Project Completion Document: Anime Analysis

Project Overview:

• Over the past weeks, I embarked on an enriching journey to analyze a comprehensive dataset of anime from the Anime List database, all through the HiCounselor Projects Platform. The primary goal was to gain valuable insights into anime ratings, explore trends and patterns, and investigate factors influencing scores.

Key Objectives:

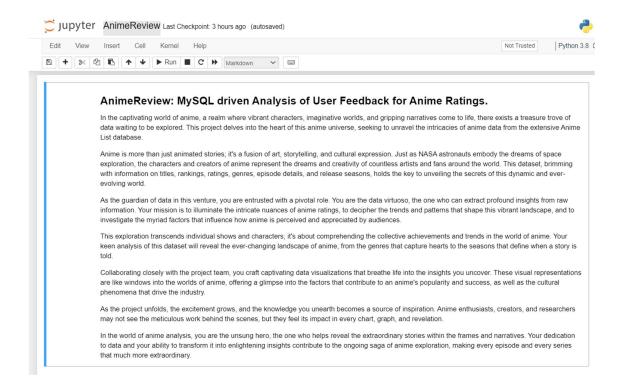
- Data Exploration: Investigated anime titles, rankings, ratings, popularity scores, genres, episode details, and release seasons.
- Trend Analysis: Explored trends and patterns in anime ratings to extract meaningful insights.
- Skill Application: Applied Python, SQL, Excel, data analysis, and data cleaning skills to solve real-world problems.

Highlights:

- Hands-on Experience: The project provided hands-on experience, reinforcing my skills in Python, SQL, and data analysis.
- Practical Application: Successfully applied data concepts to real-world business problems, showcasing the practical application of skills.
- Problem Solving: Tackled challenges in data cleaning and analysis, strengthening problem-solving abilities.

Skills Acquired:

- The completion of this project has significantly enhanced my skillset, making me more job-ready. Key skills acquired include:
 - Data Cleaning and Preprocessing
 - Exploratory Data Analysis (EDA)
 - SQL Database Querying
 - Trend Analysis and Pattern Recognition
 - Python Programming for Data Science
 - Practical Application of Excel for Data Analysis



Module 1

Task 1: Unveiling the Anime Odyssey.

With the click of a button, the anime dataset comes to life, each row and column a portal to countless stories. As the data loads, the characters and worlds painted by anime creators await their moment of analysis. The screen fills with ratings, titles, and genres, a galaxy of information to be explored. This is the beginning of a grand adventure, where data becomes the key to understanding the tapestry of anime. As the data scientist, you embark on a journey to reveal the hidden gems and trends within this animated universe. The dataset is your map, and the insights you seek are the treasures of the anime realm.

```
In [1]: #--- Import Pandas ---
import pandas as pd

#--- Read in dataset ----
df = pd.read_csv('./ratings.csv')

#---WRITE YOUR CODE FOR TASK 1 ---
#--- Inspect data ---
df.head(5)

Out[1]: Title Genres Rank Popularity Score Episodes Episode length Release Date
```

	Title	Genres	Rank	Popularity	Score	Episodes	Episode length	Release Date
0	Kaguya-sama wa Kokurasetai: Ultra Romantic	Comedy,Suspense	13	404	8.99	12	23 min. per ep.	Spring 2022
1	3-gatsu no Lion 2nd Season	Drama,Slice of Life	14	533	8.96	22	25 min. per ep.	Fall 2017
2	Gintama Movie 2: Kanketsu-hen - Yorozuya yo Ei	Action,Comedy,Sci-Fi	19	939	8.93	1	1 hr. 50 min.	NaN
3	Gintama.	Action,Comedy,Sci-Fi	12	735	8.99	12	24 min. per ep.	Winter 2017
4	Koe no Katachi	NaN	15	19	8.96	1	2 hr. 10 min.	NaN

Task 2: Navigating the Data Abyss.

As the data scientist delves deeper into the anime dataset, a new quest emerges—to uncover the enigmatic null values. With a few lines of code, the landscape of missing information begins to take shape. Each null value represents a mystery, a missing piece of the anime puzzle waiting to be solved. In this journey, the data virtuoso's mission is clear: to understand the gaps in the data, revealing where insights are yet to be found. The null values are not roadblocks; they are invitations to further exploration, guiding the way to a more complete understanding of the anime universe.

```
In [2]: # --- WRITE YOUR CODE FOR TASK 2 ---
        null_values = df.isnull().sum()
        #--- Inspect data ---
        null_values
Out[2]: Title
                          10327
        Genres
        Popularity
                              0
                           6324
         Score
        Episodes
                              9
        Episode length
        Release Date
                          15256
        dtype: int64
```

Task 3: Refining the Anime Constellation.

Out[3]:

In the quest to enhance the anime dataset, the data virtuoso works their magic. With astute data manipulation, missing genres, scores, and release dates are given new life. The anime landscape becomes more vibrant and cohesive as genres find their place, scores harmonize, and release dates align. Data is not just a static canvas; it's a living tapestry that gains depth and completeness under the skilled hand of the data scientist. With these transformations, the anime dataset is a step closer to revealing its full splendor, where every anime's essence shines with clarity and precision.

```
In [3]: #--- WRITE YOUR CODE FOR TASK 3 ---
df['Genres'].ffill(inplace = True)
df['Release Date'].ffill(inplace = True)
df['Score'].fillna(round(df['Score'].mean(), 2), inplace=True)
df.head()
#--- Inspect data ---
```

	Title	Genres	Rank	Popularity	Score	Episodes	Episode length	Release Date
0	Kaguya-sama wa Kokurasetai: Ultra Romantic	Comedy,Suspense	13	404	8.99	12	23 min. per ep.	Spring 2022
1	3-gatsu no Lion 2nd Season	Drama,Slice of Life	14	533	8.96	22	25 min. per ep.	Fall 2017
2	Gintama Movie 2: Kanketsu-hen - Yorozuya yo Ei	Action, Comedy, Sci-Fi	19	939	8.93	1	1 hr. 50 min.	Fall 2017
3	Gintama.	Action,Comedy,Sci-Fi	12	735	8.99	12	24 min. per ep.	Winter 2017
4	Koe no Katachi	Action,Comedy,Sci-Fi	15	19	8.96	1	2 hr. 10 min.	Winter 2017

Task 4: Preserving the Anime Chronicles.

With the final flourish, the data virtuoso encapsulates the enriched anime dataset into a new artifact - 'anime.csv.' This digital tome holds within it the culmination of exploration, analysis, and refinement. It's a testament to the dedication and expertise of the data scientist, ensuring that the insights and revelations are not lost but perpetuated for others to uncover. 'anime.csv' is more than a file; it's a time capsule, preserving the evolving tales of anime for future researchers, enthusiasts, and dreamers. The data scientist's work endures, a legacy that continues to inspire and guide those who seek to navigate the boundless world of anime

```
In [4]: # ...WRITE YOUR CODE FOR TASK 4 ...

#export the cleaned data
anime = df.to_csv(index = False)
anime
#--- Inspect data ---

Out[4]: 'Title, Genres, Rank, Popularity, Score, Episodes, Episode length, Release Date\nKaguya-sama wa Kokurasetai: Ultra Romantic, "Comedy,
Suspense", 13,404,8.99,12,23 min. per ep., Spring 2022\n3-gatsu no Lion 2nd Season, "Drama, Slice of Life", 14,533,8.96,22,25 min.
per ep., Fall 2017\nGintama Movie 2: Kanketsu-hen - Yorozuya yo Eien Nare, "Action, Comedy, Sci-Fi", 19,939,8.93,1,1 hr. 50 min., F
all 2017\nGintama., "Action, Comedy, Sci-Fi", 12,735,8.99,12,24 min. per ep., Winter 2017\nKoe no Katachi, "Action, Comedy, Sci-Fi", 1
5,19,8.96,1,2 hr. 10 min., Winter 2017\nMo Dao Zu Shi: Wanjie Pian, "Action, Drama, Supernatural", 49,2505,8.7,12,24 min. per ep.,
Winter 2017\nCode Geass: Hangyaku no Lelouch, "Action, Drama, Sci-Fi", 50,17,8.7,25,24 min. per ep., Fall 2006\nGintama, "Action, Co
medy, Sci-Fi", 16,127,8.95,201,24 min. per ep., Spring 2006\nGinga Eiyuu Densetsu, "Drama, Sci-Fi", 11,709,9.03,110,26 min. per e
p., Spring 2006\nVinland Saga, "Action, Adventure, Drama", 43,102,8.73,24,24 min. per ep., Spring 2006\nVinland Saga, "Action, Adventure, Drama", 43,102,8.73,24,24 min. per ep., Fall 2004\nMain fare properties, 111. Spring Song, "Action, Fantasy, Supernatural", 45,724,8.72,1,2 hr. 2 min., Fall 2000\nMain fare properties, 12,23 min. per ep., Fall 2021\nMain fare properties, 12,23 min. per ep., Fall 2021\nMain fare properties, 12,24 min. per ep., Fall 2021\nMain fare Story, "Drama
a, Romance, Supernatural", 18,105,8.94,24,24 min. per ep., Fall 2020\nMa
```

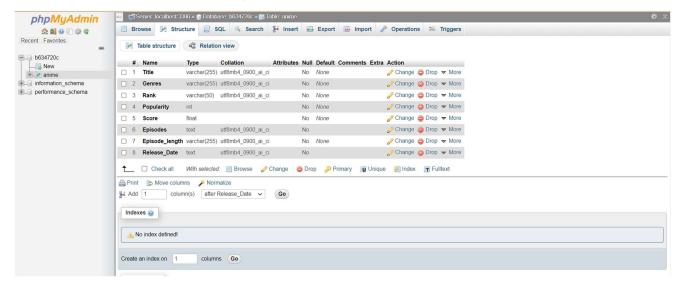
Task 5: Data Download, Import, and Database Connection.

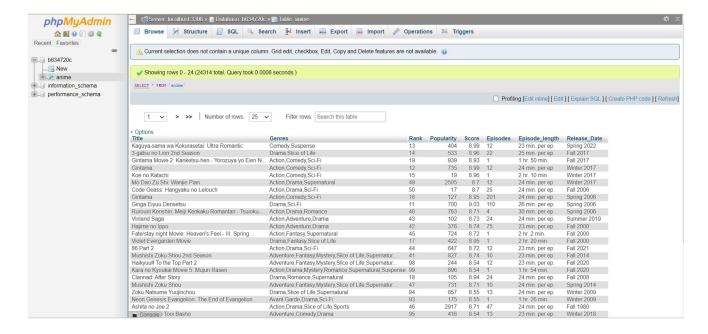
```
In [1]: # -- Load the sql extention ----
%load_ext sql

# --- Load your mysql db using credentials from the "DB" area ---
%sql mysql+pymysql://b634720c:Cab#22se@localhost/b634720c

Out[1]: 'Connected: b634720c@b634720c'
```

To connect with data, we have used phpMyAdmin web-based application with provided log-in credentials. Where I have created table in our database and imported data from csv file to 'anime' table.





Task 2: The Quest for Anime Excellence.

Our mission transcends data manipulation; it's a journey to honor anime's finest. Sorting titles, ranks, and scores showcases the pinnacle of creativity and storytelling. We do this to celebrate the shows that deeply connect with audiences, fostering inspiration and discussion. This task is about recognizing the exceptional artistry that defines anime, igniting a passionate conversation about the medium's masterpieces. As we unveil the top 10, we pay tribute to the creators who have crafted stories that touch hearts and transcend borders.

In [3]: %%sql
 SELECT Title, `Rank`, Score FROM anime ORDER BY Score DESC LIMIT 10;

* mysql+pymysql://b634720c:***@localhost/b634720c
10 rows affected.

Out[3]:

Title	Rank	Score
Fullmetal Alchemist: Brotherhood	1	9.14
Gintama°	5	9.08
Spy x Family	2	9.08
Shingeki no Kyojin Season 3 Part 2	3	9.08
Steins;Gate	4	9.08
Hunter x Hunter (2011)	8	9.05
Gintama: The Final	7	9.05
Gintama'	6	9.05
Fruits Basket: The Final	9	9.04
Gintama': Enchousen	10	9.04

Task 3: Spotlight on Anime Singularities.

This task isn't merely about sifting data; it's a pursuit to acknowledge anime's singular gems. We select titles, scores, and episode counts to uncover those one-episode wonders that have left a lasting impact. Our aim is to showcase the exceptional quality of anime that can be experienced in a single episode. By celebrating these unique creations, we aim to highlight the diversity and innovation within the medium. This endeavor invites exploration and appreciation of the concise storytelling and artistry that make these short stories stand out in the world of anime.

In [8]: %%sql
SELECT Title, Score, Episodes FROM anime WHERE Episodes = '1' ORDER BY Score DESC LIMIT 10;

* mysql+pymysql://b634720c:***@localhost/b634720c 10 rows affected.

Out[8]:

Title	Score	Episodes
Gintama: The Final	9.05	1
Koe no Katachi	8.96	1
Koe no Katachi	8.96	1
Koe no Katachi	8.96	1
Violet Evergarden Movie	8.95	1
Violet Evergarden Movie	8.95	1
Violet Evergarden Movie	8.95	1
Gintama Movie 2: Kanketsu-hen - Yorozuya yo Eien Nare	8.93	1
Gintama Movie 2: Kanketsu-hen - Yorozuya yo Eien Nare	8.93	1
Gintama Movie 2: Kanketsu-hen - Yorozuya yo Eien Nare	8.93	1

Task 4: Championing Multi-Episode Marvels.

Beyond data manipulation, this task celebrates the depth and brilliance of multi-episode anime. We select titles, scores, and episode counts to spotlight the series that have woven intricate narratives and left an enduring impact. Our goal is to recognize the artistry and storytelling prowess that thrives within longer anime series. This endeavor encourages us to appreciate the evolving characters, complex plots, and world-building that these multi-episode marvels offer. As we unveil the top 10, we invite enthusiasts to explore and engage with the expansive stories that have captivated audiences and contributed to the anime landscape's rich tabestry.

In [9]: %%sql SELECT Title, Score, Episodes FROM anime WHERE Episodes <> '1' ORDER BY Score DESC LIMIT 10;

Out[9]:

Title	Score	Episodes
Fullmetal Alchemist: Brotherhood	9.14	64
Steins;Gate	9.08	24
Gintama°	9.08	51
Shingeki no Kyojin Season 3 Part 2	9.08	10
Spy x Family	9.08	12
Gintama'	9.05	51
Hunter x Hunter (2011)	9.05	148
Fruits Basket: The Final	9.04	13
Gintama': Enchousen	9.04	13
Ginga Eiyuu Densetsu	9.03	110

Task 5: Exploring the Anime Frontrunner.

Our task transcends data manipulation; it's a voyage to uncover anime's most sought-after and highly rated gems. We handpick titles, popularity scores, and user ratings to spotlight the shows that have captured the hearts of audiences. This mission is about acknowledging the powerful connection between popularity and quality in the world of anime. We strive to showcase the series that have not only garnered a massive following but have also earned commendable scores. By unveiling the top 10, we invite fans and enthusiasts to delve into these remarkable stories, sparking discussions about what makes an anime truly exceptional.

In [10]: %%sql

SELECT Title, Popularity, Score FROM anime WHERE Score <> 0 ORDER BY Popularity DESC LIMIT 10;

* mysql+pymysql://b634720c:***@localhost/b634720c 10 rows affected.

Out[10]:

Title	Popularity	Score
Zuoshou Shanglan	20349	6.45
Zui Ke Ai De Ren (Movie)	20348	6.45
Zui Ke Ai De Ren	20347	6.45
Zui Ke Ai De Ren	20347	6.45
Zoku Zou no Senaka: Baton Touch	20346	6.45
Zhongkui Chuanqi: Sui Han Sanyou	20345	6.45
Zhonghua Xun Lixian Ji	20344	6.45
Zhonghua Xun Lixian Ji	20344	6.45
Zhonghua Meide	20343	6.45
Zhonghua Meide	20343	6.45

Task 6: Celebrating Short-Form Anime Brilliance.

Beyond data manipulation, our task is an homage to the captivating world of single-episode anime. We carefully select titles, considering both popularity and user ratings, to showcase the finest examples of concise storytelling. This endeavor shines a spotlight on the remarkable quality that these compact narratives can deliver. It's about recognizing the artistry and creativity that thrives within the constraints of a single episode. By unveiling the top 10, we invite fans to explore these bite-sized masterpieces and spark conversations about the art of telling a compelling story in a limited timeframe.

```
In [11]: 
%%sql
SELECT Title, Popularity, Score, Episodes
FROM anime
WHERE episodes > '1' AND score <> 0
ORDER BY popularity DESC
LIMIT 10;
```

* mysql+pymysql://b634720c:***@localhost/b634720c 10 rows affected.

Out[11]:

Title	Popularity	Score	Episodes
Zuoshou Shanglan	20349	6.45	13
Zui Ke Ai De Ren	20347	6.45	14
Zui Ke Ai De Ren	20347	6.45	14
Zhonghua Xun Lixian Ji	20344	6.45	26
Zhonghua Xun Lixian Ji	20344	6.45	26
Zhonghua Meide	20343	6.45	99
Zhonghua Meide	20343	6.45	99
Zhonghua Chuanshuo Meide Gushi	20342	6.45	39
Zhongguo Nanhai: Hong Zhanhui	20341	6.45	12
Zhongdian Fuzi	20340	6.45	26

Task 7: Elevating Multi-Episode Anime Excellence.

This task goes beyond data manipulation; it's a celebration of the depth and quality found in multi-episode anime. We meticulously curate titles, taking into account popularity and user ratings, to spotlight the cream of the crop. Our aim is to recognize the storytelling and artistic brilliance that thrives in longer anime series. This mission encourages viewers to explore the intricate plots, character development, and world-building that these multi-episode marvels offer. As we unveil the top 10, we invite enthusiasts to engage with these expansive narratives and ignite discussions about the diversity and innovation within the realm of longer anime series.

```
In [12]: %%sql
SELECT Title, Popularity, Score, Episodes FROM anime
WHERE Episodes > '1' and Score <> 0
ORDER BY Popularity DESC LIMIT 10;
```

* mysql+pymysql://b634720c:***@localhost/b634720c 10 rows affected.

Out[12]:

Title	Popularity	Score	Episodes
Zuoshou Shanglan	20349	6.45	13
Zui Ke Ai De Ren	20347	6.45	14
Zui Ke Ai De Ren	20347	6.45	14
Zhonghua Xun Lixian Ji	20344	6.45	26
Zhonghua Xun Lixian Ji	20344	6.45	26
Zhonghua Meide	20343	6.45	99
Zhonghua Meide	20343	6.45	99
Zhonghua Chuanshuo Meide Gushi	20342	6.45	39
Zhongguo Nanhai: Hong Zhanhui	20341	6.45	12
Zhongdian Fuzi	20340	6.45	26

Task 9: Exploring Comedy in Extended Anime Narratives.

Our task transcends data manipulation; it's a journey into the world of comedy in multi-episode anime. We meticulously select titles, scores, episodes, and genres, focusing on the laughter-inducing genre of comedy. This mission seeks to shine a spotlight on the laughter-rich narratives that evolve over multiple episodes. It's about celebrating the humor, character dynamics, and storytelling that flourish in longer anime series with a comedic touch. As we unveil the top 10, we invite fans to delve into these entertaining and laughter-infused stories, sparking discussions on the art of sustaining humor over extended narratives in the anime world.

In [14]:
%%sql
SELECT Title, Score, Episodes, Genres FROM anime
WHERE Episodes <> '1' and Genres LIKE '%Comedy%'
ORDER BY Score DESC LIMIT 10;

* mysql+pymysql://b634720c:***@localhost/b634720c 10 rows affected.

Out[14]:

Title	Score	Episodes	Genres
Gintama°	9.08	51	Action,Comedy,Sci-Fi
Spy x Family	9.08	12	Action,Comedy
Gintama'	9.05	51	Action,Comedy,Sci-Fi
Gintama': Enchousen	9.04	13	Action, Comedy, Sci-Fi
Gintama.	8.99	12	Action,Comedy,Sci-Fi
Kaguya-sama wa Kokurasetai: Ultra Romantic	8.99	12	Comedy,Suspense
Gintama.	8.99	12	Action,Comedy,Sci-Fi
Kaguya-sama wa Kokurasetai: Ultra Romantic	8.99	12	Comedy,Suspense
Kaguya-sama wa Kokurasetai: Ultra Romantic	8.99	12	Comedy,Suspense
Gintama.	8.99	12	Action,Comedy,Sci-Fi

Task 10: Quest for Anime's Pinnacle.

Beyond data manipulation, our mission is a journey to uncover the highest-rated anime. We select titles, focusing on the top scores, to identify the crown jewel of the anime world. This task is about recognizing and celebrating the artistry and storytelling that have earned the highest acclaim from audiences. It's an opportunity to shed light on the creations that have touched hearts and set a standard for excellence. By unveiling the top-rated anime, we invite fans to explore these exceptional stories, sparking conversations about what makes these titles truly exceptional.

```
In [15]: %%sql
SELECT Title, Score AS top_score FROM anime
WHERE Score = (SELECT MAX(Score) FROM anime);

* mysql+pymysql://b634720c:***@localhost/b634720c
1 rows affected.
```

Out[15]:

Title top_score
Fullmetal Alchemist: Brotherhood 9.14

Task 11: Unveiling Anime's Hidden Gems.

Our mission goes beyond data manipulation; it's a quest to reveal the anime world's lesser-known treasures. We select titles with the lowest scores, acknowledging that even amidst the diversity of anime, there are unique and underrated gems. This task is about recognizing the creativity and efforts put into these stories, often overshadowed by more prominent titles. It's an opportunity to shine a light on these hidden narratives that hold their own charm and appeal. By unveiling the lowest-rated anime, we invite fans to explore these underappreciated stories, sparking discussions about the diverse spectrum of anime and what makes each title special in its own way.

```
In [16]: %%sql
SELECT Title, Score as lowest_score FROM anime
WHERE Score = (SELECT MIN(Score) FROM anime WHERE Score > 0 );

* mysql+pymysql://b634720c:***@localhost/b634720c
1 rows affected.
```

Out[16]:

Title lowest_score

Tenkuu Danzai Skelter+Heaven 1.84

Task 12: Unraveling the Tapestry of Multi-Episode Anime.

Our task extends beyond data manipulation; it's a voyage into the intricate web of multi-episode anime. We gather information on episode counts and the number of titles to understand the landscape of longer narratives. This mission is about recognizing the diversity in episode lengths and the preferences of anime creators and audiences. It's an opportunity to unravel the patterns that emerge from this analysis, shedding light on the rich tapestry of multi-episode anime. By showcasing the top 10 episode counts, we invite enthusiasts to explore the variations in storytelling lengths, sparking discussions about the depth and breadth of narrative possibilities in the anime world.

```
In [17]: %%sql
           SELECT Episodes, COUNT(Title) as no_of_animes FROM anime WHERE Episodes <> '1' and Episodes <> 'Unknown' GROUP BY Episodes ORDER BY COUNT(Title) DESC
            LIMIT 10;
             * mysql+pymysql://b634720c:***@localhost/b634720c
            10 rows affected.
Out[17]: Episodes no_of_animes
             12 2035
                     2
                                  1614
                   13
                                 1100
                    26
                                  1009
                    3
                                   766
                     4
                                   580
                   52
                                   509
                     6
                                   447
                   24
                                   360
                    10
                                   321
```