

# NBA Player Statistics & Performance Over the Course of the Modern NBA



By:  
Stephen Crimi  
Reed Daniel  
Jason Klein  
Kelly McClelland



### Topic:

The relationship between NBA player statistics and performance over the course of the modern NBA

### Reason for selecting the topic:

We would like to explore statistical data to determine if the NBA has a baseline for its athletes, how that baseline weighs against the top performers throughout the decades, and if you can predict the number of games a player would play per season.



# Source Data

- Our source data is from Kaggle and includes NBA Players Stats since 1950 in the form of .csv files.
- For purpose of our analysis we will only focus on years 1980 to present as that is when the “modern” NBA began.

File Name	Number of Rows	Number of Columns
Player.csv	3,922	8
Season_Stats.csv	24,690	52
Player_Data.csv	4,550	8
NBA_Players_AllStars_All.csv	943	7

File Name	Number of Rows	Number of Columns
player_df.shape	3,919	7
seasons_df_shape	18,297	52
per_game_df.shape	18,297	11

# Questions to Answer







- Does the NBA look the same decade by decade in terms of performance?
- What does a prototypical player look like in each decade?
  - How has that change over time?
- Can you predict whether an NBA player can be an All Star?
  - What are the most important stats in terms of determining an NBA All Star?



# Project and Communication Protocols

- Our group maintains a steady flow of communication through Slack and our Teams page to stay updated on the status of the project deliverables as well as to coordinate additional meetings outside of class time.
- GitHub
  - The main branch should include:
    - All code necessary to perform exploratory analysis
    - Some code necessary to complete the machine learning portion of the project
    - README.md must include:
      - Description of the communication protocols
      - Outline of the project (this may include images, but should be easy to follow and digit). (Note: The descriptions and explanations required in all other project deliverables should also be in your README.md as part of your outline, unless otherwise noted.
  - Individual Branches
    - Description of the communication protocols
    - At least one branch for each team member
    - Each team member has at least four commits for the duration of the second segment (eight total commits per person)

# Project and Communication Protocols (con'td)

Responsibility Legend	Symbol	Team Member	Systems	Phase
The team member in the square role will be responsible for the repository		K. McClelland	GitHub Teams Calendar	I & II
The member in the triangle role will create a mockup of a machine learning model. This can even be a diagram that explains how it will work concurrently with the rest of the project steps.		S. Crimi	Machine Learning	I & II
The member in the circle role will create a mockup of a database with a set of sample data, or even fabricated data. This will ensure the database will work seamlessly with the rest of the project		J. Klein	Database	I & II
The member(s) in the X Roll will decide which technologies will be used for each step of the project		S. Crimi J. Klein R. Daniel	Machine Learning Database Dashboard	I & II I & II II



# Technologies Used

- Data Clean & Analysis
  - Pandas will be used to clean and perform exploratory analysis
  - Other Python dependencies like Numpy, Plotly, hvplot, Matplotlib will be imported as needed to assist in our data cleaning
- Database Storage
  - PostresSQL: is where we will be storing our database
- Machine Learning
  - We used PowerBI for our preliminary data inspection to assist in determining our input variables and our desired output
  - SciKitLearn library will be used to create a classifier
    - Balanced Accuracy Score
    - Confusion Matrix
    - Classification Report Imbalanced
  - Tensorflow
- Dashboard
  - The dashboard will be comprised of HTML, CSS, & JavaScript components to create an interactive dashboard allowing users to select player information from drop down fields
  - Results & visualizations will be displayed on GitHub pages and a Tableau dashboard if further storytelling is needed

# Tools

- Tools used to clean and analyze data

Systems	Tools
Data Cleaning & Analysis	Pandas Numpy MathPlotLib
ERD & FSD	Power BI Lucid Charts Quick Database Diagrams
Database	PgAdmin4 AWS
Machine Learning	SciKitLearn Jupyter Notebook Supervised Learning Model
Dashboard	HTML CSS JavaScript Tableau GitHub Pages





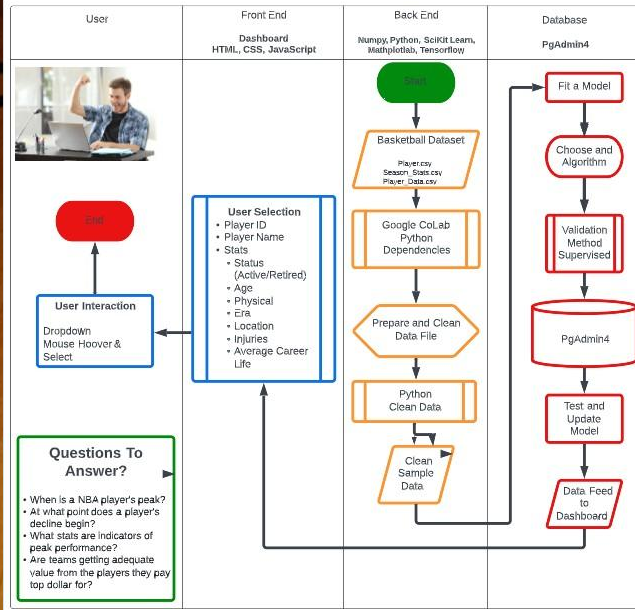
ERD & FSD



# FSD & ERD

## High-Level FSD

### High-Level System Design





Database



# Database (cont'd)

- Perform data cleaning and output ETL documents
- Load created schema into Postgres via pgAdmin
- Upload data into PostgreSQL database
- Create AWS RDS and connect to PostgreSQL
- Connect AWS RDS to Python with SQLAlchemy

## Clean data and end of ETL

```
9 # merge_players DF to csv file
10 merge_players.to_csv('players_clean.csv')
11 #files.download('players_clean.csv')
12 # modern_season_stats DF to csv file
13 merge_modern_season_stats.to_csv('modern_season_stats_clean.csv')
14 files.download('modern_season_stats_clean.csv')
15
16 # per_game_stats DF to csv file
17 per_game_stats.to_csv('per_game_stats_clean.csv')
18 files.download('per_game_stats_clean.csv')
```

```
1 merge_players.sample(2)
```

	Player	height	weight	born	year_start	year_end	position
3187	Brandon Roy	78	214	1984	2007.0	2013.0	G
1447	Billy Reid	77	190	1957	1981.0	1981.0	G

```
1 merge_modern_season_stats.head(2)
```

	Year	Player	Pos	Age	Tm	G	GS	MP	PER	TS%	...	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	id	AllStar
0	1980	Kareem Abdul-Jabbar	C	32	LAL	82	0	3143	25.3	0.639	...	696	886	371	81	280	297	216.0	2034	0	1
1	1980	Tom Abernethy	PF	25	GSW	67	0	1222	11.0	0.511	...	129	191	87	35	12	39	118.0	362	1	0

rows x 52 columns

```
1 per_game_stats.head(2)
```

	Year	Player	Pos	MP(Game)	Points(Game)	Assists(Game)	Rebounds(Game)	Steals(Game)	Blocks(Game)	Turnover(Game)	id
5727	1980	Kareem Abdul-Jabbar	C	38.3	24.8	4.5	10.8	1.0	3.4	3.6	0
5728	1980	Tom Abernethy	PF	18.2	5.4	1.3	2.9	0.5	0.2	0.6	1

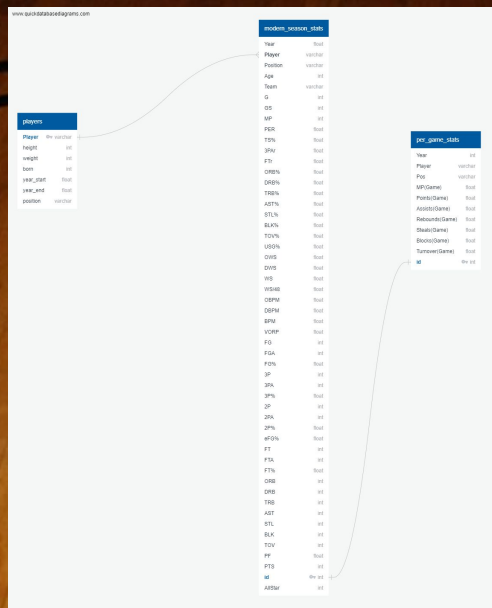


# Database (cont'd)

# Database ERD

- Load created schema into Postgres via pgAdmin

## Sample of schema uploaded into Postgres



## schema ERD

# Database (cont'd)

- Upload data into PostgreSQL database

## Players data SQL table

Query Editor Query History Scratch Pad

```
1 SELECT * FROM players WHERE position != 'null';
```

Data Output Explain Messages Notifications

	Player PK character varying	height integer	weight integer	bom integer	year_start double precision	year_end double precision	position character varying
1	Jim Parsons	78	198	1992	1990	G	
2	James Bailey	81	218	1957	1980	F	
3	Larry Bird	81	218	1956	1980	F	
4	Lawrence Boston	80	225	1956	1980	F	
5	Dudley Bradley	78	194	1957	1980	G	
6	Bill Cartwright	85	245	1957	1995	C	
7	Jeff Cook	82	214	1956	1980	F	
8	Hollis Copeland	78	179	1955	1980	F	
9	John Coughran	79	225	1951	1980	F	
10	Terry Crosby	76	194	1957	1980	G	
11	Pat Cummings	81	229	1956	1980	F	
12	Paul Dawkins	77	190	1957	1980	F	
13	Greg Deane	76	190	1957	1980	G	
14	Larry Dennis	81	225	1957	1980	F	
15	Terry Dueroed	74	179	1956	1980	G	
16	Earl Evans	80	201	1955	1980	F	
17	Mike Evans	73	170	1955	1980	G	
18	Gary Garland	76	179	1957	1980	G	
19	Dave Greenwood	81	220	1957	1980	F	
20	Roy Hamilton	74	179	1957	1980	G	
21	Gerald Henderson	74	174	1956	1980	G	
22	Johnny High	75	183	1957	1980	G	
23	Brad Holland	75	179	1956	1980	G	
24	Phil Hubbard	80	214	1956	1980	F	
25	Geoff Huston	74	174	1957	1980	G	
26	Abdul Jawant	80	209	1954	1980	F	
27	Vinny Johnson	74	188	1956	1980	G	
28	Major Jones	81	225	1953	1980	F	

✓ Successfully run. Total query runtime: 375 msec. 2353 rows affected.

## Players data SQL table without null values

Query Editor Query History Scratch Pad

```
1 SELECT * FROM players WHERE position != 'null';
```

Data Output Explain Messages Notifications

	Player PK character varying	height integer	weight integer	bom integer	year_start double precision	year_end double precision	position character varying
1	Jim Parsons	78	198	1992	1990	G	
2	James Bailey	81	218	1957	1980	F	
3	Larry Bird	81	218	1956	1980	F	
4	Lawrence Boston	80	225	1956	1980	F	
5	Dudley Bradley	78	194	1957	1980	G	
6	Bill Cartwright	85	245	1957	1990	C	
7	Jeff Cook	82	214	1956	1980	F	
8	Hollis Copeland	78	179	1955	1980	F	
9	John Coughran	79	225	1951	1980	F	
10	Terry Crosby	76	194	1957	1980	G	
11	Pat Cummings	81	229	1956	1980	F	
12	Paul Dawkins	77	190	1957	1980	F	
13	Greg Deane	76	190	1957	1980	G	
14	Larry Dennis	81	225	1957	1980	F	
15	Terry Dueroed	74	179	1956	1980	G	
16	Earl Evans	80	201	1955	1980	F	
17	Mike Evans	73	170	1955	1980	G	
18	Gary Garland	76	179	1957	1980	G	
19	Dave Greenwood	81	220	1957	1980	F	
20	Roy Hamilton	74	179	1957	1980	G	
21	Gerald Henderson	74	174	1956	1980	G	
22	Johnny High	75	183	1957	1980	G	
23	Brad Holland	75	179	1956	1980	G	
24	Phil Hubbard	80	214	1956	1980	F	
25	Geoff Huston	74	174	1957	1980	G	
26	Abdul Jawant	80	209	1954	1980	F	
27	Vinny Johnson	74	188	1956	1980	G	
28	Major Jones	81	225	1953	1980	F	

✓ Successfully run. Total query runtime: 375 msec. 2353 rows affected.



# Database (cont'd)

- Upload data into PostgreSQL database

modern\_season\_stats SQL table

Query Editor Query History Scratch Pad

```
1 SELECT * FROM modern_season_stats;
```

Data Output Explain Messages Notifications

Year	Player	Position	Age	Team	G	GS	MP	PER	TS%	SPR	FT%	ORB%	DRB%
integer	character varying	character varying	integer	character varying	integer	integer	integer	double precision	double precision	double precision	double precision	double precision	double precision
1	1980	Kareem Abdul-Jabbar	C	32	LAL	82	0	3143	25.3	0.639	0.001	0.344	7.2
2	1980	Tom Abernethy	PF	25	GSW	67	0	1222	11	0.511	0.003	0.298	5.4
3	1980	Alvan Adams	C	25	PHO	75	0	2168	19.2	0.571	0.002	0.27	8.2
4	1980	Tiny Archibald	PG	31	BOS	80	80	2864	15.3	0.574	0.023	0.548	2.3
5	1980	Dennis Awrey	C	31	CHI	26	0	560	7.4	0.524	0	0.833	6
6	1980	Gus Bailey	SG	28	WSB	20	0	180	9.3	0.467	0.029	0.371	3.3
7	1980	James Bailey	PF	22	SEA	67	0	726	12.3	0.495	0	0.373	10.2
8	1980	Oreg Ballard	SF	25	WSB	82	0	2438	18.1	0.532	0.043	0.206	9.8
9	1980	Mike Bartoni	SF	28	IND	77	0	2330	13.7	0.533	0.004	0.275	8.3
10	1980	Marvin Barnes	SF	27	SOU	20	0	287	8.2	0.432	0	0.533	12.4
11	1980	Rick Barry	PF	35	HOU	72	0	1816	14.8	0.517	0.287	0.198	3.1
12	1980	Tim Bassett	PF	28	TOT	12	0	164	6.7	0.419	0	0.441	7.1
13	1980	Tim Bassett	PF	28	NLN	7	0	92	7.9	0.44	0	0.545	7.9
14	1980	Tim Bassett	PF	28	SAS	5	0	72	5.3	0.375	0	0.25	6.1
15	1980	Billy Ray	SG	23	POR	16	0	235	24.1	0.552	0.13	0.267	6
16	1980	Ron Bateman	PF	29	WSB	6	0	64	10.2	0.449	0	0.261	9.3
17	1980	Kent Benson	C	25	TOT	73	0	1891	12.1	0.513	0.008	0.228	7.3
18	1980	Kent Benson	C	25	MIL	56	0	1389	13.3	0.519	0.002	0.225	7.7
19	1980	Kent Benson	C	25	DET	17	0	502	12.4	0.499	0.021	0.235	6.2
20	1980	Del Beshore	PG	23	CHI	68	0	869	6.4	0.423	0.104	0.348	2.1
21	1980	Henry Bibby	PG	30	PHI	82	0	2035	11.1	0.491	0.083	0.457	3.6
22	1980	Larry Bird	PF	23	BOS	82	82	2955	20.5	0.538	0.098	0.246	8.3
23	1980	Otis Birdsong	SG	24	HOU	82	0	2885	18.5	0.538	0.023	0.266	6.1
24	1980	Ron Boone	SG	33	TOT	81	0	2392	12.1	0.501	0.055	0.214	2.7
25	1980	Ron Boone	SG	6	0	108	0	0	0	0.295	0	0.175	4.5
26	1980	Ron Boone	SG	33	LAL	75	0	2286	12.5	0.506	0.057	0.216	2.7
27	1980	Lawrence Boston	PF	23	WSB	13	0	125	12.6	0.499	0	0.266	2.6
28	1980	Tom Bowell	PF	26	TOT	79	0	2077	14.3	0	0	0	0

Successfully run. Total query runtime: 737 msec. 18927 rows affected.

per\_game\_stats SQL table

Query Editor Query History Scratch Pad

```
1 SELECT * FROM per_game_stats;
```

Data Output Explain Messages Notifications

Year	Player	Pos	MP(Game)	Points(Game)	Assists(Game)	Rebounds(Game)	Steals(Game)	Blocks(Game)	Turnovers(Game)	Id
integer	character varying	character varying	double precision	double precision	double precision	double precision	double precision	double precision	double precision	[P] integer
1	1980	Kareem Abdul-Jabbar	C	38.3	24.6	4.5	10.8	1	3.4	3.6
2	1980	Tom Abernethy	PF	18.2	5.4	1.3	2.9	0.5	0.2	0.6
3	1980	Alvan Adams	C	28.9	14.9	4.3	8.1	1.4	0.7	2.9
4	1980	Tiny Archibald	PG	35.8	14.1	8.4	2.5	1.3	0.1	3
5	1980	Dennis Awrey	C	21.5	3.3	1.5	4.4	0.5	0.6	1
6	1980	Gus Bailey	SG	9	1.9	1.3	1.4	0.4	0.2	0.6
7	1980	James Bailey	PF	10.8	4.7	0.4	2.9	0.3	0.8	1.2
8	1980	Oreg Ballard	SF	29.7	15.6	1.9	7.8	1.1	0.4	1.6
9	1980	Mike Bartoni	SF	30.3	11.8	3.6	5.9	1.1	0.6	2.5
10	1980	Marvin Barnes	SF	14.4	3.2	0.9	3.8	0.2	0.6	0.9
11	1980	Rick Barry	PF	25.2	12	3.7	3.3	1.1	0.4	2.1
12	1980	Tim Bassett	PF	13.7	2.8	1.2	2.8	0.7	0	0.8
13	1980	Tim Bassett	PF	13.1	3.4	0.6	2.6	0.7	0	0.6
14	1980	Tim Bassett	PF	14.4	2	2	3	0.6	0	1
15	1980	Billy Ray	SG	14.7	11.2	1.9	1.8	0.9	0.1	1.2
16	1980	Ron Bateman	PF	10.7	3.8	1.2	2.3	0	0.7	0.7
17	1980	Kent Benson	C	25.9	9.6	2.4	6.2	1	1.3	2.2
18	1980	Kent Benson	C	24.8	8.8	2.3	5.9	1	1.3	1.9
19	1980	Kent Benson	C	29.5	12.1	3	7.1	1.1	1.1	3
20	1980	Del Beshore	PG	12.8	3.6	2	0.9	0.9	0.1	1.5
21	1980	Henry Bibby	PG	24.8	9	3.7	2.5	0.8	0.1	1.8
22	1980	Larry Bird	PF	36	21.3	4.5	10.4	1.7	0.6	3.2
23	1980	Otis Birdsong	SG	35.2	22.7	2.5	4	1.7	0.3	2.2
24	1980	Ron Boone	SG	29.5	12.4	3.8	2.8	1.2	0	2.4
25	1980	Ron Boone	SG	17.7	5.7	1.2	1.8	0.8	0	2.2
26	1980	Ron Boone	SG	30.5	12.9	4	2.9	1.2	0	2.5
27	1980	Lawrence Boston	PF	14.6	4.3	0.2	3	0.3	0.3	0.3
28	1980	Tom Bowell	PF	26.3	11.4	2	5.6	0.4	0	0

Successfully run. Total query runtime: 494 msec. 18927 rows affected.

# Database (cont'd)

- PostgreSQL database tables with joins

Joining modern\_season\_stats and per\_game\_stats SQL table

Query Editor Query History Scratch Pad

```
1 SELECT * FROM modern_season_stats AS mss
2 JOIN per_game_stats as pgs ON mss.id = pgs.fid;
3
```

Data Output Explain Messages Notifications

Year	Player	Position	Age	Team	G	GS	MP	PER	TS%	3P%	FT%	ORB%	DRB%
double precision	character varying	character varying	integer	character varying	integer	integer	integer	double precision	double precision	double precision	double precision	double precision	double precision
1	1980 Kareem Abdul-Jabbar	C	32	LAL	82	0	3143	25.3	0.639	0.001	0.344	7.2	22.0
2	1980 Tom Abernethy	PF	25	GSW	67	0	1222	11	0.511	0.003	0.258	5.4	21.0
3	1980 Alvan Adams	C	25	PHO	75	0	2168	19.2	0.571	0.002	0.27	8.2	16.0
4	1980 Tiny Archibald	PG	31	BOS	80	80	2864	15.3	0.574	0.023	0.548	2.3	5.0
5	1980 Dennis Awrey	C	31	CHI	26	0	560	7.4	0.524	0	0.833	6	16.0
6	1980 Gus Bailey	SG	28	WSB	20	0	180	9.3	0.467	0.029	0.371	3.3	12.0
7	1980 James Bailey	PF	22	SEA	67	0	726	12.3	0.495	0	0.373	10.2	16.0
8	1980 Greg Ballard	SF	25	WSB	82	0	2438	18.1	0.532	0.043	0.206	9.8	16.0
9	1980 Mike Bantom	SF	28	IND	77	0	2330	13.7	0.533	0.004	0.275	8.3	12.0
10	1980 Marvin Barnes	PF	27	SOC	20	0	287	8.2	0.432	0	0.553	12.4	16.0
11	1980 Rick Barry	SF	35	HOU	72	0	1816	14.8	0.517	0.287	0.198	3.1	11.0
12	1980 Tim Bassett	PF	28	TOT	12	0	164	6.7	0.419	0	0.441	7.1	12.0
13	1980 Tim Bassett	PF	28	NJN	7	0	92	7.9	0.44	0	0.545	7.9	12.0
14	1980 Tim Bassett	PF	28	SAS	5	0	72	5.3	0.375	0	0.25	6.1	10.0
15	1980 Billy Ray	SG	23	POR	16	0	235	24.1	0.552	0.13	0.267	6	7.0
16	1980 Ron Behagen	PF	29	WSB	8	0	64	10.2	0.449	0	0.261	9.3	12.0
17	1980 Kent Benson	C	25	TOT	73	0	1891	13.1	0.513	0.008	0.228	7.3	16.0
18	1980 Kent Benson	C	25	MIL	56	0	1389	13.3	0.519	0.002	0.225	7.7	10.0
19	1980 Kent Benson	C	25	DET	17	0	502	12.4	0.499	0.021	0.235	6.2	16.0
20	1980 Del Beshore	PG	33	CHI	68	0	869	6.4	0.423	0.104	0.348	2.1	5.0
21	1980 Henry Bibby	PG	30	PHI	82	0	2035	11.1	0.491	0.083	0.457	3.6	7.0
22	1980 Larry Bird	PF	33	BOS	82	82	2955	20.5	0.538	0.098	0.266	8.3	23.0
23	1980 Otis Bivinsong	SG	24	KCK	82	0	2885	18.5	0.538	0.023	0.266	6.1	6.0
24	1980 Ron Boone	SG	33	TOT	81	0	2392	12.1	0.501	0.055	0.214	2.7	6.0
25	1980 Ron Boone	SG	33	LAL	6	0	106	3	0.395	0	0.175	4.5	7.0
26	1980 Ron Boone	SG	33	UTA	75	0	2286	12.5	0.506	0.057	0.216	2.7	6.0
27	1980 Lawrence Boston	PF	23	WSB	13	0	125	12.6	0.48	0	0.44	7.1	12.0
28	1980 Tom Boswell	PF	26	TOT	79	0	2077	14.3	0.61	0	0.44	7.1	12.0

✓ Successfully run. Total query runtime: 369 msec. 18927 rows affected.

Joining modern\_season\_stats and players SQL table

Query Editor Query History Scratch Pad

```
1 SELECT * FROM modern_season_stats AS mss
2 JOIN per_game_stats as pgs ON mss.fid = pgs.fid;
3
4 SELECT * FROM modern_season_stats as mss
5 JOIN players ON mss."Player" = players."Player";
```

Data Output Explain Messages Notifications

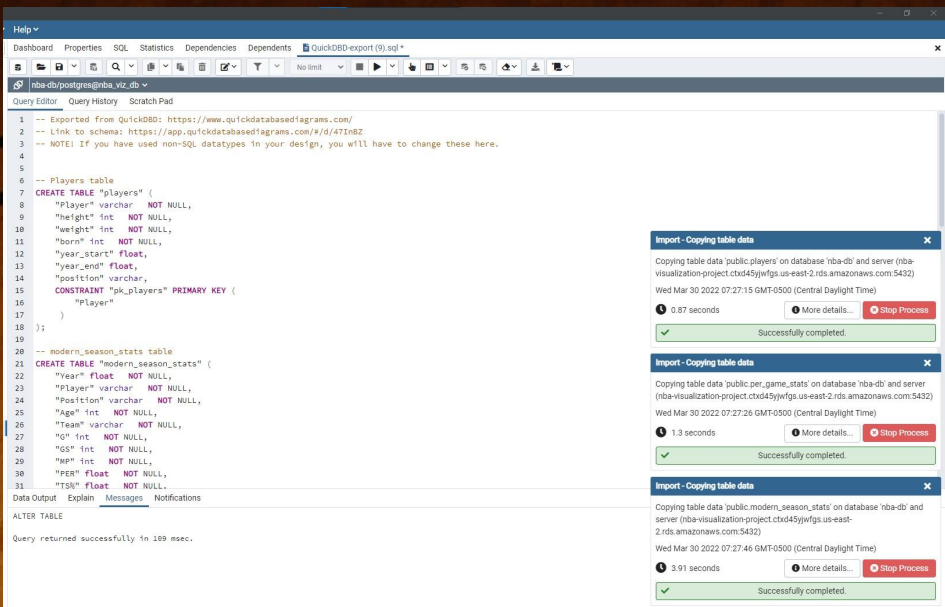
Year	Player	Position	Age	Team	G	GS	MP	PER	TS%	3P%	FT%	ORB%	DRB%
double precision	character varying	character varying	integer	character varying	integer	integer	integer	double precision	double precision	double precision	double precision	double precision	double precision
1	1980 Kareem Abdul-Jabbar	C	32	LAL	82	0	3143	25.3	0.639	0.001	0.344	7.2	22.0
2	1980 Tom Abernethy	PF	25	GSW	67	0	1222	11	0.511	0.003	0.258	5.4	21.0
3	1980 Alvan Adams	C	25	PHO	75	0	2168	19.2	0.571	0.002	0.27	8.2	16.0
4	1980 Tiny Archibald	PG	31	BOS	80	80	2864	15.3	0.574	0.023	0.548	2.3	5.0
5	1980 Dennis Awrey	C	31	CHI	26	0	560	7.4	0.524	0	0.833	6	16.0
6	1980 Gus Bailey	SG	28	WSB	20	0	180	9.3	0.467	0.029	0.371	3.3	12.0
7	1980 James Bailey	PF	22	SEA	67	0	726	12.3	0.495	0	0.373	10.2	16.0
8	1980 Greg Ballard	SF	25	WSB	82	0	2438	18.1	0.532	0.043	0.206	9.8	16.0
9	1980 Mike Bantom	SF	28	IND	77	0	2330	13.7	0.533	0.004	0.275	8.3	12.0
10	1980 Marvin Barnes	PF	27	SOC	20	0	287	8.2	0.432	0	0.553	12.4	16.0
11	1980 Rick Barry	SF	35	HOU	72	0	1816	14.8	0.517	0.287	0.198	3.1	11.0
12	1980 Tim Bassett	PF	28	TOT	12	0	164	6.7	0.419	0	0.441	7.1	12.0
13	1980 Tim Bassett	PF	28	NJN	7	0	92	7.9	0.44	0	0.545	7.9	12.0
14	1980 Billy Ray	SG	23	POR	16	0	235	24.1	0.552	0.13	0.267	6	7.0
15	1980 Ron Behagen	PF	29	WSB	8	0	64	10.2	0.449	0	0.261	9.3	12.0
16	1980 Kent Benson	C	25	TOT	73	0	1891	13.1	0.513	0.008	0.228	7.3	16.0
17	1980 Kent Benson	C	25	MIL	56	0	1389	13.3	0.519	0.002	0.225	7.7	10.0
18	1980 Kent Benson	C	25	DET	17	0	502	12.4	0.499	0.021	0.235	6.2	16.0
19	1980 Del Beshore	PG	33	CHI	68	0	869	6.4	0.423	0.104	0.348	2.1	5.0
20	1980 Henry Bibby	PG	30	PHI	82	0	2035	11.1	0.491	0.083	0.457	3.6	7.0
21	1980 Larry Bird	PF	33	BOS	82	82	2955	20.5	0.538	0.098	0.266	8.3	23.0
22	1980 Otis Bivinsong	SG	24	KCK	82	0	2885	18.5	0.538	0.023	0.266	6.1	6.0
23	1980 Ron Boone	SG	33	TOT	81	0	2392	12.1	0.501	0.055	0.214	2.7	6.0
24	1980 Ron Boone	SG	33	LAL	6	0	106	3	0.395	0	0.175	4.5	7.0
25	1980 Ron Boone	SG	33	UTA	75	0	2286	12.5	0.506	0.057	0.216	2.7	6.0

✓ Successfully run. Total query runtime: 518 msec. 18927 rows affected.



# Database (cont'd)

- Create AWS RDS and connect to PostgreSQL



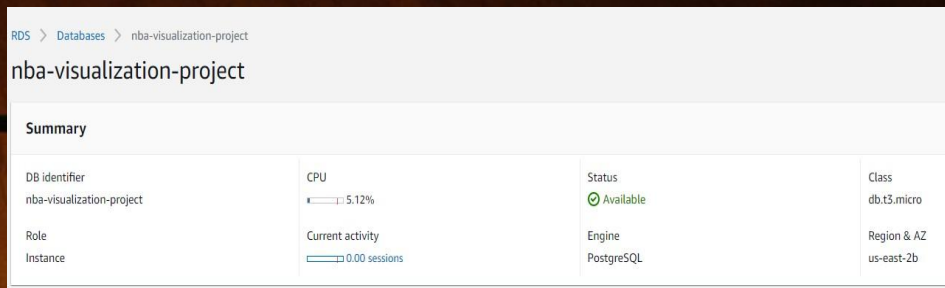
The screenshot shows a SQL client interface with a query editor on the left and a console window on the right. The query editor contains SQL code to create two tables: 'players' and 'modern\_season\_stats'. The console window shows three successful import operations for these tables.

```
1 -- Exported from QuickDBD: https://www.quickdatabasediagrams.com/
2 -- Link to schema: https://app.quickdatabasediagrams.com/#/id/471mZ
3 -- NOTE! If you have used non-SQL datatypes in your design, you will have to change these here.
4
5
6 -- Players table
7 CREATE TABLE "players" (
8   "Player" varchar NOT NULL,
9   "height" int NOT NULL,
10  "weight" int NOT NULL,
11  "born" int NOT NULL,
12  "year_start" float,
13  "year_end" float,
14  "position" varchar,
15  CONSTRAINT "pk_players" PRIMARY KEY (
16    "Player"
17  )
18 );
19
20 -- modern_season_stats table
21 CREATE TABLE "modern_season_stats" (
22   "Year" float NOT NULL,
23   "Player" varchar NOT NULL,
24   "position" varchar NOT NULL,
25   "Age" int NOT NULL,
26   "Team" varchar NOT NULL,
27   "GP" int NOT NULL,
28   "GS" int NOT NULL,
29   "MP" int NOT NULL,
30   "PER" float NOT NULL,
31   "TS" float NOT NULL
32 );
```

Import - Copying table data  
Copying table data 'public.players' on database 'nba-db' and server 'nba-visualization-project.ctxd45jwfgs-us-east-2.rds.amazonaws.com:5432'  
Wed Mar 30 2022 07:27:15 GMT-0500 (Central Daylight Time)  
0.87 seconds More details... Stop Process  
Successfully completed.

Import - Copying table data  
Copying table data 'public.per\_game\_stats' on database 'nba-db' and server 'nba-visualization-project.ctxd45jwfgs-us-east-2.rds.amazonaws.com:5432'  
Wed Mar 30 2022 07:27:26 GMT-0500 (Central Daylight Time)  
1.3 seconds More details... Stop Process  
Successfully completed.

Import - Copying table data  
Copying table data 'public.modern\_season\_stats' on database 'nba-db' and server 'nba-visualization-project.ctxd45jwfgs-us-east-2.rds.amazonaws.com:5432'  
Wed Mar 30 2022 07:27:46 GMT-0500 (Central Daylight Time)  
3.91 seconds More details... Stop Process  
Successfully completed.



The screenshot shows the AWS RDS console for the 'nba-visualization-project' database instance. The summary table provides key details about the instance.

Summary			
DB identifier nba-visualization-project	CPU 5.12%	Status Available	Class db.t3.micro
Role Instance	Current activity 0.00 sessions	Engine PostgreSQL	Region & AZ us-east-2b

Database created in AWS

PostgreSQL database connected to AWS

# Database (cont'd)

- Connect AWS RDS to Python with SQLAlchemy

```
[ ] 1 import sqlalchemy
    2 from sqlalchemy.ext.automap import automap_base
    3 from sqlalchemy.orm import Session
    4 from sqlalchemy import create_engine, func
    5 import psycopg2
    6 import pandas as pd
    7 from psycopg2 import sql

[ ] 1 # SQLAlchemy create_engine('postgresql://username:password@host/db_name)
    2 engine = create_engine('postgresql://postgres:password@visualization-project.ctxd45yjjwfgs.us-east-2.rds.amazonaws.com/nba-db')

[ ] 1 Base = automap_base()
    2 Base.prepare(engine, reflect=True)

[ ] 1 # query syntax: df = pd.read_sql(query.statement, connection)
    2
    3 ## players DF from players table
    4 players_df = pd.read_sql('SELECT * FROM players', engine)
    5 ## seasons_df from modern_season_stats table
    6 seasons_df = pd.read_sql('SELECT * FROM modern_season_stats', engine)
    7 ## per_game_df from per_game_stats table
    8 per_game_df = pd.read_sql('SELECT * FROM per_game_stats', engine)

[ ] 1 per_game_df.where(per_game_df.Player == 'Michael Jordan').dropna()
```

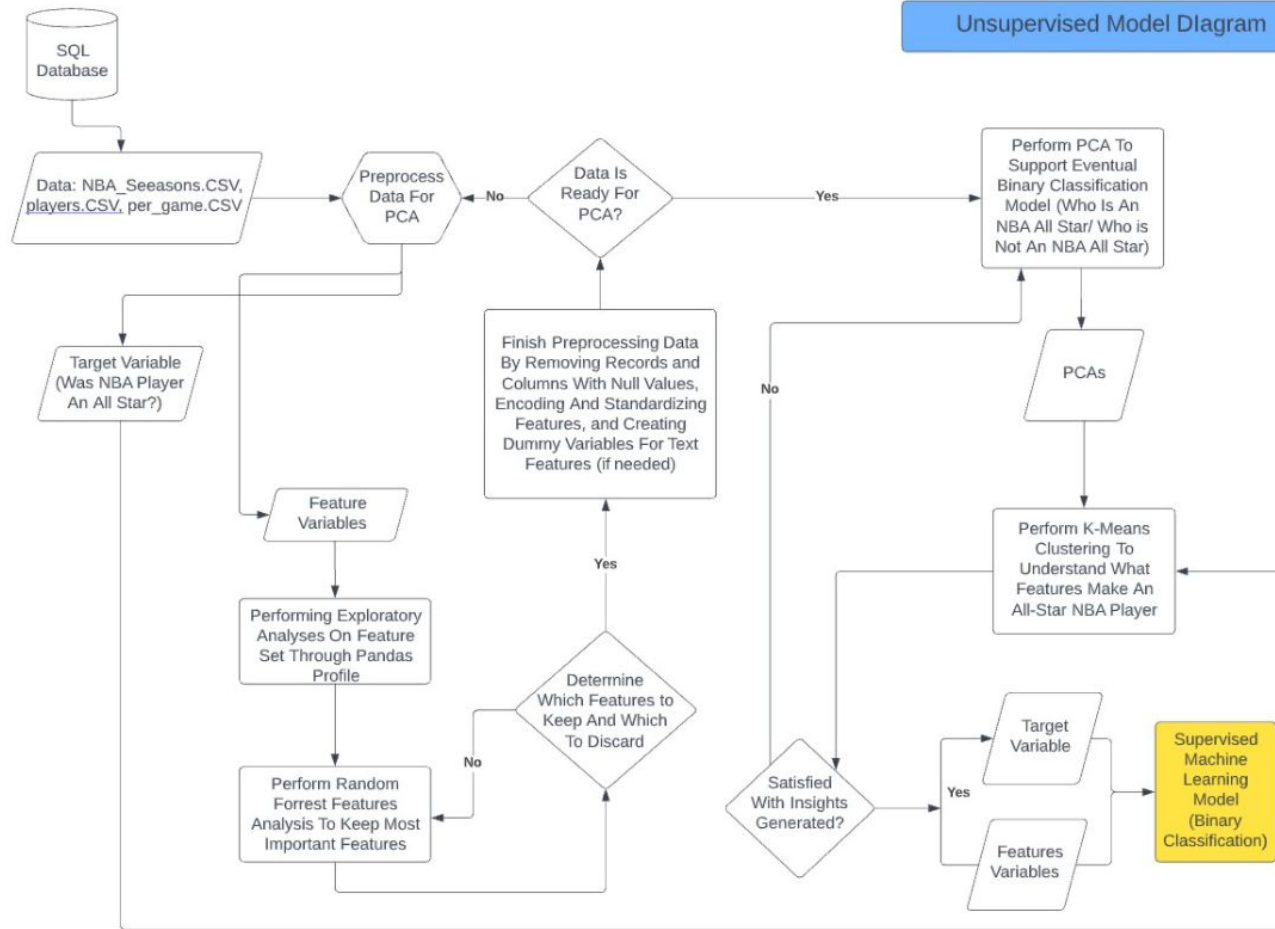
	Year	Player	Pos	MP(Game)	Points(Game)	Assists(Game)	Rebounds(Game)	Steals(Game)	Blocks(Game)	Turnover(Game)	id
1984	1985.0	Michael Jordan	SG	38.3	28.2	5.9	6.5	2.4	0.8	3.5	1984.0
2361	1986.0	Michael Jordan	SG	25.1	22.7	2.9	3.6	2.1	1.2	2.5	2361.0
2735	1987.0	Michael Jordan	SG	40.0	37.1	4.6	5.2	2.9	1.5	3.3	2735.0
3134	1988.0	Michael Jordan	SG	40.4	35.0	5.9	5.5	3.2	1.6	3.1	3134.0
3561	1989.0	Michael Jordan	SG	40.2	32.5	8.0	8.0	2.9	0.8	3.6	3561.0
4010	1990.0	Michael Jordan	SG	39.0	33.6	6.3	6.9	2.8	0.7	3.0	4010.0



# Machine Learning

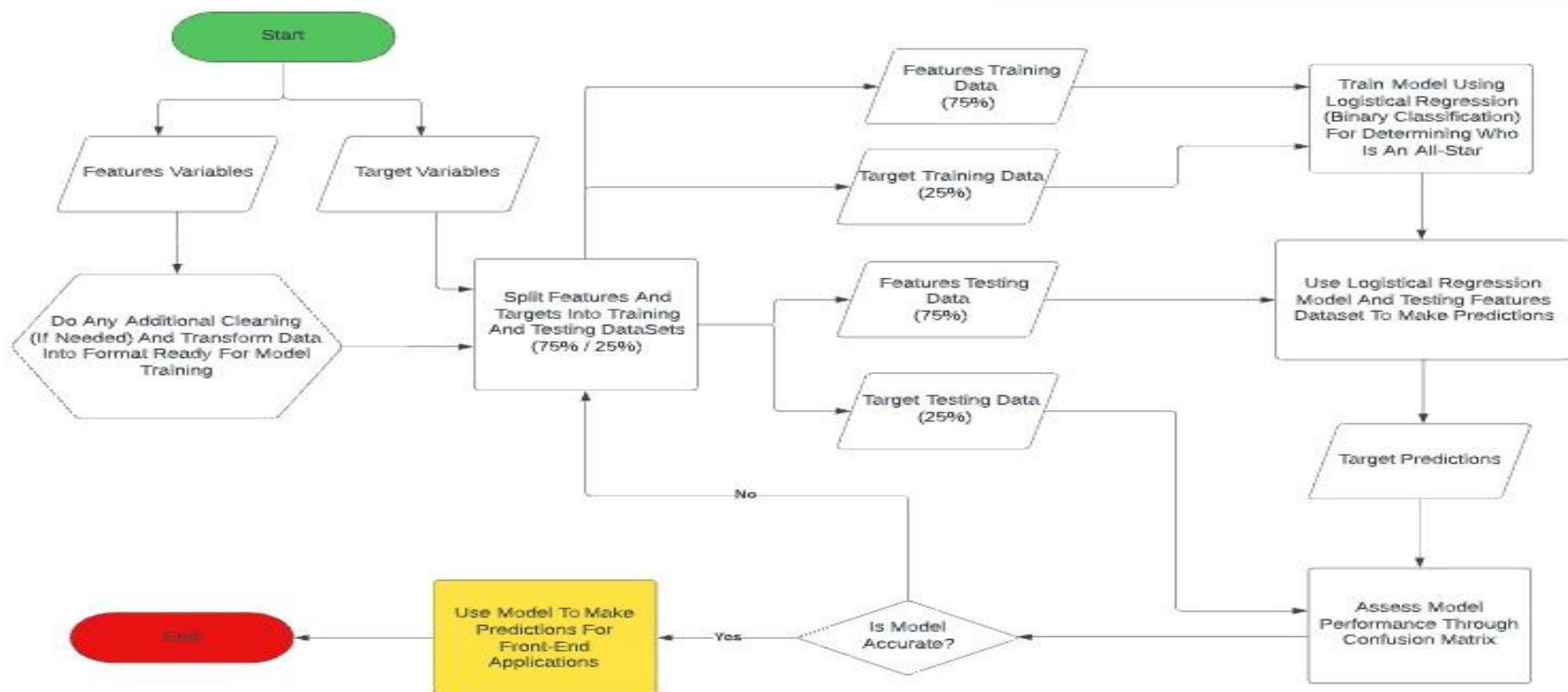


# Unsupervised Model Diagram





## Supervised Machine Learning

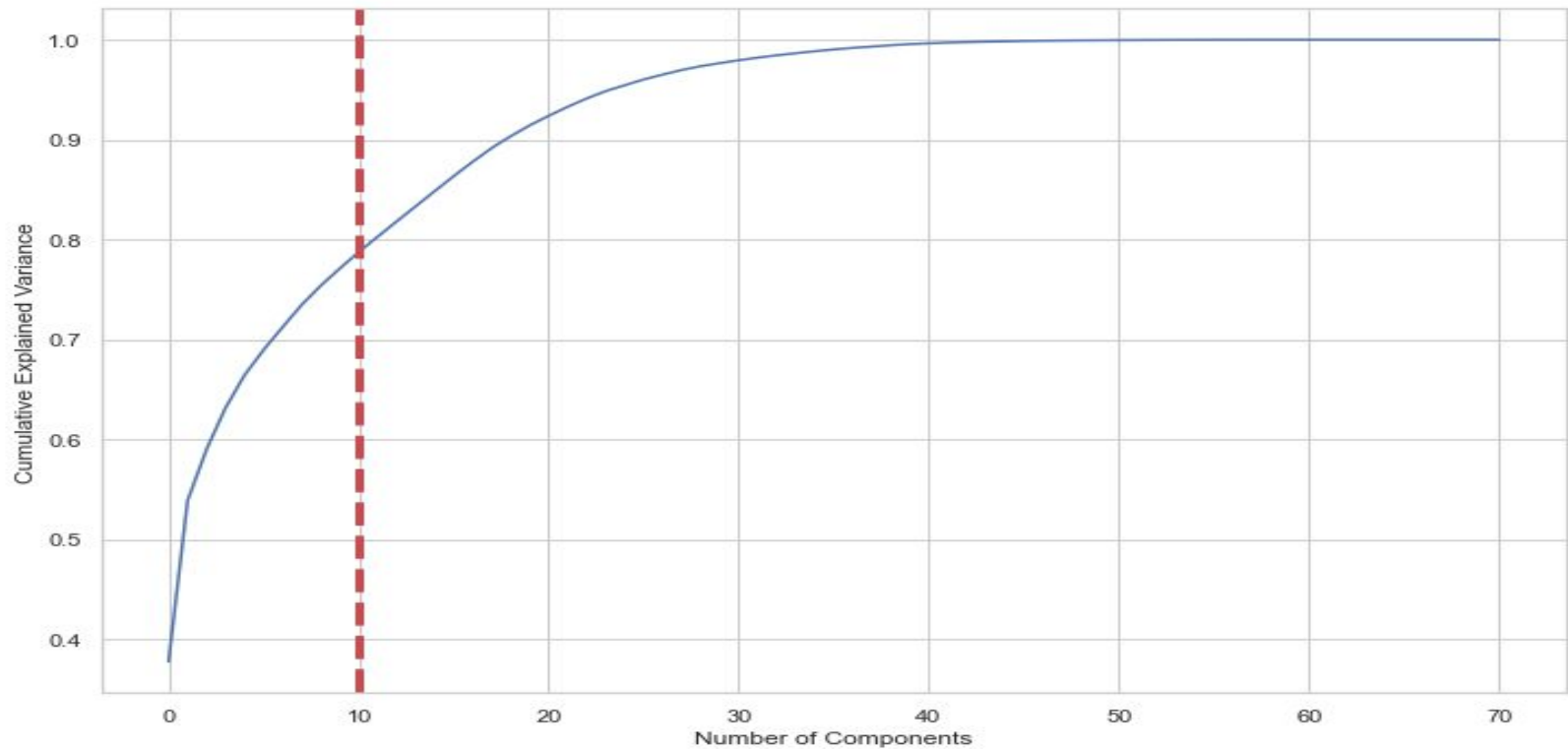


# Pre-Processing

- Games  $\Rightarrow$  NBA AllStar
- Initially ranked by feature importance through random forests without PCA.
- Saw overrepresentation issues with non-NBA AllStar players so performed over- and under-sampling.
  - Decided to go with under-sampling and applied it as a pre-processing step for ALL OF OUR MODELS.
- Performed other cleaning tasks for the features.
- Target was whether player would be an NBA AllStar (1) or not (0).



# PCA: Going From 68 to 10!



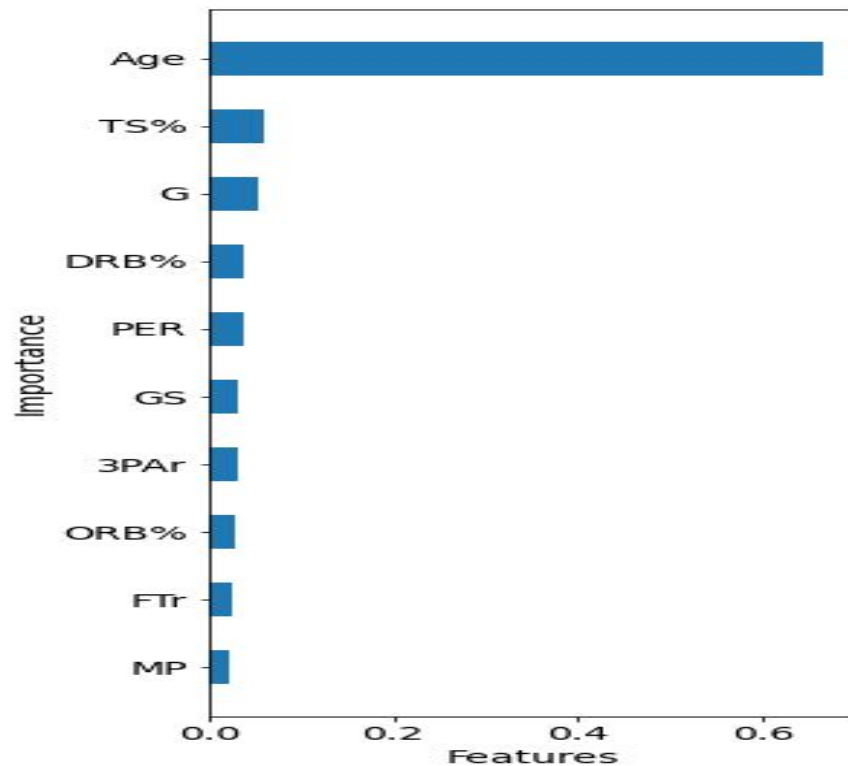
# PCA: Top Ten Features Explanatory Power

Out[25]:

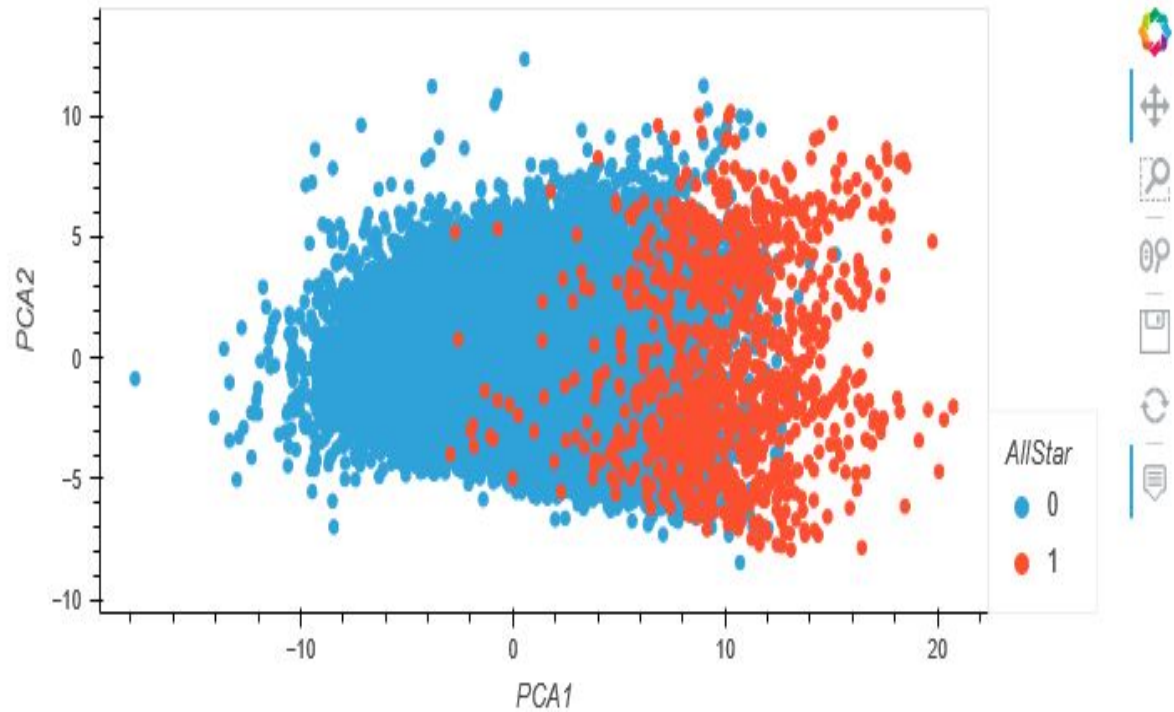
	Cumulative Variance Ratio	Explained Variance Ratio
0	0.377207	0.377207
1	0.538460	0.161253
2	0.589985	0.051525
3	0.631495	0.041511
4	0.664259	0.032763
5	0.689535	0.025276
6	0.712278	0.022743
7	0.734474	0.022196
8	0.753694	0.019219
9	0.770883	0.017190



# PCA: Top Ten Features



# K-Means



# Supervised Models

- Metrics of Interests: Accuracy, Recall, and F-1 score.
- Binary Classification!!!
- Models:
  - a.) Logistic Regression
  - b.) Support Vector Machine
  - c.) Decision Tree
  - d.) Random Forests (with top ten variables this time!)
  - e.) Boosting



# Summary of Supervised Performances!

Metric	Logistical Regression	SVM	Decision Tree	Random Forest	Boosting
Testing Accuracy	91.1%	90.7%	89.3%	92.4%	91.9%
Recall (for predicting non-AllStars)	91%	91%	90%	92%	92%
Recall (for predicting AllStars)	94%	94%	85%	94%	92%
F-1 Scores (for predicting AllStars)	51%	50%	44%	55%	53%

# Dashboard Storyboards & Mockups





# Storyboards & Dashboard

- A blueprint for the dashboard is created and includes all of the following:
  - Storyboard on Google Slide(s)
  - Description of tool(s) that will be used to create final dashboard
  - Description of interactive element(s)
- Tools used :
  - HTML
  - .css
  - JavaScript
  - GitHub Pages



# Dashboard - Page 1

## Storyboard Layout

### NBA Prime Time!

Navigation Bar

Page Header: NBA Final Project

Scope of Project

Questions we'd like to answer

Stat Range Sliders

Filter For Decade & Percentiles

Dropdown for Player Info

Table for Player stat data

## Mockup Page

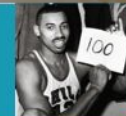
### NBA Prime Time!



Performance in the League:  
How do players measure up  
against the greats?

- Does performance in the NBA look the same decade by decade?
- What does a prototypical player look like in each decade?
- How has that changed over time?

- Does the decade have an affect on the number of games a player will play in a season?



LET'S FIND OUT!

Filter Search

Era Search

Enter Name

Michael Jordan

Enter Birth Year

1991

Enter Position

C

Enter Height(in)

72

Enter Weight(lbs)

200

Enter Stat

points

Player Name:

Player Info

Enter Decade

90s

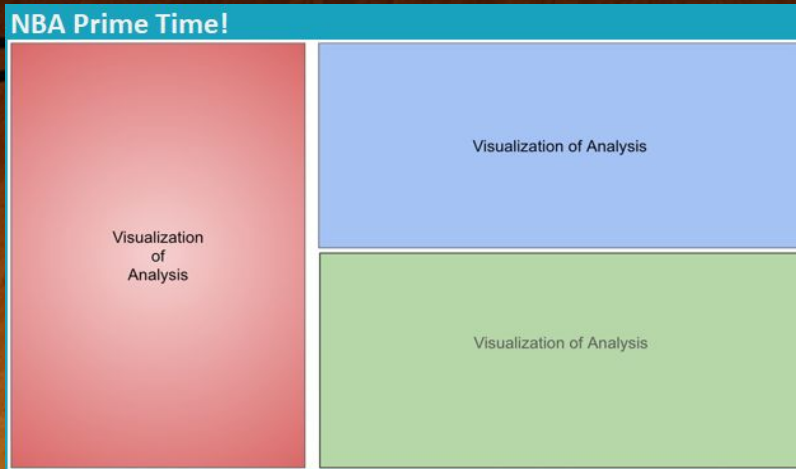
Enter Percentile

Top 25%

Date	Position	2P	2PA	2P%	3P	3PA	3P%	Allstar
1995 SG		928	1573	0.524	9	52	0.173	Yes
1996 SG		147	310	0.474	3	18	0.167	Yes
1997 SG		1086	2233	0.491	42	144	0.182	Yes
1998 SG		1062	1948	0.544	7	53	0.132	Yes
1999 SG		939	1697	0.553	27	99	0.276	Yes
2000 SG		942	1719	0.548	42	245	0.376	Yes
1991 SG		961	1744	0.551	29	93	0.312	Yes
1992 SG		916	1719	0.533	27	100	0.27	Yes
1993 SG		911	1733	0.514	41	239	0.352	Yes
1994 SG		150	272	0.493	16	32	0.5	No
1996 SG		805	1590	0.506	111	269	0.427	Yes
1997 SG		809	1595	0.507	111	297	0.374	Yes
1998 SG		851	1757	0.482	30	126	0.235	Yes
2002 SF		541	1271	0.425	10	53	0.189	Yes
2003 SF		443	1472	0.45	16	55	0.291	Yes

# Dashboard - Page 2

## Storyboard Layout



## Mockup Page





## Dashboard - Page 3

### Storyboard Layout

