Does Draft Position Predict NBA Player Success

Introduction & Question

In the NBA, teams select new players to add to their roster every year through the league's draft. The earlier a player is picked in the draft, the higher the expectations of future performance. Making the right picks in the NBA Draft is imperative to having a successful team. Because draft position reflects how teams value prospects, it is natural that one might ask: "Does draft position accurately predict an NBA player's success throughout their career?"

To evaluate these players' performance in a uniform and comparable way, we use the box plus-minus (BPM), an advanced statistic that measures a player's contribution relative to an average player. BPM measures the points a player has contributed to their team, per 100 possessions they spent on the court, above or below the league average.

Data Description & Sources

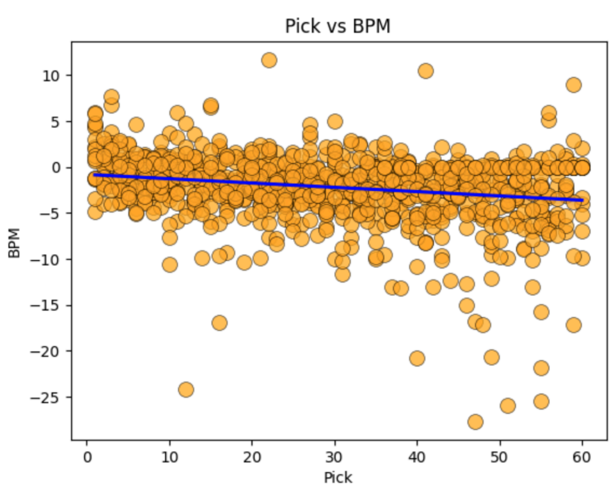
The data utilized for this project were sourced from "Sports Reference," which provides users with sports statistics and analytics for various professional sports leagues, from both current and past seasons.

The data collected for our research question consisted of two primary variables. The first being the draft position. This is a numerical discrete variable because, inherently, NBA Draft picks are integers ranging from 1 to 60. The second variable that we collected data on in seeking to answer our question is the box score plus-minus. This is numerical continuous data because the box score plus minus can take on any numerical value. This is because this statistic is calculated by measuring the points per 100 possessions a player contributed above a league-average player, translated to an average team (Basketball Reference, 2025).

Methodology

To test whether draft position predicts player success, we estimated a simple General Linear Model (GLM) using linear regression, where draft pick number served as the predictor variable and Box Plus-Minus (BPM) served as the outcome variable. In this model, the intercept represents the expected BPM for a hypothetical first overall pick, while the slope coefficient (B1) captures how much BPM changes for each increase in draft position. A negative slope indicates that later draft picks tend to have lower BPM values. The GLM assumes a linear relationship between variables, normally distributed residuals, and equal variance across draft positions; violations of these assumptions—such as nonlinear patterns or unmeasured confounders like team context, injuries, or minutes played—represent limitations of the model. Despite these limitations, the GLM provides a clear and effective framework for evaluating whether draft position meaningfully predicts long-term NBA performance.

Results & Analysis



The analysis from the tests that were conducted reveals that there is a negative correlation between draft pick (1-60) and box score plus minus (BPM); this means that as players get drafted in later picks their BPM becomes worse. This is shown through the regression line plotted on the graph because the slope of the line (B1) is -0.0468. This means that for every one-unit increase in the draft position, their BPM decreases by 0.0468 points.

This data is confirmed to be significant because the p-values for both the intercept and slope are nearly zero. The p-value for the slope of 0.000 allows us to reject the null hypothesis that there is no coordination between BPM and draft pick position in the NBA Draft. The same goes for the intercept of the data (B0) whose p-value is 0.001. From this P-value we could conclude that if this data didn't have any correlation (Null Hypothesis), it would be nearly impossible to see this data. This provides evidence that the negative relationship between draft pick and BPM is not random.

Conclusions

In conclusion, there is a strong negative correlation between NBA draft pick position and player success as measured by Box Plus-Minus (BPM). This means that the later a player is drafted, the lower their BPM tends to be over their career. Using data collected from Sports Reference, we found a clear negative slope in our regression and p-values near 0.000, indicating that this relationship is statistically significant. These results suggest that NBA teams generally make effective evaluations on draft day, as earlier draft picks tend to become more successful players than those selected later.

References

Basketball Reference (2025). *NBA draft picks (2010–2025).* Sports Reference LLC. https://www.basketball-reference.com/draft/

Google (2025). *Gemini (2025) [Google Colab Assistant].* https://gemini.google.com/.

Appendix

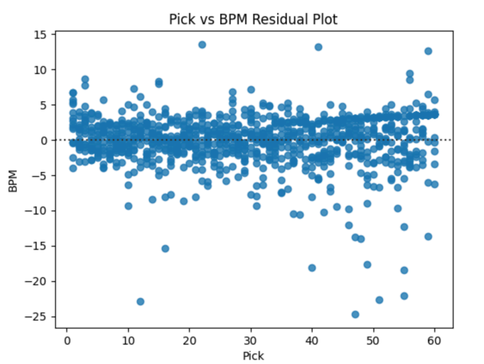
Model

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    BPM   R-squared:                       0.046
Model:                            OLS   Adj. R-squared:                  0.045
Method:                 Least Squares   F-statistic:                     46.43
Date:                Tue, 02 Dec 2025   Prob (F-statistic):           1.68e-11
Time:                        19:45:02   Log-Likelihood:                -2612.9
No. Observations:                 960   AIC:                             5230.
Df Residuals:                     958   BIC:                             5240.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.8051      0.241     -3.344      0.001      -1.278      -0.333
Pick          -0.0468      0.007     -6.814      0.000      -0.060      -0.033
==============================================================================
Omnibus:                      412.109   Durbin-Watson:                   1.904
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3433.230
Skew:                          -1.752   Prob(JB):                         0.00
Kurtosis:                      11.576   Cond. No.                         71.1
==============================================================================
```

Note.
B0 is equal to -0.8051. This means that the average rookie BPM is -0.8051
B1 is equal to -0.0468 - Thai means that for every +1 increase in draft position, a player's plus minus decreases by -0.0468

- The difference between the 1st pick and the 60th pick is almost 3 points

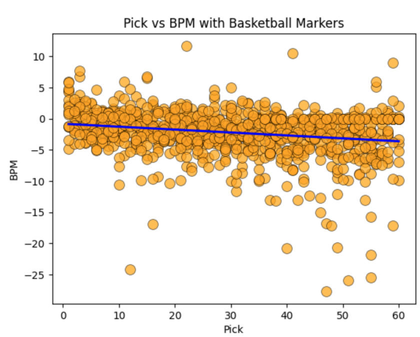
Residual



Note.
Looks homoscedastic because the residuals are evenly spread across picks 1-60

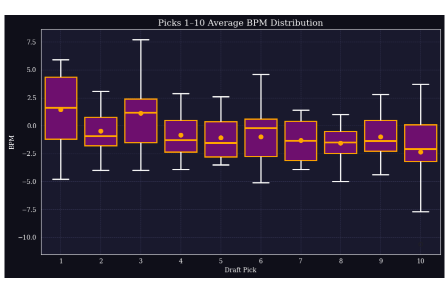
- We do see some heteroscedasticity in the later picks, but this results from later picks not playing as much, so their BPM stats are more extreme

Regression



Note.
Shows a negative correlation overall
p<0.001
Pick 0.00
There's a p <0.001 percent chance that we would see this data if there was no correlation between draft pick position and BPM

Box & Whisker Correlation



Note. Shows an overall negative relationship between draft pick position and Box Plus-Minus (BPM). The relationship is statistically significant (*p* < .001). As draft pick number increases, average BPM decreases, indicating that later draft selections tend to be less successful. The very small *p*-value suggests that the probability of observing this pattern due to random chance alone is extremely low.

Coding

Residual Plot & Pick vs BPM regression

```
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns

# BPM Regression
sns.regplot(x='Pick', y='BPM', data=data)
plt.title('Correlation between Pick Number and Average BPM')
plt.xlabel('Pick Number')
plt.ylabel('Average BPM')
plt.ylim(-40, 20) # Extended y-axis range
plt.show()

# Residual Plot
X = data['Pick']
X = sm.add_constant(X)   # adds intercept
y = data['BPM']
model = sm.OLS(y, X).fit()
fitted_vals = model.fittedvalues
residuals = model.resid
plt.figure(figsize=(8, 5))
sns.scatterplot(x=fitted_vals, y=residuals)
plt.axhline(0)
```

Box and Whisker Plot

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


plt.rcParams['font.family'] = 'serif'
```

```
picks_1_10 = list(range(1, 11))
sub = data[data['Pick'].isin(picks_1_10)].copy()

plt.figure(figsize=(10, 6), facecolor="#0f0f1a")
ax = plt.gca()
ax.set_facecolor("#1a1a2e")

#box and whisker plot set up/color it

sns.boxplot(
   data=sub,
   x='Pick',
   y='BPM',
   order=picks_1_10,
   color='purple',
   showmeans=True,
   meanprops={
       "marker": "o",
       "markerfacecolor": "orange",
       "markeredgecolor": "orange",
       "markersize": 7
   },
   medianprops={
       "linewidth": 3,
       "color": "orange"
   },
   whiskerprops={
       "linewidth": 2,
       "color": "white"
   },
   capprops={
       "linewidth": 2,
       "color": "white"
   },
   boxprops={
       "edgecolor": "orange",
       "linewidth": 2
   }
)
```

```
plt.xlabel('Draft Pick', color="white")
plt.ylabel('BPM', color="white")
plt.title("Picks 1-10 Average BPM Distribution", color="white",
fontsize=14)


ax.tick_params(colors="white")

#background change
ax.grid(color="#3a3a5a", linestyle="--", linewidth=0.6)

plt.tight_layout()
plt.show()
```