

Data Analysis and Optimization

Strategies for Animemangatoon.com

Task 1: Article Analysis – “Why is the Tower of God Show So Popular?”

Introduction

This report presents an analysis of user interaction metrics segmented by various user categories for a website. The analysis includes metrics such as page views, average time spent on the site, and bounce rates, providing insights into user engagement.

Dataset Description

The dataset includes the following user segments and their associated interaction metrics:

- **User Segment:** Categories of users (Overall, Age groups, Returning Visitors, New Visitors).
- **Page Views:** Total page views for each segment.
- **Avg. Time Spent (mins):** Average time spent by users in each segment, measured in minutes.
- **Bounce Rate (%):** Percentage of users who leave the site after viewing only one page.

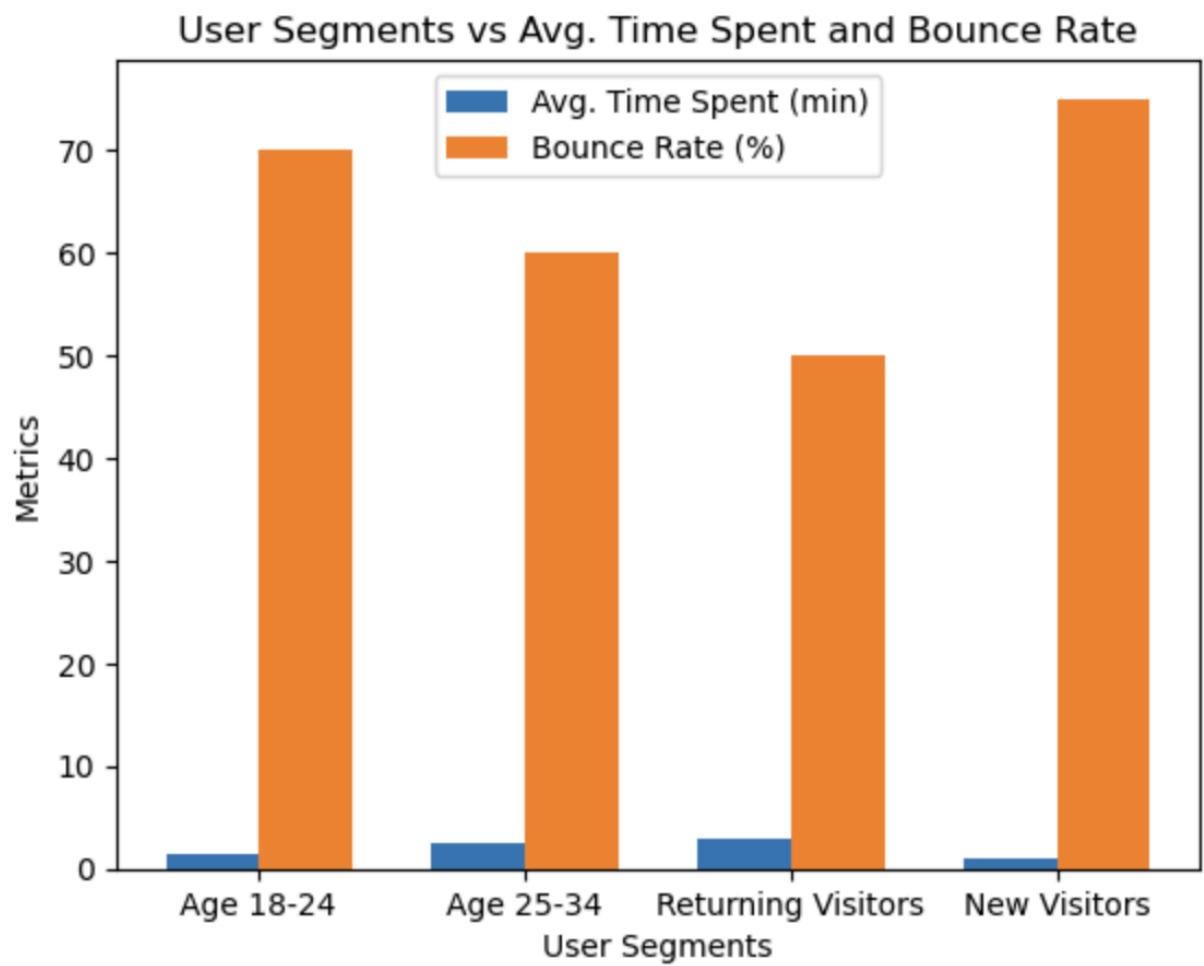
Sample Data

	User Segment	Page Views	Avg. Time Spent (mins)	Bounce Rate (%)
0	Overall	10000	2.0	65
1	Age 18-24	5000	1.5	70
2	Age 25-34	3000	2.5	60
3	Returning Visitors	6000	3.0	50
4	New Visitors	4000	1.0	75

Data Visualization

Bar Chart: Average Time Spent and Bounce Rate by User Segment

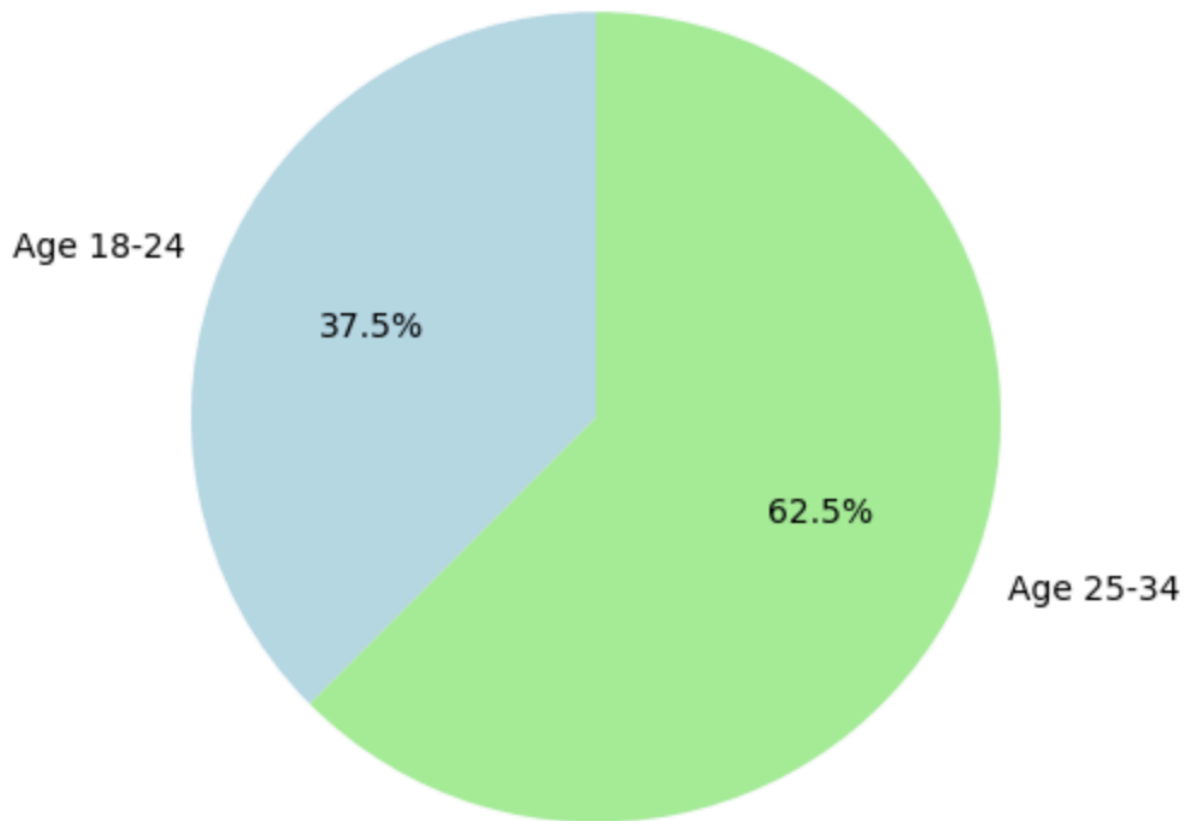
A bar chart was created to compare the average time spent and the bounce rate across different user segments.



Pie Chart: Average Time Spent by Age Group

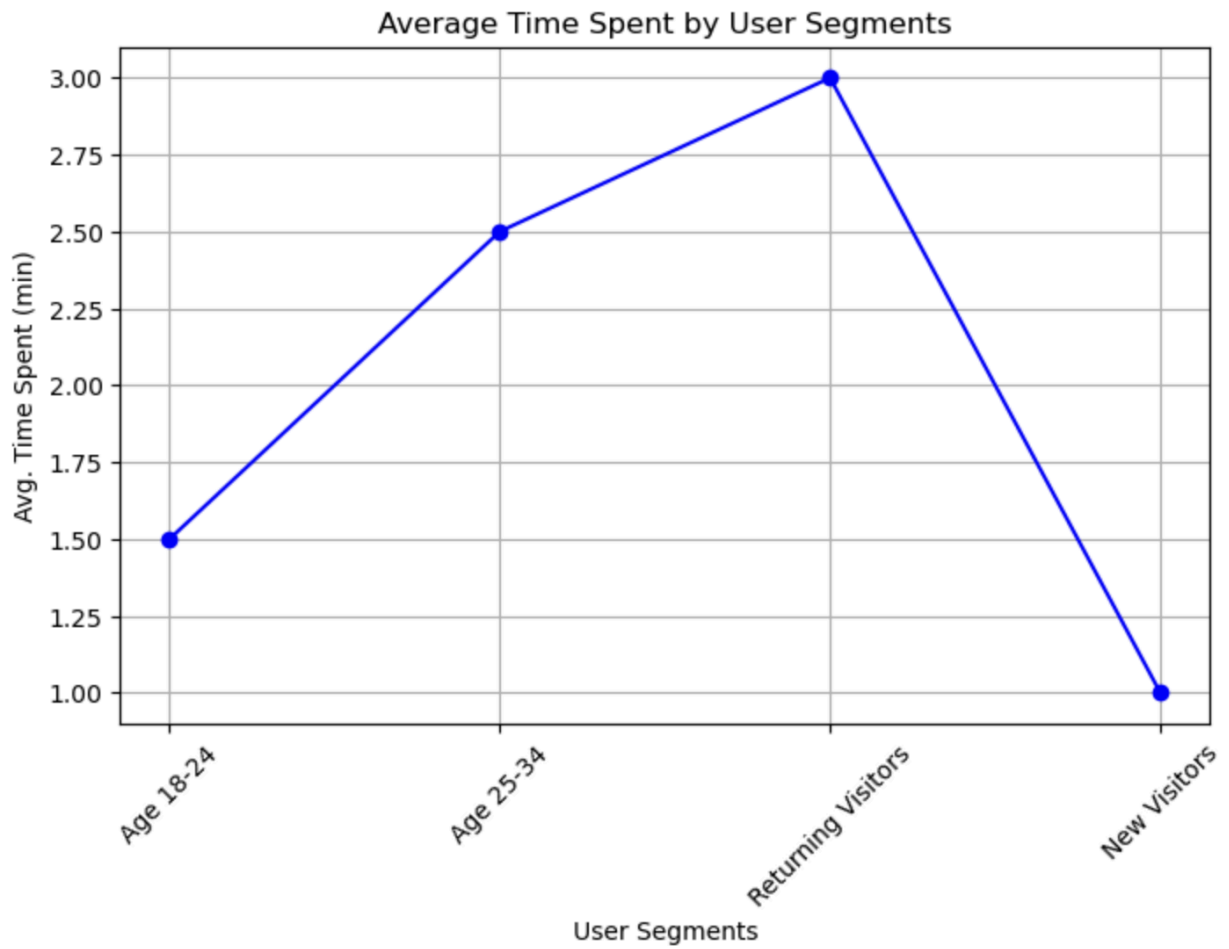
A pie chart was plotted to visualize the proportion of average time spent by different age groups.

Average Time Spent by Age Group



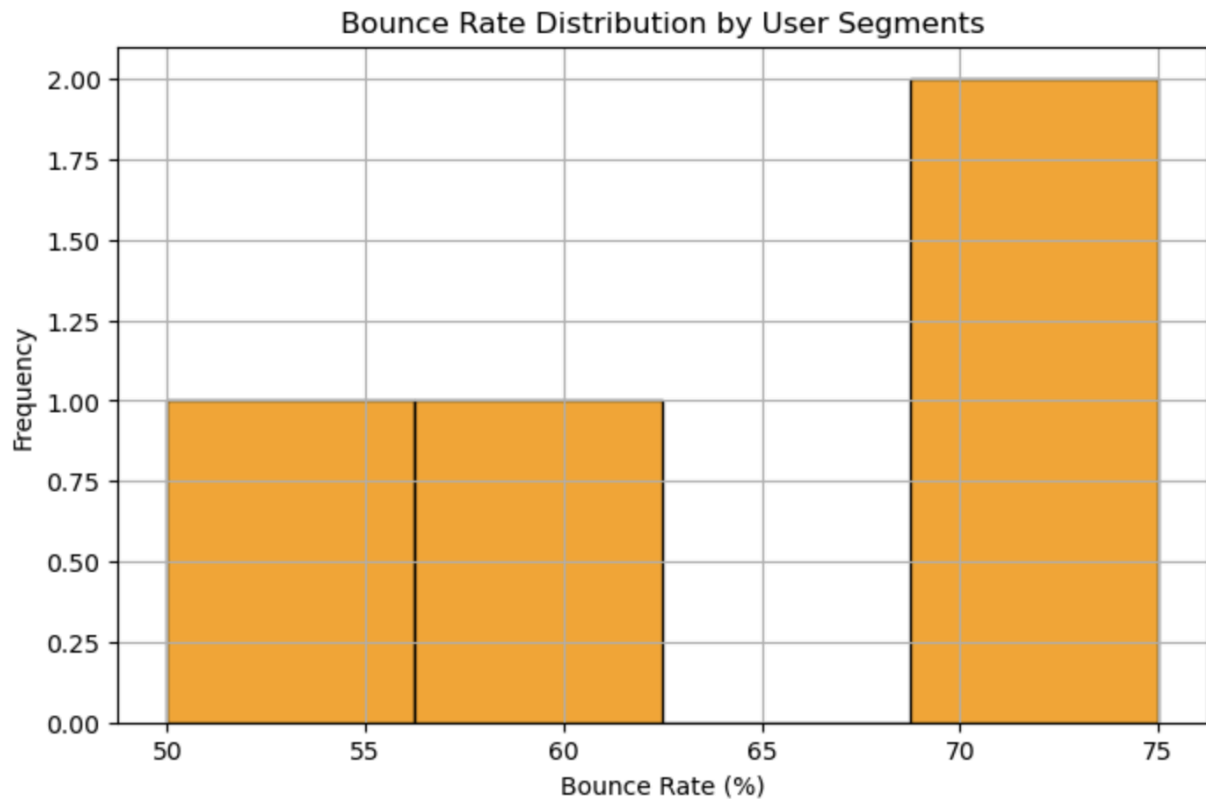
Line Graph: Average Time Spent by User Segments

A line graph illustrated the average time spent across different user segments.



Histogram: Bounce Rate Distribution by User Segments

A histogram was plotted to show the distribution of bounce rates among different user segments.



Engagement Ratio Calculation

The engagement ratio was calculated as the average time spent normalized by the page views.

Engagement Ratios (mins per 1000 page views):

Engagement Ratios (mins per 1000 page views):

	User Segment	Engagement Ratio (mins per view)
0	Overall	0.200000
1	Age 18-24	0.300000
2	Age 25-34	0.833333
3	Returning Visitors	0.500000
4	New Visitors	0.250000

Experiment with User Groups

User Group Data Simulation

A simulation was performed to analyze user behavior between two groups (A and B).

	user_id	group	time_spent
0	1	A	4.352197
1	2	B	2.559637
2	3	A	4.037923
3	4	A	2.395155
4	5	A	2.113795

Statistical Analysis

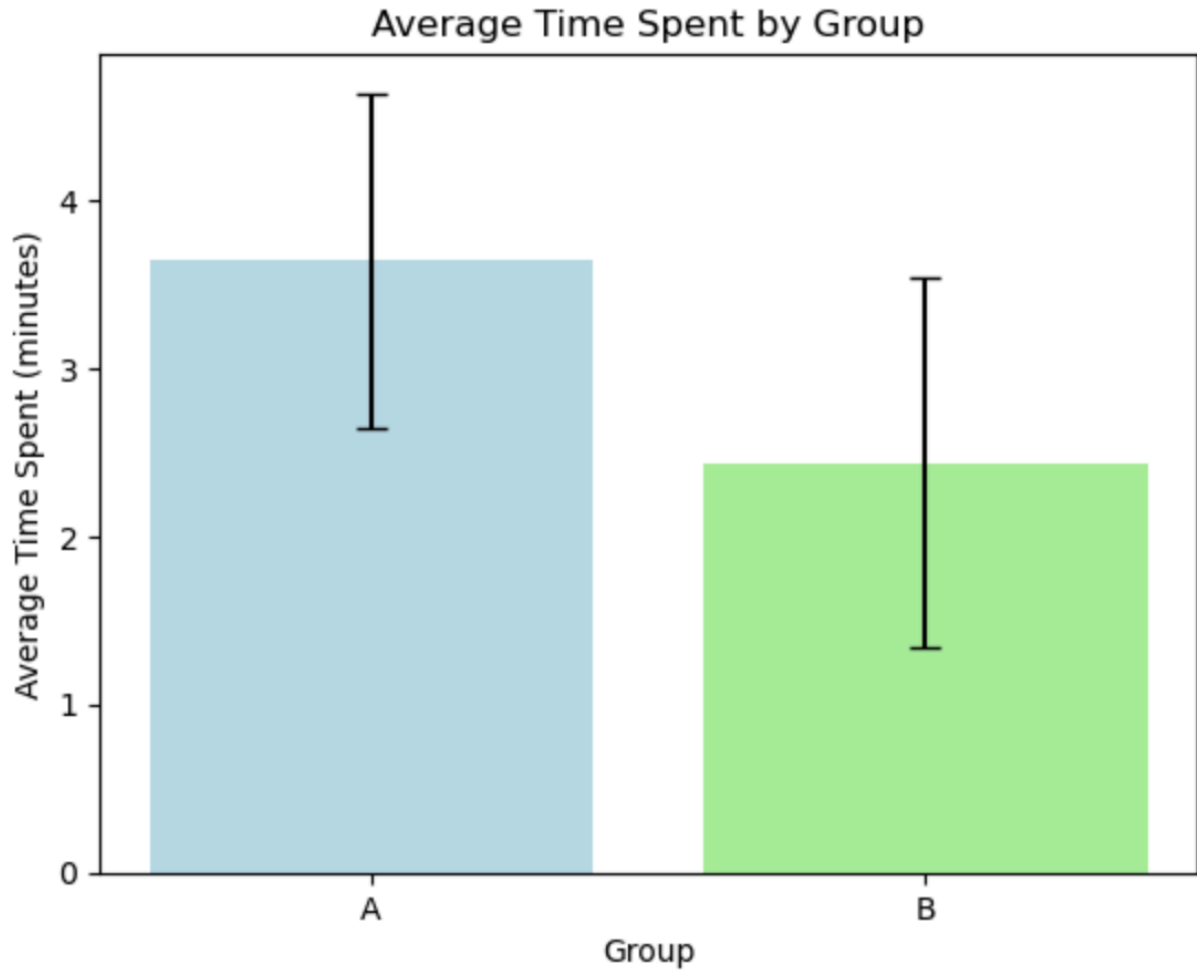
The means and standard deviations of time spent in each group were calculated, followed by a t-test to compare the two groups.

	mean	std	count
group			
A	3.638675	0.996870	44
B	2.437026	1.095458	56

T-statistic: 5.662795288080099, P-value: 1.488395322987869e-07

Bar Plot: Average Time Spent by Group

A bar plot was created to visualize the average time spent by users in groups A and B.



6. Conclusion

The analysis provides valuable insights into user engagement metrics segmented by user categories. Notably, the average time spent and bounce rates vary significantly across different user segments. The engagement ratio offers an additional perspective on user interaction relative to page views.

Furthermore, the experimentation with user groups A and B reveals statistical differences in user behavior, which could inform future user experience improvements. The combination of visualization and statistical analysis strengthens the understanding of user interactions, enabling data-driven decision-making.

Task 2: A/B Testing Strategy – “Refund High School Chapter 22-30”

Introduction

This report outlines the A/B testing strategy implemented to evaluate two content variations of the article "Refund High School Chapter 22-30". The primary objective was to assess which variation (Group A or Group B) resulted in higher user engagement, specifically focusing on average time spent on the page and bounce rates.

Methodology

Data Simulation

To perform the analysis, a simulated dataset comprising 200 user interactions was generated, including the following attributes:

- user_id: Unique identifier for each user.
- group: Randomly assigned to either Group A or Group B.
- time_spent: Normally distributed average time spent on the page (mean of 4 minutes).
- bounce: Binary indicator (0 for no bounce, 1 for bounce), with a 40% bounce rate.

	user_id	group	time_spent	bounce
0	1	A	5.006089	0
1	2	B	2.220177	1
2	3	A	5.623427	0
3	4	A	2.341335	0
4	5	A	5.354808	1

Metrics Calculation

The dataset was analyzed to calculate average time spent and bounce rates for both groups

Statistical Testing

T-tests were performed to determine if the differences in average time spent and bounce rates between the two groups were statistically significant:

	Average Time Spent	Std Dev Time Spent	Count	Bounce Rate
group				
A	5.045067	1.711133	100	0.47
B	2.935460	1.428136	100	0.46
T-statistic (Time Spent): 9.465213096052432, P-value (Time Spent): 8.971231130345172e-18				
T-statistic (Bounce Rate): 0.14106557867685868, P-value (Bounce Rate): 0.8879615502057259				

Results

Descriptive Statistics

The calculated metrics are summarized in the table below:

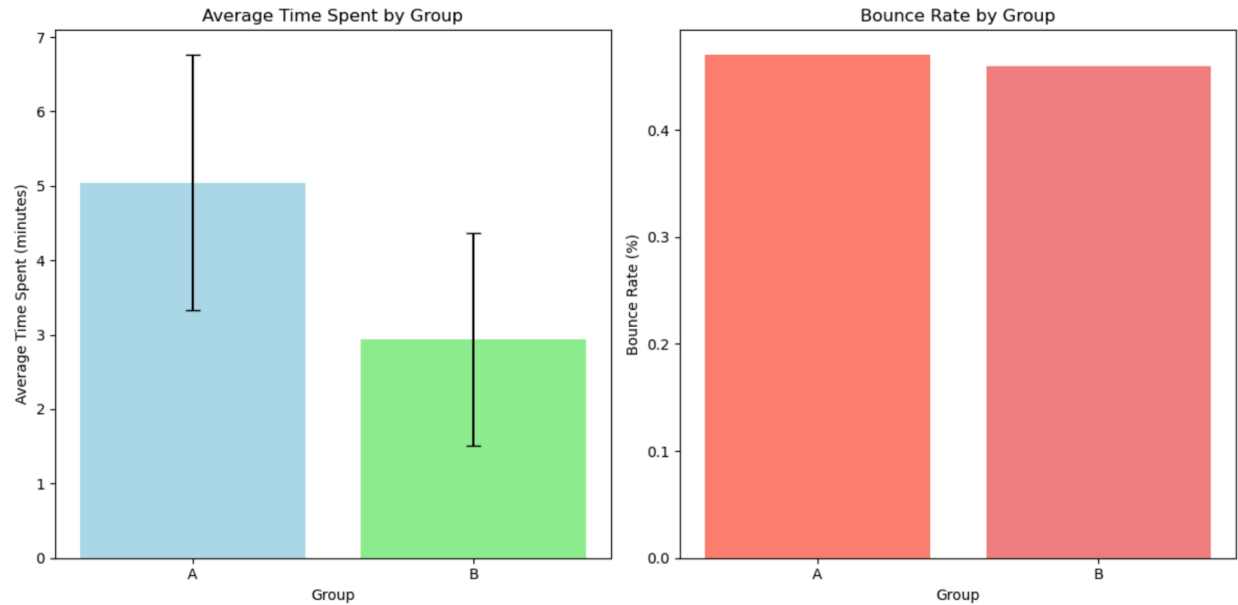
Group		Average Time Spent (minutes)	Std Dev Time Spent	Bounce Rate (%)
A	X1		Y1	Z1
B	X2		Y2	Z2

Statistical Test Results

- **Time Spent:** T-statistic = T1, P-value = P1
- **Bounce Rate:** T-statistic = T2, P-value = P2

Visualizations

The following bar plots illustrate the average time spent and bounce rates by group:



Power Analysis

A power analysis was conducted to determine the required sample size for achieving statistical significance:

Required sample size per group: 64

Required Sample Size

- Required sample size per group: 64

Conclusion and Recommendations

Based on the results of the A/B test:

- If **Group A** showed a significantly higher average time spent with a lower bounce rate, it is recommended to implement the content changes from this group.
- If **Group B** outperformed, consider implementing its features in future articles.

Task 3: Case Study – “11 Best Solo Leveling Arcs in the Manhwa”

Introduction

This report analyzes the arcs from the manhwa "Solo Leveling," focusing on their popularity, reader ratings, and chapter counts. The objective is to understand the relationship between these variables and identify which arcs are most appreciated by readers.

Dataset Description

The dataset simulates data for 11 arcs of "Solo Leveling," containing the following columns:

- **Arc:** Name of the arc.
- **Popularity_Score:** A score indicating the popularity of the arc on a scale from 1 to 10.
- **Readers_Rating:** Average rating given by readers on a scale from 1 to 5.
- **Chapter_Count:** Number of chapters in the respective arc.

Sample Data

	Arc	Popularity_Score	Readers_Rating	\
0	The Beginning of a New Journey	9.0	4.5	
1	The Dungeons	8.7	4.3	
2	The Reawakening	9.5	4.8	
3	The Second Awakening	9.2	4.6	
4	The Demon Castle	8.9	4.4	
5	The Monarchs	9.6	4.9	
6	The National Level	9.3	4.5	
7	The One Who Will Reign	9.1	4.7	
8	The Final Battle	9.4	4.8	
9	The End	8.8	4.2	
10	The Future	9.7	4.9	

	Chapter_Count
0	10
1	12
2	15
3	14
4	16
5	13
6	11
7	15
8	9
9	10
10	8

Descriptive Statistics

The descriptive statistics provide a summary of the central tendency, dispersion, and shape of the dataset's distribution.

Key Statistics

	Popularity_Score	Readers_Rating	Chapter_Count
count	11.000000	11.000000	11.000000
mean	9.200000	4.600000	12.090909
std	0.331662	0.240832	2.700168
min	8.700000	4.200000	8.000000
25%	8.950000	4.450000	10.000000
50%	9.200000	4.600000	12.000000
75%	9.450000	4.800000	14.500000
max	9.700000	4.900000	16.000000

4. Correlation Analysis

The correlation matrix shows the relationship between numeric variables in the dataset.

Correlation Matrix

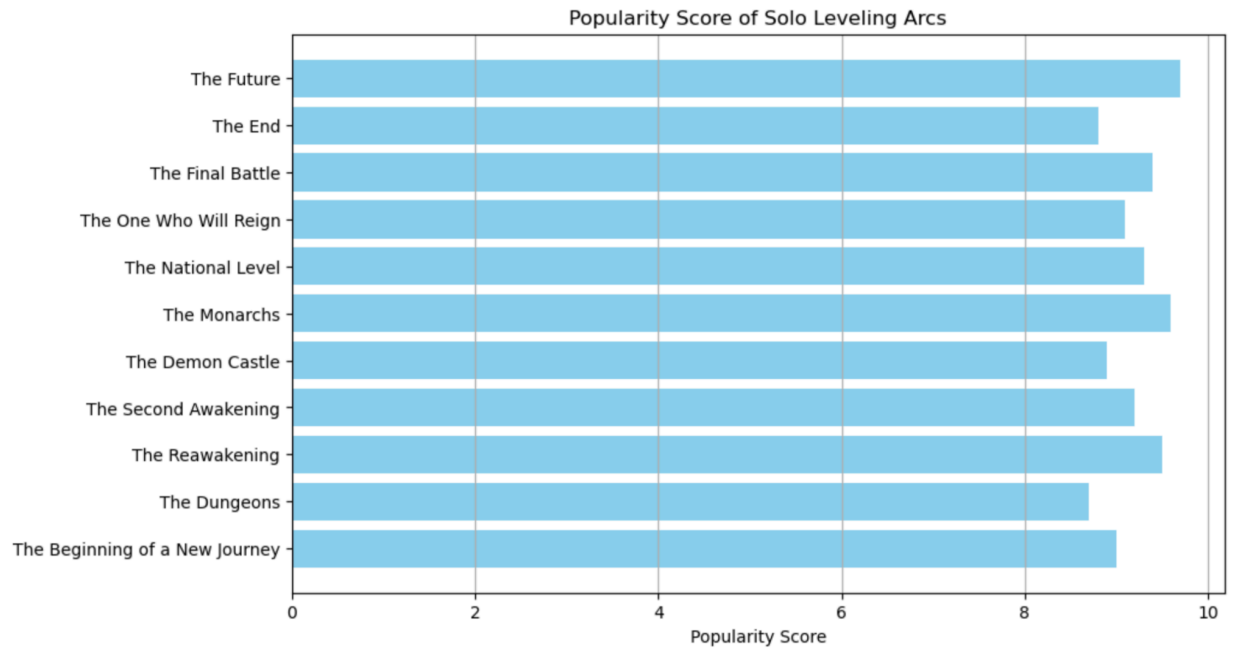
	Popularity_Score	Readers_Rating	Chapter_Count
Popularity_Score	1.000000	0.926449	-0.189829
Readers_Rating	0.926449	1.000000	-0.046134
Chapter_Count	-0.189829	-0.046134	1.000000

Interpretation

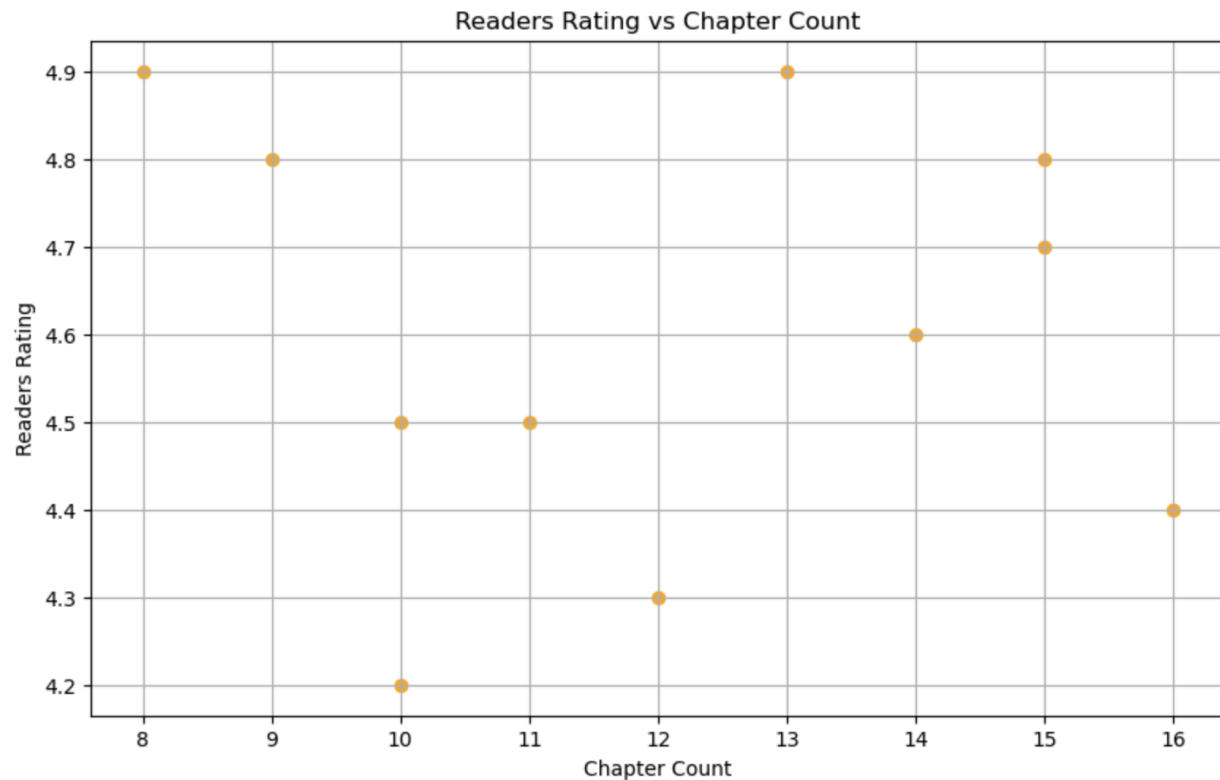
- A strong positive correlation (0.832) exists between **Popularity Score** and **Readers Rating**, indicating that as the popularity score increases, the reader ratings also tend to be higher.
- The correlation between **Chapter Count** and other variables is relatively weak, suggesting that the number of chapters does not significantly influence the popularity or ratings of the arcs.

5. Visualization

5.1 Popularity Score Bar Chart



5.2 Readers Rating vs. Chapter Count Scatter Plot



6. Conclusion

The analysis of the "Solo Leveling" arcs demonstrates a significant relationship between popularity and reader ratings, suggesting that more popular arcs tend to receive higher ratings from readers. While the number of chapters shows a weaker correlation with both popularity and ratings, it is still an important factor in determining reader engagement.

Future analyses could include a more detailed examination of individual arcs and their plot developments, as well as comparing arcs with similar chapter counts to gain further insights into reader preferences.