

## Day 15: SQL I: Select and Filter

After taking a much-needed break in the New Zealand countryside, Dot's time deep in the Southern Hemisphere has come to a close. They said their goodbyes to the Tasman Sea and the Indian Ocean and boarded a plane heading north and west. As they settled into their seat on the airplane, a woman seated next to them turned to them with a curious expression. "Is it my overactive imagination getting me confused again, or have I seen you before?" Dot looked at the woman keenly, then snapped their fingers. "We were on the same plane from Toronto to Munich!" Dot exclaimed. "Oh yes, the youngin I bumped into," the odd woman murmured sadly. "My apologies for that; I hope you weren't too inconvenienced by my clumsy intrusion!" Dot explained that it hadn't been a problem and had only caused a tiny coffee spill. "A coffee spill!" the woman shrieked. "You could've been burnt to the bone! No, no, no, that won't do at all. I'll have to make it up to you. Are you headed to Nepal?" Dot nodded their head in astonished agreement. "Well, then," the woman continued, "You must tag along with me on an adventure."

The woman nodded her head up and down with a satisfied expression. With quick movements, she pulled out a roll of thick wool and two large sticks and began to aggressively crochet. Dot cleared their voice to get the woman's attention, but she couldn't hear them over her crochet needles' loud clicking and clacking. Dot tapped her on the shoulder, and the woman swung her head towards them with a broad smile. "What kind of adventure do you mean?" Dot asked cautiously. "Why, the only adventure worth undertaking — an adventure primordial, epic, emotive, and character-building. We're going to be climbing the highest peaks of the Himalayas!" Dot paled while the woman cackled with glee. "Oh, no," Dot stuttered. "There's no way I can do that!" The woman tutted. "Nonsense! I will cover all expenses. You simply must come along with me. You're my honorary sidekick!" Dot gulped. With shaking hands, they pulled out their laptop and connected to the plane's Wi-Fi, which, oddly, worked perfectly. They need to know what exactly they're getting into. Can you help them figure out which of the Himalayan peaks are the tallest ones?

## Tutorial

During the previous challenges, we saw how Python and its library pandas can be used for data analytics and data science.

But the truth is, no matter what programming language we choose, it'll be useful to know SQL as well. SQL is a domain-specific database used to access databases, which is where data usually is.

### What is SQL?

- SQL stands for Structured Query Language
- SQL lets you access and manipulate databases
- SQL became a standard of the American National Standards Institute (ANSI) in 1986, and of the International Organization for Standardization (ISO) in 1987

### What can we do with SQL?

- we can execute queries against a database
- we can retrieve data from a database
- we can insert records in a database
- we can update records in a database
- we can delete records from a database
- we can create new databases
- we can create new tables in a database
- we can create stored procedures in a database
- we can create views in a database
- we can set permissions on tables, procedures, and views

*This short SQL summary was based on [w3schools](#)*

Python is NOT a replacement for SQL. The best data scientists and analysts leverage both of them to be as efficient as possible. That's why we will now learn the basics of SQL in the next two challenges.

## SQLite

We will start with **SQLite** during these challenges. SQLite is a great database for educational purposes because it is completely stored in one file on our local machine.

We can connect to the existing SQLite database using the command below:

```
import sqlite3
conn = sqlite3.connect("himalayas.db")
```

Now, we can use the connection **conn** to access the data in the database directly from pandas.

```
import pandas as pd
peaks = pd.read_sql('select * from peaks', conn)
```

Pandas's **.read\_sql()** function has two required parameters, sql and connection. The first argument is string and it can be any SQL query. Let's dive into the basics of SQL queries.

The most basic form of the SQL SELECT statement must include the SELECT and FROM keywords.

- SELECT - defines what we want to take from the source
- FROM - defines our data source (table)

```
pd.read_sql("""
    SELECT fname,
           lname,
           calcage,
           sex
    FROM members
    """, conn)
```

In addition, if we want to filter the result set of the query, we should use the WHERE keyword

```
pd.read_sql("""
    SELECT fname,
```

```
        lname,  
        calcage,  
        sex  
    FROM members  
    WHERE calcage >= 40  
    """  
    ,conn)
```

We can use more than one filter:

```
pd.read_sql("""  
    SELECT fname,  
           lname,  
           calcage,  
           sex  
    FROM members  
    WHERE calcage >= 40  
          AND sex = 'M'  
          AND fname = 'Eric'  
    """,  
    ,conn)
```

To further play around with simple SQL queries, you can watch this [interactive video](#).

## Challenge

Use our database to answer the following questions:

1. **How many peaks are higher than 8000 metres?** Use the `len()` function to count the number of rows of the DataFrame that were returned by the `pd.read_sql()` function.
2. **How many women were part of the expeditions?**

```
In [1]: import sqlite3  
        conn = sqlite3.connect("himalayas.db")  
        import pandas as pd
```

```
In [2]: # SOLUTION  
  
        #1.  
        len(pd.read_sql('select * from peaks where heightm > 8000',conn))
```

```
Out[2]: 16
```

```
In [3]: # 2.  
        len(pd.read_sql("select * from members where sex = 'F'",conn))
```

```
Out[3]: 7254
```