

```
In [1]: import pandas as pd
import matplotlib as plt
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2
from sklearn.model_selection import train_test_split
!pip install duckdb
import duckdb
```

Requirement already satisfied: duckdb in /Users/ameliagrevin/Downloads/CS577_Assignment1/new_venv/lib/python3.11/site-packages (1.4.2)

```
In [23]: %load_ext sql
```

```
In [24]: %sql sqlite:///data/basic_examples.db --alias sqlite_ex
```

Connecting to 'sqlite_ex'

```
In [25]: %reload_ext sql
```

```
In [26]: %sql duckdb:///data/nba_stats.db --alias statDB
```

Connecting and switching to connection 'statDB'

```
In [ ]: %%sql statDB
CREATE OR REPLACE TABLE merged USING AS
SELECT *
FROM totals
LEFT JOIN per_game USING(player_id, season)
LEFT JOIN shooting USING(player_id, season)
LEFT JOIN season_info USING(player_id, season)
LEFT JOIN award_shares USING(player_id, season)
LEFT JOIN career USING(player_id);
```

```
Out[ ]: Count
```

```
In [ ]: %%sql statDB
SELECT player, player_id, mp, g, gs, fg, fga, fg_percent,
       x3p, x3pa, x3p_percent, x2p, x2pa, x2p_percent,
       e_fg_percent, ft, fta, ft_percent
FROM merged
WHERE season >= 1947 AND season <= 2025;
```

Out[]:

player	player_id	mp	g	gs	fg	fga	fg_percent	x3p	xs
Bam Adebayo	adebaba01	2416.0	71	71.0	530	1017	0.521	15.0	4
Grayson Allen	allengr01	2513.0	75	74.0	340	682	0.499	205.0	44
Jarrett Allen	allenja01	2442.0	77	77.0	519	819	0.634	0.0	
Jose Alvarado	alvarjo01	1028.0	56	0.0	142	345	0.412	78.0	20
Giannis Antetokounmpo	antetgi01	2567.0	73	73.0	837	1369	0.611	34.0	12
Deni Avdija	avdijde01	2257.0	75	75.0	406	803	0.506	88.0	23
Bogdan Bogdanović	bogdab001	2401.0	79	33.0	471	1100	0.428	240.0	62
Bojan Bogdanović	bogdab002	1478.0	57	27.0	309	681	0.454	133.0	33
Bojan Bogdanović	bogdab002	921.0	28	27.0	199	425	0.468	86.0	20
Bojan Bogdanović	bogdab002	557.0	29	0.0	110	256	0.43	47.0	12

Truncated to [displaylimit](#) of 10.