

# CAREER TRACK ANALYSIS WITH SQL AND TABLEAU PROJECT

By Luis Bernardo Davila Mendes

Exploring Student  
Enrollments and  
Completions in Data-  
Related Career Tracks

# Introducción

One of the functionalities the 365 company introduced in a 2021 platform release included the option for student enrollment in a career track. The tracks represent an ordinal sequence of courses that eventually lead to obtaining the skills for one of three job titles: data scientist, data analyst, or business analyst. Completing a career track on the platform is a challenging task. To acquire a corresponding career certificate, a student must pass nine course exams (seven compulsory and two elective courses) and sit for a career track exam encompassing topics from all seven required courses.

In this Career Track Analysis with SQL and Tableau project, it will be analyzed the career track enrollments and achievements of 365's students. Will be needed to retrieve the necessary information from an SQL database. Afterward this information will be sent to Tableau, visualizing the results, and finally interpreting them.

It was receive an SQL database with information of the Year 2022 (from January 1 to October 20) with the followings databases:

- **career\_track\_info:**
  - **track\_id** – the unique identification of a track, which serves as the primary key to the table
  - **track\_name** – the name of the track (Data science, Business analyst or Data analyst).
- **career\_track\_student\_enrollments:**
  - **student\_id** – the unique identification of a student
  - **track\_id** – the unique identification of a track. Together with the previous column, they make up the primary key to the table—i.e., each student can enroll in a specific track only once. But a student can enroll in more than one career track.
  - **date\_enrolled** – the date the student enrolled in the track. A student can enroll in more than one career track.
  - **date\_completed** – the date the student completed the track. If the track is not completed, the field is NULL.

And it will be tried to answer the following questions:

1. What is the number of enrolled students monthly? Which is the month with the most enrollments? Speculate about the reason for the increased numbers.
2. Which career track do students enroll most in?
3. What is the career track completion rate? Can you say if it's increasing, decreasing, or staying constant with time?

4. How long does it typically take students to complete a career track? What type of subscription is most suitable for students who aim to complete a career track: monthly, quarterly, or annual?
5. What advice and suggestions for improvement would you give the 365 team to boost engagement, increase the track completion rate, and motivate students to learn more consistently?

# Preprocessing in SQL

We will start creating a dataset containing more relevant information, we will create a dataset with the following columns:

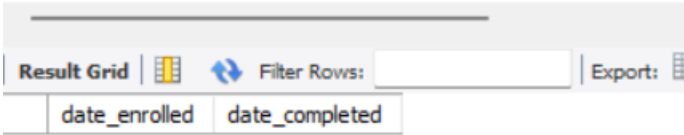
- **student\_track\_id** – this serves as an identification for each student-track pair. That is, each row of the resulting table should be uniquely labeled. Do this in an ordinal manner.
- **student\_id** – the unique identification of a student
- **track\_name** – the name of the track
- **date\_enrolled** – the date the student enrolled in the track
- **track\_completed**
  - 0 – the track is not completed (the completion date is NULL)
  - 1 – the track is completed (the completion date is not NULL)
- **days\_for\_completion** – the difference in days between the completion date and the enrollment date
- **completion\_bucket** – the bucket a student falls into based on the number of days it took them to complete a track (if they have done so).

The buckets are as follows:

- Same day – the **days\_for\_completion** field is equal to 0
- 1 to 7 days – the **days\_for\_completion** field is between 1 and 7 days, inclusive
- 8 to 30 days – the **days\_for\_completion** field is between 8 and 30 days, inclusive
- 31 to 60 days – the **days\_for\_completion** field is between 31 and 60 days, inclusive
- 61 to 90 days – the **days\_for\_completion** field is between 61 and 90 days, inclusive
- 91 to 365 days – the **days\_for\_completion** field is between 91 and 365 days, inclusive
- 366+ days – the **days\_for\_completion** field is more extensive than 365 days

Note: We check first if exist some rows were the enroll date is after the finish date (should be impossible), and this as the result (didn't exist any rows):

```
12 | select date_enrolled , date_completed
13 | from career_track_student_enrollments
14 | where date_enrolled>date_completed
```



The screenshot shows a SQL query editor interface. At the top, the query is displayed with line numbers 12, 13, and 14. Below the query, there is a toolbar with buttons for 'Result Grid', 'Filter Rows', and 'Export'. The 'Result Grid' button is active, and a table with two columns, 'date\_enrolled' and 'date\_completed', is visible below the toolbar.

date_enrolled	date_completed
---------------	----------------

So it was created the new dataset using MySQL using the following code:

```
8  -- Final table to export
9  with firsts_columns as (select row_number() over(order by cs.student_id, ct.track_name) as student_track_id ,
10                          cs.student_id,
11                          ct.track_name,
12                          cs.date_enrolled,
13                          case when cs.date_completed is null then 0 else 1 end as track_completed,
14                          case when cs.date_completed is null then 'Not completed yet'
15                              else DATEDIFF(cs.date_completed, cs.date_enrolled) end as days_for_completion
16                          from career_track_student_enrollments cs join career_track_info ct
17                          on cs.track_id = ct.track_id)
18  select student_track_id , student_id, track_name, date_enrolled, track_completed, days_for_completion,
19         case when days_for_completion = 'Not completed yet' then 'Not completed yet'
20             when days_for_completion between 1 and 7 then '1 to 7 days'
21             when days_for_completion between 8 and 30 then '8 to 30 days'
22             when days_for_completion between 31 and 60 then '31 to 60 days'
23             when days_for_completion between 61 and 90 then '61 to 90 days'
24             when days_for_completion between 91 and 365 then '91 to 365 days'
25             when days_for_completion > 365 then '366+'
26             else 'Same day' end as completion_bucket
27  from firsts_columns;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

	student_track_id	student_id	track_name	date_enrolled	track_completed	days_for_completion	completion_bucket
▶	1	6	Business Analyst	2022-09-16	0	Not completed yet	Not completed yet
	2	6	Data Scientist	2022-09-16	0	Not completed yet	Not completed yet
	3	20	Data Analyst	2022-02-25	0	Not completed yet	Not completed yet
	4	31	Data Analyst	2022-04-02	0	Not completed yet	Not completed yet
	5	209	Business Analyst	2022-06-02	0	Not completed yet	Not completed yet
	6	209	Data Scientist	2022-01-28	0	Not completed yet	Not completed yet

With this Dataset we export it as CSV to use it in tableau.

# Tableau

Now to analyze and visualize the information that we had exported in SQL we use tableau to create a dashboard containing the following graphs:

- A bar chart with the enrollment of students by date, and the percentage of completion by date in a line chart.
- A bar chart with the number of students belonging to each completion bucket.
- A card with the number of students.
- A card with the number of students who completed a track.
- A card with the number of tracks per student.
- A bar chart with the number of students by career track and a line chart with the percentage of completion by career track.
- Finally a bar chart with the average time to complete each career track.

Each of the graph are colored by a career track (blue is business analyst, orange is data analyst and red is data science).

Obtaining the following dashboard:



## Analyzing the results + Conclusion

This part consist in answering some of the questions made at the beginning, referring also other important insights. Lest start answering:

1. What is the number of enrolled students monthly? Which is the month with the most enrollments? Speculate about the reason for the increased numbers.

In 2022 the number of enrolled students by month were the followings:

- January: 1284;
- February: 1021;
- March: 1216;
- April: 978;
- May: 898;
- June: 1192;
- July: 958;
- August: 1653 (the higher month);
- September: 815;
- October: 426 (the lower month);
- (We don't have information of November nor December).

The number from January to July tend to be stable (ranging from approximately 900-1200 students enrolled per moth), then reached peak in august and fall until October; this patter could be because since the begging of the year until august people think that they have time to complete a career course, but when the end of the year is near they don't want to start something that could be left behind in the Christmas holidays.

An important thing to see is that the month with higher number of students enrolled are the month with the lower percentage of completion, except for august that is high on both.

2. Which career track do students enroll most in?

Students enrolls most in Data analyst, them follows Data science and finally comes business analyst, being the completion rate inverse to the number of students enrolls (business analyst have the higher completion rate and data analyst the lower one), this could be because the concepts taught in Data analyst are more complex than the one of a Business analyst.

3. What is the career track completion rate? Can you say if it's increasing, decreasing, or staying constant with time?

The career track completion rate follows a patter that takes about 4-5 months, for example starts in January with a rate of 1,4% them peak reaching a local minimum in March (with a 0.7% completion rate), them start raising until reach a local maximum in May (with a 1.5% completion rate) and them the cycle is repeated with lower intervals and more variability.

4. How long does it typically take students to complete a career track? What type of subscription is most suitable for students who aim to complete a career track: monthly, quarterly, or annual?

From the people that had completed a career track we could say that the most common thing that happen is completing it in 91 to 365 days, but is also true that if we sum up all the people that have completing a career course before 60 days, we reach almost the same amount of numbers.

If we group by career track we see that the average time that takes from a business analyst the complete a track is 105 days, a data analyst it takes 134 days, and finally the one that takes on average more time is the data science career track with an average of 150 days.

So, the most suitable subscription type would be the monthly (for the fast people) and the quarterly (for the majority) one for people that aim to complete a career track.

5. What advice and suggestions for improvement would you give the 365 team to boost engagement, increase the track completion rate, and motivate students to learn more consistently?

To conclude with this final question is important to mention that in that time frame that were analyzed, were information of 9019 unique students and only 123 had completed a career track, and only a small amount of people is enroll in more than 1 career track, these are very bad conversion rates that should be improved.

Knowing that right now exist the gamification of the platform (using a career path, leaderboard, XP's and coins), also the daily goals and reminders by email, my solution is reaching and targeting a different type of audience, one that could have more available time or working in related jobs (for example software engineer) because they might want to change to the new technologies.

If you want to have a more precise and trustworthy target audience, my solution is analyze the type of person that finished a career track and see the patter to know the answer (since the initial dataset didn't have personal information about the student is impossible to give an actual target audience with the information available).