 **📚 Lesson 04**  | Aggregating Data

[Aggregations](#xnsi1outhwno)

[GROUP BY](#1q7592ixm6dr)

[ALIASES & AS](#ci0zwilamvaq)

 **Aggregations Practice**

— Aggregating Data

PROMPT: Let’s practice using your newly-learned aggregation functions on a variety of datasets.

1. How many titanic passengers are recorded in the **titanic\_passengers** table?

|  |
| --- |
| select count(\*)  from titanic\_passengers |

|  |
| --- |
| 1309 |

1. How long would it take to go on every ride in the **disney\_rides** data? Hint: Look at the ride\_duration column.

|  |
| --- |
| select sum(ride\_duration)  from disney\_rides |

|  |
| --- |
| 281.75 |

1. What are the minimum, maximum, and average of the quoted list price of all the cars in the **sales\_data\_cars** dataset? Is the average closer to the minimum price or the maximum price?

|  |
| --- |
| select  min(quoted\_list\_price),  max(quoted\_list\_price),  avg(quoted\_list\_price)  from sales\_data\_cars |

|  |
| --- |
| Min: 4300 Max: 58800 Avg: 16692.9976 |

 **GROUP BY Practice**

— Aggregating Data

PROMPT: Grouping and aggregation go hand in hand with one another. Practice using the GROUP BY keyword with these questions on the Lyft Bay Wheels bikeshare dataset (**lyft\_baywheels**).

1. Warm up: What is the total number (count) of bike rides recorded in the table?

|  |
| --- |
| select count(\*)  from lyft\_baywheels |

|  |
| --- |
| 999385 |

1. Group the count of trips by the type of rider (member\_casual). Are more rides made by system members, or by casual users making standalone transactions?

|  |
| --- |
| SELECT  member\_casual,  count(\*)  FROM lyft\_baywheels  GROUP BY 1 |

|  |
| --- |
| casual 589619  member 409766 |

1. Group the count of trips by the type of bike used (**rideable\_type**). Are traditional dock-based bikes more popular, or are the newer electric bikes more popular?

|  |
| --- |
| SELECT  rideable\_type,  count(\*)  FROM lyft\_baywheels  GROUP BY 1 |

|  |
| --- |
| docked\_bike 309286  electric\_bike 690099 |

1. Group the trip count by both the user type and bike type. Is the ratio of electric vs. traditional bike usage for each individual type of rider similar to the overall usage pattern?

|  |
| --- |
| SELECT  member\_casual,  rideable\_type,  count(\*)  FROM lyft\_baywheels  GROUP BY 1,2 |

|  |
| --- |
| casual docked\_bike 158018  casual electric\_bike 431601  member docked\_bike 151268  member electric\_bike 258498 |

 **Aliases and AS Practice**

— Aggregating Data

PROMPT: Let’s practice with a few more examples of adding aliases to queries to demonstrate how they can clarify your queries, using the AirBnB listings dataset (**airbnb\_listings**).

1. Write a query that returns the average and highest price of any listing, grouped by city. Is the city with the highest individual listing price the same as the one with the highest average price?

|  |
| --- |
| select  city,  avg(price),  max(price)  from airbnb\_listings  group by 1 |

|  |
| --- |
| Highest individual listing city: San Mateo County  Highest avg listing: Clark County |

1. Add the average number of reviews (at time of listing) to the query output. Do cities with more reviews per listing tend to have higher or lower prices? HINT: Use the DB Browser’s built-in plot creator to build a scatter plot to help you analyze the data!

|  |
| --- |
| select  city,  avg(price),  max(price),  avg(number\_of\_reviews)  from airbnb\_listings  group by 1  order by 4 DESC |

|  |
| --- |
|  |

|  |
| --- |
|  |