

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:

- 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
- 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 the years 1980 to the 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 changed 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?

Technologies Used

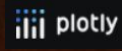
Data Clean & Analysis



Pandas



Google
Colab



Matplotlib



Numpy

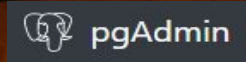


Pickle

FSD & ERD



Database Storage



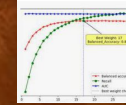
Machine Learning



Google
CoLab



Classification
Report
Imbalanced



		Confusion Matrix	
		Actually Positive (P)	Actually Negative (N)
Predicted Positive (P)	True Positives (TP)	True Positives (TP)	False Positives (FP)
Predicted Negative (N)	False Negatives (FN)	False Negatives (FN)	True Negatives (TN)



Imblearn
MarkupSafe
Pycopg2
Boto3
json

Dashboard



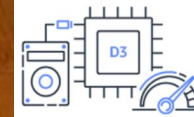
Lambda



Bootstrap



S3



Results & Visualizations



Google Slides

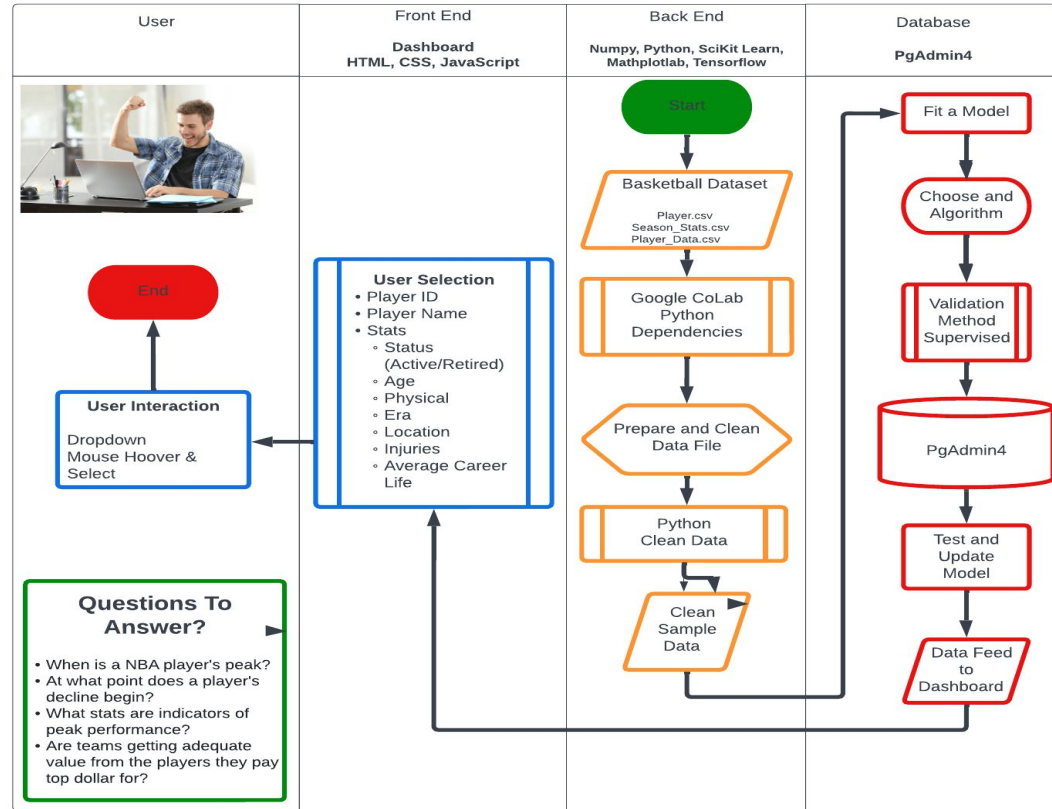


GitHub

READMEs

High-Level FSD

High-Level System Design



Database



Database (cont'd)

- Perform data cleaning and output ETL documents
- Load created schema into PostgreSQL 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

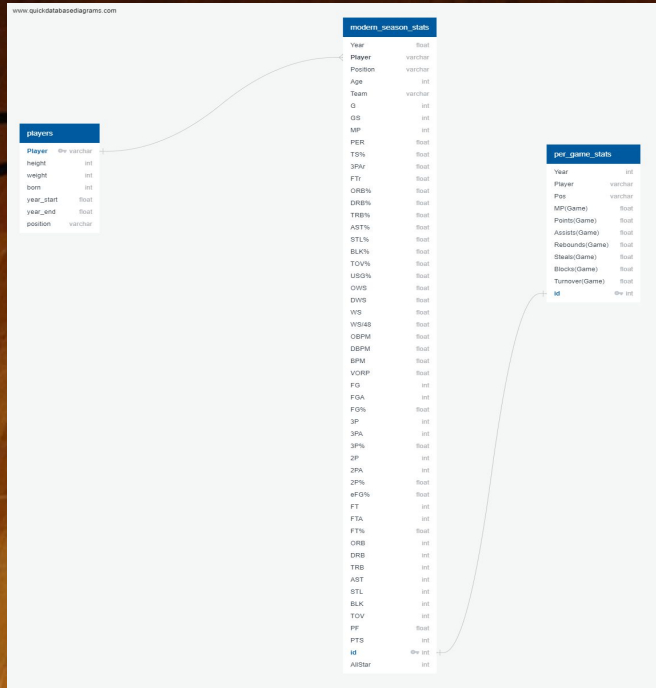
```
1 per_game_stats.head(2)
```

	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



schema ERD

Sample of schema uploaded into Postgres

```
1 -- Exported from QuickDBD: https://www.quickdatabasediagrams.com/
2 -- Link to schema: https://app.quickdatabasediagrams.com/#/id/1234567
3 -- NOTE! If you have used non-SQL datatypes in your table, use
4 -- `CAST()` to convert them. Example: `CAST('foo' AS TEXT)`
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 NOT NULL,
13    "year_end" float NOT NULL,
14    "position" varchar NOT NULL,
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    "G" int NOT NULL,
28    "GS" int NOT NULL,
29    "MP" int NOT NULL,
30    "PER" float NOT NULL,
31    "TSS" float NOT NULL,
32    "3PM" float NOT NULL,
33    "FT" float NOT NULL,
34    "ORB%" float NOT NULL,
35    "DRB%" float NOT NULL,
36    "TRB%" float NOT NULL,
37    "AST%" float NOT NULL,
38    "STL%" float NOT NULL,
39    "BLK%" float NOT NULL,
40    "TOV%" float NOT NULL,
41    "USG%" float NOT NULL,
42    "OWS" float NOT NULL,
43    "DWS" float NOT NULL,
44    "WS" float NOT NULL,
45    "WS/48" float NOT NULL,
46    "OBPM" float NOT NULL,
47    "DBPM" float NOT NULL,
48    "BPM" float NOT NULL,
49    "VORPM" float NOT NULL,
50
55    "3PA" float NOT NULL,
56    "2P" int NOT NULL,
57    "2PA" int NOT NULL,
58    "2P%" float NOT NULL,
59    "eFG%" float NOT NULL,
60    "FT" int NOT NULL,
61    "FTA" int NOT NULL,
62    "FT%" float NOT NULL,
63    "ORB" int NOT NULL,
64    "DRB" int NOT NULL,
65    "TRB" int NOT NULL,
66    "AST" int NOT NULL,
67    "STL" int NOT NULL,
68    "BLK" int NOT NULL,
69    "TOV" int NOT NULL,
70    "PF" float NOT NULL,
71    "PTS" int NOT NULL,
72    "id" int NOT NULL,
73    "Allstar" int NOT NULL,
74    CONSTRAINT "pk_modern_season_stats" PRIMARY KEY (
75        "id"
76    )
77 );
78
79 -- per_game_stats table
80 CREATE TABLE "per_game_stats" (
81    "Year" int NOT NULL,
82    "Player" varchar NOT NULL,
83    "Pos" varchar NOT NULL,
84    "MP(Game)" float NOT NULL,
85    "Points(Game)" float NOT NULL,
86    "Assists(Game)" float NOT NULL,
87    "Rebounds(Game)" float NOT NULL,
88    "Steals(Game)" float NOT NULL,
89    "Blocks(Game)" float NOT NULL,
90    "Turnover(Game)" float NOT NULL,
91    "id" int NOT NULL,
92    CONSTRAINT "pk_per_game_stats" PRIMARY KEY (
93        "id"
94    )
95 );
96
97 ALTER TABLE "modern_season_stats" ADD CONSTRAINT "fk_modern_season_stats_player" FOREIGN KEY("Player")
98 REFERENCES "players" ("Player");
99
100 ALTER TABLE "modern_season_stats" ADD CONSTRAINT "fk_modern_season_stats_id" FOREIGN KEY("id")
101 REFERENCES "per_game_stats" ("id");
```

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	born integer	year_start double precision	year_end double precision	position character varying
1	Jim Paxson	78	198	1932	1980	1990	G
2	James Bailey	81	218	1957	1980	1988	F
3	Larry Bird	81	218	1956	1980	1992	F
4	Lawrence Boston	80	225	1956	1980	1980	F
5	Dudley Bradley	78	194	1957	1980	1989	G
6	Bill Cartwright	85	245	1957	1980	1995	C
7	Jeff Cook	82	214	1956	1980	1988	F
8	Hollis Copeland	78	179	1955	1980	1982	F
9	John Coughran	79	225	1951	1980	1980	F
10	Terry Crosby	76	194	1957	1980	1980	G
11	Pat Cummings	81	229	1956	1980	1991	F
12	Paul Dawkins	77	190	1957	1980	1980	F
13	Greg Deane	76	190	1957	1980	1980	G
14	Larry Demic	81	225	1957	1980	1982	F
15	Terry Duero	74	179	1956	1980	1983	G
16	Earl Evans	80	201	1955	1980	1980	F
17	Mike Evans	73	170	1955	1980	1988	G
18	Gary Garland	76	179	1957	1980	1980	G
19	Dave Greenwood	81	220	1957	1980	1991	F
20	Roy Hamilton	74	179	1957	1980	1981	G
21	Gerald Henderson	74	174	1956	1980	1992	G
22	Johnny High	75	183	1957	1980	1984	G
23	Brad Holland	75	179	1956	1980	1982	G
24	Phil Hubbard	80	214	1956	1980	1989	F
25	Geoff Huston	74	174	1957	1980	1987	G
26	Abdul Jawani	80	209	1954	1980	1981	F
27	Vinnie Johnson	74	198	1956	1980	1992	G
28	Major Jones	81	225	1953	1980	1985	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	born integer	year_start double precision	year_end double precision	position character varying
1	Jim Paxson	78	198	1932	1980	1990	G
2	James Bailey	81	218	1957	1980	1988	F
3	Larry Bird	81	218	1956	1980	1992	F
4	Lawrence Boston	80	225	1956	1980	1980	F
5	Dudley Bradley	78	194	1957	1980	1989	G
6	Bill Cartwright	85	245	1957	1980	1995	C
7	Jeff Cook	82	214	1956	1980	1988	F
8	Hollis Copeland	78	179	1955	1980	1982	F
9	John Coughran	79	225	1951	1980	1980	F
10	Terry Crosby	76	194	1957	1980	1980	G
11	Pat Cummings	81	229	1956	1980	1991	F
12	Paul Dawkins	77	190	1957	1980	1980	F
13	Greg Deane	76	190	1957	1980	1980	G
14	Larry Demic	81	225	1957	1980	1982	F
15	Terry Duero	74	179	1956	1980	1983	G
16	Earl Evans	80	201	1955	1980	1980	F
17	Mike Evans	73	170	1955	1980	1988	G
18	Gary Garland	76	179	1957	1980	1980	G
19	Dave Greenwood	81	220	1957	1980	1991	F
20	Roy Hamilton	74	179	1957	1980	1981	G
21	Gerald Henderson	74	174	1956	1980	1992	G
22	Johnny High	75	183	1957	1980	1984	G
23	Brad Holland	75	179	1956	1980	1982	G
24	Phil Hubbard	80	214	1956	1980	1989	F
25	Geoff Huston	74	174	1957	1980	1987	G
26	Abdul Jawani	80	209	1954	1980	1981	F
27	Vinnie Johnson	74	198	1956	1980	1992	G
28	Major Jones	81	225	1953	1980	1985	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%
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
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	SF	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	KCK	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	0.494	0	0.288	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
2	1980 Tom Abernethy	PF	25	GSW	67	0	1222	11	0.511	0.003	0.258	5.4	21
3	1980 Alvan Adams	C	25	PHO	75	0	2168	19.2	0.571	0.002	0.27	8.2	21
4	1980 Tiny Archibald	PG	31	BOS	80	80	2864	15.3	0.574	0.023	0.548	2.3	5
5	1980 Dennis Awrey	C	31	CHI	26	0	560	7.4	0.524	0	0.833	6	16
6	1980 Gus Bailey	SG	28	WSB	20	0	180	9.3	0.467	0.029	0.371	3.3	12
7	1980 James Bailey	PF	22	SEA	67	0	726	12.3	0.495	0	0.373	10.2	16
8	1980 Greg Ballard	SF	25	WSB	82	0	2438	18.1	0.532	0.043	0.206	9.8	16
9	1980 Mike Bantom	PF	28	IND	77	0	2330	13.7	0.533	0.004	0.275	8.3	12
10	1980 Marvin Barnes	PF	27	SOC	20	0	287	8.2	0.482	0	0.553	12.4	16
11	1980 Rick Barry	SF	35	HOU	72	0	1816	14.8	0.517	0.287	0.198	3.1	11
12	1980 Tim Bassett	PF	28	TOT	12	0	164	6.7	0.419	0	0.441	7.1	12
13	1980 Tim Bassett	PF	28	NJN	7	0	92	7.9	0.44	0	0.545	7.9	12
14	1980 Tim Bassett	PF	28	SAS	5	0	72	5.3	0.375	0	0.25	6.1	1
15	1980 Billy Ray	SG	23	POR	16	0	235	24.1	0.552	0.13	0.267	6	7
16	1980 Ron Behagen	PF	29	WSB	8	0	64	10.2	0.449	0	0.261	9.3	12
17	1980 Kent Benson	C	25	TOT	73	0	1891	13.1	0.513	0.008	0.228	7.3	16
18	1980 Kent Benson	C	25	MIL	56	0	1389	13.3	0.519	0.002	0.225	7.7	1
19	1980 Kent Benson	C	25	DET	17	0	502	12.4	0.499	0.021	0.235	6.2	1
20	1980 Del Beshore	PG	33	CHI	68	0	869	6.4	0.423	0.104	0.348	2.1	5
21	1980 Henry Bibby	PG	30	PHI	82	0	2035	11.1	0.491	0.083	0.457	3.6	7
22	1980 Larry Bird	PF	33	BOS	82	82	2955	20.5	0.538	0.098	0.266	8.3	21
23	1980 Otis Bissong	SG	24	KCK	82	0	2885	18.5	0.538	0.023	0.266	6.1	6
24	1980 Ron Boone	SG	33	TOT	81	0	2392	12.1	0.501	0.055	0.214	2.7	6
25	1980 Ron Boone	SG	33	LAL	6	0	106	3	0.395	0	0.175	4.5	7
26	1980 Ron Boone	SG	33	UTA	75	0	2286	12.5	0.506	0.057	0.216	2.7	6
27	1980 Lawrence Boston	PF	23	WSB	13	0	125	12.6	0.48	0	0.44	7.1	12
28	1980 Tom Boswell	PF	26	TOT	79	0	2077	14.3	0.61	0	0.344	7.2	21

✓ 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
2	1980 Tom Abernethy	PF	25	GSW	67	0	1222	11	0.511	0.003	0.258	5.4	21
3	1980 Alvan Adams	C	25	PHO	75	0	2168	19.2	0.571	0.002	0.27	8.2	21
4	1980 Tiny Archibald	PG	31	BOS	80	80	2864	15.3	0.574	0.023	0.548	2.3	5
5	1980 Dennis Awrey	C	31	CHI	26	0	560	7.4	0.524	0	0.833	6	16
6	1980 Gus Bailey	SG	28	WSB	20	0	180	9.3	0.467	0.029	0.371	3.3	12
7	1980 James Bailey	PF	22	SEA	67	0	726	12.3	0.495	0	0.373	10.2	16
8	1980 Greg Ballard	SF	25	WSB	82	0	2438	18.1	0.532	0.043	0.206	9.8	16
9	1980 Mike Bantom	SF	28	IND	77	0	2330	13.7	0.533	0.004	0.275	8.3	12
10	1980 Marvin Barnes	PF	27	SOC	20	0	287	8.2	0.482	0	0.553	12.4	16
11	1980 Rick Barry	SF	35	HOU	72	0	1816	14.8	0.517	0.287	0.198	3.1	11
12	1980 Tim Bassett	PF	28	TOT	12	0	164	6.7	0.419	0	0.441	7.1	12
13	1980 Tim Bassett	PF	28	NJN	7	0	92	7.9	0.44	0	0.545	7.9	12
14	1980 Billy Ray	SG	23	POR	16	0	235	24.1	0.552	0.13	0.267	6	7
15	1980 Ron Behagen	PF	29	WSB	8	0	64	10.2	0.449	0	0.261	9.3	12
16	1980 Kent Benson	C	25	TOT	73	0	1891	13.1	0.513	0.008	0.228	7.3	16
17	1980 Kent Benson	C	25	MIL	56	0	1389	13.3	0.519	0.002	0.225	7.7	1
18	1980 Kent Benson	C	25	DET	17	0	502	12.4	0.499	0.021	0.235	6.2	1
19	1980 Del Beshore	PG	33	CHI	68	0	869	6.4	0.423	0.104	0.348	2.1	5
20	1980 Henry Bibby	PG	30	PHI	82	0	2035	11.1	0.491	0.083	0.457	3.6	7
21	1980 Larry Bird	PF	33	BOS	82	82	2955	20.5	0.538	0.098	0.266	8.3	21
22	1980 Otis Bissong	SG	24	KCK	82	0	2885	18.5	0.538	0.023	0.266	6.1	6
23	1980 Ron Boone	SG	33	TOT	81	0	2392	12.1	0.501	0.055	0.214	2.7	6
24	1980 Ron Boone	SG	33	LAL	6	0	106	3	0.395	0	0.175	4.5	7
25	1980 Ron Boone	SG	33	UTA	75	0	2286	12.5	0.506	0.057	0.216	2.7	6

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

Database (cont'd)

- Create AWS RDS and connect to PostgreSQL

The screenshot shows a PostgreSQL Query Editor with the following SQL code:

```
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  "pborn" 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   "GO" int NOT NULL,
28   "GS" int NOT NULL,
29   "MP" int NOT NULL,
30   "PER" float NOT NULL,
31   "TS" float NOT NULL
22 );
```

Below the code, a summary of three successful imports is shown:

- 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'. Completed at Wed Mar 30 2022 07:27:15 GMT-0500 (Central Daylight Time). Duration: 0.87 seconds. Status: 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'. Completed at Wed Mar 30 2022 07:27:26 GMT-0500 (Central Daylight Time). Duration: 1.3 seconds. Status: 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'. Completed at Wed Mar 30 2022 07:27:46 GMT-0500 (Central Daylight Time). Duration: 3.91 seconds. Status: Successfully completed.

At the bottom, it states: "Query returned successfully in 169 msec."

The screenshot shows the AWS RDS console for the 'nba-visualization-project' database instance. The instance is named 'nba-visualization-project' and is in the 'Available' state. The engine is 'PostgreSQL' and the instance class is 'db.t3.micro'. The CPU usage is 5.12% and the current activity is 0.00 sessions.

DB identifier	CPU	Status	Class
nba-visualization-project	5.12%	Available	db.t3.micro
Role	Current activity	Engine	Region & AZ
Instance	0.00 sessions	PostgreSQL	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: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)

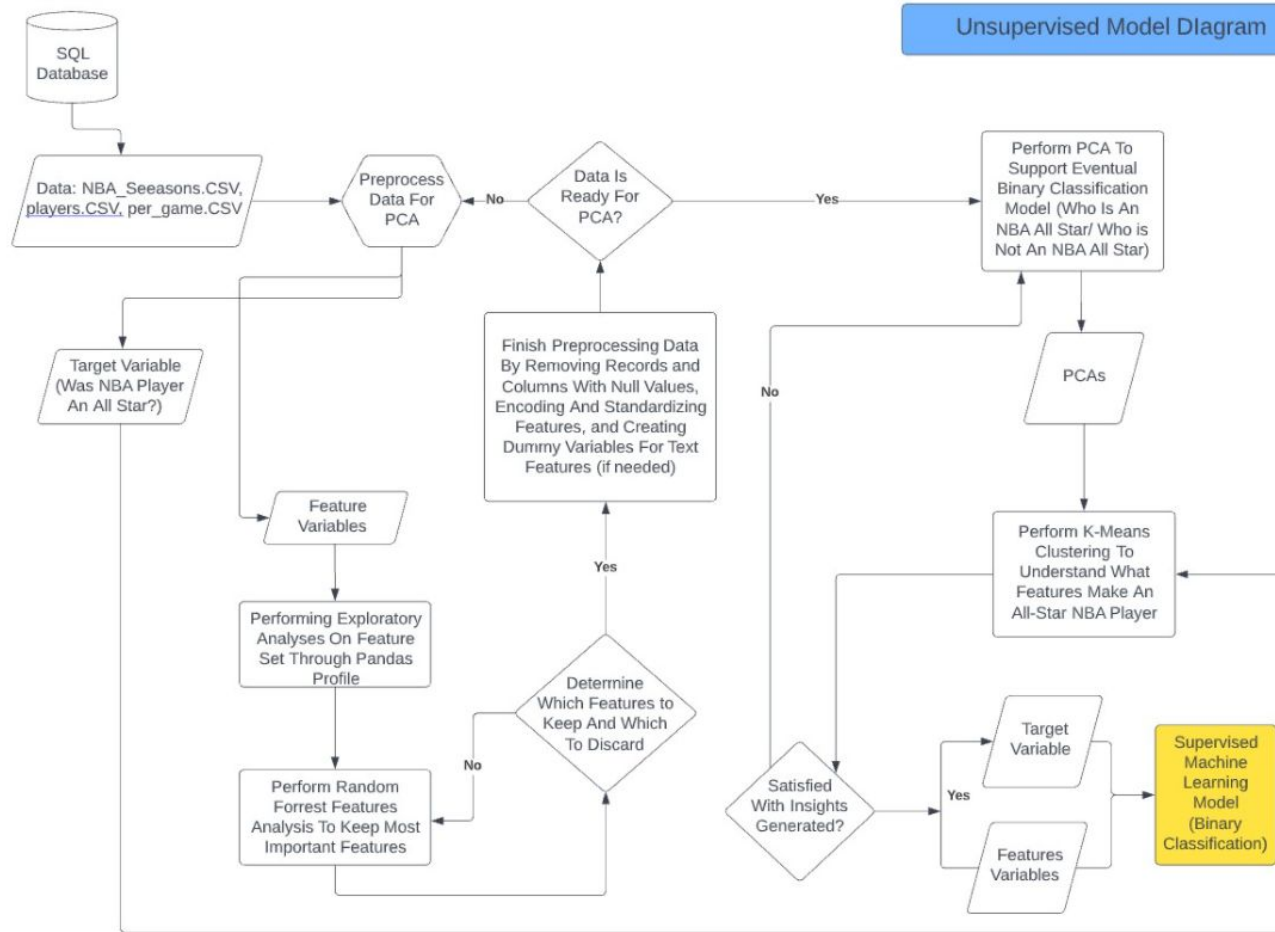
[ ] 1 per_game_df.where(per_game_df.Player == 'Michael Jordan').dropna()
```

	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

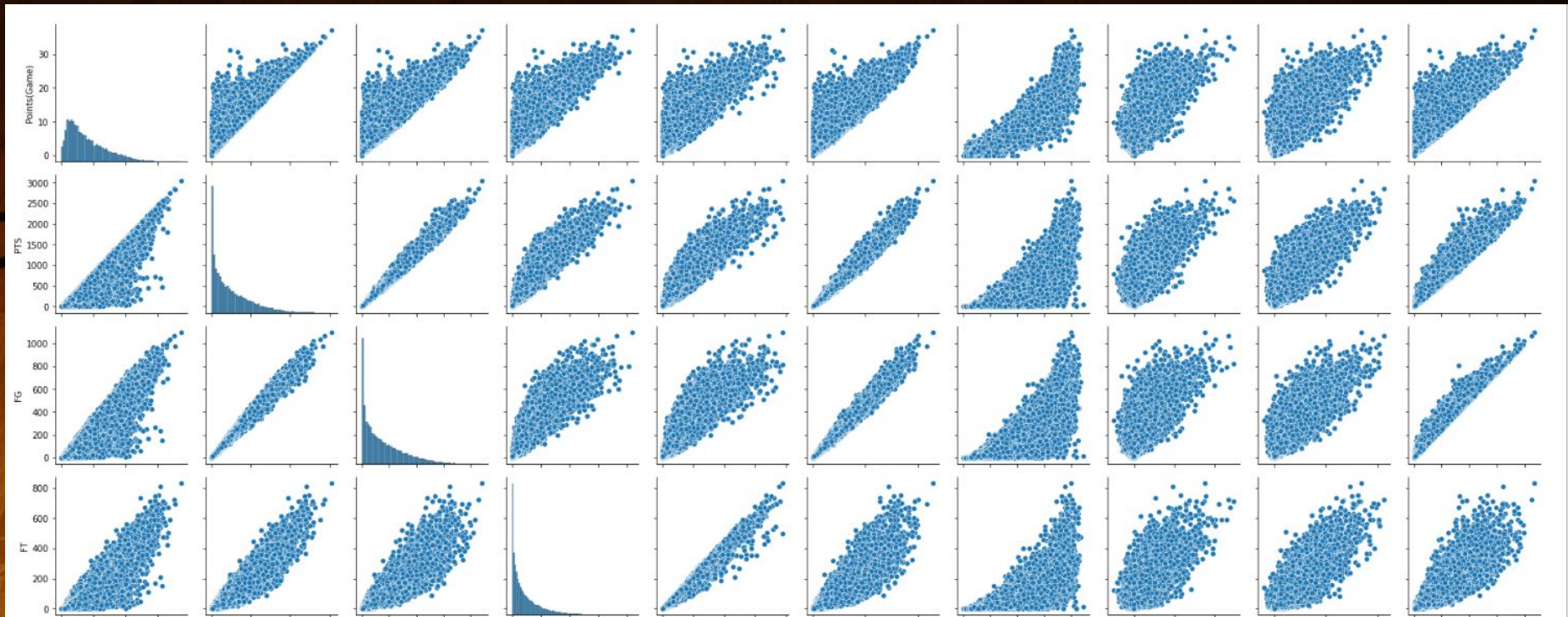
Machine Learning



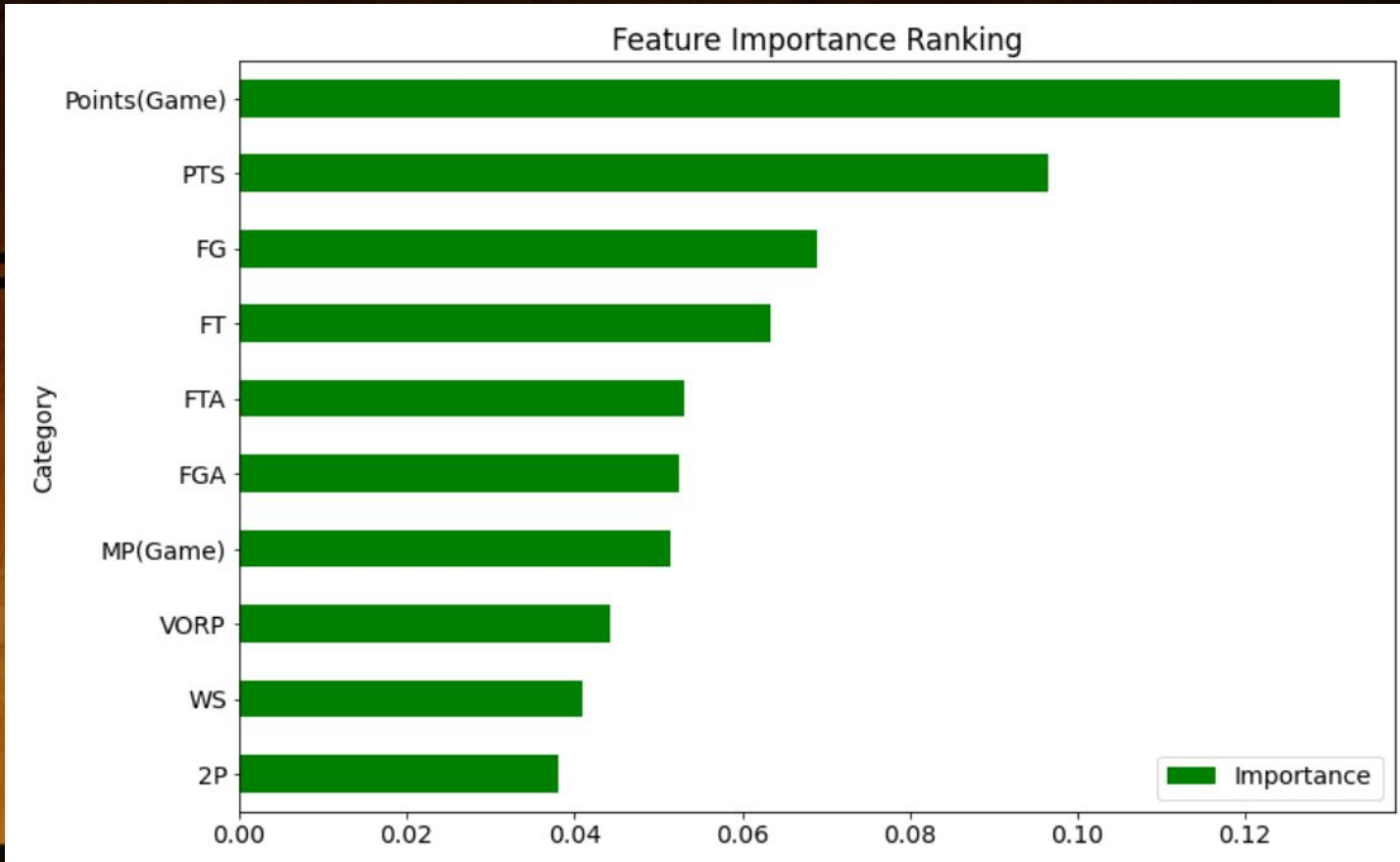
Unsupervised Model Diagram



Pre-Processing



PCA: Top Ten Features Explanatory Power



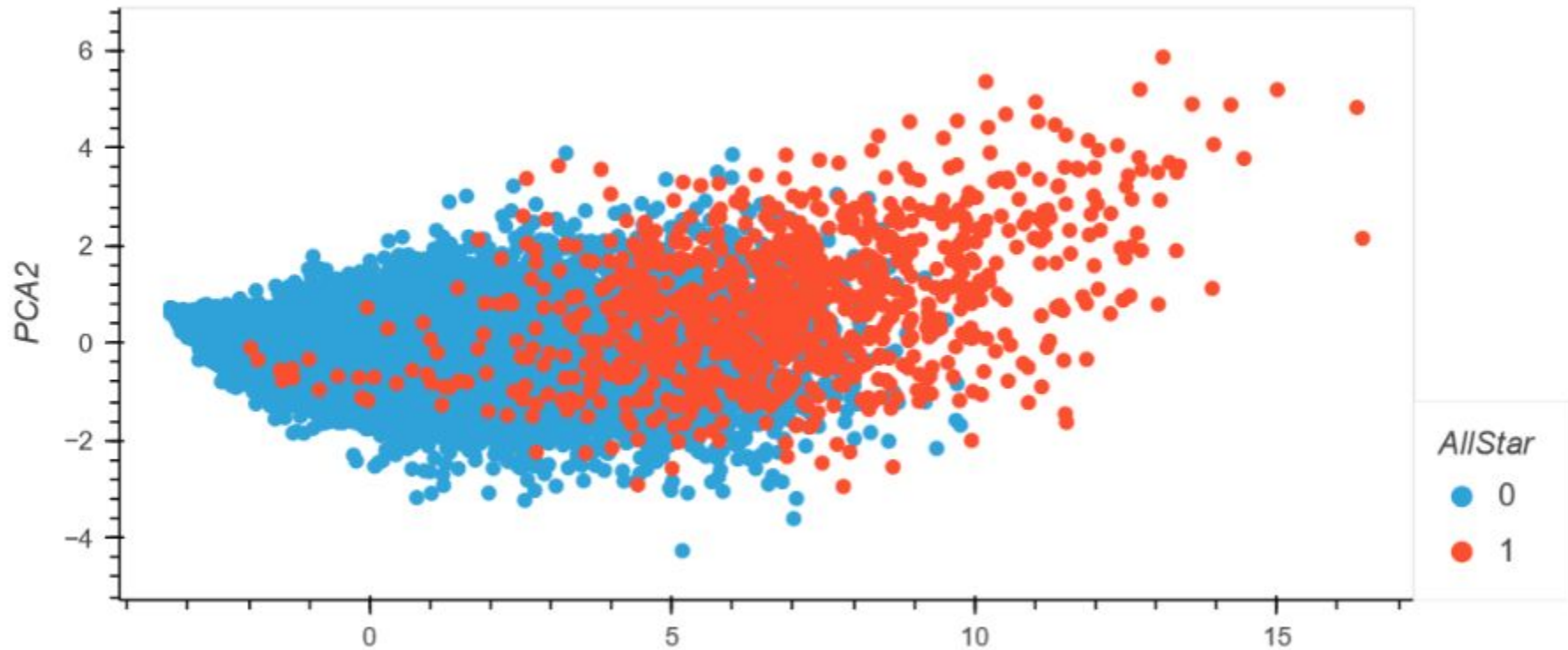
PCAs And Features

	Points(Game)	PTS	FG	FT	FTA	FGA	MP(Game)	VORP	WS	2P
PCA1	0.320261	0.332911	0.328974	0.319601	0.318029	0.324943	0.294795	0.289979	0.310322	0.319707
PCA2	-0.235294	-0.104315	-0.174816	0.072865	0.076965	-0.244932	-0.290454	0.683014	0.489685	-0.203242

CVR and EVR

	Cumulative Variance Ratio	Explained Variance Ratio
PCA1	0.883441	0.883441
PCA2	0.928994	0.045553

K-Means



Logistic Regression!

jupyter Logistical_Regression Last Checkpoint: an hour ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run

```
In [20]: # Displaying results
print("Confusion Matrix")
display(cm_df)
print(f"Accuracy Score : {acc_score}")
print("Classification Report")
print(classification_report(y_test, predictions))
```

Confusion Matrix

	Predicted 0	Predicted 1
Actual 0	4140	356
Actual 1	11	225

Accuracy Score : 0.9224429416737109

Classification Report

	precision	recall	f1-score	support
0	1.00	0.92	0.96	4496
1	0.39	0.95	0.55	236
accuracy			0.92	4732
macro avg	0.69	0.94	0.75	4732
weighted avg	0.97	0.92	0.94	4732

In []:

Summary of Supervised Performances!

Phase II					
Metric	Logistical Regression	SVM	Decision Tree	Random Forest	Boosting (L=5)
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%
Phase III					
Metric	Logistical Regression	SVM	Decision Tree	Random Forest	Boosting (L=1)
Testing Accuracy	92.2%	91.9%	90.9%	91.1%	91.4%
Recall (for predicting non-AllStars)	92%	92%	91%	91%	91%
Recall (for predicting AllStars)	95%	95%	93%	95%	93%
F-1 Scores (for predicting AllStars)	55%	54%	50%	52%	52%

Dashboard Storyboards & Mockups



Storyboards & Dashboard

- Storyboard
 - Page 1 - See our data and filter it yourself
 - Page 2 - Brief overview of our process
 - Page 3 - Meet our Team

Dashboard - Page 1

Storyboard Layout/Page

Navigation Bar


Jumbotron NBA Final Project

Scope of Project

Questions we'd like to answer

Player Search Filter

Table for Player stat data



The mockup shows a dashboard with a header navigation bar containing 'Home', 'Player Overview', and 'About the Data'. The main content area features a large image of NBA players, a section titled 'Performance in the League: How do players measure up against each other over the decades?' with three bullet points, and a 'PLAYER SEARCH' section with a table of player statistics.

Performance in the League: How do players measure up against each other over the decades?

- Does performance in the NBA levels the same decade by decade?
- What does a prototypical player look like in each decade?
- Can you predict an All-Star?

LET'S FIND OUT

PLAYER SEARCH

Name	Season Year	Position	Height (in)	Weight (lbs)	Points	Reb
Michael Jordan	1996	SG	77	220	497	46
Kobe Bryant	2010	SG	76	215	250	12

Dashboard - Page 1 Continued

Storyboard Layout/Page



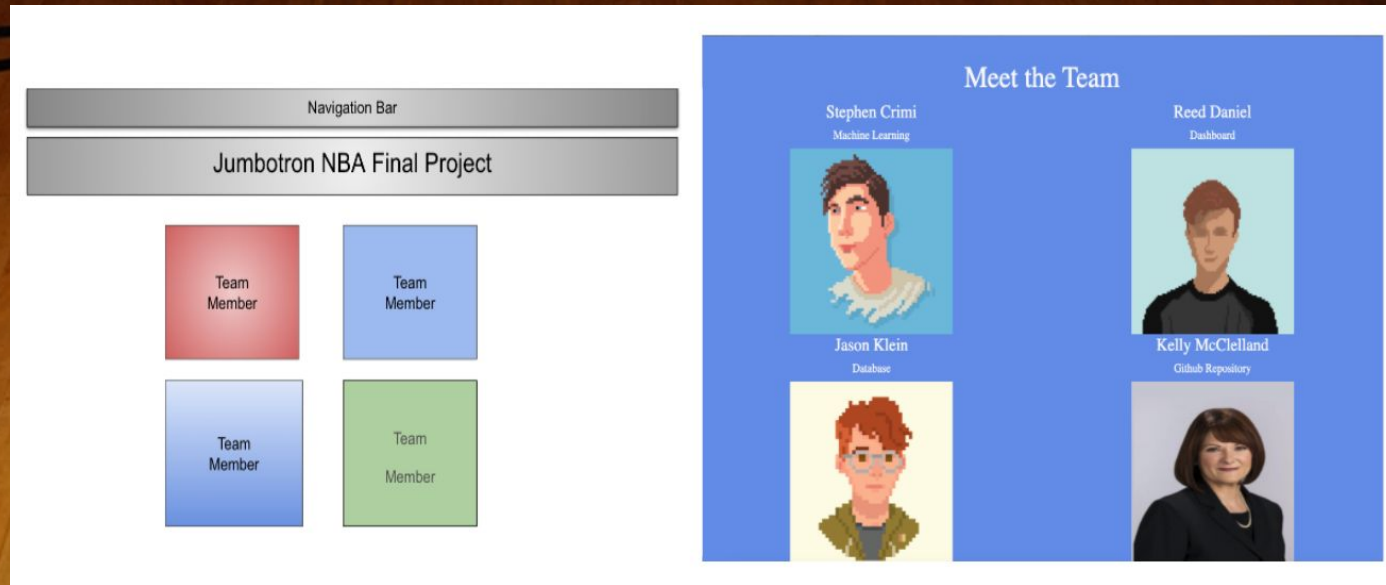
Dashboard - Page 2

Storyboard Layout/Page



Dashboard - Page 3

Storyboard Layout/Page



Summary of Results Found to Questions



Q - Does the NBA look the same decade by decade in terms of performance?

A -

Q - What does a prototypical player look like in each decade?

A -

Q - How has that changed over time?

A -

Q - Can you predict whether an NBA player can be an All-Star?

A -

Q - What are the most important stats in terms of determining an NBA All-Star?

A -

Lessons Learned

- Have redundancies between team members in case of illness or emergencies.
- Clearly articulate the work in process, challenges, and path forward.
- Identify all dependencies upfront and identify each other strengths and weaknesses.
- Validating and Collaborating information and sources need to complete the project.
- Share lesson learned with others.