BrainStation

Prepared by Anastassia Troitskaia

August 25 2024

Information available in audio.

NBA Salary Prediction & Analysis

Deep diving into our data with clear direction

Content

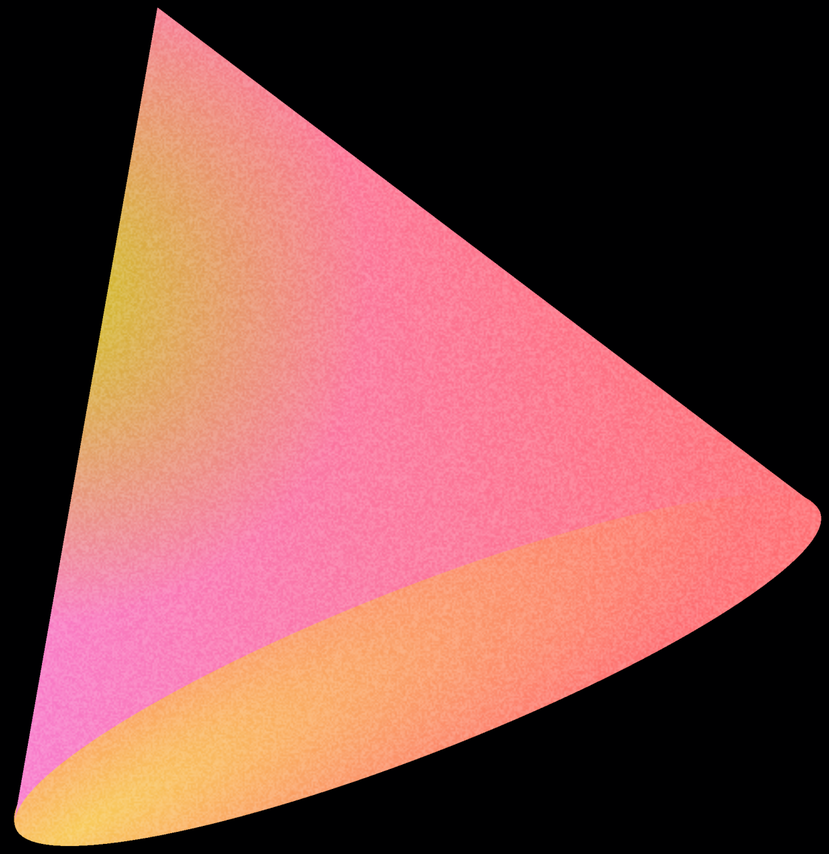
Introduction

Purpose:

To predict NBA salaries for college basketball players based on their performance statistics and historical data, aiding NBA teams in draft and salary decisions.

Description:

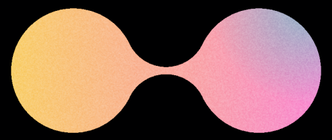
This project tackles the challenge of translating college performance into professional success, using machine learning and comprehensive data from current and former NBA players to improve player evaluation and financial planning.



Back to Content Page

Data Resources & Collection

- College Basketball Player's Performance Data
- Historical Performance Data
- NBA Draft Combine Data
- Salary Data



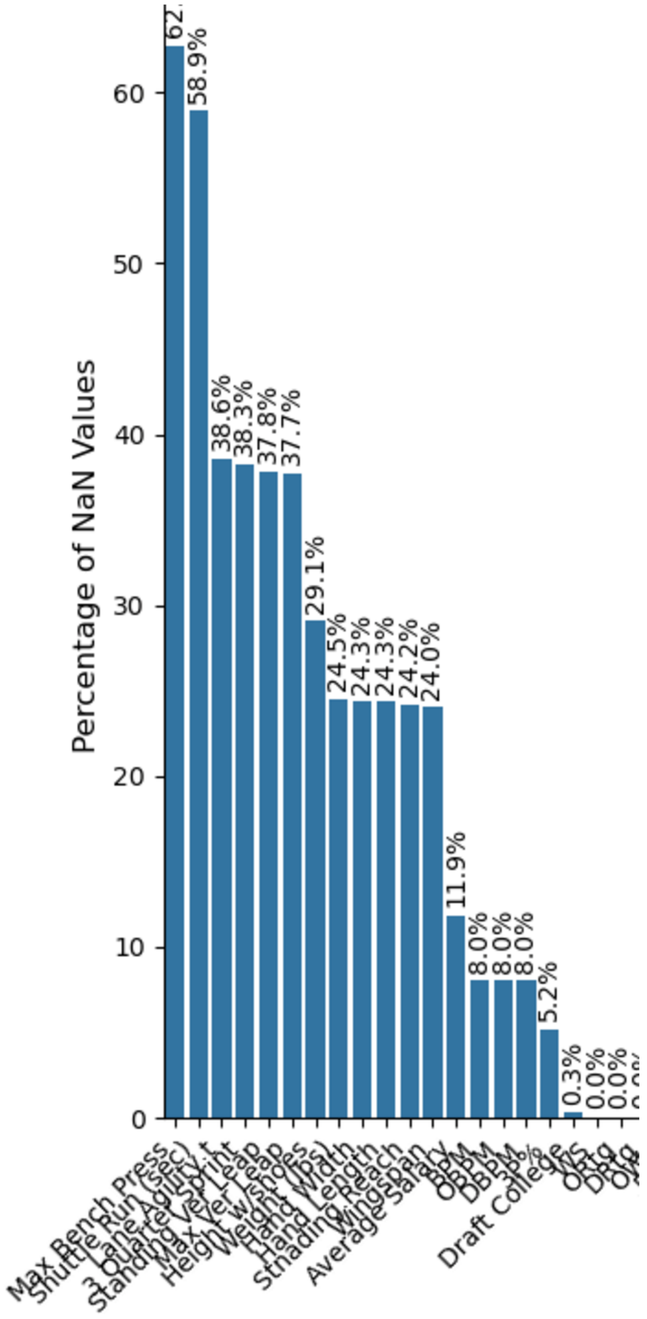
Back to Content Page

Process:

- Data time frame 2009-2024
- College Basketball Performance Data was obtained from https://www.sports-reference.com/
- NBA Draft Combined was collected using NBA Api from https://www.nba.com/stats/draft/combine https://www.nba.com/stats/draft/combine
- Salary Data was obtained from HoopsHype using Selenium web scraper https://hoopshype.com/salaries/

Issues During Collection & Cleaning:

- Six different csv files to combine
- All data is in different format. Data types had to be changed
- NBA Draft Combine data missing a lot of values.
- Salary Data has missing values due to the issue with web scraper

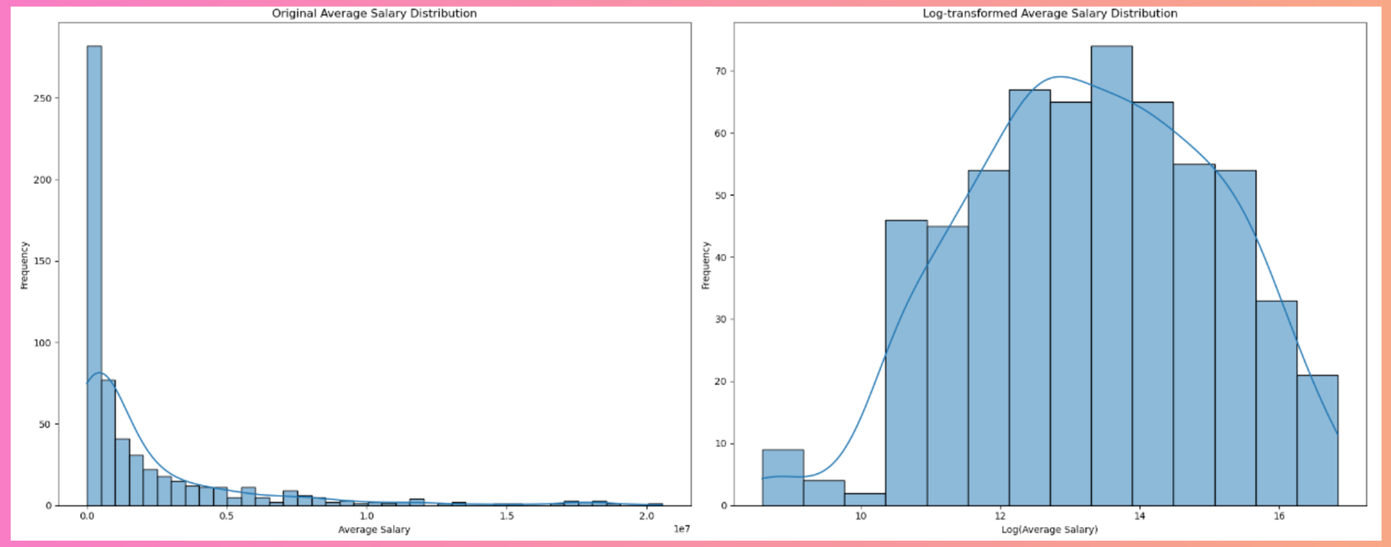


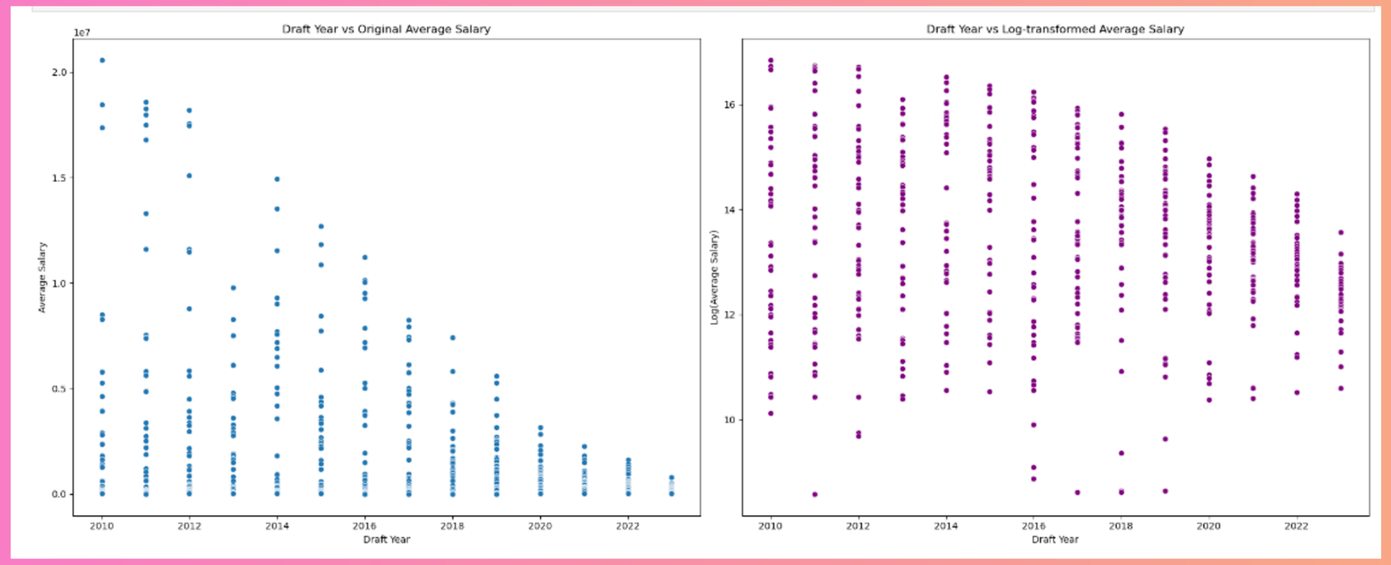
Draft College	2
3P%	35
OBPM	54
DBPM	54
BPM	54
Hand Length	164
Hand Width	164
Height w/shoes	196
Stnading Reach	163
Weight (lbs)	165
Wingspan	162
Lane Agility t	260
3 Quarter Sprint	258
Standing Ver Leap	255
Max Ver Leap	254
Average Salary	80

Missing Values: Drop or Impute

- Very high percentage of values for Max Bench Press and Shuttle Run
- Draft College fixed by adding data
- NBA Draft Combine
- Average Salary

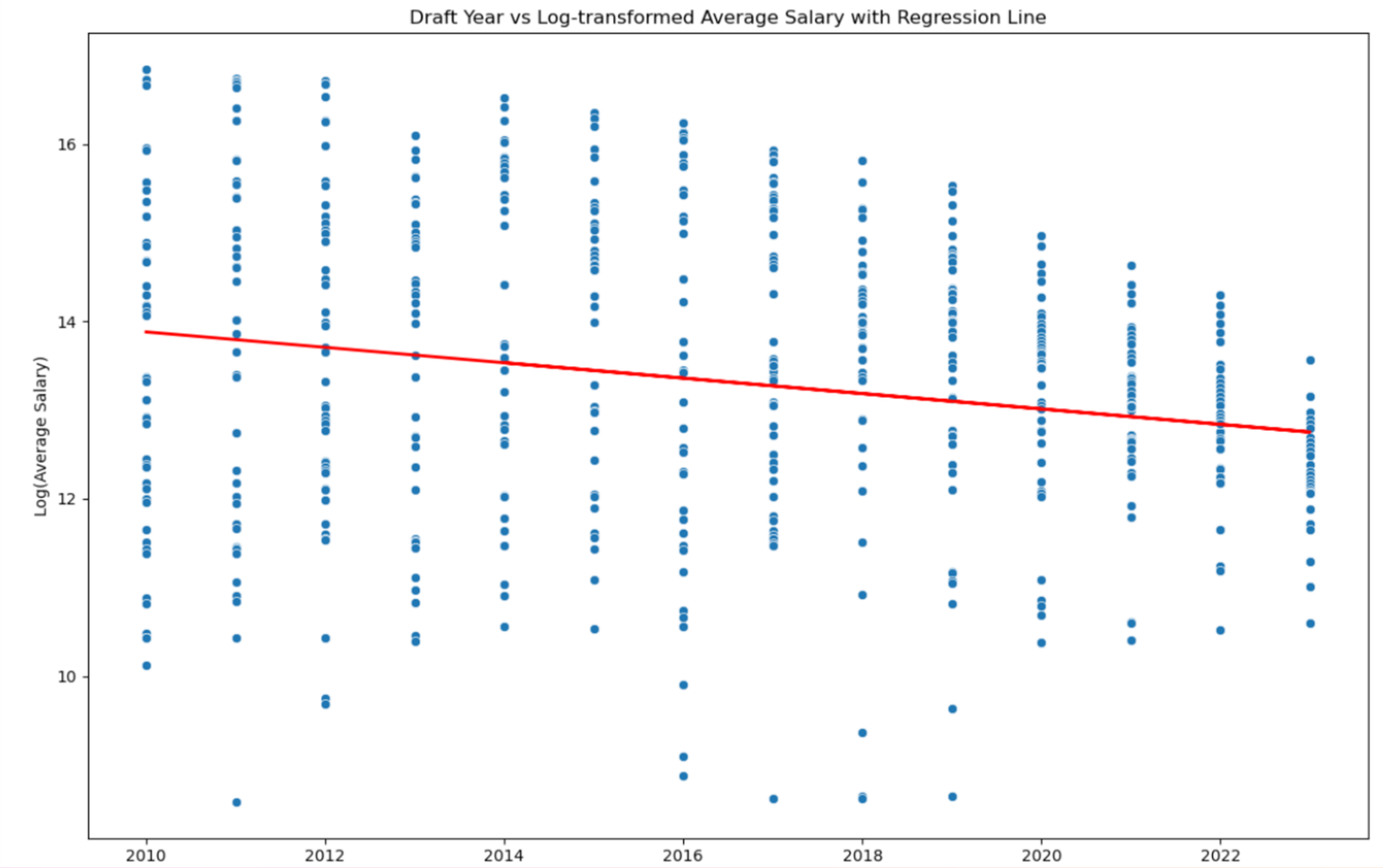
Average Salary: Imputation





- Before Log Transformation
  - Skewed to the right
  - Skewness 2.82
  - High salary outlier
- After Log Transformation
  - Data more normally distributed
  - Skeweness -0.17
  - The difference between salaries is more proportional

Average Salary: Linear Regression Draft Year and Salary



- **Observations and Conclusions:**
  - **Pearson correlation coefficient is -0.2106**
  - **P-value is small 0.0000 < 0.0001**
  - **The weak correlation suggests that we can not just use Draft Year alone to predict players salaries.**
  - **Other factors should be also looked at**

Next Steps:

- Finish cleaning my dataset. Impute other missing values. Hot-encode categorical variables.
- Feature Selection/Engineering:
  - Identify key college performance metrics that might influence NBA salary (e.g., points per game, rebounds, assists, shooting percentages).
- Create new features if needed (e.g., composite scores, efficiency ratings).
- Model selection: Multiple Regression, Ridge Regression, Random Forest, Neural Networks.

Back to Content Page

Source: Add your references here.

Questions?

Back to Content Page