.40120882	0.839684436	0.192102961	-0.038602826
PER	1.00000000	0.7019886500	0.67981745	0.622883120	0.590810290	0.193216934
PTS	0.70198865	1.0000000000	0.55823654	0.683043476	0.210252630	-0.097353385
TRB	0.67981745	0.5582365429	1.00000000	0.532154899	0.374209135	0.506925980
WS	0.62288312	0.6830434758	0.53215490	1.000000000	0.203844341	0.038856563
eFG	0.59081029	0.2102526299	0.37420914	0.203844341	1.000000000	0.262726567
height	0.19321693	-0.0973533850	0.50692598	0.038856563	0.262726567	1.000000000
position	-0.26206423	0.0477374904	-0.19798477	-0.005197819	-0.114141109	-0.094468250
shoots	-0.14167976	-0.1132685847	-0.11843646	-0.074815278	-0.059279446	0.002968970
weight	0.27429812	-0.0001830801	0.56071078	0.159940491	0.252361052	0.811746460
team	0.03851069	0.0153304730	0.01463674	0.010995065	0.038753088	0.004541374

	position	shoots	weight	team
age	0.103350976	-0.03804782	0.0165369670	0.0006707831

Linear Regression Models

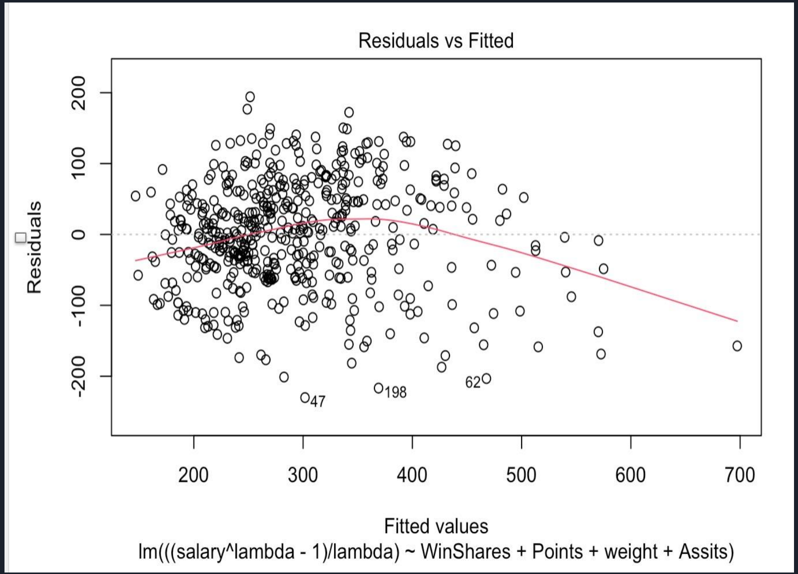
Initial Model : Linear model fitted on all parameters.

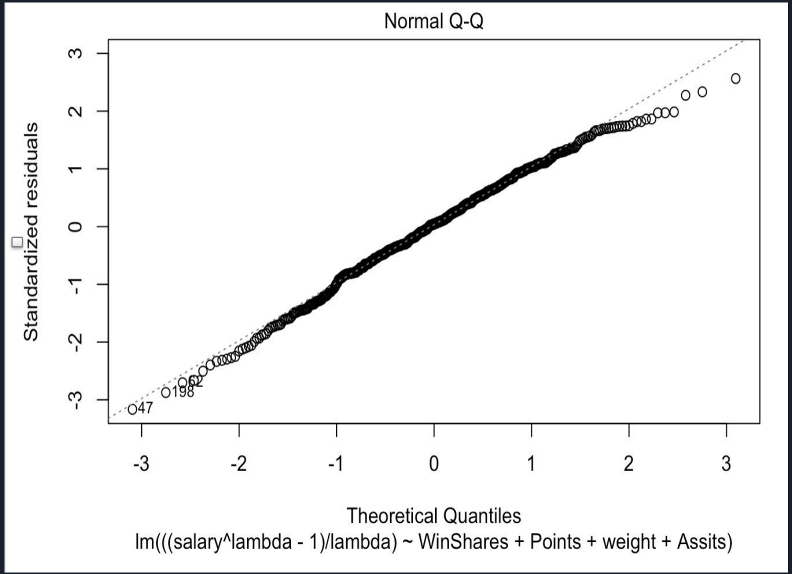




Linear Regression Models

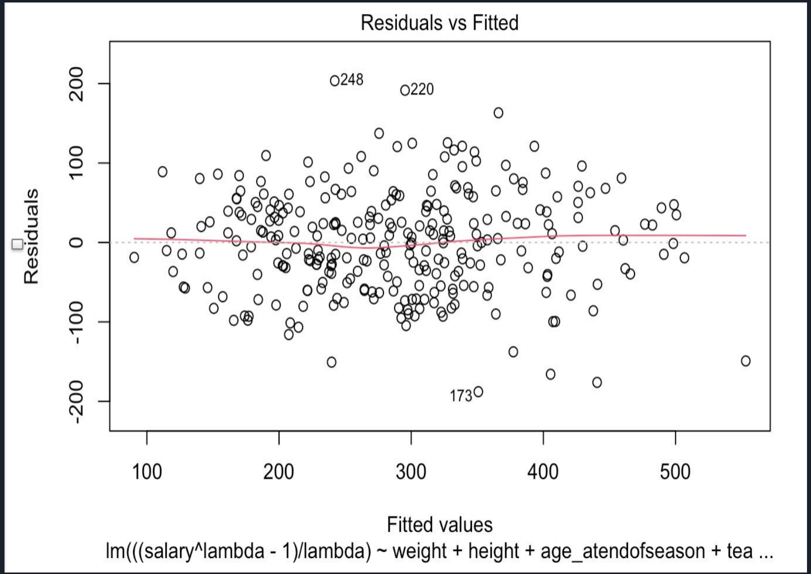
Intermediate Model : Linear model fitted on WinShares, Points Scored, Weight and Assists.

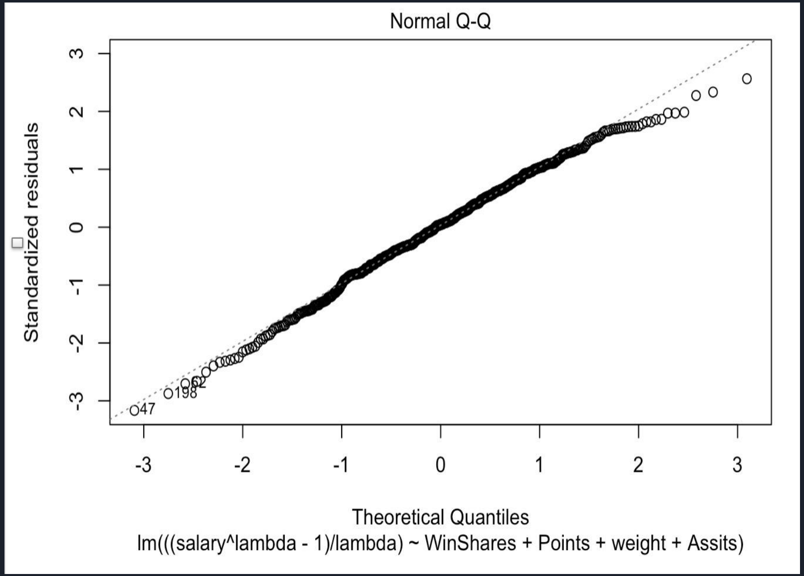




Linear Regression Models

Accepted Model : Improved the intermediate model by **removing the outliers.**





Model Performances

Models	R Squared Values	Shapiro - Wilk Test(wilk normality Test) p values	Durbin Watson Test(Variance Test)
Initial Model	0.5779	4.285e-10	0.761
Intermediate Model	0.5341	0.02219	0.1354
Accepted Model	**0.6651**	0.704	0.7627

Conclusion

- We think that the analysis we performed on the **players' on court performances** with respect to their salaries has a major significance in **potentially determining player** contracts.
- However, there are far too **many immeasurable qualities** a player has that **data alone cannot capture**.
- Things like **being a popular** player for the team could **significantly boost jersey sales** even though they may not be performing as well as everyone else on the court **isn't taken** into account in **our analysis**.
- Even though a player **isn't performing as good as he used** to, he clearly has a use within the team to help everyone get better in **practice and helping rookies** and younger players **develop their skills.**

Anything more you wanted to see?