```
import pandas as pd

# Step 1: Data Collection
# Assuming you have a CSV file named 'baseball_data.csv'
data = pd.read_csv('baseball_data.csv')

# Step 2: Data Preprocessing
# Perform any necessary data cleaning and transformation here
# Handle missing values, outliers, data type conversions, etc.

# Step 3: Exploratory Data Analysis (EDA)
# Example code for EDA
# Display the first few rows of the dataset
print(data.head())

# Check the dimensions of the dataset (rows, columns)
print(data.shape)

# Summary statistics of the dataset
print(data.describe())

# Correlation matrix
correlation_matrix = data.corr()
print(correlation_matrix)

# Example visualization
import matplotlib.pyplot as plt

# Scatter plot of two variables
plt.scatter(data['Variable1'], data['Variable2'])
plt.xlabel('Variable1')
```

```
plt.ylabel('Variable2')
plt.title('Scatter Plot')
plt.show()

# Histogram of a variable
plt.hist(data['Variable3'], bins=20)
plt.xlabel('Variable3')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.show()

# Step 4: Perform further analysis based on your research questions, such as player performance analysis or predictive modeling.

# ... (Continue with your specific analysis steps)

# Step 10: Conclusion
# Summarize your findings and insights.
```