7. Summary

Contents

[1. Summary 1](#_Toc2374)

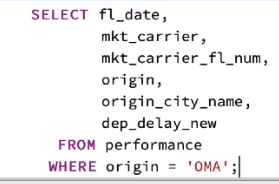
# 1. Summary

=>slides: Pg. 1

In this module, we'll walk together through a simple business case that illustrates how to use the skills that you've learned in this course.

=>slides: Pg. 2

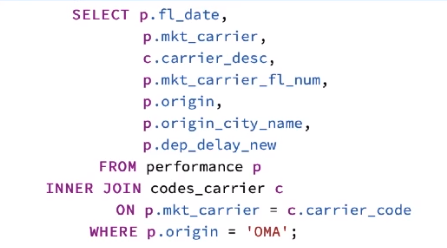
Let's say that you are the manager of the Omaha, Nebraska airport and that you want to better understand departure delays at your airport.



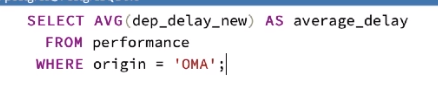
First, let's look at the performance table, which contains data about all commercial flights conducted by US carriers in the country during the month of January. To see all fields for flights where Omaha was the departure city, or city of origin, we can write this query using the airport code for Omaha, OMA. Since we know we are interested primarily in delay information, we'll also specify the columns we want our query to return. When we run this query, Postgres returns information for 2061 flights in our database that originated in Omaha.

=>slides: Pg. 2

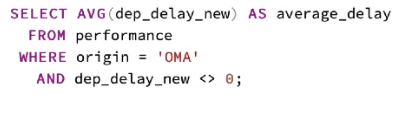
We can also see from the data that multiple airlines operate flights from Omaha. Let's join our data to the codes\_carrier lookup table to help us interpret the market carrier code.



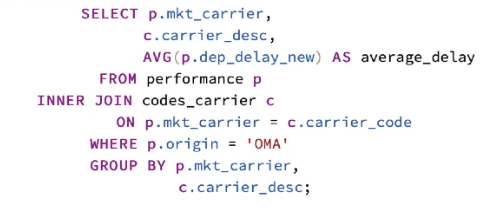
We simply alias the performance table as p and inner joined the codes\_carrier table based upon the common carrier code. We add the carrier\_desc field from the codes\_carrier table to our result set. When we run this query, we see the results that are identical to our first query, however, the market carrier code now is accompanied by an additional column that specifies the air carrier name. This is good information, but we may want to delve further into airline performance.



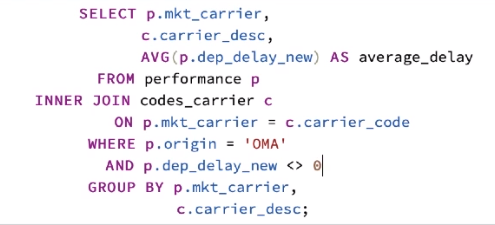
A useful metric would be the average delay for flights departing the Omaha airport. We can find that quite simply by using the following query. The result tells us that the average delay for flights departing the Omaha airport was an average delay of just over 13 minutes. The dep\_delay column indicates 0 for all flights that do not experience a departure delay. Maybe we are curious how long the average delay is when a flight is delayed. The flights that take off on time naturally lower the average delay.



We can expand our query to exclude those flights that did not experience a delay. Note that we did not use the special IS NOT NULL keyword here because 0 is in fact a value. When we run this query, we find that when a departure delay does occur, it averages about 49 minutes in duration. Another great question might be to explore the delay occurrences for each airline that operates at the airport. To do this, we'll add the carrier field from the performance table and return to an earlier inner join on the codes\_carrier table.



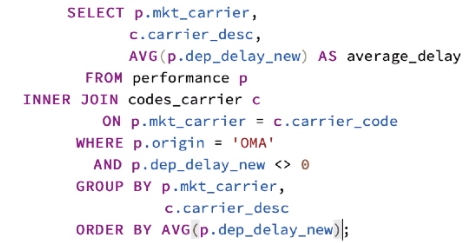
This query returns a listing of airlines that operate from the Omaha airport, along with their average overall delay.



Again, we can exclude those flights that depart on time to better understand the average length of delay, only for flights where a delay occurs. We do this by adding the WHERE clause.

=>slides: Pg. 2

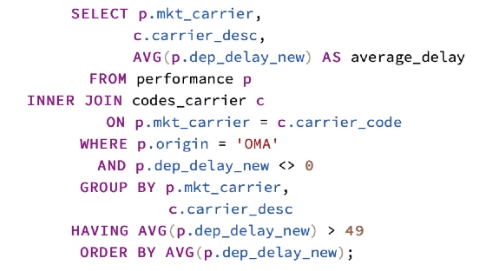
We can also add an ORDER BY clause to our query to return the results ordered in ascending order of delay length.



When we run this query, we can see that Southwest Airlines experienced the shortest average delay while United Airlines experienced the longest average delay.

=>slides: Pg. 2

Remember from our earlier query, the average length of delay in Omaha when a delay occurred was about 49 minutes.



We can add a HAVING clause to our query to limit the results of our aggregate function, average, to only those that exceed 49 minutes in length. Note that the HAVING clause must always immediately follow the GROUP BY clause. Running this query, we can see that American, Delta, and United have delays exceeding the average of 49 minutes when a delay occurs. This can help us understand what airlines we may want to focus on when addressing or further exploring delays at the Omaha airport.

=>slides: Pg. 3

I hope that this course has provided you with a solid starting point for building your own queries to select data from SQL, to limit your results, and to aggregate information.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*