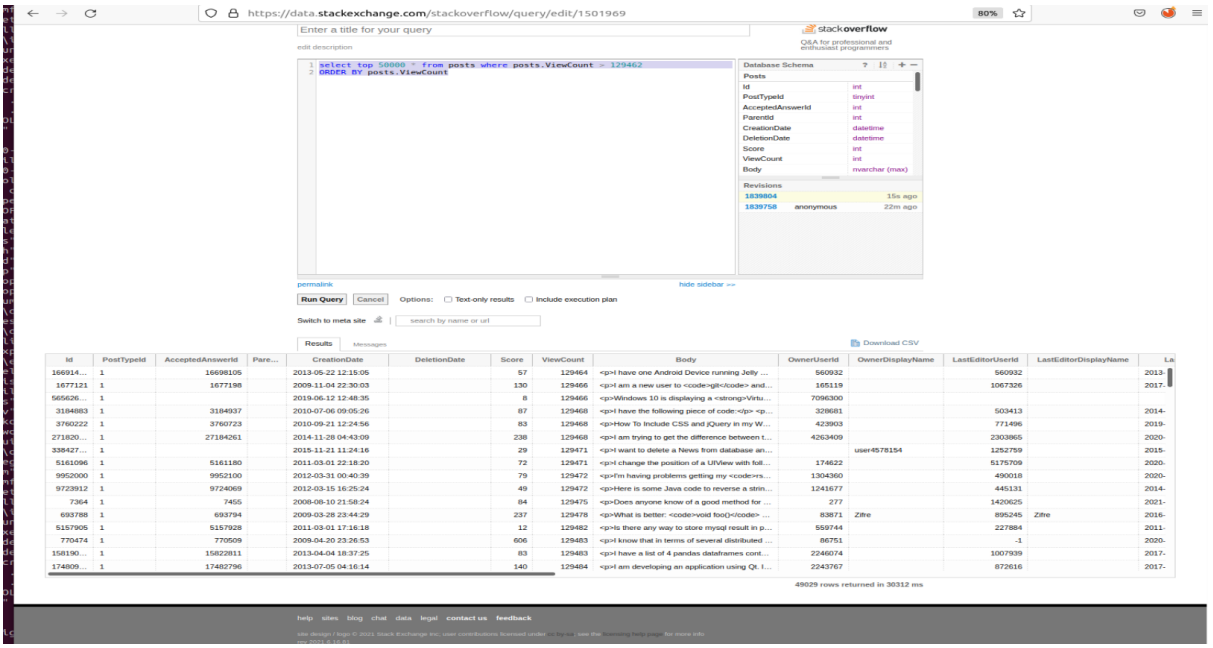
Supporting Documentation: Task 1
Student: Úna McGinn
CA675 Assignment 1

Following trial and error in an attempt to obtain the top close to 50,000 records, the following query was tried.

```
select top 50000 * from posts where posts.ViewCount > 129462
ORDER BY posts.ViewCount
```



From the above, I can see that the lowest ViewCount in the dataset is 129464. Therefore, for the 1st dataset I will select the records with Viewcount > 129464 to remove the need to account for the possibility that there may be multiple videos with 129464 views.

Dataset Part 1:

```
select top 50000 * from posts where posts.ViewCount > 129464
ORDER BY posts.ViewCount
```

Result: 49028 rows

Dataset Part 2:

select top 50000 * from posts where posts.ViewCount <= 129464 and posts.ViewCount > 77000
ORDER BY posts.ViewCount
Result: 47385 rows

From the above, I can see that the lowest ViewCount in the dataset is 77001. There are less than 50,000 records in the dataset so all values in this range are in the dataset.

Dataset Part 3:

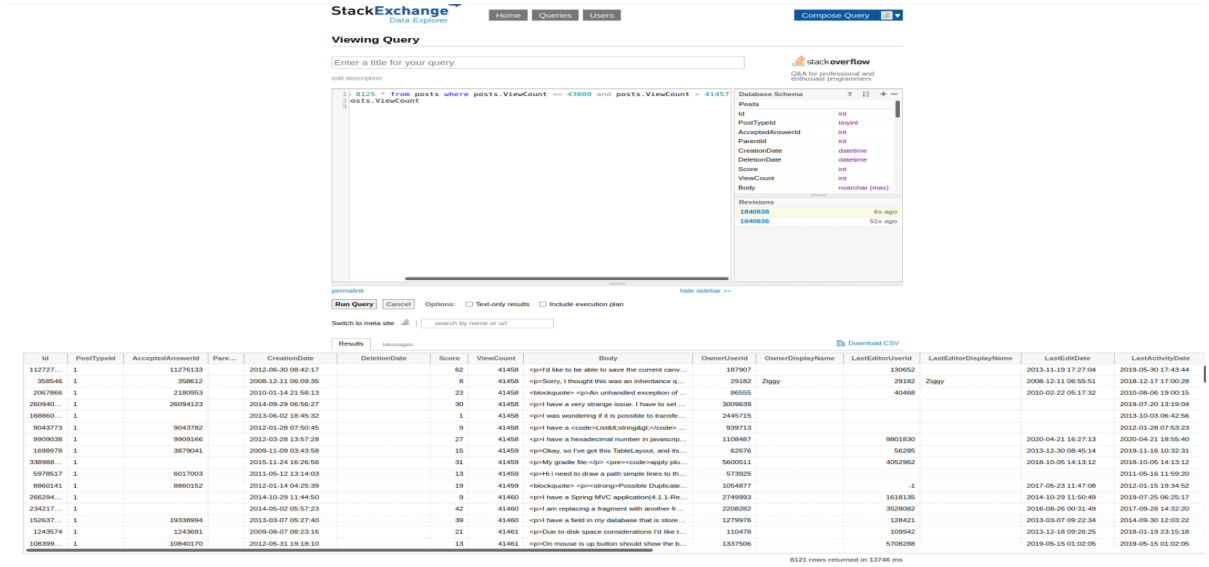
select top 50000 * from posts where posts.ViewCount <= 77000 and posts.ViewCount > 55500
ORDER BY posts.ViewCount

Result: 46963 rows

Dataset Part 4:
select top 50000 * from posts where posts.ViewCount <= 55500 and posts.ViewCount > 43000
ORDER BY posts.ViewCount

Result: 48499 rows

Dataset Part 5:
Tried the following: select top 8125 * from posts where posts.ViewCount <= 43000 and posts.ViewCount > 35000
ORDER BY posts.ViewCount



By trial and error, used the following interval to obtain the next 8125 records.

select top 8125 * from posts where posts.ViewCount <= 43000 and posts.ViewCount > 41457
ORDER BY posts.ViewCount

Result: 8125 rows

For the final subset I take the top 8125 records i.e.(the number of records required to make up 200,000 records). These 5 batches have 49028, 47385, 46963, 48499 and 8125 records respectively, making up the total dataset of 200,000.