This project focuses on analyzing baseball statistics of currently active players (those whose final game was from 2020 to the present). First, I created a database using DB Browser (SQLite). This database was then connected to Jupyter Notebook for analysis using the sqlite3 library. I also imported other necessary libraries, including pandas, numpy, matplotlib, seaborn, and datetime.

After establishing the connection, I created a cursor to execute SQL queries to get the data on active players who have played at least 50 games. The result was converted into a pandas data frame. I then created a new column (Age) to calculate the players' ages using the datetime function and another column (FullName) by concatenating their first and last names. Afterward, all name and birth-related columns along with any rows containing null values were droped. The data was ready for analysis and the following questions were answered:

1. Which player had the most RBIs from 2015-2018?
2. How many times did Albert Pujols ground into double plays in 2016?
3. Create a histogram of the total number of triples each year.
4. Create a scatter plot of triples (3B) vs. stolen bases (SB).

Additionally, I explored further by answering the following:

1. Which year recorded the most home runs among active players, and what was the total amount?
2. Which player had the most triples (3B) in a single season, and in which year did it occur?
3. Which city produced the most currently active players?
4. What is the correlation between games played, home runs, triples, stolen bases, RBIs, and base on balls (BB)?

This concludes my analysis.