**Executive summary**

**Dataset Overview:**

This analysis is based on a dataset obtained from Kaggle, which includes statistics on NBA games spanning from 2003 to 2022. We selected this dataset due to our passion for basketball and the NBA. The dataset allowed us to explore statistical trends of different NBA teams and players over the years.

**Database Schema:**

The database schema was designed to derive meaningful analytics and correspond to Cassandra’s strengths in fast writes and reads.. It comprises of:

* **player\_stats**: Includes averages of Points, Assists and Rebounds of players over the years. The partition key is the player name, and cluster column is the season.
* **seasonal\_performance:** Includes averages of Points, Assists and Rebounds of players over the years. The partition key is the season and cluster column is the team\_id.
* **game\_outcome\_performance:** Includes statistical data of games. The partition key is both season and outcome, and cluster column is team\_id

**Key Insights and Queries:**

Our analysis yielded several insights:

* Player Performance Trends: We inspected Deni Avdija, which is our only Israeli in the NBA. The stats revealed an overall upward trajectory in average points, rebounds, and assists per season, showcasing that Deni is getting used to the NBA and starting to gain confidence in his play.
* Shooting Efficiency: Correlation analysis highlighted a strong relationship between field goal percentages and points scored, as well as strong relationship between rebounds and team wins.
* Team Performance: The 2022 season analysis demonstrated that certain teams, like the Boston Celtics, outperformed others in scoring, which could be correlated with their successful run in the season.
* NBA Performance: We saw there is a steady increase of points scored per game each year, showing that players are getting better and the game style is constantly changing.

Supporting queries focused on calculating average seasonal performance metrics, identifying correlation coefficients between different statistics, and ranking teams based on their average points per game.

**Challenges and Resolutions:**

We encountered challenges related to data preprocessing, particularly in handling missing values and data type misalignment. This was resolved with rigorous checks of the data before inserting to the database.

Another challenge was optimizing queries for performance, which we addressed by refining our database schema and leveraging indexed columns.





