

## SQL Analysis

**Project Title:** Enhancing Student Performance & Retention in Schools using Data Analytics & Predictive Modeling

**Tool:** Snowflake

### **1. CREATING SCHEMA, DATA PREVIEW & CHECKING DATA TYPES**

#### **-- Creating Schema**

```
CREATE SCHEMA school_data;
```

#### **-- Data Preview**

```
SELECT * FROM ACTIVITIES  
LIMIT 1000;
```

```
SELECT * FROM ATTENDANCE1  
LIMIT 1000;
```

```
SELECT * FROM CLASSES  
LIMIT 1000;
```

```
SELECT * FROM DISCIPLINARY_RECORDS  
LIMIT 1000;
```

```
SELECT * FROM EXAM_SCORES  
LIMIT 1000;
```

```
SELECT * FROM STUDENTS  
LIMIT 1000;
```

```
SELECT * FROM TEACHERS  
LIMIT 1000;
```

#### **-- Checking Data Types**

```
SELECT  
    TABLE_NAME,  
    COLUMN_NAME,  
    DATA_TYPE  
FROM  
    INFORMATION_SCHEMA.COLUMNS  
WHERE  
    TABLE_SCHEMA = 'SCHOOL_DATA'  
AND  
    TABLE_NAME = 'TEACHERS';
```

	TABLE_NAME	COLUMN_NAME	DATA_TYPE
1	TEACHERS	FIRST_NAME	TEXT
2	TEACHERS	LAST_NAME	TEXT
3	TEACHERS	YEARS_EXPERIENCE	NUMBER
4	TEACHERS	SUBJECT	TEXT
5	TEACHERS	TEACHER_RATING	NUMBER
6	TEACHERS	TEACHER_ID	TEXT
7	TEACHERS	QUALIFICATION	TEXT

## 2. DATA QUALITY AUDIT

-- Student table

-- Checking null values

SELECT COUNT(DISTINCT STUDENT\_ID) AS TOTAL\_STUDENTS  
FROM STUDENTS;

	TOTAL_STUDENTS
1	10000

SELECT  
COUNT(\*) AS TOTAL\_COUNT,  
COUNT(\*) - COUNT(STUDENT\_ID) AS NULL\_STUDENTID,  
COUNT(\*) - COUNT(FIRST\_NAME) AS NULL\_FIRSTNAME,  
COUNT(\*) - COUNT(LAST\_NAME) AS NULL\_LASTNAME,  
COUNT(\*) - COUNT(GENDER) AS NULL\_GENDER,  
COUNT(\*) - COUNT(DOB) AS NULL\_DOB,  
COUNT(\*) - COUNT(GRADE\_LEVEL) AS NULL\_GRADELEVEL,  
COUNT(\*) - COUNT(ENROLLMENT\_DATE) AS NULL\_ENROLLMENTDATE,  
COUNT(\*) - COUNT(PARENT\_EDUCATION\_LEVEL) AS NULL\_PARENTEDUCATIONLEVEL,  
COUNT(\*) - COUNT(FAMILY\_INCOME) AS NULL\_FAMILYINCOME,  
COUNT(\*) - COUNT(TRANSPORT\_MODE) AS NULL\_TRANSPORTMODE  
FROM  
STUDENTS;

...	TOTAL_COUNT	NULL_STUDENTID	NULL_FIRSTNAME	NULL_LASTNAME	NULL_GENDER	NULL_DOB	NULL_GRADELEVEL	NULL_ENROLLMENTDATE	NULL_PARENTEDUCATIONLEVEL	NULL_FAMILYINCOME	NULL_TRANSPORTMODE
	10000	0	0	0	0	0	0	0	0	0	0

-- Checking for duplicate values

SELECT  
STUDENT\_ID, COUNT(\*)  
FROM  
STUDENTS

```
GROUP BY
  STUDENT_ID
HAVING COUNT(*) > 1;
```

STUDENT_ID	COUNT(*)
------------	----------

Query produced no results

-- Checking for data ranges

```
SELECT
  MIN(DOB) AS MIN_DOB, MAX(DOB) AS MAX_DOB,
  MIN(GRADE_LEVEL) AS MIN_GRADE, MAX(GRADE_LEVEL) AS MAX_GRADE,
  MIN(FAMILY_INCOME) AS MIN_INCOME, MAX(FAMILY_INCOME) AS MAX_INCOME
FROM
  STUDENTS;
```

MIN_DOB	...	MAX_DOB	MIN_GRADE	MAX_GRADE	MIN_INCOME	MAX_INCOME
2006-09-02		2019-08-31	1	12	50073	999962

-- Teachers table

-- Checking null values

```
SELECT COUNT(DISTINCT TEACHER_ID) AS TOTAL_COUNT
FROM TEACHERS;
```

...	TOTAL_COUNT
	500

```
SELECT
  COUNT(*) AS TOTAL_COUNT,
  COUNT(*) - COUNT(TEACHER_ID) AS NULL_TEACHERID,
  COUNT(*) - COUNT(FIRST_NAME) AS NULL_FIRSTNAME,
  COUNT(*) - COUNT(LAST_NAME) AS NULL_LASTNAME,
  COUNT(*) - COUNT(SUBJECT) AS NULL_SUBJECT,
  COUNT(*) - COUNT(YEARS_EXPERIENCE) AS NULL_EXPERIENCE,
  COUNT(*) - COUNT(QUALIFICATION) AS NULL_QUALIFICATION,
  COUNT(*) - COUNT(TEACHER_RATING) AS NULL_TEACHERRATING,
FROM
  TEACHERS
```

...	TOTAL_COUNT	NULL_TEACHERID	NULL_FIRSTNAME	NULL_LASTNAME	NULL_SUBJECT	NULL_EXPERIENCE	NULL_QUALIFICATION	NULL_TEACHERRATING
	500	0	0	0	0	0	0	0

### -- Checking for duplicate entries

```
SELECT
  TEACHER_ID, COUNT(*)
FROM
  TEACHERS
GROUP BY
  TEACHER_ID
HAVING
  COUNT(*) > 1;
```

TEACHER_ID	COUNT(*)
Query produced no results	

### --Checking data ranges

```
SELECT
  MIN(YEARS_EXPERIENCE) AS MIN_EXP, MAX(YEARS_EXPERIENCE) AS MAX_EXP,
  MIN(TEACHER_RATING) AS MIN_RATING, MAX(TEACHER_RATING) AS MAX_RATING
FROM
  TEACHERS;
```

MIN_EXP	...	MAX_EXP	MIN_RATING	MAX_RATING
1		35	2.0	5.0

### -- Classes table

#### -- Checking null values

```
SELECT COUNT(DISTINCT CLASS_ID) AS TOTAL_COUNT
FROM CLASSES;
```

	...	TOTAL_COUNT
1		15000

```
SELECT
  COUNT(*) AS TOTAL_COUNT,
  COUNT(*) - COUNT(CLASS_ID) AS NULL_CLASSID,
  COUNT(*) - COUNT(STUDENT_ID) AS NULL_STUDENTID,
  COUNT(*) - COUNT(TEACHER_ID) AS NULL_TEACHERID,
  COUNT(*) - COUNT(ACADEMIC_YEAR) AS NULL_YEAR,
  COUNT(*) - COUNT(CLASS_AVERAGE_SCORE) AS NULL_SCORE
FROM
  CLASSES;
```

TOTAL_COUNT	...	NULL_CLASSID	NULL_STUDENTID	NULL_TEACHERID	NULL_YEAR	NULL_SCORE
15000		0	0	0	0	0

## -- Referential Integrity

```
SELECT C.STUDENT_ID
FROM CLASSES C
LEFT JOIN STUDENTS S ON C.STUDENT_ID = S.STUDENT_ID
WHERE S.STUDENT_ID IS NULL;
```

STUDENT\_ID

Query produced no results

```
SELECT C.TEACHER_ID
FROM CLASSES C
LEFT JOIN TEACHERS T ON C.TEACHER_ID = T.TEACHER_ID
WHERE T.TEACHER_ID IS NULL;
```

TEACHER\_ID

Query produced no results

## --Checking for duplicate entries

```
SELECT CLASS_ID, COUNT(*)
FROM CLASSES
GROUP BY CLASS_ID
HAVING COUNT(*) > 1;
```

CLASS\_ID

COUNT(\*)

Query produced no results

## -- Exam\_Scores table

### -- Checking null values

```
SELECT COUNT(DISTINCT EXAM_ID) AS TOTAL_COUNT
FROM EXAM_SCORES;
```

...

TOTAL\_COUNT

360394

```
SELECT
COUNT(*) AS TOTAL_COUNT,
COUNT(*) - COUNT(EXAM_ID) AS NULL_EXAMID,
COUNT(*) - COUNT(STUDENT_ID) AS NULL_STUDENTID,
COUNT(*) - COUNT(SUBJECT) AS NULL_SUBJECT,
COUNT(*) - COUNT(EXAM_DATE) AS NULL_DATE,
COUNT(*) - COUNT(SCORE) AS NULL_SCORE
FROM
```

EXAM\_SCORES;

TOTAL_COUNT	...	NULL_EXAMID	NULL_STUDENTID	NULL_SUBJECT	NULL_DATE	NULL_SCORE
360394		0	0	0	0	0

--Checking data range

SELECT SUBJECT, MIN(SCORE) AS MIN\_SCORE, MAX(SCORE) AS MAX\_SCORE  
FROM EXAM\_SCORES  
GROUP BY SUBJECT;

SUBJECT	...	MIN_SCORE	MAX_SCORE
Arts		0	100
Science		0	100
History		0	100
English		0	100
Computer		0	100
Math		0	100

-- Attendance table

-- Checking null values

SELECT  
COUNT(\*) AS TOTAL\_COUNT,  
COUNT(\*) - COUNT(ATTENDANCE\_ID) AS NULL\_ATTENDANCEID,  
COUNT(\*) - COUNT(STUDENT\_ID) AS NULL\_STUDENTID,  
COUNT(\*) - COUNT(STATUS) AS NULL\_STATUS  
FROM ATTENDANCE19;

...	TOTAL_COUNT	NULL_ATTENDANCEID	NULL_STUDENTID	NULL_DATE	NULL_STATUS
	130000	0	0	0	0

--Checking valid status

SELECT DISTINCT STATUS FROM ATTENDANCE19;

STATUS
Present
Late
Absent

-- Activities table

-- Checking null values

SELECT  
COUNT(\*) AS TOTAL\_COUNT,  
COUNT(\*) - COUNT(ACTIVITY\_ID) AS NULL\_ACTIVITYID,  
COUNT(\*) - COUNT(STUDENT\_ID) AS NULL\_STUDENTID,

```

COUNT(*) - COUNT(ACTIVITY_TYPE) AS NULL_ACTIVITY,
COUNT(*) - COUNT(HOURS_PER_WEEK) AS NULL_HOURS,
COUNT(*) - COUNT(ACHIEVEMENT_LEVEL) AS NULL_ACHIEVEMENT
FROM ACTIVITIES;

```

...	TOTAL_COUNT	NULL_ACTIVITYID	NULL_STUDENTID	NULL_ACTIVITY	NULL_HOURS	NULL_ACHIEVEMENT
	50000	0	0	0	0	0

### --Checking data range

```

SELECT
    ACTIVITY_TYPE,
    MIN(HOURS_PER_WEEK) AS MIN_HOURS,
    MAX(HOURS_PER_WEEK) AS MAX_HOURS
FROM
    ACTIVITIES
GROUP BY
    ACTIVITY_TYPE;

```

ACTIVITY_TYPE	...	MIN_HOURS	MAX_HOURS
Debate		0	15
Dance		0	15
Drama		0	15
Sports		0	15
Art		0	15
Music		0	15
STEM Club		0	15

### -- Disciplinary Records table

#### -- Checking null values

```

SELECT
    COUNT(*) AS TOTAL_COUNT,
    COUNT(*) - COUNT(RECORD_ID) AS NULL_RECORDID,
    COUNT(*) - COUNT(STUDENT_ID) AS NULL_STUDENTID,
    COUNT(*) - COUNT(INCIDENT_DATE) AS NULL_DATE,
    COUNT(*) - COUNT(INCIDENT_TYPE) AS NULL_TYPE,
    COUNT(*) - COUNT(SEVERITY) AS NULL_SEVERITY
FROM DISCIPLINARY_RECORDS;

```

...	TOTAL_COUNT	NULL_RECORDID	NULL_STUDENTID	NULL_DATE	NULL_TYPE	NULL_SEVERITY
	30000	0	0	0	0	0

### --Checking valid incident types

```

SELECT DISTINCT INCIDENT_TYPE FROM DISCIPLINARY_RECORDS;

```

INCIDENT_TYPE
Misconduct
Other
Cheating
Late Submission
Bullying

--Checking valid severity

SELECT DISTINCT SEVERITY FROM DISCIPLINARY\_RECORDS;

SEVERITY
High
Low
Medium

### 3. DATA ANALYSIS

#### A. Student Performance Analysis

--What is the overall pass/fail rates across grades and subjects?

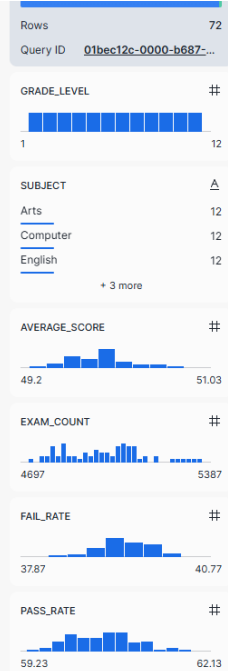
```

SELECT
    S.GRADE_LEVEL,
    E.SUBJECT,
    ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE,
    COUNT(*) AS EXAM_COUNT,
    ROUND(100*SUM(CASE WHEN E.SCORE < 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS FAIL_RATE,
    ROUND(100*SUM(CASE WHEN E.SCORE >= 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS PASS_RATE
FROM
    EXAM_SCORES E
JOIN
    STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID
GROUP BY
    S.GRADE_LEVEL, E.SUBJECT
ORDER BY
    S.GRADE_LEVEL, E.SUBJECT;

```



GRADE_LEVEL	SUBJECT	AVERAGE_SCORE	EXAM_COUNT	FAIL_RATE	PASS_RATE
1	Arts	50.55	4930	38.92	61.08
1	Computer	50.18	5121	39.74	60.26
1	English	50.24	5064	39.10	60.90
1	History	50.21	5033	39.26	60.74
1	Math	49.80	5086	40.23	59.77
1	Science	49.65	5094	40.24	59.76
2	Arts	50.34	4804	39.68	60.32
2	Computer	50.59	4804	39.49	60.51
2	English	50.08	4790	38.60	61.40
2	History	50.36	4870	38.64	61.36
2	Math	50.36	4795	39.25	60.75
2	Science	50.12	4823	39.31	60.69
3	Arts	49.70	5089	40.68	59.32
3	Computer	50.39	5022	39.03	60.97
3	English	50.07	5089	39.61	60.39
3	History	49.75	5043	39.96	60.04
3	Math	50.61	5061	38.67	61.33
3	Science	49.70	5084	39.42	60.58
4	Arts	49.58	5065	40.51	59.49
4	Computer	49.34	5017	40.50	59.50
4	English	49.47	4981	39.95	60.05
4	History	51.03	4972	38.70	61.30
4	Math	50.06	5017	39.05	60.95
4	Science	50.22	4996	39.13	60.87
5	Arts	49.92	5073	40.21	59.79



#### Analyzed 360394 records from Exam\_Scores table:

- Fail rate is higher in Grade 1 Science (40.24%) and Math (40.23%).
- A high fail rate of 40.68% can be observed in Grade 3 Arts.
- Fail rate in Grade 4 is highest in subjects like Arts (40.51%) and Computer (40.5%)
- A fail rate of 40.21% can be seen in Grade 5 Arts and 40.05% in Grade 6 Math.
- Higher fail rates are observed in Grade 7 subjects like Arts (40.25%), Math (40.01%) and Science (40.77%).
- In Grade 8, a fail rate of 40.46% and 40.35% can be observed in subjects like History and Science.
- A fail rate of 40.05% can be seen in Grade 9 Science and 40.36% in Grade 9 Computer.
- We can see higher fail rates in Grade 11 subjects like Arts (40.02%), Computer (40.23%), History (40.03%) and Science (40.36%)
- In Grade 12, a higher fail rate of 39.91% & 39.81% can be seen in Math & Arts.

#### -- Checking overall pass and fail rate across all grades

```

SELECT
  S.GRADE_LEVEL,
  ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE,
  COUNT(*) AS EXAM_COUNT,
  ROUND(100*SUM(CASE WHEN E.SCORE < 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS FAIL_RATE,
  ROUND(100*SUM(CASE WHEN E.SCORE >= 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS PASS_RATE
FROM
  EXAM_SCORES E
JOIN
  STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID

```

GROUP BY  
S.GRADE\_LEVEL  
ORDER BY  
S.GRADE\_LEVEL;

GRADE_LEVEL	...	AVERAGE_SCORE	EXAM_COUNT	FAIL_RATE	PASS_RATE
1		50.10	30328	39.59	60.41
2		50.31	28886	39.16	60.84
3		50.03	30388	39.56	60.44
4		49.95	30048	39.64	60.36
5		50.20	30391	39.35	60.65
6		50.14	28922	39.29	60.71
7		49.76	31948	40.02	59.98
8		50.09	30739	39.46	60.54
9		50.22	28660	39.31	60.69
10		49.90	29514	39.31	60.69
11		49.75	29557	39.96	60.04
12		49.82	31013	39.47	60.53

- The student in Grade 7 witnessed the higher rate of 40.02% while other grades have fail rate in the range of 39%

-- Checking overall pass and fail rate across all subjects

```
SELECT
  E.SUBJECT,
  ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE,
  COUNT(*) AS EXAM_COUNT,
  ROUND(100*SUM(CASE WHEN E.SCORE < 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS FAIL_RATE,
  ROUND(100*SUM(CASE WHEN E.SCORE >= 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS PASS_RATE
FROM
  EXAM_SCORES E
JOIN
  STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID
GROUP BY
  E.SUBJECT
ORDER BY
  E.SUBJECT;
```

SUBJECT	...	AVERAGE_SCORE	EXAM_COUNT	FAIL_RATE	PASS_RATE
Arts		50.05	59938	39.71	60.29
Computer		50.02	60124	39.59	60.41
English		50.12	60037	39.12	60.88
History		50.08	60092	39.59	60.41
Math		49.99	60201	39.35	60.65
Science		49.86	60002	39.73	60.27

- The overall failure rate tends to be in the range of 39% across all the subjects.

#### -- Performance by Gender

```
SELECT
  S.GENDER,
  E.SUBJECT,
  ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE,
  ROUND(100*SUM(CASE WHEN SCORE < 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS FAIL_RATE
FROM
  EXAM_SCORES E
JOIN
  STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID
GROUP BY
  S.GENDER, E.SUBJECT
ORDER BY
  S.GENDER, E.SUBJECT;
```

	GENDER	...	SUBJECT	AVERAGE_SCORE	FAIL_RATE
1	Female		Arts	50.19	39.61
2	Female		Computer	49.95	39.69
3	Female		English	49.96	39.26
4	Female		History	50.09	39.76
5	Female		Math	49.75	39.54
6	Female		Science	49.71	40.17
7	Male		Arts	49.92	39.82
8	Male		Computer	50.08	39.50
9	Male		English	50.27	38.98
10	Male		History	50.08	39.42
11	Male		Math	50.22	39.16
12	Male		Science	50.00	39.30

- Females showed poorer performance with a fail rate of 40.17% in subject like science while Males showed poor performance in arts at 39.82%.

```
SELECT
  S.GENDER,
  ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE,
  ROUND(100*SUM(CASE WHEN SCORE < 40 THEN 1 ELSE 0 END)/COUNT(*), 2) AS FAIL_RATE
FROM
  EXAM_SCORES E
JOIN
  STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID
```

```
GROUP BY
  S.GENDER
ORDER BY
  S.GENDER;
```

- The overall fail rate tends to be similar with 39.67% in females and 39.36% in males.

	GENDER	AVERAGE_SCORE	...	FAIL_RATE
1	Female	49.94		39.67
2	Male	50.10		39.36

#### -- Parental education impact on performance

```
SELECT
  S.PARENT_EDUCATION_LEVEL,
  ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE,
  COUNT(*) AS EXAM_COUNT
FROM
  EXAM_SCORES E
JOIN
  STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID
GROUP BY
  S.PARENT_EDUCATION_LEVEL
ORDER BY
  AVERAGE_SCORE DESC;
```

	PARENT_EDUCATION_LEVEL	...	AVERAGE_SCORE	EXAM_COUNT
1	Master		50.15	74321
2	PhD		50.10	70133
3	High School		49.98	70096
4	Bachelor		49.94	72872
5	None		49.92	72972

- There isn't much a major difference but still we can see that parent holding masters or PhD degrees tends to influence children to study more and perform better.
- We can say that students with more educated parents slightly perform better than others.

#### -- Family Income vs. Performance

```
SELECT
  CASE
    WHEN S.FAMILY_INCOME < 200000 THEN 'LOW INCOME'
    WHEN S.FAMILY_INCOME BETWEEN 200000 AND 500000 THEN 'MID INCOME'
```

```

        ELSE 'HIGH INCOME'
    END AS INCOME_BRACKET,
    ROUND(AVG(SCORE), 2) AS AVERAGE_SCORE
FROM
    EXAM_SCORES E
JOIN
    STUDENTS S ON E.STUDENT_ID = S.STUDENT_ID
GROUP BY
    INCOME_BRACKET
ORDER BY
    AVERAGE_SCORE DESC;

```

	INCOME_BRACKET	...	AVERAGE_SCORE
1	HIGH INCOME		50.04
2	LOW INCOME		50.04
3	MID INCOME		49.97

- We can observe that income bracket doesn't correlate with better performance as students whose parents have low income tend to perform better than students with mid income parents.

## B. ATTENDANCE INSIGHTS

--Creating view of all the attendance tables

```

CREATE OR REPLACE VIEW ATTENDANCE AS
    SELECT * FROM ATTENDANCE1
    UNION ALL
    SELECT * FROM ATTENDANCE2
    UNION ALL
    SELECT * FROM ATTENDANCE3
    UNION ALL
    SELECT * FROM ATTENDANCE4
    UNION ALL
    SELECT * FROM ATTENDANCE5
    UNION ALL
    SELECT * FROM ATTENDANCE6
    UNION ALL
    SELECT * FROM ATTENDANCE7
    UNION ALL
    SELECT * FROM ATTENDANCE8
    UNION ALL
    SELECT * FROM ATTENDANCE9
    UNION ALL
    SELECT * FROM ATTENDANCE10

```

```
UNION ALL
SELECT * FROM ATTENDANCE11
UNION ALL
SELECT * FROM ATTENDANCE12
UNION ALL
SELECT * FROM ATTENDANCE13
UNION ALL
SELECT * FROM ATTENDANCE14
UNION ALL
SELECT * FROM ATTENDANCE15
UNION ALL
SELECT * FROM ATTENDANCE16
UNION ALL
SELECT * FROM ATTENDANCE17
UNION ALL
SELECT * FROM ATTENDANCE18
UNION ALL
SELECT * FROM ATTENDANCE19;
```

**-- Attendance percentage per student**

```
SELECT
    S.STUDENT_ID,
    S.GRADE_LEVEL,
    ROUND(100*SUM(CASE WHEN A.STATUS = 'Present' THEN 1 ELSE 0 END)/COUNT(*), 2) AS
ATTENDANCE_PERCENTAGE
FROM
    STUDENTS S
JOIN
    ATTENDANCE A ON S.STUDENT_ID = A.STUDENT_ID
GROUP BY
    S.STUDENT_ID, S.GRADE_LEVEL
ORDER BY
    ATTENDANCE_PERCENTAGE DESC;
```

	STUDENT_ID	...	GRADE_LEVEL	ATTENDANCE_PERCENTAGE
1	S1141		4	89.38
2	S6147		2	88.72
3	S2855		9	88.72
4	S10794		6	88.72
5	S1234		9	88.72
6	S8915		10	88.61
7	S1694		1	88.61
8	S6920		6	88.61
9	S9369		10	88.61
10	S10643		8	88.50
11	S4404		8	88.50
12	S9040		7	88.50
13	S2347		8	88.39
14	S9569		6	88.39
15	S4527		1	88.39
16	S3302		2	88.39
17	S4694		4	88.39
18	S4641		11	88.39
19	S6437		3	88.39

	STUDENT_ID	...	GRADE_LEVEL	ATTENDANCE_PERCENTAGE
1	S9516		1	80.50
2	S3557		11	80.61
3	S1813		4	81.16
4	S6796		4	81.27
5	S10132		7	81.27
6	S3407		1	81.27
7	S8624		5	81.38
8	S10334		1	81.38
9	S6296		8	81.38
10	S10828		1	81.49
11	S3965		7	81.49
12	S9501		12	81.49
13	S9796		3	81.49
14	S6097		9	81.60
15	S1387		11	81.60
16	S6378		2	81.60
17	S10742		7	81.60

- Students tend to be regular for all the classes as we can see attendance percentage of more than 80% for all the students.
- The student with the ID: S1141 has the highest attendance whereas the lowest percentage stands at 80.5% for the student with ID: S9516

## -- Attendance vs. Exam Performance

```
SELECT
  S.STUDENT_ID,
```

```

ROUND(100*SUM(CASE WHEN A.STATUS = 'Present' THEN 1 ELSE 0 END)/COUNT(*), 2) AS
ATTENDANCE_PERCENTAGE,
ROUND(AVG(SCORE),2) AS AVERAGE_SCORE
FROM
STUDENTS S
JOIN
ATTENDANCE A ON S.STUDENT_ID = A.STUDENT_ID
JOIN
EXAM_SCORES E ON S.STUDENT_ID = E.STUDENT_ID
GROUP BY
S.STUDENT_ID;

```

	STUDENT_ID	ATTENDANCE_PERCENTAGE	AVERAGE_SCORE
1	S2163	84.99	54.35
2	S2169	87.51	49.67
3	S2260	87.08	56.06
4	S2405	84.67	56.39
5	S2414	86.31	47.06
6	S2386	85.65	51.90
7	S2504	85.21	45.06
8	S2512	86.42	48.54
9	S2568	85.32	48.94
10	S2574	84.45	59.53
11	S3786	87.51	52.15
12	S3854	85.87	48.13
13	S4046	86.09	46.55
14	S4047	83.79	44.05
15	S4121	84.56	51.18
16	S4125	85.54	62.73
17	S4189	85.76	41.58
18	S3329	84.88	42.07
19	S3381	84.45	58.53
20	S3384	87.62	51.08
21	S3419	85.10	50.26
22	S3427	85.32	51.74
23	S3506	86.53	48.71
24	S3508	84.23	50.97
25	S3641	85.32	44.91

- As we can observe from the table that higher attendance percentage is not correlated to better performance in exams.



- There are students with higher attendance but average score is less.
- This shows that performance in exams solely depends on how much each has prepared well for the exams and also, depends on their intellectual capability.

### C. TEACHER EFFECTIVENESS

-- Teachers/subjects showing consistently high/low performance

```
SELECT
    T.TEACHER_ID,
    CONCAT(T.FIRST_NAME, T.LAST_NAME) AS TEACHER_NAME,
    T.SUBJECT,
    T.YEARS_EXPERIENCE,
    T.TEACHER_RATING,
    ROUND(AVG(E.SCORE), 2) AS AVERAGE_STUDENT_SCORE
FROM
    TEACHERS T
JOIN
    CLASSES C ON T.TEACHER_ID = C.TEACHER_ID
JOIN
    EXAM_SCORES E ON C.STUDENT_ID = E.STUDENT_ID
GROUP BY
    T.TEACHER_ID, TEACHER_NAME, T.SUBJECT, T.YEARS_EXPERIENCE, T.TEACHER_RATING
ORDER BY
    AVERAGE_STUDENT_SCORE DESC;
```

	TEACHER_ID	TEACHER_NAME	SUBJECT	YEARS_EXPERIENCE	TEACHER_RATING	AVERAGE_STUDENT_SCORE
1	T508	Alyssa Carpenter	English	20	3.0	52.70
2	T288	Shelley Lopez	Science	23	4.6	52.43
3	T243	Martha Patel	History	11	2.4	52.26
4	T422	Tammie Cohen	Science	13	4.7	52.25
5	T252	Timothy Hernandez	Sports	31	4.1	52.18
6	T450	Derek Baker	History	1	4.2	52.11
7	T145	Virginia Campbell	Computer	14	4.3	52.11
8	T369	Cassie Coleman	Arts	11	3.5	52.07
9	T328	Heather Hamilton	Sports	1	4.0	52.04
10	T171	Eric Rogers	Sports	32	4.3	52.01
11	T265	Gregory Palmer	Arts	20	2.1	51.95
12	T510	Melissa Ballard	Math	22	3.4	51.92
13	T504	Martin Garner	Science	21	2.1	51.90
14	T483	Nancy Webb	Science	29	2.2	51.90
15	T269	Jeffrey Parks	Science	22	4.7	51.86
16	T349	David Fitzpatrick	Math	26	4.7	51.85
17	T300	Vickie Bernard	Arts	35	4.4	51.82
18	T215	Joseph Mitchell	English	5	4.8	51.81
19	T541	Justin Flores	English	33	3.6	51.80
20	T201	Elizabeth Thomas	History	3	2.1	51.74
21	T577	Chelsey Norman	Science	33	4.9	51.74
22	T526	Cheryl Burke	Science	27	4.2	51.73
23	T237	Tony Smith	Math	26	2.0	51.63
24	T432	Judith Lee	English	14	2.8	51.63
25	T580	Eric Taylor	English	6	4.6	51.60
26	T182	Richard Ayers	Computer	6	4.9	51.60
27	T105	Joseph Tapia	Science	11	3.4	51.60
28	T582	Kathy Odom	Computer	9	2.8	51.60
29	T313	Rachel Martin	Arts	21	4.4	51.56
30	T216	Valerie Bryant	Math	14	4.0	51.49

- After reviewing the data, it looks like teacher experience and teacher ratings don't have a strong impact on student performance.
- For example, in Science, students scored an average of 52.43 with a highly rated teacher (rating 4.6), while another group scored 51.9 with a teacher who had a much lower rating (2.2).
- This suggests that student performance depends more on how well they prepare and how much effort they put in (like following a proper study plan or attending extra classes), rather than just the teacher's experience or rating.

## -- Teacher qualification impact

```

SELECT
    T.QUALIFICATION,
    ROUND(AVG(E.SCORE), 2) AS AVERAGE_STUDENT_SCORE
FROM
    TEACHERS T
JOIN
    CLASSES C ON T.TEACHER_ID = C.TEACHER_ID
JOIN
    EXAM_SCORES E ON C.STUDENT_ID = E.STUDENT_ID
GROUP BY
    T.QUALIFICATION

```

ORDER BY

AVERAGE\_STUDENT\_SCORE DESC;

	QUALIFICATION	...	AVERAGE_STUDENT_SCORE
1	B.Ed		50.05
2	PhD		50.01
3	M.Ed		49.99

- The average student scores are almost identical regardless of whether the teacher holds a B.Ed, M.Ed, or PhD (all hover around 50).
- This indicates that teacher qualification level does not significantly impact student exam performance in this dataset.
- Students taught by teachers with a PhD scored about the same as those taught by teachers with only a B.Ed.
- Teacher qualifications don't seem to make a noticeable difference in student exam results. Whether a teacher has a B.Ed, M.Ed, or even a PhD, the average scores remain nearly the same. This suggests that other factors such as student study habits, or access to additional learning resources, may play a bigger role in academic performance than formal qualifications alone.

#### D. EXTRACURRICULAR ACTIVITIES

-- Performance difference between participants and non-participants

SELECT

CASE WHEN A.STUDENT\_ID IS NOT NULL THEN 'Participant' ELSE 'Non-Participant' END AS  
ACTIVITY\_STATUS,

ROUND(AVG(E.SCORE), 2) AS AVERAGE\_SCORE

FROM

EXAM\_SCORES E

LEFT JOIN

ACTIVITIES A ON E.STUDENT\_ID = A.STUDENT\_ID

GROUP BY

ACTIVITY\_STATUS;

	ACTIVITY_STATUS	...	AVERAGE_SCORE
1	Participant		50.03
2	Non-Participant		50.58

- While it is good to involve yourself in extra curricular activities, but as per this dataset, participating in extra curricular activities doesn't show any impact on students' performance.
- Students involved in activities have an average score of 50.03 while non-participating students have an average score of 50.58.

## -- Activities with the highest academic correlation

```
SELECT
  A.ACTIVITY_TYPE,
  ROUND(AVG(E.SCORE), 2) AS AVERAGE_STUDENT_SCORE,
  ROUND(AVG(HOURS_PER_WEEK), 2) AS AVERAGE_ACTIVITY_HOURS
FROM
  ACTIVITIES A
JOIN
  EXAM_SCORES E ON A.STUDENT_ID = E.STUDENT_ID
GROUP BY
  A.ACTIVITY_TYPE
ORDER BY
  AVERAGE_STUDENT_SCORE DESC;
```

	ACTIVITY_TYPE	AVERAGE_STUDENT_SCORE	AVERAGE_ACTIVITY_HOURS
1	Dance	50.09	7.57
2	Sports	50.06	7.50
3	Debate	50.06	7.53
4	Drama	50.05	7.44
5	STEM Club	50.02	7.56
6	Music	49.99	7.56
7	Art	49.93	7.53

- Students who participated in activities like Dance or Sports tend to perform better in academics as per the above dataset.
- While difference is very small as compared to other activities, still we can see a better average score.
- This is also proven in one of the research studies which was conducted extensively like for 18 months where participants who were made to dance showed better improvement in cognition and neuroplasticity development as compared to other activities like exercises or sports.
- Students should be encouraged to be participate in extra curricular activities for their overall development even though that may not directly translate into better performance in academics.

## 5. DISCIPLINARY IMPACT

### -- Disciplinary records vs. average score

```
SELECT
  D.INCIDENT_TYPE,
  D.SEVERITY,
  ROUND(AVG(E.SCORE), 2) AS AVERAGE_STUDENT_SCORE
```

```

FROM
  DISCIPLINARY_RECORDS D
JOIN
  EXAM_SCORES E ON D.STUDENT_ID = E.STUDENT_ID
GROUP BY
  D.INCIDENT_TYPE, D.SEVERITY
ORDER BY
  AVERAGE_STUDENT_SCORE;

```

	INCIDENT_TYPE	SEVERITY	...	AVERAGE_STUDENT_SCORE
1	Late Submission	Medium		49.85
2	Bullying	Medium		49.91
3	Bullying	Low		49.94
4	Bullying	High		49.96
5	Misconduct	Medium		49.96
6	Misconduct	Low		49.99
7	Cheating	High		50.00
8	Other	High		50.01
9	Misconduct	High		50.01
10	Cheating	Medium		50.02
11	Late Submission	High		50.03
12	Other	Medium		50.04
13	Late Submission	Low		50.11
14	Cheating	Low		50.11
15	Other	Low		50.11

- The average student scores are almost the same (between 49.8 and 50.1) across all incident types (Bullying, Misconduct, Cheating, Late Submission, Other).
- Even when the severity increases (Low → Medium → High), the scores barely change.
- This means that disciplinary incidents don't appear to have a major impact on student exam performance in your dataset.
- Similarly, the severity of the incident (low, medium, or high) does not significantly lower average scores.
- Students with disciplinary issues (like bullying, cheating, or misconduct) scored almost the same as students with other types of incidents. The seriousness of the incident also doesn't seem to affect their scores much. This suggests that in this dataset, disciplinary actions are not strongly linked to academic performance.

## -- Disciplined vs. Non-disciplined

```
SELECT
  CASE WHEN D.STUDENT_ID IS NOT NULL THEN 'Non-Disciplined' ELSE 'Disciplined' END AS
  DISCIPLINARY_STATUS,
  ROUND(AVG(E.SCORE), 2) AS AVERAGE_STUDENT_SCORE
FROM
  EXAM_SCORES E
LEFT JOIN
  DISCIPLINARY_RECORDS D ON E.STUDENT_ID = D.STUDENT_ID
GROUP BY
  DISCIPLINARY_STATUS;
```

	DISCIPLINARY_STATUS	...	AVERAGE_STUDENT_SCORE
1	Disciplined		50.33
2	Non-Disciplined		50.00

- The data shows that students with disciplinary records (average score: 50.33) perform almost the same as those without any records (average score: 50.00).
- This suggests that disciplinary actions are more related to behavior and environment rather than academic ability.
- In other words, having a disciplinary issue doesn't mean a student is less intelligent or performs worse in exams, it may simply reflect their surroundings, habits or peer influence.

## CONCLUSION:

- The comprehensive SQL analysis conducted using Snowflake explored various facets of student performance, attendance, teacher effectiveness, extracurricular activities, and disciplinary records within the school dataset. This investigation involved thorough data quality audits, including checking for null values, duplicates, and data ranges across multiple tables like STUDENTS, TEACHERS, CLASSES, EXAM\_SCORES, ATTENDANCE1, ACTIVITIES, and DISCIPLINARY\_RECORDS.
- Key findings from the Student Performance Analysis revealed varied pass/fail rates across grades and subjects, with Grade 7 exhibiting the highest overall fail rate at 40.02%. While subject-specific differences in performance by gender were observed (females in Science, males in Arts), overall fail rates between genders were similar. Parental education level showed a slight, positive influence on student performance. Intriguingly, family income did not correlate with better performance, with students from low-income brackets sometimes outperforming those from mid-income backgrounds.
- Insights from Attendance data indicated high student regularity, with all students maintaining over 80% attendance. However, a critical observation was that higher attendance did not

correlate with better exam performance; rather, individual preparation and intellectual capability appeared to be more significant determinants of success.

- The analysis of Teacher Effectiveness yielded noteworthy results: teacher experience, ratings, and formal qualifications (B.Ed, M.Ed, or PhD) did not significantly impact student exam scores. Average student scores remained consistent regardless of these teacher attributes, suggesting that student performance is more profoundly influenced by their personal effort, study habits, and access to learning resources.
- Regarding Extracurricular Activities, participation generally did not show a major impact on academic performance, although students involved in Dance or Sports recorded slightly higher average scores. The analysis supports encouraging participation for overall student development, even if a direct academic correlation is not always evident.
- Finally, the Disciplinary Impact analysis demonstrated that disciplinary incidents, their specific type, or severity did not significantly affect student exam performance. Students with disciplinary records performed academically similar to those without, implying that disciplinary issues are more related to behavioral aspects, environment, habits, or peer influence rather than academic ability.
- In conclusion, this comprehensive SQL analysis highlights that while various factors contribute to the school environment, student academic performance within this dataset is primarily driven by individual preparation, consistent study habits, and inherent intellectual capability, with many commonly assumed influences such as teacher qualifications or disciplinary records showing minimal direct correlation to exam results.