 **🧪 (Optional) LevelUp Lab** | Construction Job Demand

**INTRODUCTION:** This is a lab for students looking to level up their SQL skills with subqueries! Subqueries allow us to write efficient queries and to perform more meaningful analysis, highlighting to employers a deeper understanding of the nuances of SQL.

In this lab, you’ll be using subqueries to analyze data regarding construction jobs and the role weather might play in driving future demand. Companies in this space need to be able to assess what future demand will look like in order to allocate resources and teams accordingly. Sending a small team to a location may mean potential customers will go with a competitor because they can get the job done in a shorter time frame due to your limited resources.

**HOW IT WORKS:** Follow the prompts in the questions below to investigate your data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write, and **blue boxes** for text-based answers.

PROMPT: Congratulations, you’ve landed an internship on the Data Analysis Team at Hover. Your manager is interested in the demand of jobs. She has a hunch that severe weather events have an effect on the number of job requests, but she’d like you to help her prove her hunch is true with data.

**—** Data Set **Description**

The data needed for this lab (hover) comes from two different sources.

The first data source (hover\_jobs) consists of historical data from a software solution provider for construction companies:

* **job\_identifier** - The unique ID for a job
* **organization\_id** - ID for the account that purchased the job
* **job\_location\_city** - The city where the job was located
* **job\_location\_region\_code** - The state where the job was located
* **job\_first\_upload\_complete\_datetime** - The date on which the customer uploaded photos for the job.
* **job\_deliverable** - the type of job that the customer requested. Either complete (full building) or roof.

The second dataset (hover\_weather) comes from the [NOAA Storm Prediction Center](https://www.spc.noaa.gov/climo/reports/) and catalogs adverse weather events across the United States:

* **comments** - A description of the weather event
* **county** - The county of the weather event
* **state** - The state of the weather event
* **location** - The address of the weather event
* **longitude** - The longitude of the weather event
* **latitude** - The latitude of the weather event
* **datetime** - The datetime of the weather event
* **composite\_key** - Unique key for the weather event, combining the timestamp, longitude, and latitude of the event

**— Task 1:** Explore the jobs data.

Your manager would like to find out more information regarding when customers make requests.

1. Write a query that returns the total number of jobs at the monthly level for each year (i.e. Sept 2016, Oct 2016, etc.). You’ll need to use the strftime function in SQL with the **job\_first\_upload\_complete\_datetime** variable. When done correctly, your output will have 31 rows.

(paste your query below 👇)

|  |
| --- |
| SELECT  strftime('%Y',job\_first\_upload\_complete\_datetime) as year,  strftime('%m',job\_first\_upload\_complete\_datetime) as month,  count(job\_identifier) as num\_jobs  from hover\_jobs  GROUP BY 1,2  ORDER BY 1,2 |

1. Use the SQL app’s built-in visualizer to graph the task in part A. Is there any seasonality to the job requests? Seasonality can be described as a pattern that repeats itself over cycles of time, such as yearly. Is there a season that has more requests than others?

|  |
| --- |
|  |

(write your **answer** below 👇)

|  |
| --- |
| Summer months. |

**— Task 2:** Explore the weather data.

In Task 1, you discovered that there are seasonal patterns to when customers submit job requests. Now you’ll investigate the weather data and see how that can help you further with your analysis.

1. The entire weather dataset consists of “adverse weather events”, e.g., tornados, fallen trees, sustained high gusts of wind, etc. Write a query that counts the number of adverse weather events for each month and year of the data (i.e. Sept 2016, Oct 2016, etc.).

(paste your query below 👇)

|  |
| --- |
| select  strftime('%Y', datetime) AS year,  strftime('%m', datetime) AS month,  count(\*)  from hover\_weather  group by 1,2  order by 1,2 |

1. Modify your query in 2A to filter out any information prior to September 2016 (the start of the job request data). Visualize the data using the built-in visualizer.

(paste your query below 👇)

|  |
| --- |
| select  strftime('%Y', datetime) AS year,  strftime('%m', datetime) AS month,  count(\*)  from hover\_weather  where datetime >= '2016-09-01'  group by 1,2  order by 1,2 |

|  |
| --- |
|  |

1. The weather data includes values for all 50 states. Modify your query once more so that it only shows information from the states that are seen in the jobs data. Visualize the filtered data again with the built-in visualizer and compare the filtered data to the original data from 2B. Did the filtering change anything about the pattern? Note: Your query should **NOT** use a JOIN clause, but instead use a **subquery**!.

(paste your query below 👇)

|  |
| --- |
| select  strftime('%Y', datetime) AS year,  strftime('%m', datetime) AS month,  count(\*)  from hover\_weather  where datetime >= '2016-09-01' and state in  (select distinct job\_location\_region\_code as state from hover\_jobs)  group by 1,2  order by 1,2 |

|  |
| --- |
|  |

1. Write a few-sentence summary describing the relationship between the job requests and weather events you observed in Tasks 1 and 2.

(write your **answer** below 👇)

|  |
| --- |
| The plots are very similar. |

**— Task 3:** Does weather affect demand for jobs?

Given the perceived relationship between the weather event graph and job requests graph, you will now investigate if there is indeed a relationship between adverse weather events and job requests.

A colleague of yours already wrote a query that returns the total number of weather events grouped at the weekly level for each state in the weather data. You’ll build off their work.

1. Write a query that performs a JOIN on the hover\_jobs table and the hover\_weekly\_weather\_events table your colleague created. Since you are only interested in the weeks that have had an adverse weather event, use an INNER JOIN to match the tables on both the week timestamp AND the state. If you don’t do this, you will end up with output crossing states, e.g. requests in TX incorrectly joined to events in KS.

SELECT the following columns from the jobs table in your query:

* **job\_deliverable**,
* **job\_location\_region\_code**,
* **job\_first\_upload\_complete\_datetime**, truncated down to the ‘week’ level. Alias this column as **job\_ts**.

And from the hover\_weekly\_weather\_events table, you will need the following:

* **n\_weather\_events**

Remember to use the function and NOT the alias job\_ts when joining to the weather\_ts column, otherwise you will get an error. If done correctly, your output table will have 25,257 rows.

(paste your query below 👇)

|  |
| --- |
| select  job\_deliverable,  job\_location\_region\_code,  strftime('%W', job\_first\_upload\_complete\_datetime) as job\_ts,  n\_weather\_events  from hover\_jobs hj  join hover\_weekly\_weather\_events he  on strftime('%Y', weather\_ts) = strftime('%Y', job\_first\_upload\_complete\_datetime)  and strftime('%W', weather\_ts) = strftime('%W', job\_first\_upload\_complete\_datetime)  and he.state = hj.job\_location\_region\_code  where job\_first\_upload\_complete\_datetime >= '2016-09-01' |

1. Use your query in 3A as a subquery in a new query that counts the total number of jobs and the total number of weather events for each state and week. The variable job\_ts can be used to count the total number of jobs. Order your output alphabetically by state. If done correctly, your output table should have 320 rows.

(paste your query below 👇)

|  |
| --- |
| with weather\_events as (  select  job\_deliverable,  job\_location\_region\_code as state,  strftime('%Y', job\_first\_upload\_complete\_datetime) as year,  strftime('%W', job\_first\_upload\_complete\_datetime) as job\_ts,  n\_weather\_events  from hover\_jobs hj  join hover\_weekly\_weather\_events he  on strftime('%Y', weather\_ts) = strftime('%Y', job\_first\_upload\_complete\_datetime)  and strftime('%W', weather\_ts) = strftime('%W', job\_first\_upload\_complete\_datetime)  and he.state = hj.job\_location\_region\_code  where job\_first\_upload\_complete\_datetime >= '2016-09-01'  )  select  state,  year,  job\_ts,  count(job\_ts) as num\_of\_jobs,  sum(n\_weather\_events)  from weather\_events  group by 1,2,3 |

1. Now you are ready to determine whether there is a relationship between adverse weather events and job requests. Using the built-in visualizer, create a scatterplot of your data. On your x-axis you should have the total number of weather events and on the y-axis you should have the total number of job requests. Is there a relationship between adverse weather events and job requests? If so, what kind of relationship is it?

|  |
| --- |
|  |

(write your **answer** below 👇)

|  |
| --- |
| There is a linear relationship between job requests and weather events with a slightly positive correlation. |