

## Deciphering Big Data – Unit 7

### Normalisation Task

#### Instructions

Below you will see a table with data in un-normalised form. You should normalise this data to 3rd Normal Form (3NF), showing each step of the process i.e., demonstrating 1NF, 2NF and 3NF. Use the table below or [download the Excel file](#).

Add your findings to your e-portfolio.

Table

Student Number	Student Name	Exam Score	Support	Date of Birth	Course Name	Exam Boards	Teacher Name
1001	Bob Baker	78	No	25/08/2001	Computer Science	BCS	Mr Jones
					Maths	EdExcel	Ms Parker
					Physics	OCR	Mr Peters
1002	Sally Davies	55	Yes	02/10/1999	Maths	AQA	Ms Parker
					Biology	WJEC	Mrs Patel
					Music	AQA	Ms Daniels
1003	Mark Hanmill	90	No	05/06/1995	Computer Science	BCS	Mr Jones
					Maths	EdExcel	Ms Parker
					Physics	OCR	Mr Peters
1004	Anas Ali	70	No	03/08/1980	Maths	AQA	Ms Parker
					Physics	OCR	Mr Peters
					Biology	WJEC	Mrs Patel
1005	Cheuk Yin	45	Yes	01/05/2002	Computer Science	BCS	Mr Jones
					Maths	EdExcel	Ms Parker
					Music	AQA	Ms Daniels

#### Unnormalized Form (UNF)

The provided table is unnormalized as it contains **repeating groups** (multiple courses, exam boards, and teacher names within the same row).

Student Number	Student Name	Exam Score	Support	Date of Birth	Course Name	Exam Boards	Teacher Name
1001	Bob Baker	78	No	25/08/2001	Computer Science	BCS	Mr Jones
					Maths	EdExcel	Ms Parker
					Physics	OCR	Mr Peters
1002	Sally Davies	55	Yes	02/10/1999	Maths	AQA	Ms Parker
					Biology	WJEC	Mrs Patel
					Music	AQA	Ms Daniels
1003	Mark Hanmill	90	No	05/06/1995	Computer Science	BCS	Mr Jones
					Maths	EdExcel	Ms Parker
					Physics	OCR	Mr Peters
1004	Anas Ali	70	No	03/08/1980	Maths	AQA	Ms Parker
					Physics	OCR	Mr Peters
					Biology	WJEC	Mrs Patel
1005	Cheuk Yin	45	Yes	01/05/2002	Computer Science	BCS	Mr Jones
					Maths	EdExcel	Ms Parker
					Music	AQA	Ms Daniels

### First Normal Form (1NF)

- Ensure each cell contains atomic values.
- Remove multivalued attributes by creating a new row for each combination of values.

1NF Table:

Student Number	Student Name	Exam Score	Support	Date of Birth	Course Name	Exam Board	Teacher Name
1001	Bob Baker	78	No	25/08/2001	Computer Science	BCS	Mr Jones
1001	Bob Baker	78	No	25/08/2001	Maths	EdExcel	Ms Parker
1001	Bob Baker	78	No	25/08/2001	Physics	OCR	Mr Peters
1002	Sally Davies	55	Yes	02/10/1999	Maths	AQA	Ms Parker
1002	Sally Davies	55	Yes	02/10/1999	Biology	WJEC	Mrs Patel
1002	Sally Davies	55	Yes	02/10/1999	Music	AQA	Ms Daniels
1003	Mark Hanmill	90	No	05/06/1995	Computer Science	BCS	Mr Jones
1003	Mark Hanmill	90	No	05/06/1995	Maths	EdExcel	Ms Parker
1003	Mark Hanmill	90	No	05/06/1995	Physics	OCR	Mr Peters
1004	Anas Ali	70	No	03/08/1980	Maths	AQA	Ms Parker
1004	Anas Ali	70	No	03/08/1980	Physics	OCR	Mr Peters
1004	Anas Ali	70	No	03/08/1980	Biology	WJEC	Mrs Patel
1005	Cheuk Yin	45	Yes	01/05/2002	Computer Science	BCS	Mr Jones
1005	Cheuk Yin	45	Yes	01/05/2002	Maths	EdExcel	Ms Parker
1005	Cheuk Yin	45	Yes	01/05/2002	Music	AQA	Ms Daniels

## 2nd Normal Form (2NF)

- Remove partial dependencies: Create separate tables to ensure all non-key attributes are fully dependent on the whole primary key.

Table 1: Student Information

Student Number	Student Name	Exam Score	Support	Date of Birth
1001	Bob Baker	78	No	25/08/2001
1002	Sally Davies	55	Yes	02/10/1999
1003	Mark Hanmill	90	No	05/06/1995
1004	Anas Ali	70	No	03/08/1980
1005	Cheuk Yin	45	Yes	01/05/2002

Table 2: Course Information

Course Name	Exam Boards	Teacher Name
Computer Science	BCS	Mr Jones
Maths	EdExcel	Ms Parker
Physics	OCR	Mr Peters
Biology	WJEC	Mrs Patel
Music	AQA	Ms Daniels

Table 3: Student-Course Enrollment

Student Number	Course Name
1001	Computer Science
1001	Maths
1001	Physics
1002	Maths
1002	Biology
1002	Music
1003	Computer Science
1003	Maths
1003	Physics
1004	Maths
1004	Physics
1004	Biology
1005	Computer Science
1005	Maths
1005	Music

### 3rd Normal Form (3NF)

- Remove transitive dependencies: Separate teacher details if necessary.

Table 1: Student Information (unchanged)

Student Number	Student Name	Exam Score	Support	Date of Birth
1001	Bob Baker	78	No	25/08/2001
1002	Sally Davies	55	Yes	02/10/1999
1003	Mark Hanmill	90	No	05/06/1995
1004	Anas Ali	70	No	03/08/1980
1005	Cheuk Yin	45	Yes	01/05/2002

Table 2: Course Information (with teacher IDs)

Course Name	Exam Boards	Teacher ID
Computer Science	BCS	T1
Maths	EdExcel	T2
Physics	OCR	T3
Biology	WJEC	T4
Music	AQA	T5

Table 3: Teacher Information

Teacher ID	Teacher Name
T1	Mr Jones
T2	Ms Parker
T3	Mr Peters
T4	Mrs Patel
T5	Ms Daniels

Table 4: Student-Course Enrollment (unchanged)

Student Number	Course Name
1001	Computer Science
1001	Maths
1001	Physics
1002	Maths
1002	Biology
1002	Music
1003	Computer Science
1003	Maths
1003	Physics
1004	Maths
1004	Physics
1004	Biology
1005	Computer Science
1005	Maths
1005	Music