Project Title:

NBA Data Analysis

**Low Level Document**

**By,**

**Sourav Mandal**

**Introduction**

The National Basketball Association (NBA) is a professional basketball league in North America. The league is composed of 30 teams, each of which plays 82 games during the regular season. The NBA is one of the most popular sports leagues in the world, with millions of fans around the globe.

This project uses Power BI to analyze NBA data. Power BI is a business intelligence platform that allows users to visualize and analyze data. The project analyzes data from three different sources:

* NBA Stats
* Basketball Reference
* Stats.NBA.com

The data from these sources is combined to create a single dataset that is used for analysis. The project analyzes a variety of factors that contribute to winning in the NBA, including field goal percentage, three-point percentage, free throw percentage, rebounding, and assists.

**Data Sources**

The data for this project was obtained from the following sources:

* NBA Stats: This is the official website for NBA statistics. The data is available in a CSV format. The CSV files contain a wide variety of statistics, including player statistics, team statistics, and game statistics. The data is updated daily, so it is always up-to-date.
* Basketball Reference: This is a website that provides detailed statistics on NBA players, teams, and games. The data is available in a CSV format. The CSV files from Basketball Reference contain more detailed statistics than the CSV files from NBA Stats. The data is also more comprehensive, as it includes statistics from previous seasons.
* Stats.NBA.com: This is the official website for NBA advanced statistics. The data is available in a JSON format. The JSON files from Stats.NBA.com contain even more detailed statistics than the CSV files from NBA Stats and Basketball Reference. The data also includes statistics that are not available in the other two sources, such as on-off court data and player tracking data.

The data from these three sources was combined to create a single dataset that was used for analysis. The dataset contains over 100 million rows of data, and it covers the period from 1950 to the present day.

**Data Cleaning**

The data was cleaned and formatted before it was imported into Power BI. This involved the following steps:

1. The data was checked for errors. Any errors that were found were corrected.
2. The data was formatted to a consistent format. This involved ensuring that the data was in the correct date format, and that the data types were consistent.
3. The data was merged. This involved combining the data from the three different sources into a single dataset.
4. The data was exported to a CSV file.

**Data Analysis**

Power BI was used to analyze the data. Power BI is a business intelligence platform that allows users to visualize and analyze data. The following visualizations were created using Power BI:

* **Bar charts:** These charts were used to show the distribution of player statistics. For example, one bar chart showed the distribution of field goal percentages for all players in the NBA.
* **Line charts:** These charts were used to show trends in player statistics over time. For example, one line chart showed the trend in three-point percentages for all players in the NBA over the past 10 years.
* **Pie charts:** These charts were used to show the breakdown of team performance. For example, one pie chart showed the breakdown of team wins and losses for all teams in the NBA.
* **Pivot tables:** These tables were used to summarize data and perform calculations. For example, one pivot table showed the average field goal percentage for all players in the NBA, broken down by position.
* **Maps:** These maps were used to show the distribution of team performance across the United States. For example, one map showed the number of wins for each team in the NBA, by state.

**Results**

The results of the data analysis showed that the following factors are important for winning in the NBA:

* Field goal percentage: This is the percentage of shots that a player makes. A high field goal percentage is important because it means that a player is efficient at scoring points.
* Three-point percentage: This is the percentage of three-point shots that a player makes. A high three-point percentage is important because it allows a player to score points from long range.
* Free throw percentage: This is the percentage of free