HOW PLAYER HEIGHT AFFECTS THE NBA

ANALYSIS TEAM:

LINDSAY REYNOLDS, NICK SHEETS

MOTIVATIONS AND HYPOTHESIS

- We are both basketball fans and interested in basketball statistics
- We wanted to explore if height has any relation to salary and stats in the form of points, rebounds, and assists
- Alternative Hypothesis:
  - If an NBA player is taller, then the points averages, rebounds averages, and salary are greater than players who are shorter
  - If an NBA player is taller, then the player's assist averages are less than players who are shorter
- Null Hypothesis:
  - If a player's height is greater, then player's salary, points, rebounds, and assists stats are not affected

QUESTIONS ASKED

- Does height impact performance (ppg, rebounds or assists per game)?
- Do taller players score more often?
- Do taller players grab more rebounds?
- Do taller players have more or fewer assists per game than shorter players?
- Does height impact salary?
- Do taller or shorter players make more money?

DATA SOURCES USED

- Websites used: Kaggle and Basketball-Reference
- Data used:
  - Kaggle: NBA Players, Justinas Cirtautas

    https://www.kaggle.com/justinas/nba-players-data
    - Find heights of players
    - Find yearly averages by player for points, rebounds and assists
    - Limitation - there is no data for blocks
  - Sports Reference LLC: Basketball-Reference.com "2020-21 NBA Player Contracts" (4/22/2021)

    https://www.basketball-reference.com/contracts/players.html
    - Used to gather data on players' salaries
  - Center for Disease Control, 2021

    https://www.cdc.gov/nchs/fastats/body-measurements.htm
    - Used to pull average male height in U.S.

DATA EXPLORATION

- Issues anticipated for the CSV files chosen for analysis
  - Only a partial season of stats for 2019-20 in Stats CSV
  - No data on blocks per game
  - Heights listed in centimeters
  - Player names are like Stephen Curry\curryst01 in Salary CSV
  - Large values for salaries

DATA CLEANUP

- CSV files read in as Pandas DataFrames
- Updates and additions
  - Listed player names consistently, filtered out unnecessary columns and renamed/reordered columns as needed
  - Narrowed scope of stats analysis to **3 recent seasons:**

    2016-17, 2017-18, and 2018-19
  - Added columns for height in inches, salary in millions

DATA CLEANUP

- Created a data frame that included player name and personal info such as height
- Averaged player stats across the three seasons in a new data frame
- Merged these together to one data frame, with one row per player, and a total of 748 different players

```
3 player_groupby = organized_stats_df.groupby(["player_name"])
4 player_groupby.mean()
```

player_name	height (in)	pts	reb	ast
AJ Hammons	84.00	2.20	1.60	0.20
Aaron Brooks	72.00	3.65	0.80	1.25
Aaron Gordon	81.00	15.43	6.80	2.63
Aaron Harrison	78.00	3.45	1.65	0.90
Aaron Holiday	73.00	5.90	1.30	1.70
...	...	...	...	...
Zach Lofton	76.00	0.00	0.00	0.00
Zach Randolph	81.00	14.30	7.45	1.95
Zaza Pachulia	83.00	5.13	4.83	1.60
Zhaire Smith	76.00	6.70	2.20	1.70

DATA CLEANUP

- Next merged the data frame with the salary data frame
  - Left merge for same set of unique players (748 expected)
- Number of rows increased to 766 following merge
- Suspected some players had duplicate rows in Salary CSV
- Example

	A	B	C	D	E
1	Rk	Player	Tm	2020-21	2021-22
324	323	Alex Len\lenal01	TOR	4032648	
383	382	Alex Len\lenal01	WAS	4032648	
510					

DATA CLEANUP

- Dropped rows for player duplicates in salary data frame (contains 748 rows again, one per player)
  - List contains 748 players once again
- Dropped players in salary data frame without 2020-21 salaries (no longer salaried players in NBA)
  - Only 324 different players across the three seasons have salary data for analysis
  - Note large change in short amount of time, as players cycle out and new ones replace them
- Cleanup complete, and three files were saved for use in the analysis notebook
    - player_stats_breakout_by_season.csv
    - player_stats.csv
    - player_stats_with_salaries.csv

DATA ANALYSIS

What is the average height of players in the NBA?

- Mean = 78.87 inches (around 6’ 6″)
- Median = 79 inches (6’ 7″)
- Mode = 81 inches (6’ 9″)
- Across the three seasons analyzed there were 81 players in all that were 81 inches tall (6’ 9″)
- Average male height in US = 5’ 9″

  (CDC, 2021)

DATA ANALYSIS

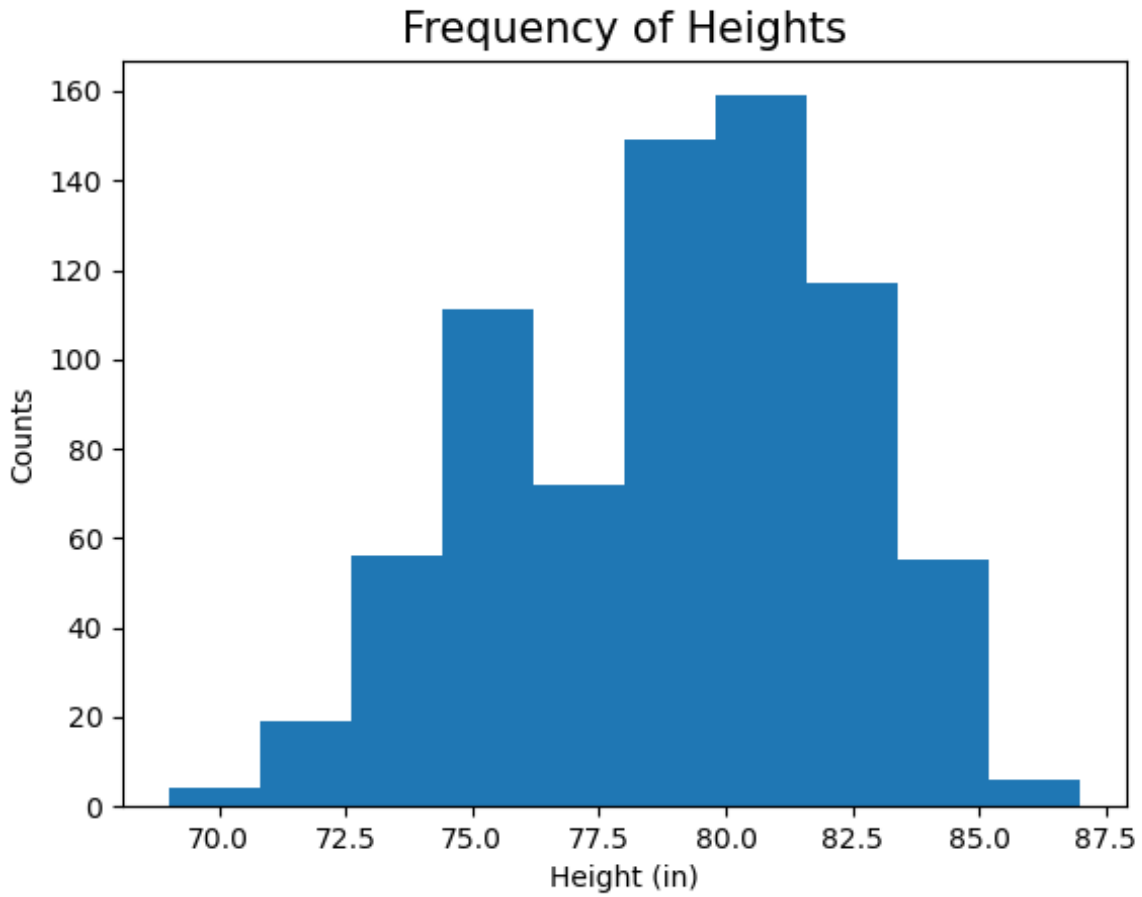
- About 40% of players in our dataset are making between 1M and 5M in 2020-21

SALARY BINNING

Salary Range	player_count	perc_of_players
< 1M	20	6.17%
1M to 2.4M	63	19.44%
2.5M to 4.9M	66	20.37%
5M to 7.4M	39	12.04%
7.5M to 9.9M	28	8.64%
10M to 14.9M	39	12.04%
15M to 19.9M	22	6.79%
> 20M	27	8.33%

DATA VISUALIZATIONS

HEIGHT HISTOGRAM



NormaltestResult(statistic=1.2860512176624757, pvalue=0.5256994544947969)

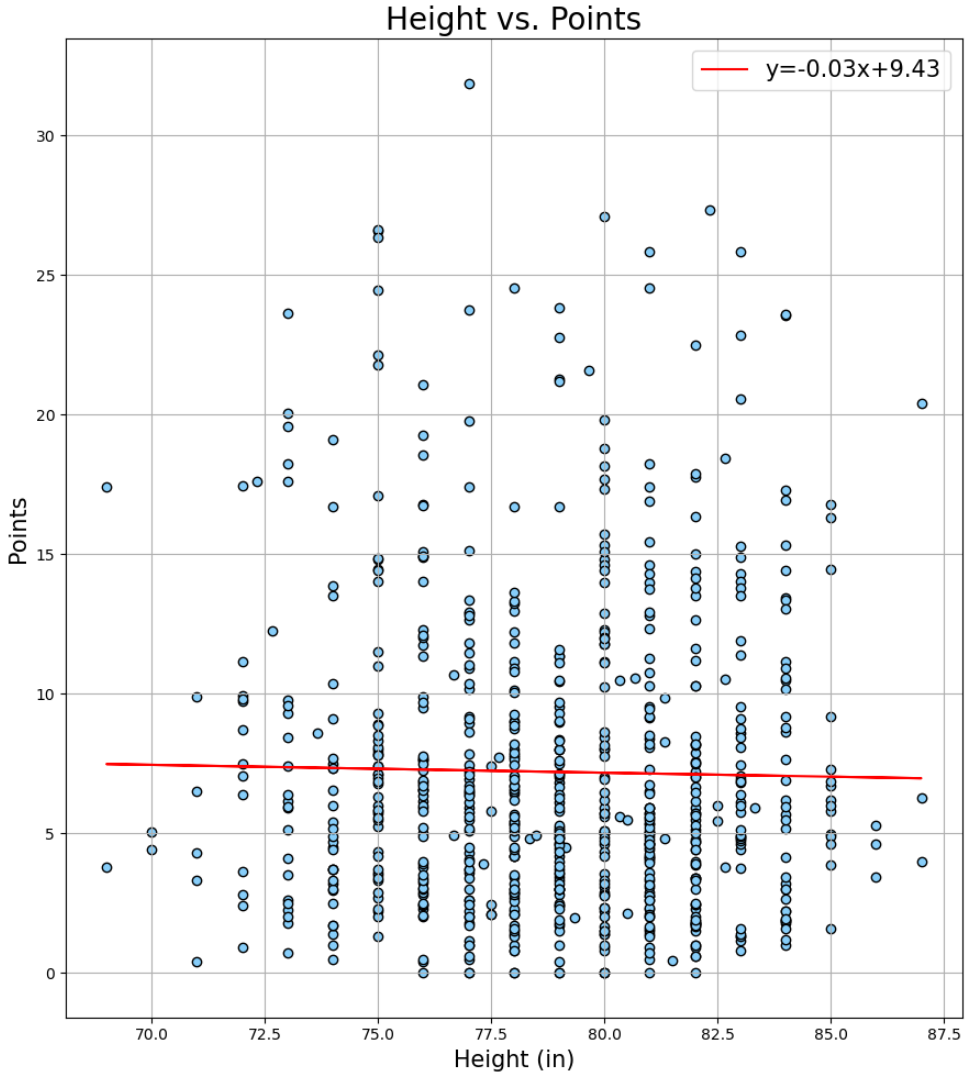
HEIGHT BINNING

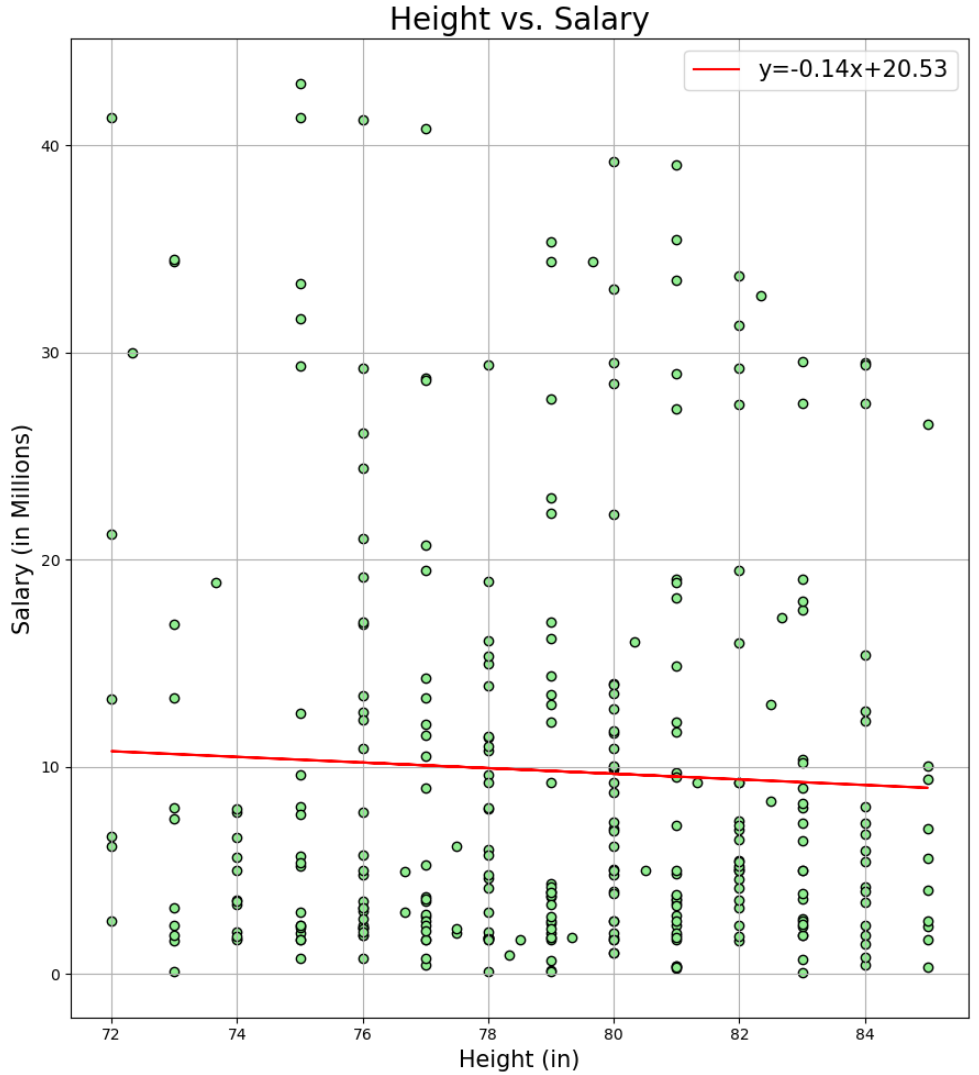
Height Range	player_count	perc_of_players
<= 6ft	22	2.94%
6'1" - 6'3"	109	14.57%
6'4" - 6'6"	202	27.01%
6'7" - 6'9"	233	31.15%
6'10" - 7'	161	21.52%
> 7ft	21	2.81%

- Nearly 1/3 of the players are 6’7″ to 6’9″
- Around 58% are 6’4″ to 6’9″

DATA VISUALIZATIONS

HEIGHT vs.
POINTS
&
HEIGHT vs.
SALARY



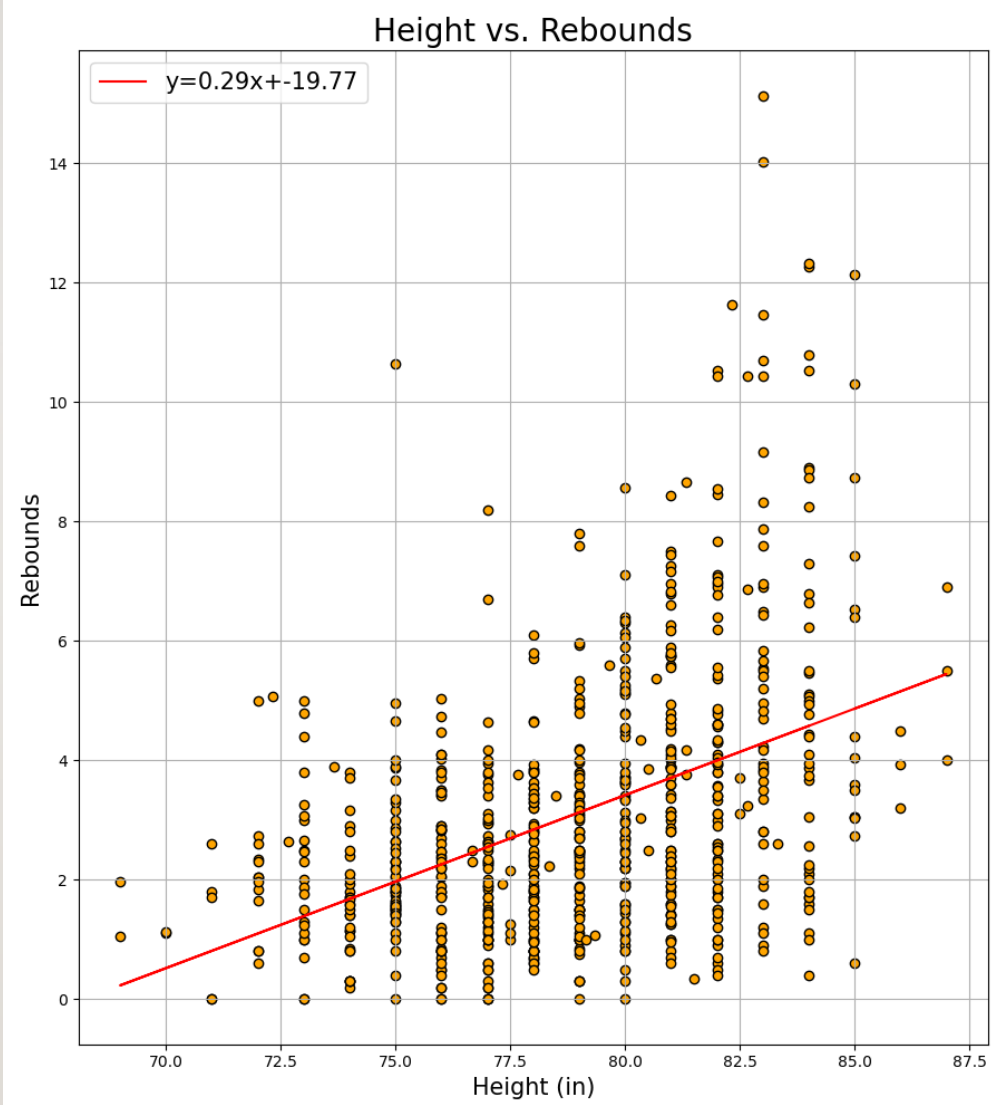


pearsonr results: (-0.017377411665314996, 0.6351390828365829) pearsonr results: (-0.044071115087400095, 0.4291784196774443)

DATA VISUALIZATIONS

HEIGHT vs. REBOUNDS

For players of all heights

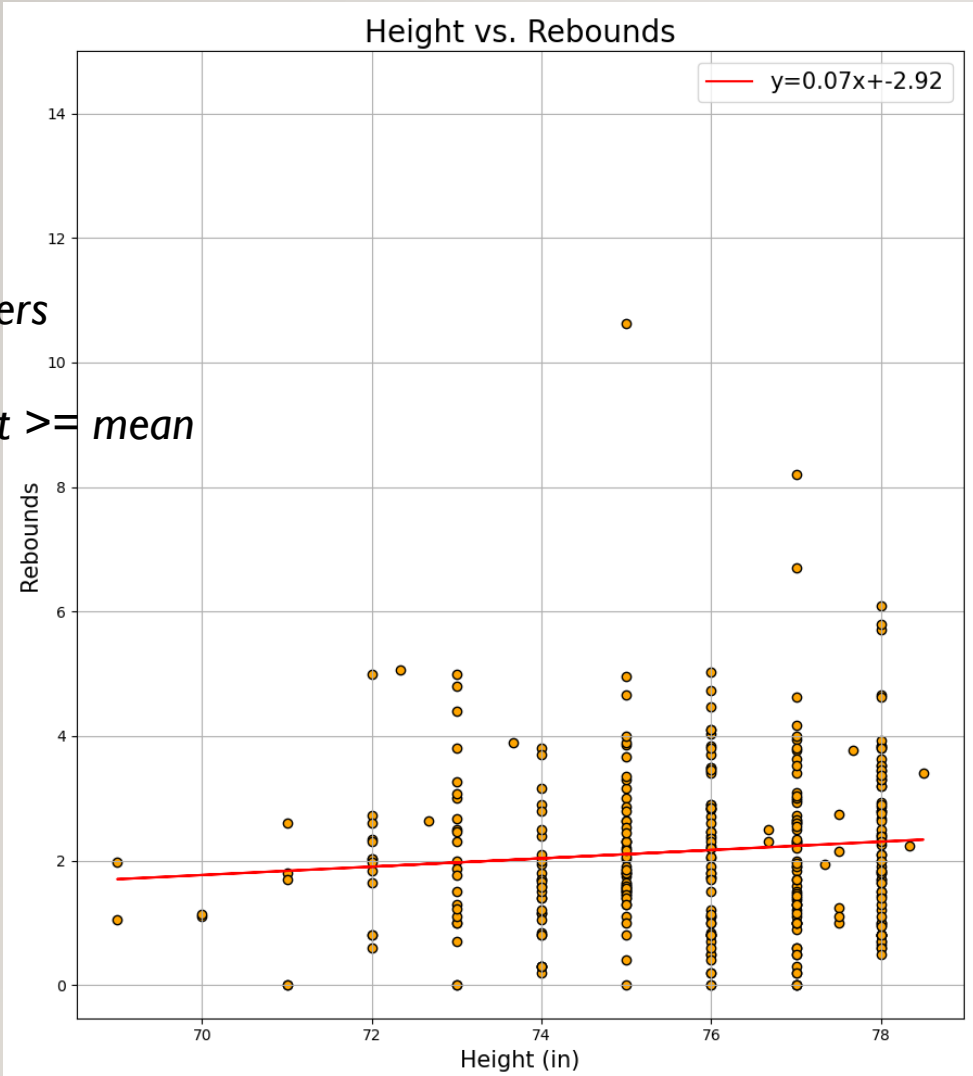


pearsonr results: (0.43453524023725276, 8.423531901409829e-36)

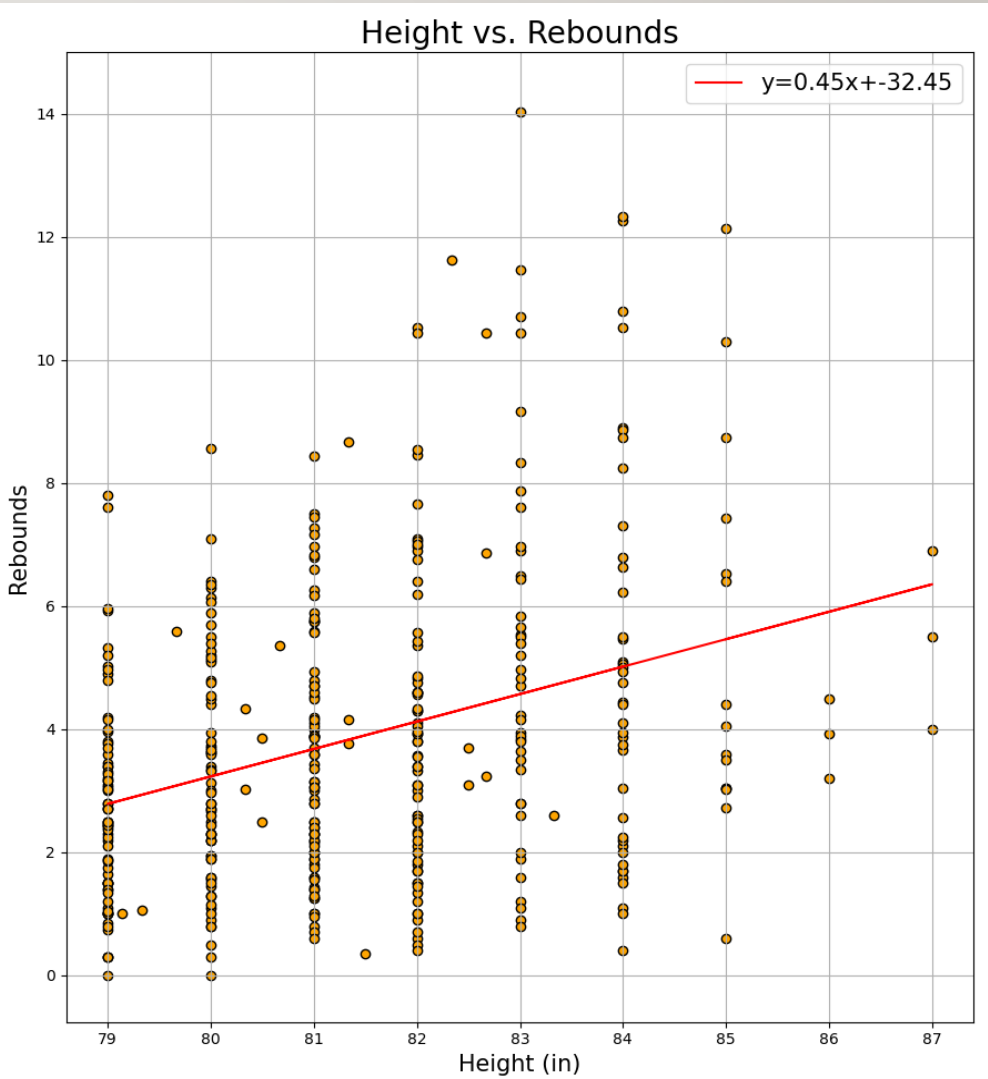
DATA VISUALIZATIONS

HEIGHT vs. REBOUNDS

Split between players of height < mean & players of height >= mean



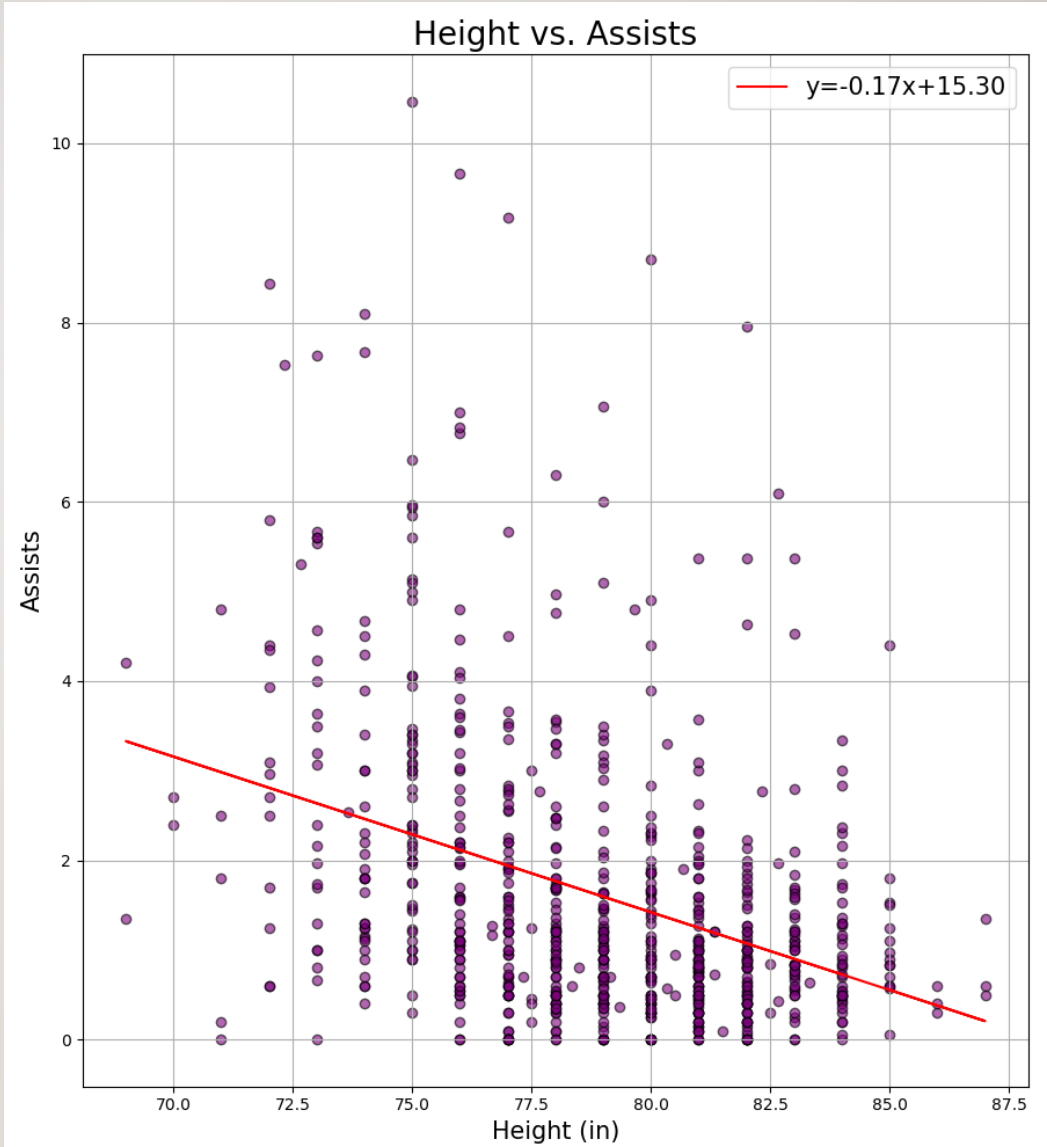
pearsonr results: (0.09624547228754901, 0.07856345203318187)



pearsonr results: (0.3216356492219352, 2.1519075913158547e-11)

DATA VISUALIZATIONS

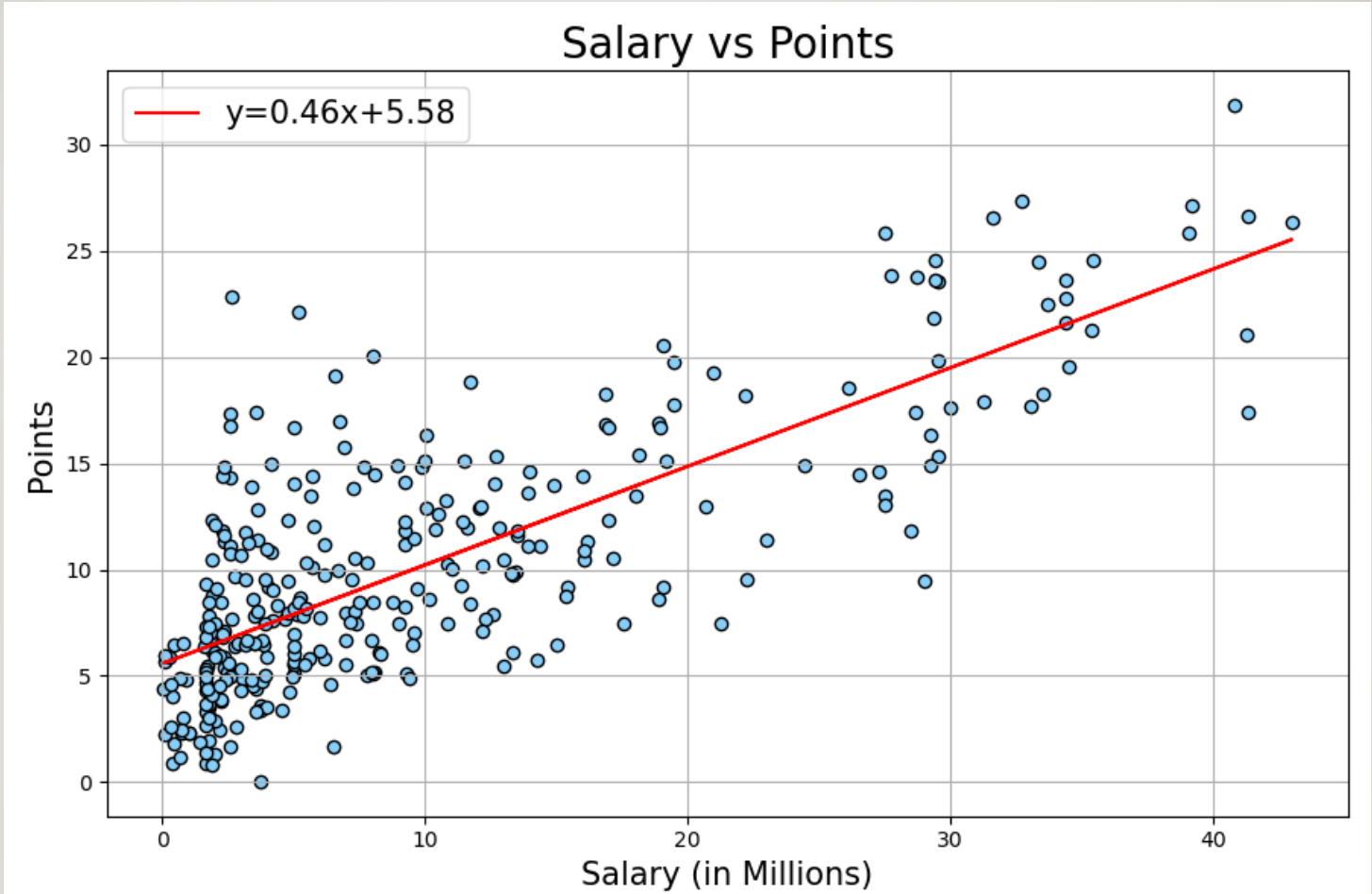
HEIGHT vs. ASSISTS



pearsonr results: (-0.3698558763535108, 1.1604091735536877e-25)

DATA VISUALIZATIONS

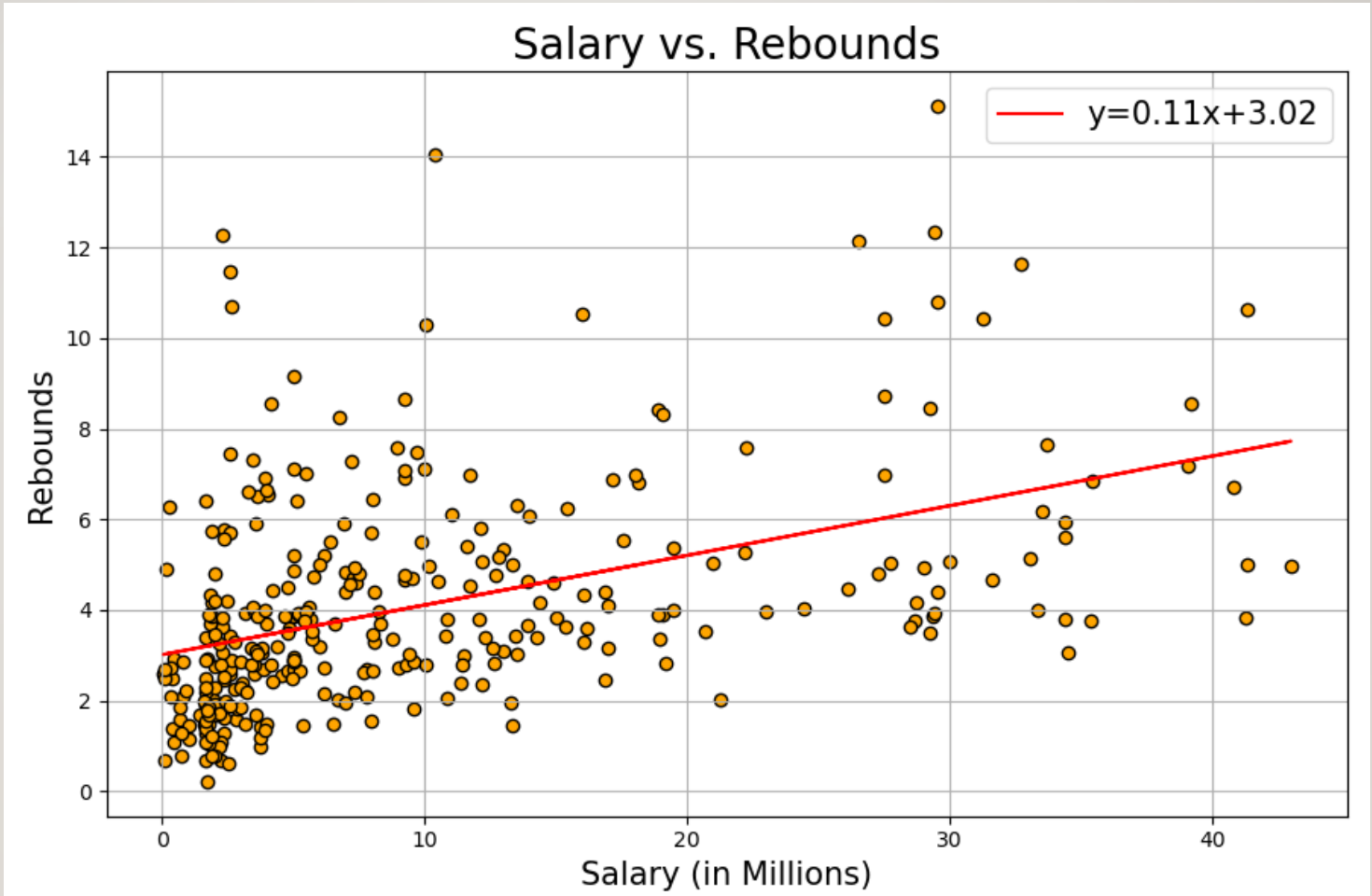
SALARY vs. POINTS



pearsonr results: (0.7696157006414956, 1.0565978011907831e-64)

DATA VISUALIZATIONS

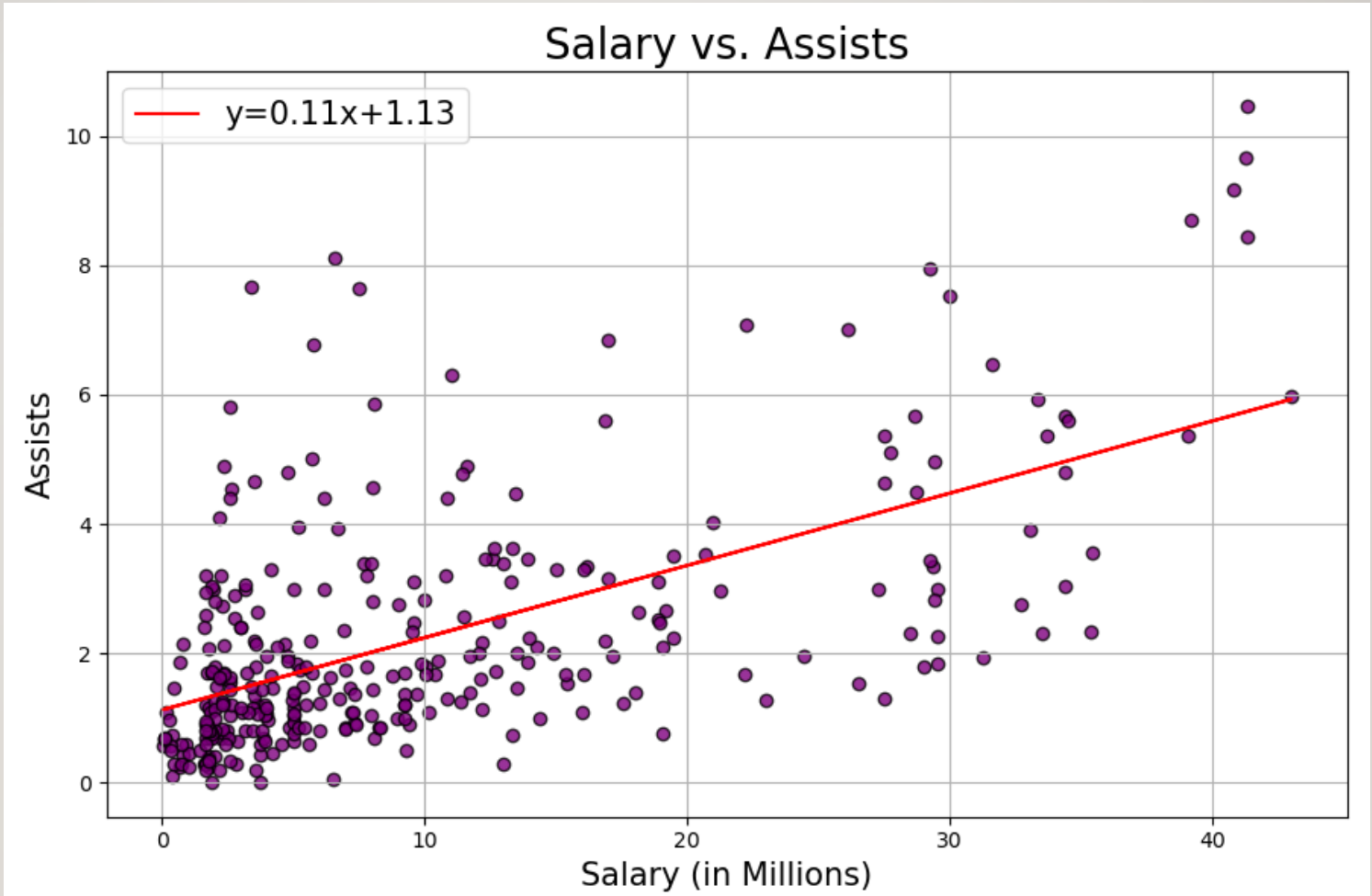
SALARY vs. REBOUNDS



pearsonr results: (0.45025171413569315, 1.405623025197144e-17)

DATA VISUALIZATIONS

SALARY vs. ASSISTS



pearsonr results: (0.601504128402172, 2.9064123213974663e-33)

SUMMARY OF FINDINGS

- Height does not affect points or salary, but can affect the amount of rebounds
- The data does not suggest taller players score more points
- The data shows taller players grab more rebounds, this to be true with a strong positive correlation
- Taller players have less assists than smaller players, showing a negative correlation
- The data shows height does not affect players' salaries
- There is no correlation with height and salary, the linear regression line is fairly flat

FINAL THOUGHTS

- Additional data that could have been analyzed
  - Blocks per game
  - Average heights for each team
    - Does height have any impact on wins and losses?
  - What is the average height of women in the WNBA as compared to average height of women in the U.S?
  - Could have looked at the pip install available to pull stats for the NBA

Q & A (PLUS AUDIENCE PARTICIPATION, IF TIME ALLOWS)

- If anyone would like to test our abilities to pull data from our Pandas DataFrames
  - Choose Player, College or Country
  - Choose All or Average – Do you want all the rows in our data for your players or only the average of them?
  - Up to 50 rows of results are displayed