Unveiling the Stars: An Exploratory Study on NASA Astronauts

In this project, you'll analyze NASA astronauts using the provided dataset. Your task involves data cleaning, exploration, and information extraction using Python and MySQL. The dataset includes astronaut details like names, birthdates, gender, education, military history, space missions, and more. Your goal is to uncover insights into astronauts' demographics, careers, and achievements through descriptive statistics and data visualization.

Project Description

NASA astronaut exploration is a captivating journey into space, fueled by individual stories and collective triumphs. These astronauts, with diverse backgrounds and expertise, embody years of dedication and knowledge pursuit. Analyzing astronaut data reveals the rich tapestry of NASA's corps, and your role as a data virtuoso is pivotal. You uncover trends and patterns, shedding light on educational journeys, military service, and space achievements. Your work inspires researchers and dreamers, making astronaut stories more extraordinary and contributing to the saga of human exploration.

Task 1: Exploring NASA's Data Universe.

Our analysis of NASA's dataset is a mission to unveil profound insights within the realm of space exploration. Beyond mere data analysis, it's a journey to harness the knowledge hidden in the stars. Through this exploration, we aim to uncover patterns that will guide future missions, enhancing NASA's cosmic endeavors. Every data point is a piece of the cosmic puzzle, fueling innovation and inspiring generations to reach for the stars. In these numbers and statistics, we find the roadmap to the next frontier of human exploration.

Task 2: Exploring Data Completeness.

In the midst of your data exploration, armed with the NASA astronaut dataset, you've arrived at a crucial juncture. You're now delving into the realm of data quality, seeking to understand the integrity of the information you're working with. With a few lines of code, you've calculated the number of missing values for each column in the dataset, and the results are illuminating.

Task 3: Data Refinement for NASA Astronaut Data.

In your ongoing journey through the NASA astronaut dataset, you've embarked on a quest for data integrity and precision. The code you've executed represents a pivotal moment in this endeavor, where you're not just analyzing data but also sculpting it to reveal the true narrative of these space pioneers.

Task 4: Preparing Data for SQL Analysis.

In the midst of your data journey through the NASA astronaut dataset, you've arrived at a crucial juncture where your focus shifts to data export and preparation for a new phase in your analysis.

The destination is clear: a CSV file named 'astronauts.csv' that will serve as the foundation for your SQL exploration. This export step ensures that the data you've curated and cleaned is ready to be loaded into a relational database, where you can perform more intricate queries and derive deeper insights.

Write SQL Queries

1. Retrieving All Data.

Select all columns in the table astronauts.

Use the SQL SELECT statement to retrieve all records denoted by \*.

Specify the table to query, which is named 'astronauts'.

1. Counting Astronauts by Status.
2. Find out the status of the astronaut.
3. Use the SQL SELECT statement to retrieve specific columns: 'Status' and the count of records denoted as 'Number'.
4. Specify the table to query, which is named 'astronauts'.
5. Group the results by the 'Status' column using the GROUP BY clause.
6. Counting Astronauts by Military Branch.
7. Find out the Military Branch of the astronaut.
8. Use the SQL SELECT statement to retrieve specific columns: 'Military\_Branch' and the count of records denoted as 'Number'.
9. Specify the table to query, which is named 'astronauts'.
10. Group the results by the 'Military\_Branch' column using the GROUP BY clause.
11. Top 5 Military Ranks Among Astronauts.
12. Find out the top 5 military ranks of the astronaut.
13. Use the SQL SELECT statement to retrieve specific columns: 'Military\_Rank' and the count of records denoted as 'Number'.
14. Specify the table to query, which is named 'astronauts'.
15. Group the results by the 'Military\_Rank' column using the GROUP BY clause.
16. Sort the results in descending order based on the 'Number' column using the ORDER BY clause.
17. Apply a limit to the results, showing only the top 5 rows using the LIMIT clause.
18. Count of Astronauts by Gender.
19. Calculate the number of male and female astronauts.
20. Use the SQL SELECT statement to retrieve specific columns: 'Gender' and the count of records denoted as 'Number'.
21. Specify the table to query, which is named 'astronauts'.
22. Group the results by the 'Gender' column using the GROUP BY clause.
23. Average Life Expectancy of Astronauts.
24. What is the average life expectancy of an astronaut?
25. Calculate the average life expectancy of astronauts.
26. Create a subquery that calculates the life expectancy for each astronaut using conditional logic based on their 'Status' (Deceased or not).
27. In the subquery:
28. -If the 'Status' is 'Deceased,' calculate life expectancy by subtracting the birth year from the death year.
29. -If the 'Status' is not 'Deceased,' calculate life expectancy by subtracting the birth year from the year 2023.
30. -Alias the calculated life expectancy column as 'life\_Expectancy' within the subquery.
31. Compute the average of these calculated life expectancies rounded to the nearest whole number.
32. The result is presented as 'Average\_Life\_Expectancy'.
33. Hint: To extract the year from a date format, you can use the YEAR() method
34. Average Life Expectancy of Female Astronauts.
35. What is the average life expectancy of a female astronaut?
36. Calculate the average life expectancy of female astronauts.
37. Create a subquery that calculates the life expectancy for each female astronaut using conditional logic based on their 'Status' (Deceased or not) and 'Gender.'
38. In the subquery:
39. -If the 'Status' is 'Deceased' and 'Gender' is 'Female,' calculate life expectancy by subtracting the birth year from the death year.
40. -If the 'Status' is not 'Deceased' and 'Gender' is 'Female,' calculate life expectancy by subtracting the birth year from the year 2023.
41. -Alias the calculated female life expectancy column as 'Female\_life\_Expectancy' within the subquery.
42. Compute the average of these calculated female life expectancies rounded to the nearest whole number.
43. The result is presented as 'Female\_Average\_Life\_Expectancy.'
44. Hint: To extract the year from a date format, you can use the YEAR() method.
45. Average Life Expectancy of Male Astronauts.
46. What is the average life expectancy of a male astronaut?
47. Calculate the average life expectancy of male astronauts.
48. Create a subquery that calculates the life expectancy for each male astronaut using conditional logic based on their 'Status' (Deceased or not) and 'Gender.'
49. In the subquery:
50. -If the 'Status' is 'Deceased' and 'Gender' is 'Male,' calculate life expectancy by subtracting the birth year from the death year.
51. -If the 'Status' is not 'Deceased' and 'Gender' is 'Male,' calculate life expectancy by subtracting the birth year from the year 2023.
52. -Alias the calculated male life expectancy column as 'Male\_life\_Expectancy' within the subquery.
53. Compute the average of these calculated male life expectancies rounded to the nearest whole number.
54. The result is presented as 'Male\_Average\_Life\_Expectancy'.
55. Hint: To extract the year from a date format, you can use the YEAR() method.
56. Top 10 Graduate Majors Among Astronauts.
57. Find out the top 10 educational backgrounds (graduates) that astronauts have.
58. Use the SQL SELECT statement to retrieve specific columns: 'Graduate\_Major' column and calculates the count of records, denoted as 'Number'.
59. Specify the table to query, which is named 'astronauts'.
60. Group the results by the 'Graduate\_Major' column using the GROUP BY clause.
61. Sort the results in descending order based on the count using the ORDER BY clause.
62. Apply a limit to the results, showing only the top 10 'Graduate\_Major' categories with the highest counts.
63. Astronaut Education Statistics.
64. Calculate the count of astronauts with undergraduate and graduate degrees.
65. Calculate the total count of astronauts and alias it as 'Number\_of\_Astronauts.'
66. Calculate the count of astronauts with undergraduate degrees and alias it as 'Astronauts\_with\_Undergraduate\_Degrees.'
67. Calculate the count of astronauts with graduate degrees and alias it as 'Astronauts\_with\_Graduate\_Degrees.'
68. Retrieve this information from the 'astronauts' table.
69. Top 5 States of Birth for Astronauts.
70. Find out the top 5 countries/states from which astronauts come.
71. Extract the state name from the 'Birth\_Place' column and alias it as 'STATE.'
72. Calculate the count of astronauts born in each state.
73. Retrieve this information from the 'astronauts' table.
74. Group the results by the 'STATE' column.
75. Sort the results in descending order based on the count.
76. Display only the top 5 states with the highest counts of astronauts.
77. Hint: To extract the state from the 'Birth\_Place' column, you can use SUBSTRING\_INDEX(Birth\_Place, ',', -1) to get the last part of the 'Birth\_Place' string, which is the state. Please note that there should be no space after the delimiter (',').