**Assignment 2: Data Collection and Curation**

**Name: Email:**

Statement:

For this assignment's preparation, the author(s) did not use any generative AI tools.

For this assignment's preparation, the author(s) have utilized [Generative AI Tool Name], a language model created by [Generative AI Tool Provider]. Within this assignment, the [Generative AI Tool Name] was used for purposes such as [e.g., brainstorming, grammatical correction, writing paraphrasing, citation, specific sections of the assignment].”

**Objective**: In this assignment, students will hone their skills in data collection, data analysis, and visualization. They'll use web scraping to gather data and then apply statistical methods to derive insights.

**Question 1: Amazon Best-Seller Rank Analysis**

**Objective**: Analyze the Amazon best-seller ranks for books written by Steven S. Skiena and recommend a book for purchase.

*Tasks*:

1. **Web Scraping**: Write a program to scrape the Amazon best-seller rank for each of Skiena's books. (Note: Ensure you adhere to Amazon’s terms of service and robots.txt when scraping their website.)
   * Tools recommended: Python, Beautiful Soup.
   * For each book, gather information on its name, ISBN, and best-seller rank.
2. **Data Analysis and Visualization**:
   * Plot the rank of all of Skiena's books over time. Use appropriate visualization tools like line graphs or scatter plots.
   * Analyze the trend and discuss any patterns you observe.
3. **Recommendation**:
   * Based on the rank trends, recommend which one of Skiena's books should be the next one to purchase.
   * Justify your recommendation.
   * Discuss if you believe any of these books would be an appropriate gift for a friend.

**Question 2: Player Ranking in Sports**

**Objective**: Devise a ranking system to determine the best players in a chosen sport based on historical statistical records.

*Tasks*:

1. **Data Collection**:
   * Choose a sport: baseball, football, basketball, cricket, or soccer.
   * Identify a reliable data set or source with historical statistical records for all major participants in your chosen sport.
   * Tools recommended: Python, pandas, web scraping libraries if necessary.
2. **Ranking System Development**:
   * Develop a system or algorithm to rank players at each position.
     + Define the criteria or metrics you'll use to rank the players (e.g., batting average for baseball, goals scored for soccer).
     + Discuss any weighting factors you'll apply to different metrics.
3. **Data Analysis and Visualization**:
   * Apply your ranking system to the data set.
   * Visualize the top players at each position using appropriate charts or graphs.
   * Discuss any interesting patterns or anomalies you observe in the rankings.
4. **Report & Discussion**:
   * Prepare a report summarizing your methodology, analysis, and conclusions.
   * Discuss the limitations of your ranking system and potential improvements.

**Submission Guidelines**:

1. **Code**: Include well-commented code for both questions. Ensure it's readable and modular.
2. **Report**: A comprehensive report discussing your methods, findings, and conclusions for both questions.
3. **Visualization**: Include relevant plots, charts, or graphs in your report.

**Assessment Criteria**:

1. **Code Quality** (20%): Readability, modularity, and efficiency.
2. **Analysis Depth** (40%): Depth and clarity of analysis, including the justification of choices made.
3. **Visualization Quality** (20%): Relevance and clarity of visualizations.
4. **Report Quality** (20%): Overall presentation, structure, and coherence of the report.