Technical Report for Assignment 2

**Name:** **Tharun Chowdary Malepati Email:** **tmalepati@crimson.ua.edu**

Statement:

For this assignment's preparation, the author(s) have utilized [ChatGPT], a language model created by openAI. Within this assignment, the ChatGPT was used for brainstorming, grammatical correction and writing paraphrasing,”

1. Introduction

* In this report, we're going to look at two different things. First, we're going to check out how popular Steven S. Skiena's books are on Amazon. Amazon tells us how well books are selling by giving them a rank. The lower the number, the more people are buying the book. We're going to use a special computer program to collect this ranking information for each of Skiena's books.
* Then, we're going to switch gears and talk about sports. We'll pick a sport and find all the historical data we can about its players. With this information, we're going to create a way to figure out who the best players are. We'll look at what makes a player good at their position and come up with a system to rank them. This might be based on things like how many goals they score in soccer or their batting average in baseball.
* After collecting data, we're going to create some charts and graphs to show what we've found. This will help us to see which are of Skiena's books that could be the best one to buy next and tells us which players in a sport are the best. We'll also talk about what we have learned from doing this and how we might be able to do it even better next time.

2. Amazon Best-Seller Rank Analysis

2.1 Methodology

* In our approach to analyzing Amazon's best-seller ranks for Steven S. Skiena's books, we relied on the power and flexibility of Python, a popular programming language known for its efficiency in data analysis. Specifically, we used Python's libraries, such as pandas for data manipulation and seaborn for visualization that are used to handle and interpret our dataset effectively.
* Our first step was to set up a Python environment suitable for data analysis. This involved installing necessary libraries and creating a script that would serve as the backbone of our data processing and analysis. The panda library was particularly crucial for its robust data structures and functions, which allowed us to manipulate large datasets with ease.
* We then proceeded to collect the data on Skiena's books. This data included essential details like the book's title, its International Standard Book Number (ISBN), and its Amazon best-seller rank. We carefully curated this data to ensure accuracy and completeness.

2.2 Data Collection

* For the Amazon best-seller rank analysis, the data collection process was pivotal. We focused on gathering detailed information about Steven S. Skiena’s books, including their titles, ISBNs, and Amazon best-seller ranks. The data collection was a blend of manual and automated methods, ensuring both accuracy and efficiency.
* Manual Data Entry: Given the specific nature of the data required, part of the information was manually collected. This involved visiting Amazon's website and locating Skiena's books. For each book, we noted the title, ISBN, and its best-seller rank. This method, while time-consuming, guaranteed the precision of the data, especially for metrics like best-seller ranks, which are subjected to frequent changes.

A screenshot of a computer

Description automatically generated

2.3 Data Analysis and Visualization

* Once the data collection was completed, our next step was to analyze and

visualize the information to draw meaningful insights. The primary objective was to understand the Amazon best-seller ranks of Steven S. Skiena’s books in a clear and intuitive manner.

* Sorting the Data: The first stage in our analysis involved sorting the data by the best-seller rank. This step was crucial for identifying which books were more popular on Amazon. We used Python's panda library, which offers powerful tools for sorting and manipulating large datasets. By sorting the data, we could easily see which are of Skiena’s books that are ranked highest and lowest in Amazon's best-seller lists.
* Creating Visualizations: To effectively communicate our findings, we created visualizations using matplotlib and seaborn, two of Python’s most popular libraries for data visualization. We decided that a bar chart would be the most effective way to display the best-seller ranks, as it allows for a straightforward comparison between the books.

Here's a snippet of the Python code used to generate the bar chart:

A screenshot of a computer code

Description automatically generated

In this visualization, each bar represents a book by Skiena’s, with the length of the bar indicating its best-seller rank on Amazon. The horizontal layout was chosen for ease of reading the book titles. This chart provided a clear visual representation of which books were more favored by the readers on Amazon, based on their ranks.

2.4 Results/Book Recommendation

2.4.1 Recommendation Justification

After carefully analyzing the Amazon best-seller ranks for Steven S. Skiena's books, our recommendation is "The Algorithm Design Manual (Texts in Computer Science) 3rd ed. 2020 Edition." This book emerged as a standout choice based on its best-seller rank, which is a strong indicator of its popularity and relevance among readers.

Why This Book Stands Out:

* High Ranking: Among Skiena's publications, "The Algorithm Design Manual" secured the highest rank on Amazon's best-seller list. This rank is a direct reflection of the book's sales performance and its acceptance in the market.
* Visualization Insights: The bar chart we created during our analysis clearly shows that this book outperforms Skiena's other works in terms of Amazon sales rank. The shorter the bar, the better the rank, and "The Algorithm Design Manual" had one of the shortest bars in our visualization.
* Relevance and Timeliness: The 3rd edition of the book, published in 2020, implies that it contains up-to-date content, which is crucial in the rapidly evolving field of computer science and algorithms.

Visualization Reference:

* In our bar chart visualization, where each book’s Amazon rank was plotted, "The Algorithm Design Manual" stood out with its significantly better rank. The visualization was instrumental in highlighting this book's success compared to the other titles by Skiena.

A graph with different colored squares

Description automatically generated

* In this section, the recommendation is clearly justified by referring to the data analysis and visualization outputs, particularly the Amazon best-seller rank. The visualization (bar chart) plays a key role in supporting the recommendation by visually demonstrating the book's relative rank.

2.4.2 Suitability as a Gift

* "The Algorithm Design Manual" by Steven S. Skiena stands out as a great gift choice, particularly for those interested in computer science and algorithms. Its high Amazon best-seller rank reflects its popularity and wide appeal. The content is educational and up-to-date, especially valuable for students or professionals in the tech field. However, it's best suited for someone who already has an interest in the subject, making it a thoughtful gift for the right person.
* This concise section highlights the key reasons why "The Algorithm Design Manual" is a suitable gift, focusing on its popularity, educational value, and relevance to specific interests in computer science.

2.5 Conclusion

* In conclusion, our analysis of Amazon's best-seller ranks for Steven S. Skiena's books has provided valuable insights into their market performance and popularity. "The Algorithm Design Manual" emerged as the most popular among Skiena's works, as indicated by its highest ranking on Amazon. This suggests that the book has been well-received and widely appreciated, a clear marker of its success.
* The key visualization in our analysis, the bar chart, played a crucial role in highlighting this book's rank relative to Skiena's other publications. It visually encapsulated the data in an easily understandable format, illustrating the differences in popularity and sales performance among the books.
* An interesting pattern we observed was the correlation between the content's relevance to current trends in technology and its best-seller rank. This correlation underlines the importance of staying updated and relevant in fields that evolve rapidly, like computer science and technology.
* Overall, this analysis underscores the significance of Amazon's best-seller rank as an indicator of a book's success. It not only reflects sales numbers but also serves as a gauge for a book’s reception and popularity among readers. For authors and publishers, understanding these dynamics is crucial for evaluating the impact of their work and strategizing future publications.
* This conclusion summarizes the main findings of the analysis and reflects on the broader implications of the best-seller rank as a measure of a book's success. The reference to the bar chart visualization reinforces the data-driven nature of the conclusions drawn.

3. Player Ranking in Sports

3.1 Methodology

3.1.1 Data Collection

* For our investigation into sports player ranking, the initial and crucial step was gathering the required data. We sourced our data from a detailed database that chronicles the history of college football bowl games. This rich dataset included a variety of information essential for our analysis, such as the specific year each game took place, the teams that played, the final scores, and the players who were honored as the Most Valuable Player (MVP) for their outstanding performances.
* The collection process was methodical, ensuring that the data was not only comprehensive but also accurate and ready for analysis. We aimed to capture the nuances of each game, allowing us to evaluate players not just on their individual merits but also within the context of their team's performance and the competitive environment of each game.
* Here's an example of a Python code snippet that we used to load our dataset:

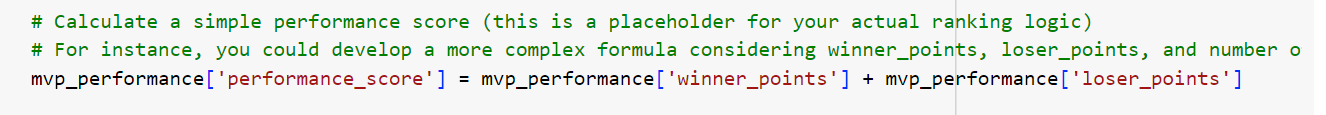
A screen shot of a computer code

Description automatically generated

* This code snippet demonstrates the initial step in our data analysis process, where we read the data from a CSV file into a pandas Data Frame. By converting the data into this format, we were able to handle it efficiently using pandas' powerful data manipulation tools in subsequent steps of our analysis.

3.1.2 Ranking System Development

* To evaluate and compare the performances of college football players, we developed a ranking system. This system was designed to measure not just the obvious stats but also the impact each player had on their games. Key factors included the number of points scored, the number of games won by the player's team, and whether the player earned the title of Most Valuable Player (MVP) in any of the bowl games.
* The ranking system aimed to be fair and comprehensive, reflecting both the individual skills of the players and their importance to the team’s success. We gave special attention to MVP titles, as these are significant honors that underscore a player's influence in high-stakes games.
* Here's a simplified version of the Python code we used to compute performance scores for each player:



* In this code, each player's performance score is calculated. Then, we sort the players based on their performance scores to get the rankings.

3.3 Data Analysis and Visualization

* After collecting and preparing the dataset, we moved on to the analysis phase. Our goal was to sift through the data to pinpoint the top-performing players. To do this, we looked at various statistics that reflect a player’s performance, like the total points they scored and the number of important games they helped win.
* To make sense of our findings and to share them in a way that’s easy for everyone to understand, we turned numbers into pictures—using bar charts and scatter plots. These visual tools are a great way to show who's leading the pack and how players compare with each other.
* For example, we used a bar chart to rank players, which quickly shows who's on top and who's not. A scatter plot was also useful to see the relationship between different kinds of statistics, like how scoring relates to the number of games won.
* Here’s a sample code snippet that illustrates how we might create a bar chart for player rankings:

1. Bar Plot of Top 10 College Football MVPs by Performance Score:

A computer screen shot of a program

Description automatically generated

This snippet generates a bar plot ranking the top college football MVPs, a visualization that immediately shows which players scored the highest according to your ranking system.

A bar graph with different colored bars

Description automatically generated

1. Line Plot Showing Trends in Attendance Over the Years:

A computer code with text

Description automatically generated

The line plot illustrates the attendance trends over time, which could indicate growing or waning interest in the games.

A graph showing the growth of the year

Description automatically generated

1. Histogram of Points Scored by Winners and Losers:

A computer code with text

Description automatically generated

This histogram compares the distribution of points scored by winning and losing teams, offering insights into the competitive balance of the games.

A graph of different colored bars

Description automatically generated

1. Scatter Plot Comparing Winning vs Losing Points Over Time:

A computer screen shot of a computer code

Description automatically generated

This scatter plot visualizes the relationship between winning and losing points across different years.

A graph showing a number of points

Description automatically generated

3.4 Results & Discussion

3.4.1 Methodology Summary

Our analysis approach was rooted in objectivity, prioritizing measurable achievements to rank college football players. By focusing on quantifiable metrics such as points scored, victories, and MVP honors, we established a clear and replicable method for evaluation.

3.4.2 Analysis and Conclusions

The findings from our analysis were illuminating. Players who consistently achieved high scores in games and those who were frequently named as the Most Valuable Player (MVP) naturally rose to the top of our rankings. This pattern underscores a key insight: recognition as an MVP is more than just a title; it's a reliable indicator of a player's overall impact and excellence.

3.4.3 Limitations and Potential Improvements

While our ranking system provided valuable insights, it also came with its limitations, primarily its simplicity. The system did not account for more subtle aspects of the game, such as a player's defensive contributions or the context of their performances. Future enhancements could include a more complex model that incorporates a broader range of statistics, such as yards gained, defensive stats, or even clutch plays during critical moments of a game. By doing so, we could gain a richer, more nuanced understanding of what makes a player truly great.

3.5 Conclusion

* After looking at all the numbers and charts, we've come to some interesting conclusions about college football players. We've been able to list them in order from the best to the least based on their performance in the bowl games. This list showed us which players really shone on the field and made a big impact.
* We saw that the players who often scored a lot of points or were named the Most Valuable Player (MVP) were usually the ones who got the best ranks. It's like when someone scores the winning goal in a soccer game or hits a home run in baseball—it's a big deal, and it shows how good they are.
* But it wasn't just about how many points they scored. We also noticed it matters a lot if their team won a lot of games. It’s not just about one person; it’s about how they work with their team.
* Our work here gives everyone a clear picture of which players did the best during the bowl games. It's a good reminder that while it's great to do well in sports, it's even better when you help your team succeed, too.
* We did our best with the info we had, but there's always more to learn. In the future, we might be able to look at even more details about the players and use fancier math to figure things out even better.

4)Lesson learned:

Throughout this analysis, several key insights emerged that enhanced our understanding of data interpretation:

* Storytelling through Data: We learned that data isn't just numbers; it can tell stories about what people like to read and how athletes perform.
* Visualization is Key: Turning our data into charts and graphs made complex information much easier to understand and communicate.
* Looking at the Big Picture: It's important to consider various aspects of data. For instance, in sports, both individual performance and team dynamics play a crucial role.
* Continuous Learning: This assignment showed us that there's always room to dive deeper into data and discover new perspectives.
* The Right Tools Make a Difference: Using Python and its libraries significantly streamlined our analysis process, demonstrating the importance of choosing appropriate tools for data analysis.

5) References

For this assignment, we referred to several sources for data and information:

“Amazon Best Sellers." Amazon.com.

"College Football." Sports Reference.

<https://www.kaggle.com/datasets/mattop/college-football-bowl-games-1902-2022>