NBA Analytics Project

"The big edge analytics gives you is just removing bad habits. Teams make bad decisions because they're based on tradition or belief, but with data, you can objectively measure what's working and what's not" - Daryl Morey

Background

Being involved with basketball for my entire lifetime, whether it be watching the NBA or playing myself, I always tend to be attracted to the data that drives this sport. I am always looking at stats, graphics, and insights that tell you a story–the details that you wouldn't just be able to see by watching games. The numbers behind the game reveal patterns and trends that can give teams a competitive edge. I find that data brings the game to life in a whole new way.

Step 1: Ask

- What are the most important statistics to winning a basketball game?
    - What quarter is the most valuable?
    - Which stat is more crucial for a win: fewer turnovers, more threes, or more rebounds?

Step 2: Prepare

- **Data:** NBA_2020_Games
- **What is in it?**: This dataset contains the team statistics of every game played from the 2019-2020 NBA season. The interesting thing with this data is that it has two rows for every game, one row for each team that played. It includes all the team stats that you would usually see on a box score.
- **Is it reliable?**:
    - (R)eliability: The dataset appears very relevant, as it focuses on NBA games from 2020. It contains essential basketball-related metrics like points per quarter, turnovers, rebounds, and three-pointers, which are useful for analyzing game outcomes
    - (O)riginality: The dataset was collected directly from ESPN's website after each game, making it highly original. Since ESPN is a leading sports media organization and a reliable source for up-to-date and accurate sports data, the originality of this dataset is strong
    - (C)omprehensiveness: The dataset includes a broad range of games from the 2020 season, which seems comprehensive. However, there is no detailed information about whether any data points are missing, such as overtime stats or context around the games played (like home-court advantage or playoff conditions).
    - (C)urrent: The dataset was collected from the 2019-2020 NBA season.

- (C)ited: Given that ESPN is one of the most authoritative sources for sports data and statistics, the dataset's credibility is high. The information they provide is widely used in both professional and amateur analysis, further enhancing the dataset's trustworthiness.

- **Dataset Obstacles:** The NBA has two teams from Los Angeles. The Los Angeles Clippers were listed as 'LA' and the Los Angeles Lakers were listed as 'Los Angeles.' I knew the difference because the Lakers made the NBA Finals this season meaning they played more games than the Clippers. I changed them to be LAL and LAC for the Lakers and Clippers respectively.

Step 3: Process

- **Data Cleaning**: I imported this .csv file into SQL and started by updating the dataset to reflect the new names for the Lakers and Clippers.

```sql
UPDATE `nba-analytics-project.NBA.NBA_2020_Games`
SET Team = CASE
    WHEN Team = 'LA' THEN 'LAC'
    WHEN Team = 'Los Angeles' THEN 'LAL'
  END
WHERE Team IN ('LA', 'Los Angeles')
```

I followed this by writing a query to tell the relationship between out-scoring your opponent in the first quarter and the percentage of the times that leads to a win. I repeated this step for all 4 quarters.

```sql
SELECT
  Team,
  COUNT(CASE WHEN Result = 'W' AND Q1_Points_For > Q1_Points_Against THEN 1 END) AS Wins_When_Outscoring_Q1,
  COUNT(CASE WHEN Q1_Points_For > Q1_Points_Against THEN 1 END) AS Games_Leading_Q1,
  ROUND(COUNT(CASE WHEN Result = 'W' AND Q1_Points_For > Q1_Points_Against THEN 1 END) / COUNT(CASE WHEN Q1_Points_For > Q1_Points_Against THEN 1 END) * 100, 2) AS Win_Percentage_When_Leading_Q1
FROM `nba-analytics-project.NBA.NBA_2020_Games`
GROUP BY Team
```

Next, I used a similar process for the three-point data, rebounding data & turnover data.

```sql
SELECT
  Team,
  COUNT(CASE WHEN Result = 'W' AND Threes_Made_For > Threes_Made_Against THEN 1 END) AS Wins_When_More_Threes,
  COUNT(CASE WHEN Threes_Made_For > Threes_Made_Against THEN 1 END) AS Games_More_Threes,
  ROUND(COUNT(CASE WHEN Result = 'W' AND Threes_Made_For > Threes_Made_Against THEN 1 END) / COUNT(CASE WHEN Threes_Made_For > Threes_Made_Against THEN 1 END) * 100, 2) AS Win_Percentage_When_More_Threes
FROM `nba-analytics-project.NBA.NBA_2020_Games`
GROUP BY Team
```

```sql
SELECT
  Team,
  COUNT(CASE WHEN Result = 'W' AND Rebounds_For > Rebounds_Against THEN 1 END) AS Wins_When_More_Rebounds,
  COUNT(CASE WHEN Rebounds_For > Rebounds_Against THEN 1 END) AS Games_More_Rebounds,
  ROUND(COUNT(CASE WHEN Result = 'W' AND Rebounds_For > Rebounds_Against THEN 1 END) / COUNT(CASE WHEN Rebounds_For > Rebounds_Against THEN 1 END) * 100, 2) AS Win_Percentage_When_More_Rebounds
FROM `nba-analytics-project.NBA.NBA_2020_Games`
GROUP BY Team
```

```sql
SELECT
  Team,
  COUNT(CASE WHEN Result = 'W' AND Total_Turnovers_For < Total_Turnovers_Against THEN 1 END) AS Wins_When_Fewer_Turnovers,
  COUNT(CASE WHEN Total_Turnovers_For < Total_Turnovers_Against THEN 1 END) AS Games_Fewer_Turnovers,
  ROUND(COUNT(CASE WHEN Result = 'W' AND Total_Turnovers_For < Total_Turnovers_Against THEN 1 END) / COUNT(CASE WHEN Total_Turnovers_For < Total_Turnovers_Against THEN 1 END) * 100, 2) AS Win_Percentage_When_Fewer_Turnovers
FROM `nba-analytics-project.NBA.NBA_2020_Games`
GROUP BY Team
```

Finally, I exported these new tables to Excel.

Step 4: Analyze:

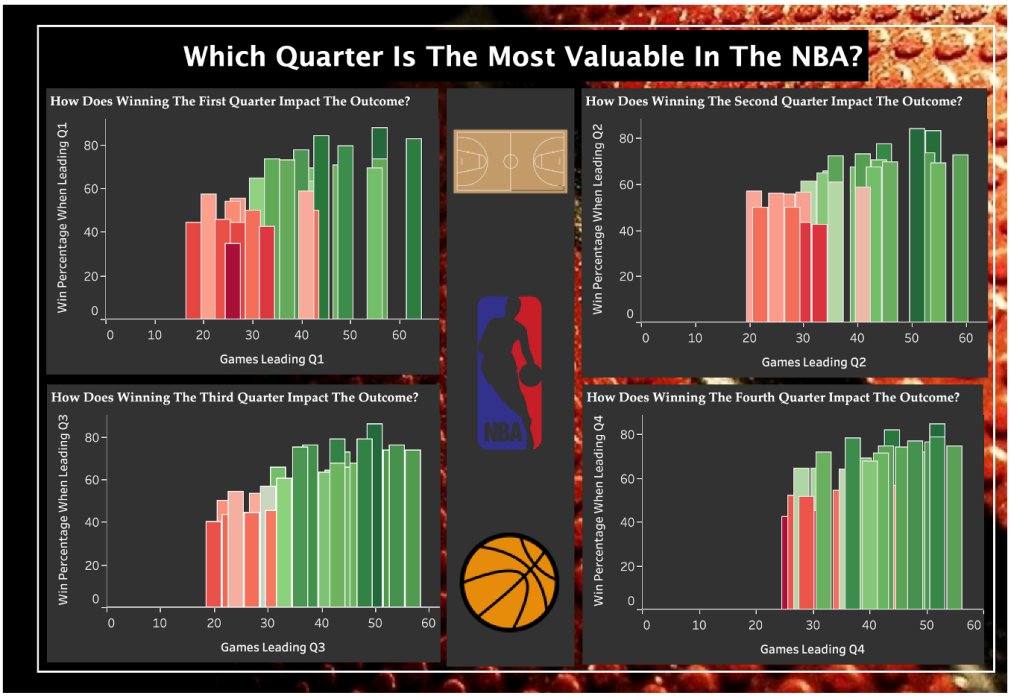
- **Patterns & Trends:**
  - Out-scoring your opponent in the first quarter led to a **62.27%** win rate on average
  - Out-scoring your opponent in the second quarter led to a **61.66%** win rate on average
  - Out-scoring your opponent in the third quarter led to a **61.96%** win rate on average
  - Out-scoring your opponent in the fourth quarter led to a **64.01%** win rate on average
  - Making more threes than your opponent led to a **61.37%** win rate on average
  - Grabbing more rebounds than your opponent led to a **65.36%** win rate on average
  - Committing fewer turnovers than your opponent led to a **52.90%** win rate on average
- Based on data from the 2019-2020 NBA season, winning the 4th quarter is the most important followed by winning the 1st quarter. This makes a lot of sense since closing out games at the end and setting the tone at the beginning of games is often stressed by coaches and players.
- Out-rebounding your opponent also leads to the highest win percentage of any statistic making it the most ideal stat to beat your opponent in.

Step 5: Share:

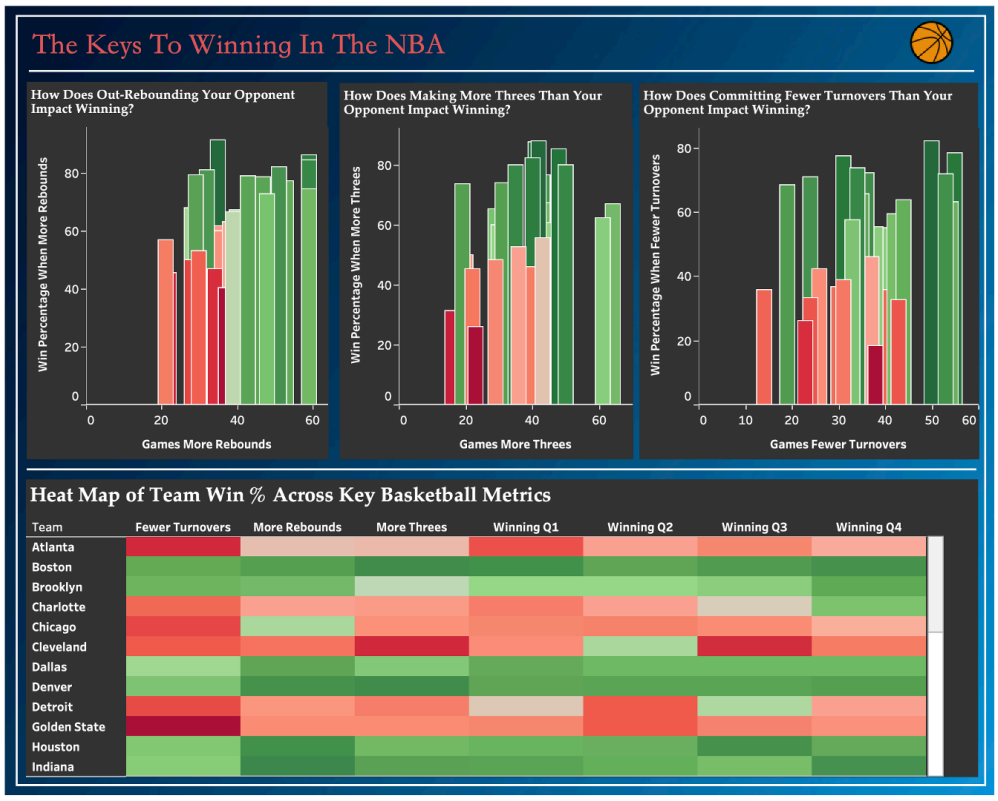
- I created two dashboards using Tableau.
  - The first dashboard consists of four visualizations that each highlight the impact of out-scoring your opponent in each quarter. The taller the bars in the graphs, the more often it results in a win.
  - The second dashboard also contains four visualizations, three of which compare the importance of threes made, rebounds, and fewer turnovers and how they relate to winning. Again, the taller the bars in the graphs, the more often it leads to winning. The fourth visualization is a heat map of every team in the NBA and how they stack up against the rest of the league in each category. The greener the

box, the more effectively the team converts their statistical advantages into actual victories.

Dashboard 1:



Dashboard 2:



Step 6: Act:

Key Takeaways:

- **Impact of Winning Quarters**:
    - The analysis reveals that winning the **fourth quarter** is the most critical factor in securing a victory, with the highest win percentage associated with leading in this period. Teams that close out games strong are better positioned to win.
    - The **first quarter** follows closely, indicating that starting strong and building early momentum sets the foundation for success.
    - The **third quarter** also plays a significant role, reflecting the importance of teams making strong adjustments after halftime.
    - Finally, winning the **second quarter** is the least impactful, though still a positive contributor to overall success.

- **Key Statistical Categories**:
    - **Rebounds** have the highest correlation to victory. Teams that win the battle on the boards, particularly by securing offensive rebounds, are more likely to control possession and scoring opportunities, leading to a higher win percentage.
    - **Three-point shooting** follows as the second most important statistic. In the modern NBA, success from beyond the arc has become a key differentiator between winning and losing teams.
    - **Turnovers** are the third most important factor. Teams that minimize mistakes and protect the ball have a significant advantage, as fewer turnovers mean more opportunities to score and less chance of giving the opponent easy points.

- **Team-Specific Insights**:
    - According to the heat map, the **L.A. Lakers** had the best overall season, performing exceptionally well across all key statistics and quarters. Their consistency in these areas was pivotal to their success.
    - On the other hand, the **Atlanta Hawks** had the worst overall performance, as indicated by all red data in the heat map. They struggled significantly across all key categories, making them the only team with consistently poor results.