# Computer Science Students Career Analysis

Created by:

Shamil Sharifzade

COMPUTER SCIENCE
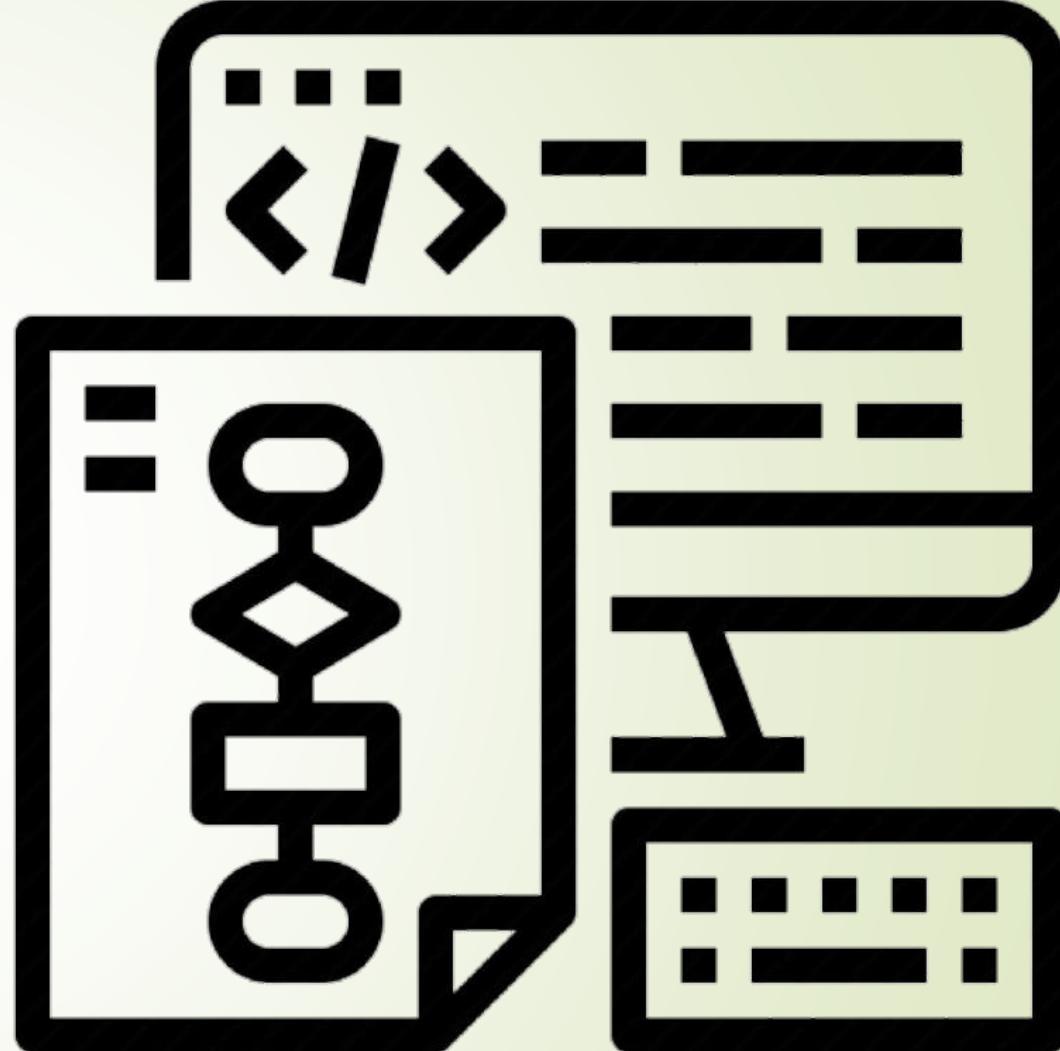
Table Columns:

Student ID: Unique identifier for each student.

Name: Name of the student.

Gender: Gender of the student.

Age: Age of the student.

GPA: Grade Point Average of the student.

Major: Field of study within computer science.

Interested Domain: Area of interest within the field of computer science.

Projects: Noteworthy projects completed by the student.

Python: Proficiency level in Python programming.

SQL: Proficiency level in SQL querying.

Java: Proficiency level in Java programming.

- **Future Career:** Intended career path or job aspiration.

Data Definition Language (DDL) operation

```sql
CREATE TABLE cs_students (
    student_id INTEGER PRIMARY KEY,
    "name" VARCHAR(50),
    gender VARCHAR(10) CHECK (gender IN ('Male', 'Female')),
    "age" INTEGER,
    gpa FLOAT,
    major VARCHAR(50),
    interested_domain VARCHAR(50),
    projects VARCHAR(50),
    future_career VARCHAR(50),
    "python" VARCHAR(10) CHECK ("python" IN ('Strong', 'Weak', 'Average')),
    "sql" VARCHAR(10) CHECK ("sql" IN ('Strong', 'Weak', 'Average')),
    "java" VARCHAR(10) CHECK ("java" IN ('Strong', 'Weak', 'Average'))
)
```

```
1  SELECT
2      gender,
3      round(avg(gpa)::NUMERIC, 3) AS avg_gpa
4  FROM public.cs_students
5  GROUP BY gender
```

gender	avg_gpa
1 Female	3.628
2 Male	3.605

```
1  SELECT
2      round(avg(gpa)::NUMERIC, 3) AS avg_gpa
3  FROM public.cs_students
```

avg_gpa
1 3.615

What is the average GPA of all students and for each gender?

What is the gender distribution among computer science students?

➡ In our analysis of computer science students, we found an interesting pattern in gender distribution. Out of the total students, **78 are female**, making up **43.33%** of the entire group. Meanwhile, **102 students are male**, representing **56.67%** of the population.

➡ This shows that while both genders are well represented in the field, **male students still slightly outnumber female students**. However, the gap isn't too wide — indicating that **more women are actively joining the computer science field**, showing a positive trend toward **greater gender balance and inclusivity** in technology education.

```
1  SELECT
2      gender,
3      count(gender) AS student_count,
4      round(
5          count(gender)::NUMERIC / (SELECT count(*) FROM public.cs_students) * 100,
6          2
7      ) AS percentage
8  FROM public.cs_students
9  GROUP BY gender
```

gender	student_count	percentage
1 Female	78	43.33
2 Male	102	56.67

What are the most common interested domains and future career among students?

```sql
(
    SELECT
        interested_domain,
        COUNT(*) AS domain_count,
        NULL AS future_career,
        NULL AS career_count
    FROM public.cs_students
    GROUP BY interested_domain
    ORDER BY domain_count DESC
    LIMIT 10
)
UNION ALL
(
    SELECT
        NULL AS interested_domain,
        NULL AS domain_count,
        future_career,
        COUNT(*) AS career_count
    FROM public.cs_students
    GROUP BY future_career
    ORDER BY career_count DESC
    LIMIT 10
)
ORDER BY
    domain_count DESC NULLS LAST,
    career_count DESC NULLS LAST
```

	interested_domain	domain_count	future_career	career_count
1	Artificial Intelligence	19	<NULL>	<NULL>
2	Web Development	19	<NULL>	<NULL>
3	Cybersecurity	18	<NULL>	<NULL>
4	Mobile App Development	17	<NULL>	<NULL>
5	Database Management	16	<NULL>	<NULL>
6	Cloud Computing	16	<NULL>	<NULL>
7	Data Science	13	<NULL>	<NULL>
8	Computer Graphics	13	<NULL>	<NULL>
9	Machine Learning	13	<NULL>	<NULL>
10	Software Development	11	<NULL>	<NULL>
11	<NULL>	<NULL>	Web Developer	19
12	<NULL>	<NULL>	Mobile App Developer	17
13	<NULL>	<NULL>	Information Security Analyst	17
14	<NULL>	<NULL>	Cloud Solutions Architect	16
15	<NULL>	<NULL>	Database Administrator	16
16	<NULL>	<NULL>	Software Engineer	14
17	<NULL>	<NULL>	Machine Learning Engineer	13
18	<NULL>	<NULL>	NLP Research Scientist	11
19	<NULL>	<NULL>	Graphics Programmer	9
20	<NULL>	<NULL>	Data Scientist	8

Which interested domain has the highest average GPA?

```sql
1   SELECT
2       interested_domain,
3       round(avg(gpa)::NUMERIC, 2) AS avg_gpa
4   FROM public.cs_students
5   GROUP BY interested_domain
6   ORDER BY avg_gpa DESC
```

	interested_domain	avg_gpa
1	Natural Language Processing	3.9
2	Distributed Systems	3.9
3	Data Privacy	3.8
4	Software Engineering	3.8
5	Quantum Computing	3.8

Is there a difference in average Python, SQL, or Java proficiency between interested domain?

```sql
SELECT
    interested_domain,
    MODE() WITHIN GROUP (ORDER BY python) AS mode_python,
    MODE() WITHIN GROUP (ORDER BY "sql") AS mode_sql,
    MODE() WITHIN GROUP (ORDER BY "java") AS mode_java
FROM public.cs_students
GROUP BY interested_domain
```
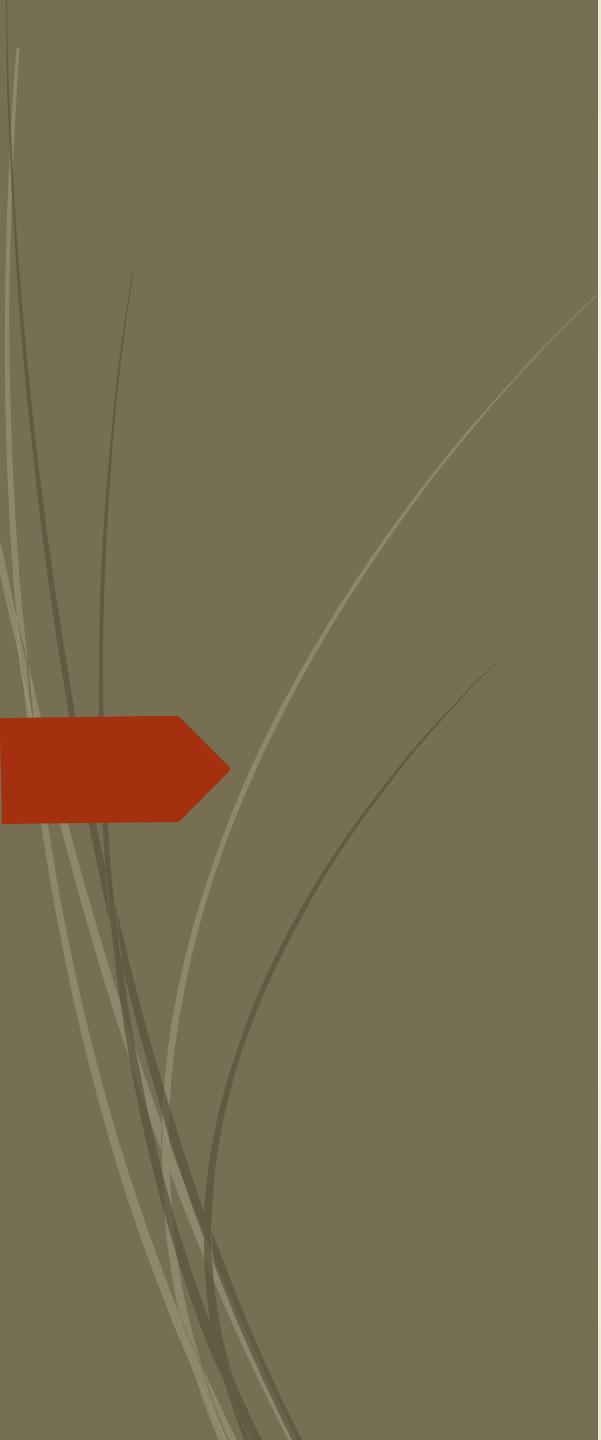
	interested_domain	mode_python	mode_sql	mode_java
1	Artificial Intelligence	Strong	Average	Weak
2	Bioinformatics	Weak	Average	Strong
3	Biomedical Computing	Weak	Average	Average
4	Blockchain Technology	Weak	Average	Strong
5	Cloud Computing	Strong	Strong	Average
6	Computer Graphics	Weak	Average	Strong
7	Computer Vision	Strong	Average	Weak
8	Cybersecurity	Average	Weak	Strong
9	Data Mining	Strong	Average	Weak
10	Data Privacy	Weak	Average	Strong
11	Data Science	Average	Strong	Weak
12	Database Management	Average	Strong	Average
13	Digital Forensics	Weak	Average	Strong

Is there a correlation between age and GPA?

```sql
1  WITH find_covariance AS (
2      SELECT
3          SUM((age - avg_age) * (gpa - avg_gpa)) / (COUNT(*) - 1) AS covariance
4      FROM (
5          SELECT
6              age,
7              gpa,
8              (SELECT AVG(age) FROM public.cs_students) AS avg_age,
9              (SELECT AVG(gpa) FROM public.cs_students) AS avg_gpa
10         FROM public.cs_students
11     ) AS subquery
12 )
13 SELECT
14     covariance / (STDDEV_SAMP(age) * STDDEV_SAMP(gpa)) AS correlation
15 FROM public.cs_students, find_covariance
16 GROUP BY covariance
```

correlation
1 0.2029246915972213

There is a weak positive linear relationship between students' age and GPA. Older students tend to have slightly higher GPAs on average, but the effect is not strong.

Thanks for your attention

Shamil Sharifzade