**NBA Top 50 Players Salary Analysis**

This project analyzes the salaries and performance metrics of the top 50 highest-paid NBA players for the 2023-24 season. It uses Python for data manipulation and visualization, aiming to uncover insights into the relationships between player salaries, performance metrics, age, experience, and team contributions.

**Project Overview**

The dataset contains the following attributes for analysis:

* **Player Information:** Name, age, draft year, and experience.
* **Salary Data:** Annual salary in millions of dollars.
* **Performance Metrics:** Points, assists, and rebounds per game.
* **Team Information:** Team name and player positions.

**Objectives**

1. Analyze top 10 salaried player and relative matrices.
2. Identify top-performing players relative to their salaries.
3. Analyze the correlation between salaries and key metrics.
4. Analyze salaries with performance metrices like assist and rebounds.
5. Analyzing salary distribution by team

**Installation**

Install packages like sqlite, pandas, matplotlib, seaborn, folium.

**Code Description**

**Statistical Analysis**

. We can see out of top 10 players

* Majority are of age of >30
* Has experience >10
* Has point >20
* Majority has forward position
* Majority has salary more than 40 Millions
* Player with highest salary and point has 6.4 assist and rebounds 11.7
* Damian Lillard has highest assit with salary 45640084.0

**Correlation Analysis**

The heatmap identifies relationships between salary metrics:

heatmap\_data = top\_salary\_player.pivot\_table(

    index='age',

    columns='experience',

    values='salary',

    aggfunc='mean'

)

A graph of different colored squares

Description automatically generated

This heatmap visualizes how closely related these variables are.

* Salaries increase with age and experience but plateau after a certain point.
* Players in their early 30s with 10–15 years of experience dominate the high salary range.
* Lower salaries (10–15 million) are observed among younger age groups (e.g., **24–26 years**) with fewer years of experience (**5–7 years**).

**Salary vs. Performance Bubble Chart**

This Bubble chart shows how player salaries compare to their performance

# Plot bubble chart

plt.scatter(

    x=top\_salary\_player['assists'],    # X-axis: Assists

    y=top\_salary\_player['rebounds'],   # Y-axis: Rebounds

    s=top\_salary\_player['salary'],  # Size of bubbles: Salary (scaled for better visualization)

    c=top\_salary\_player['salary'],     # Color by salary

    cmap='coolwarm',                   # Color map

    alpha=0.9,                         # Transparency of bubbles

    edgecolors="w",                    # White border around bubbles

)

A diagram of a graph

Description automatically generated

* The largest and darkest bubbles (highest salaries, exceeding 40 million) occur in the range of **6–8 assists** and **9–12 rebounds**. These players demonstrate a strong overall performance, contributing significantly in both metrics.
* Smaller, lighter bubbles (salaries around 10–15 million) are clustered at lower levels of assists (1–4) and rebounds (2–6), indicating that minimal contributions in these areas correlate with lower earnings.
* Players excelling in both assists and rebounds command the highest salaries.
* Some bubbles represent outliers, where a high salary exists despite moderate assists or rebounds, possibly due to unique skillsets or reputation.

**Team Distribution by Location in map and Salalry Analysis**

The map visualizes the location of teams those pays highest to their players

# Assuming 'team\_coordinates' is a DataFrame with team names and their geographic coordinates (latitude, longitude)

# Example:

team\_coordinates = pd.DataFrame({

    'team\_name': ['Milwaukee Bucks', 'Los Angeles Clippers', 'Portland Trail Blazers', 'Golden State Warriors', 'Utah Jazz', 'Los Angeles Lakers'],

    'latitude': [43.0731, 34.0522, 45.5122, 37.7749, 40.7608, 34.0522],

    'longitude': [-87.9161, -118.2437, -122.6587, -122.4194, -111.8910, -118.2437]

})

# Get the highest-paid player salary per team (already available in your top\_salary\_player DataFrame)

highest\_paid\_player = top\_salary\_player.groupby('team\_name').agg({'salary': 'max'}).reset\_index()

# Merge the team coordinates and the highest salary data

team\_map\_data = pd.merge(team\_coordinates, highest\_paid\_player, on='team\_name', how='left')

Check Map on this link below

[file:///C:/Users/FID%20bebshai%20forhad/Downloads/nba\_team\_salary\_map%20(1).html](C://Users/FID%20bebshai%20forhad/Downloads/nba_team_salary_map%20(1).html)  
  
A map of the united states

Description automatically generated

* Teams located in major metropolitan areas, such as Los Angeles (e.g., Lakers and Clippers), have higher salaries compared to some other regions.
* This aligns with the trend that large-market teams often have the financial flexibility to pay top-tier players.
* Coastal cities like San Francisco (Golden State Warriors) and Portland (Trail Blazers) show relatively high salaries.
* Midwestern teams like the Milwaukee Bucks also feature prominently, suggesting strong financial investments in players in that region.

**How to Use**

1. Load the database available in <https://www.kaggle.com/code/agilesifaka/historic-nba-drafting-game-and-player-analysis?select=basketball.sqlite> containing tables with information of player salary and performance data and using SQL query get the tables necessary. This project merges 3 tables player, player\_attributes and player\_salary to get the dataset to finally work on.
2. Execute the Python script to generate visualizations.
3. Interpret the insights from the heatmap, scatter plot, and bar chart.

**Results**

The analysis provides insights into:

* Teams located in major metropolitan areas, such as Los Angeles.
* Players with more experience get paid more.
* Players with more assists and rebounds get paid more

**Future Enhancements**

* Integrate historical data for trend analysis.
* Add advanced player metrics like Player Efficiency Rating (PER).
* Explore additional visualizations for draft year and age trends.