STUDENT PERFORMANCE ANALYSIS:

Most Crucial Aspect To Understand Their Progress

Index

ABSTRACT

- This project was undertaken to provide key insights into student performance and topic popularity within an educational platform.

Stakeholder Needs and Objectives

- Understanding Student Distribution:

  They wanted to know the geographical distribution of students and how students were spread across different grades. This information was crucial for resource planning and ensuring equitable access to educational materials.

- Measuring Student Success:

  They aimed to track student progress and identify factors influencing success, such as topic completion rates and test scores.

Stakeholder Needs and Objectives

- Evaluating Topic Effectiveness:

  Stakeholders needed to assess the popularity and difficulty of different topics. This would help them identify areas where students struggled and where curriculum adjustments might be necessary.

- Optimizing Resource Allocation:

  By analyzing student enrollment trends over time, stakeholders could make informed decisions about staffing, budgeting, and facility planning.

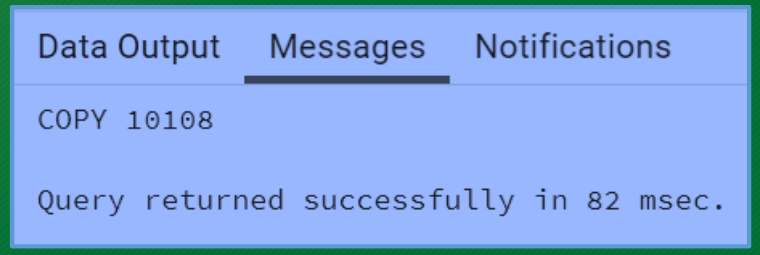
Data and Methodology

The analysis was conducted using SQL queries against a database named "student information." This database contained several tables, including:

- **date_date:** Information about dates (day, month, quarter, year).
- **city_region:** Details about student locations (city, state).
- **topic_topic:** Information about available topics (topic ID, topic name, grade).
- **student_topic:** Records of student interactions with topics (student ID, topic ID, test cleared status, percentage mark).
- **Student_student:** Student demographics (student ID, student name, grade, class start date, test date).

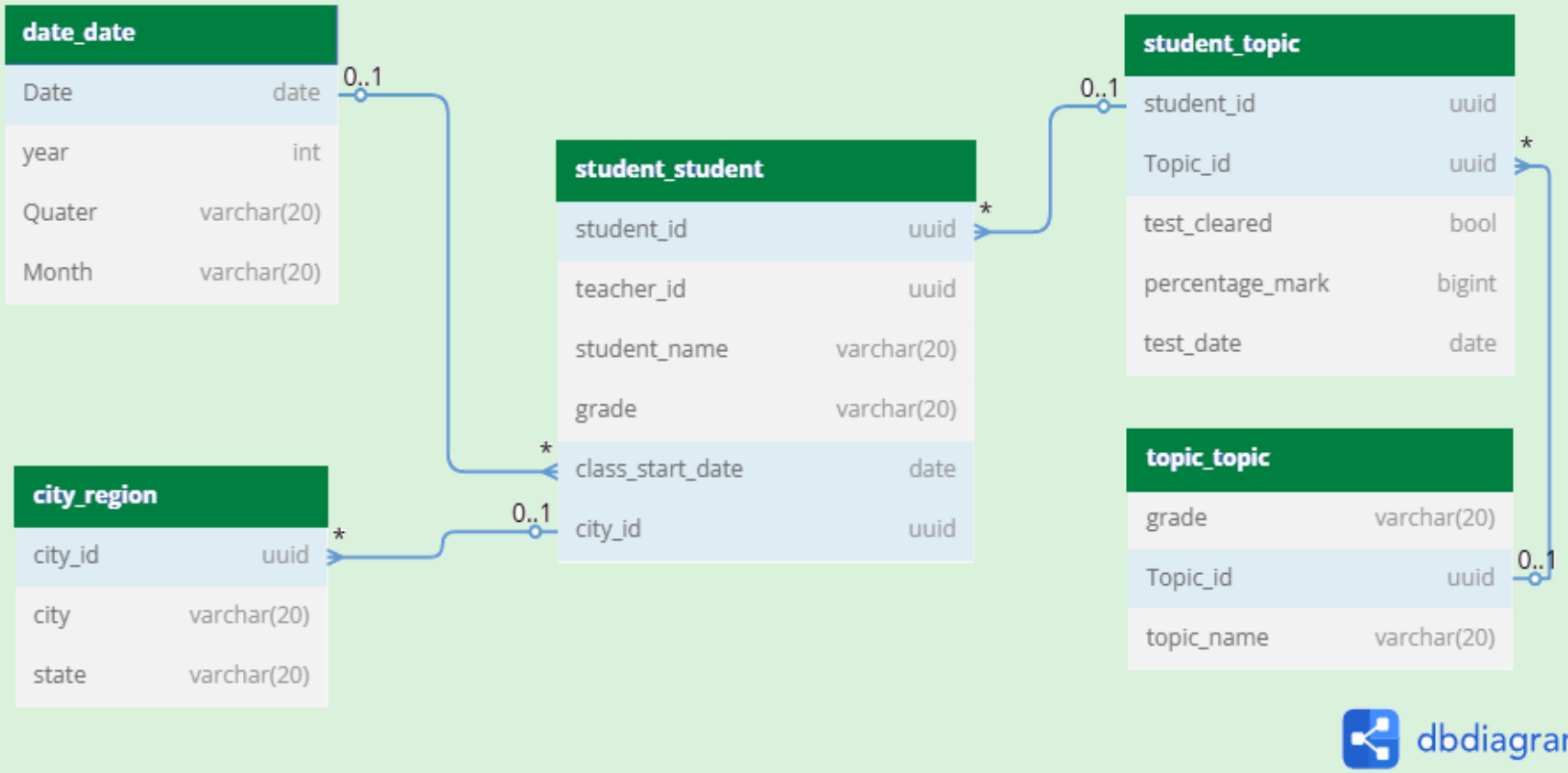
Import data to SQL database

1. Prepare csv file
2. Create tables in SQL
3. Import csv file into SQL



ER- DIAGRAM

ERD is the backbone of any well-structured database design. It helps visualize how different entities (tables) interact with each other, ensuring data integrity, efficiency, and scalability.



Structure & Content of Table


```
4    'date_date'
```

Data Output | Messages | Notifications

	column_name name	data_type character varying
1	date	date
2	year	integer
3	quarter	character varying
4	month	character varying

	date date	year integer	quarter character varying (20)	month character varying (20)
1	2023-10-15	2023	Quarter 4	October
2	2024-01-20	2024	Quarter 1	January
3	2024-03-05	2024	Quarter 1	March
4	2024-07-12	2024	Quarter 3	July
5	2024-12-28	2024	Quarter 4	December
6	2023-11-01	2023	Quarter 4	November
7	2024-02-14	2024	Quarter 1	February
8	2024-04-22	2024	Quarter 2	April
9	2024-08-08	2024	Quarter 3	August
10	2023-09-30	2023	Quarter 3	September

Structure & Content of Table


```
4   'city_region'
```

Data Output | Messages | Notifications

	column_name name	data_type character varying
1	city_id	integer
2	city	character varying
3	state	character varying

	city_id integer	city character varying (20)	state character varying (20)
1	10001	New York	NY
2	10002	Los Angeles	CA
3	10003	Chicago	IL
4	10004	Houston	TX
5	10005	Philadelphia	PA
6	10006	Phoenix	AZ
7	10007	San Antonio	TX
8	10008	San Diego	CA
9	10009	Dallas	TX
10	10010	San Jose	CA
11	10011	Austin	TX

Structure & Content of Table


```
4   'student_student'
```

Data Output | Messages | Notifications

	column_name name	data_type character varying
1	student_id	integer
2	teacher_id	integer
3	class_start_date	date
4	city_id	integer
5	student_name	character varying
6	grade	character varying

	student_id integer	teacher_id integer	student_name character varying (20)	grade character varying (20)	class_start_date date	city_id integer
1	50001	50002	Alice Smith	A	2023-09-01	10001
2	50002	50003	Bob Johnson	B	2023-09-01	10002
3	50003	50004	Carol Williams	C	2023-09-01	10003
4	50004	50005	David Brown	A	2023-09-01	10004
5	50005	50001	Eve Davis	B	2023-09-01	10005
6	50006	50007	Frank Miller	C	2023-09-01	10006
7	50007	50008	Grace Wilson	A	2023-09-01	10007
8	50008	50009	Henry Moore	B	2023-09-01	10008
9	50009	50010	Irene Taylor	C	2023-09-01	10009
10	50010	50006	Jack Anderson	A	2023-09-01	1001

Structure & Content of Table


```
4   'student_topic'
```

Data Output | Messages | Notifications

	column_name name	data_type character varying
1	student_id	integer
2	topic_id	integer
3	test_cleared	boolean
4	percentage_mark	bigint
5	test_date	date

	student_id integer	topic_id integer	test_cleared boolean	percentage_mark bigint	test_date date
1	50001	20001	true	90	2023-10-20
2	50002	20002	true	85	2023-10-21
3	50003	20003	false	60	2023-10-22
4	50004	20004	true	95	2023-10-23
5	50005	20005	true	88	2023-10-24
6	50006	20006	true	92	2023-10-25
7	50007	20007	true	87	2023-10-26
8	50008	20008	false	55	2023-10-27
9	50009	20009	true	93	2023-10-28
10	50010	20010	true	89	2023-10-29
11	50011	20011	true	86	2023-10-30
12	50012	20012	false	62	2023-10-3

Structure & Content of Table


```
4    'topic_topic'
```

Data Output | Messages | Notifications

	column_name name	data_type character varying
1	topic_id	integer
2	grade	character varying
3	topic_name	character varying

	topic_id integer	grade character varying (20)	topic_name character varying (50)
1	20001	A	Math
2	20002	B	Science
3	20003	C	History
4	20004	A	English
5	20005	B	Art
6	20006	C	Music
7	20007	A	Physics
8	20008	B	Chemistry
9	20009	C	Geography
10	20010	A	Computer Science
11	20011	B	Biology
12	20012	C	Economics
13	20013	A	Literature

SUBQUERY

How many topics are available in each grade?

Display grade, number of topics in each grade.

/4/2025 2:31:41 PM | 3 | 330 msec

Date | Rows affected | Duration

Copy | Copy to Query Editor

```
Select
        grade,
        COUNT(DISTINCT Topic_id) As Number_of_Topics
From topic_topic As t
Group by grade;
```

Messages

Successfully run. Total query runtime: 330 msec. 3 rows affected.

Data Output | Messages | Notifications

	grade character varying (20)	number_of_topics bigint
1	A	7
2	B	7
3	C	6

Grades A and B offer a slightly broader range of topics compared to Grade C.

Show the distribution of students across grades.

Display the grade and number of students in each grade, % students in each grade.

Date	Rows affected	Duration
/4/2025 1:01:11 PM	3	232 msec

Copy | Copy to Query Editor

```
Select grade, Count(*) As number_of_students,
    (Count(*) * 100) / (Select Count(*) From student_student) as
From student_student As ss
Group by grade
Order by Count(*) desc;
```

Messages

Successfully run. Total query runtime: 232 msec. 3 rows affected.

Data Output | Messages | Notifications

	grade character varying (20)	number_of_students bigint	percentage_of_students bigint
1	B	43	35
2	A	41	33
3	C	37	30

Grade B has the highest student enrollment, followed closely by Grade A, indicating a relatively even distribution across all three grades

JOINS

List the top 5 cities in terms of number of students.

Display the names of cities, states and number of students in each city.

/3/2025 8:36:48 PM | 5 | 222 msec

Date | Rows affected | Duration

Copy | Copy to Query Editor

```
Select c.city, c.state, Count(ss.student_id) As num_of_students
From student_student As ss
Join city_region as c
On ss.city_id = c.city_id
Group by c.city, c.state
Order by Count(ss.student_id) desc
Limit 5;
```

Messages

Successfully run. Total query runtime: 222 msec. 5 rows affected.

Data Output | Messages | Notifications

	city character varying (20)	state character varying (20)	num_of_students bigint
1	New York	NY	20
2	Los Angeles	CA	20
3	Chicago	IL	20
4	Houston	TX	16
5	San Diego	CA	6

Student enrollment is heavily concentrated in major US cities, particularly New York, Los Angeles, and Chicago.

List the top 3 topic in terms of difficulty level.

/4/2025 5:26:33 PM | 3 | 1 secs 495 msec
Date | Rows affected | Duration

Copy | Copy to Query Editor

```
Select t.topic_name
From topic_topic As t
Join student_topic As st
ON t.topic_id = st.topic_id
Group by t.topic_name
Order by AVG(st.percentage_mark) asc
Limit 3;
```

Messages

Successfully run. Total query runtime: 1 secs 495 msec. 3 rows affected.

Data Output | Messages | Noti

	topic_name character varying (50)
1	Statistics
2	Economics
3	Data Science

Statistics, Economics, and Data Science present the greatest academic challenges based on student performance.

These topics may require focused attention and resources to improve student understanding and outcomes.

CTE

List our most successful quarter in terms of new students added

Display all quarters of 2017, count and % students who started class.

/2025 2:02:53 PM | 4 | 214 msec
Date | Rows affected | Duration

Copy | Copy to Query Editor

```
With sdtc_QuaterWise As
(
  Select d.Quarter, Count(*) as number_of_students
  From student_student As ss
  Join date_date As d
  on ss.class_start_date = d.date
  Where d.year = 2017
  Group by d.Quarter
)

Select Quarter, number_of_students,
       ROUND((number_of_students * 100) /
       (Select SUM(number_of_students) From sdtc_QuaterWise),2)
       as percentage_students
From sdtc_QuaterWise
Order by number_of_students desc;
```

Messages

Successfully run. Total query runtime: 214 msec. 4 rows affected.

Data Output | Messages | Notifications

	quarter character varying (20)	number_of_students bigint	percentage_students numeric
1	Quarter 1	16	47.06
2	Quarter 2	14	41.18
3	Quarter 4	3	8.82
4	Quarter 3	1	2.94

Student enrollment peaked in the first quarter of 2017, significantly outpacing all other quarters

What percentage of students completed all topics in the grade?

Display grade, number of topics number of students who have completed all topics.


```
ith Stdt_TopicCompletion As
(
    Select ss.grade As new_grade,
        Count(Distinct tt.topic_id) As total_topics,
        Count(Distinct Case
            When st.test_cleared = TRUE THEN st.student_
                                    As completed_student
    From student_student As ss
    Join student_topic As st
    On ss.student_id = st.student_id
    Join topic_topic As tt
    On st.topic_id = tt.topic_id
    Group by ss.grade
)

Select new_grade,
        total_topics,
        completed_students,
        (completed_students * 100) /
            total_topics As completed_percent
From Stdt_TopicCompletion;
```

Messages

Successfully run. Total query runtime: 392 msec. 3 rows affected.

Data Output | Messages | Notifications

	new_grade character varying (20)	total_topics bigint	completed_students bigint	completed_percent bigint
1	A	19	10	52
2	B	19	18	94
3	C	18	15	83

Grade B students show the highest completion rate of all topics, significantly outpacing Grades A and C.

While all grades have a similar number of total topics, completion percentages vary widely, indicating potential differences in student engagement or topic difficulty.

Identify topics that have least scores.

Display all topic names, mean, median, 25th and 75th percentile marks for each.

2025 5:19:03 PM | 20 | 865 msec
ate | Rows affected | Duration

Copy | Copy to Query Editor

```
With TopicScores As
 (
    Select t.topic_name As name ,
           st.percentage_mark As percent_marks
    From topic_topic As t
    Join student_topic As st
    ON t.topic_id = st.topic_id
)
Select name,
       ROUND(AVG(percent_marks),2) AS mean_score,
       PERCENTILE_CONT(0.5)  Within Group (Order by perce
       PERCENTILE_CONT(0.25) Within Group (Order by perce
       PERCENTILE_CONT(0.75) Within Group (Order by perce
From TopicScores
Group by name
Order by AVG(percent_marks) asc;
```

Messages

Successfully run. Total query runtime: 865 msec. 20 rows affected.

Data Output | Messages | Notifications

	name character varying (50)	mean_score numeric	median_score double precision	score_25ptile double precision	score_75ptile double precision
1	Statistics	21.00	21	20.5	21.5
2	Economics	35.67	35	34.5	36.5
3	Data Science	39.25	28.5	27.25	40.5
4	Geography	44.83	45.5	27.25	63.75
5	Calculus	45.50	38	36.75	46.75
6	Biology	47.00	47.5	18.25	76.75
7	English	47.33	50	21.75	71.5

Statistics and Economics show the lowest average scores, indicating potential areas for curriculum review or additional student support.

Conversely, Psychology and Programming exhibit the highest average scores, suggesting strong student performance in these topics.

Impact

DIAGRAM

- **Uneven geographical distribution:** A significant concentration of students in major urban centers (New York, Los Angeles, Chicago) may lead to disparities in resource allocation and access to educational opportunities.

- **Varied topic completion and performance:** Differences in topic completion rates and test scores across grades and topics suggest potential challenges in student engagement, curriculum effectiveness, or access to support.

- **Challenging topics:** Statistics, Economics, and Data Science emerge as particularly challenging topics, indicating a need for focused attention and intervention.

- **Fluctuating enrollment:** Enrollment trends, such as the peak observed in the first quarter of 2017, highlight the need for dynamic resource allocation and planning to accommodate changing student numbers.

Recommendations

DIAGRAM

- **Prioritize equitable resource allocation:** Ensure that resources, including staff, materials, and technology, are distributed equitably to address the needs of students in all locations, with particular attention to urban centers with high student density.
- **Enhance student support:** Implement targeted intervention programs, additional tutoring, and academic support resources to address the challenges faced by students in specific grades or with particular topics.
- **Review and adjust curriculum:** Conduct a thorough review of the curriculum, especially for topics with low completion rates or poor test performance, to identify areas for improvement in clarity, engagement, and alignment with student needs.
- **Optimize resource utilization:** Leverage enrollment trend data to inform staffing, budgeting, and facility planning decisions, ensuring that resources are aligned with student needs and enrollment patterns.
- **Promote data-driven decision making:** Encourage the use of data analysis and insights to guide decision-making across all levels of the institution, fostering a culture of continuous improvement and responsiveness to student needs.

Q&A

Thanks!

Do you have any questions?

mk204009@gmail.com
+8368853313