A RESEARCH ON DATABASE TESTING

DATABASE:

1. A database is an organized collection of data stored electronically and managed by a Data Base Management System (DBMS).
2. A DBMS manages and controls access to a DB. Examples include MongoDB, MariaDB, SQL.

DATABASE TESTING:

1. Database testing is the process of verifying the accuracy, reliability and performance of a Data Base system.
2. It can be performed manually, automatically and in a hybrid way (combined procedure of automation and manual).
3. It is also called as backend testing as it mainly focuses on the server side which isn't visible to the user.

IMPORTANCE OF DATABASE TESTING:

Database testing saves data loss and aborted transaction data

Transaction here refers to the access and retrieval of data and transaction follows a few properties like:

1. ATOMICITY: If any transaction is performed on data, it should happen completely or shouldn't be implemented.
2. CONSISTENCY: Database must be valid and preserved after the transaction is complete.
3. DURABILITY: Modifications will be kept without any fail when the transaction is committed.
4. ISOLATION

Database testing helps identify and mitigate security vulnerabilities, such as SQL injection attacks, which could compromise sensitive data.

Early detection and resolution of database issues during testing can significantly reduce the cost of fixing problems after deployment.

TYPES OF DATABASE TESTING:

There are three types of database testing. They include:

1. STRUCTURAL:

   This type of database testing mainly focuses on the internal structure of the Data Base. Problems tested here include:

   i) Data Type issues
   ii) Constraint Testing (primary key, foreign key)

iii) Referential integrity checks

2. FUNCTIONAL:

This type of database testing is mainly used for checking whether the Data Base works according to the specified requirements. Problems tested here include:

i) CRUD [Create-Read-Update-Delete]
ii) Data manipulation through application interface

3. NON-FUNCTIONAL:

This type of database testing helps in measuring performance, scalability, reliability and security of the Data Base. Types include:

i) PERFORMANCE TEST- response time for each query
ii) LOAD TEST- handling large volume of data/ users
iii) SECURITY TEST- finds all possible loopholes/ weaknesses.
iv) STRESS TEST- checks whether DB works beyond normal workload
v) COMPATIBILITY TEST- aims to check the developed software application functions on many software and hardware platforms.

The Non-Functional database testing is mainly automated due to its complexity whereas the structural and functional testing can be done in both ways (automated and manual too).

TOOLS THAT ARE USED FOR DATABASE TESTING:

i) MANUAL:

1)MySQL Work Bench 2) Oracle SQL Developer 3) pG Admin

ii) Automated functions and structured DB Testing:

a) NoSQLUnit- Used for NoSQL databases.
b) pgTAP- SQL Unit Testing Framework
c) SQLTest: Automated testing of queries, procedures and result completion.

Tools for Automated Non-Functional Testing:
i) Apache JMeter: For load and performance testing
ii) LoadRunner- Load testing app
iii) OWASP-ZAP: Used for security testing

SELF HEALING:

Self-healing is the ability of automated tests to automatically adapt changes in the database schema/ data without manual intervention.

It involves using AI and ML techniques to identify, correct issues arising from Data Base updates.

They are needed in database testing as:

i) Renaming of column names
ii) Table structure change
iii) Data format change

Self-Healing improves efficiency by accelerating the error detection by automatically detecting and correcting errors.

HOW TO AUTOMATE DATA-BASE TESTING:

STEP1: Define objectives

STEP2: Create good test cases based on our objectives. The test cases include:

i) Data Integrity Test:

a) Primary Key Constraints
b) Foreign key Constraints
c) Unique Constraints

ii) Data Type Validations:

a) Range Checks
b) Format Checks

STEP3: Choose the best automation tool like Apache JMeter, Selenium etc.

STEP4: Setup the test environment.

STEP5: Integrate it with CI/CD.

STEP6: Execute the test.

STEP7: VALIDATE RESULTS AND MAINTAIN TEST SCRIPTS

STEP8: Monitor Performance

MY APPROACH TOWARDS SOLVING THIS PROBLEM:

MY APPROACH IS TO USE OLLAMA which is a large language model released by META AI on the month of FEB 2023.

It is an open-source model unlike GPT which is a proprietary model and is highly customizable, strong in code generation.

This will be run locally and then I will have a database having details of few players.

Then, I will be using Python and in Python I am using the unittest module which is used for automating the data validation logic and mysql.connector for connecting python with MySQL.

In this problem LLAMA will be used to heal the broken SQL queries and create a text file to store the error conditions.

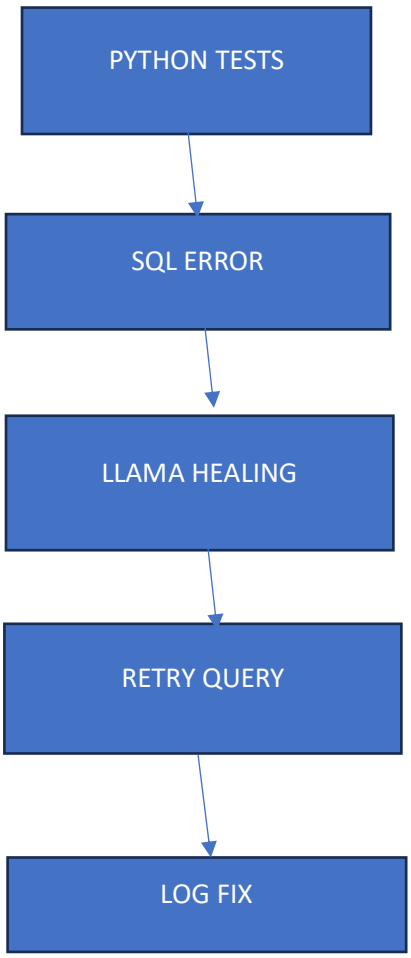


FIG1: ANALYSIS PATTERN OF SELF HEALING IN OUR DATABASE PROBLEM.

WE ALSO HAVE A FEW DATA VALIDATION RULES TO CHECK FOR WRONG TYPES, DUPLICATES.

STEPS INVOLVED IN MY PROCEDURE TO TEST SELF HEALING IN A DATABASE:

Step1: Software installations:

1. IDLE Python- Have the latest version installed (3.13.5). While installing just make sure that you add python.exe to the path which helps in proper installations of modules.

2.MySQL: MySQL is an open-source RDBMS which is fast, secure, reliable and easy to use when compared to the other database managements systems. Work with caution while installing it in your system as the installation is a long procedure.

3. After installing MySQL and Python install the modules which are necessary for completion of our project. Some of the modules which are used here for successful completion of project include:

i) mysql.connector : mysql.connector is a Python library that enables python programs to interact with MySQL databases. In our project one of the important modules is mysql.connector as our project is about self-healing in databases.
ii) Unittest: The unittest module is Python's built in unit testing framework. Along with this, it provides a structured, object-oriented approach to writing and running tests for Python codes.
iii) Datetime: We use the datetime module for mentioning the date at when the database testing takes place and the time at which the testing occurs. So, it provides us in having a log on having a note at the errors which occur in the database and tests the validity of the database.
iv) Json: Json stands for Java Script Object Notation. It is mainly used for transmitting data between a server and web application and also for storing structured data in the form of key- value pair.
v) re: The re module provides support for regular expressions. Regular expressions are powerful tools that enable tasks like searching, splitting, validating text-based patterns.

4. The tool which we use for achieving self-healing of a database along with Python and MySQL is Ollama. Ollama is a tool that is used for running and managing large language models locally on our computer. It simplifies the process of downloading, setting local LLM's and makes them accessible for research, code generation and automation of testing of DB. SO, this tool plays a very major role in our project.

5. Before we proceed into the testing part, let us firstly create a database which consist of details about the CSK Players who are from Australia. Create another database named original which is not used for testing and used only for viewing and knowing details about it.

6. Along with this we, have used a software called Memcached, which is a high performance, distributed memory object caching system. It acts as a short-term memory for applications. Here, the values are stored in the form of key-value pairs which makes it easy for storing and retrieving data. We use this for checking whether the datatype of a particular column has been changed when compared to the original database and if yes, we use self-healing for returning to the datatype of the memcached original database.

Then, here in this project to see the healed query and results of the tests conducted we have a notepad file named healing_log which logs them. For a better viewing we have created a UI for viewing them. For that, we have used the flask module. Flask is a microframework which

is used for web development as it offers a minimal core with optional extensions allowing developers to choose and add functionality.

The flask module is built on top of two powerful libraries:

1. Jinja2
2. Werkzeug

FOLDER STRUCTURE

templates- This sub folder inside the folder sql_self_healing_project contains all the HTML frontend websites which are used for the UI. Inside it we have web-based files used for viewing the original, testing DB, viewing the datatypes of each column of the table present in the database, functions like viewing, updating, deleting and inserting values into the database and a web for viewing the test results that have been logged.

static- This sub-folder has the CSS file which is used for enhancing the web UI.

app.py- This python script serves as the backend for the UI which we have created. It helps in the creation of website by defining its functionality

cache_column_datatypes.py: This python script plays a major role in caching the datatypes of each column present in the database. Used for Memcaching the columns.

db_test_config.py: This python script has details regarding the host, user, password and the database name. It is used to connect the database with python for further testing's along with the mysql.connector module.

schema_healer.py: This python file plays a major role helping the UI application created to dynamically select the table and salary column if the name of the table and salary column changes.

data_validation_test.py; This python folder plays a major role in defining the test cases and perform the automation testing in the database. If any test fails we get an Assertion Error.

healing_test.py: This python script is like data_validation_test.py but here the addition is the healing part. If any of the test fails, the test is taken to LLAMA via Ollama which we have installed and for achieving this, we have another python file named healer.py

The healer.py python file has the prompt which must be sent to LLAMA via Ollama for working of self-healing.

TEST CASES THAT WE WORK ON:

There are four test cases which we test in our self-healing project:

1. Salary being positive
2. Nationality to be ‘Australian’ unlike others as our sole purpose of the database is to have data of players from Australia.

3. Duplicates to be removed.
4. Ipl Team must be Chennai Super Kings unlike the other teams such as SRH, RCB etc.
5. If the column name changes it must get healed automatically and work accordingly These 4 test cases are checked and before running the validation tests, we must ensure that all the source codes required for the testing is kept in the same directory and at the same time make sure that all the modules are being installed for its successful completion of testing.
6. If the datatype of a particular column changes ex: int to float manually, the Ollama present must identify the change in the datatype due to caching of the columns at the datatype which is due to memcaching process done and get converted into the original datatype as per the original one.
7. Suppose the name of the table changes, the Ollama present for the self-healing must identify it and then proceed with the tests.

OUTPUT SCREENSHOT

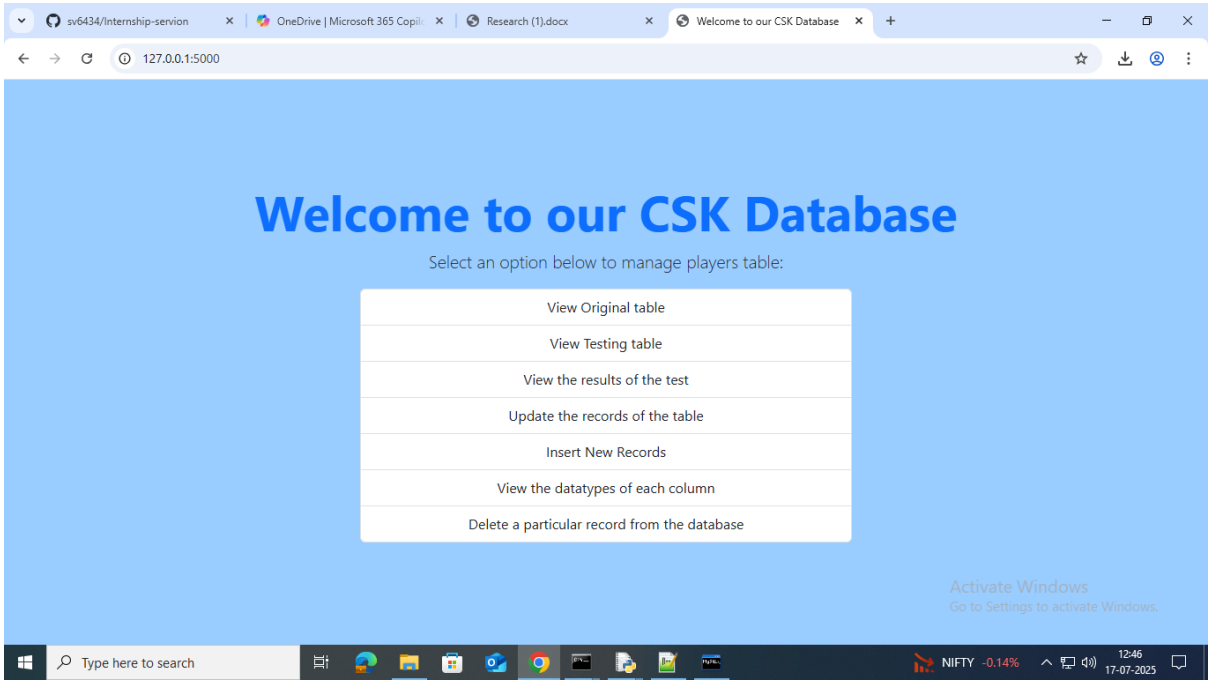


Fig1: The welcome page for our UI application

Fig2: The original database contents. Used only for read purpose and compare it with the testing database

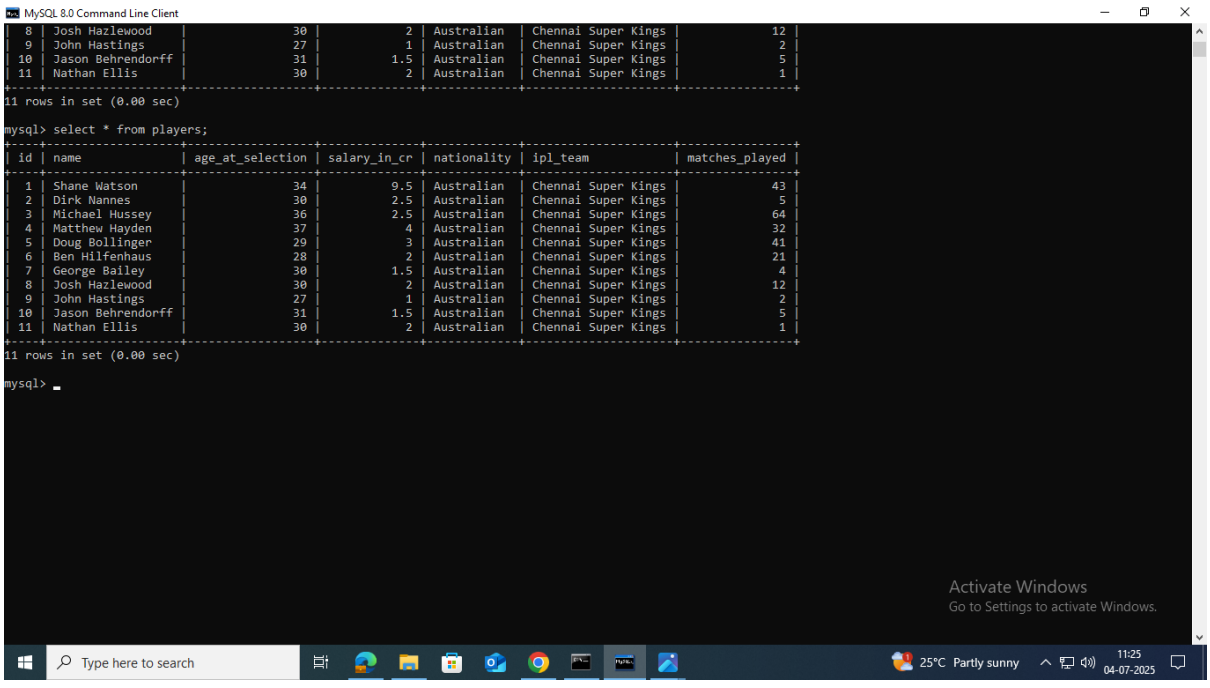


Fig3: The original database in MySQL.

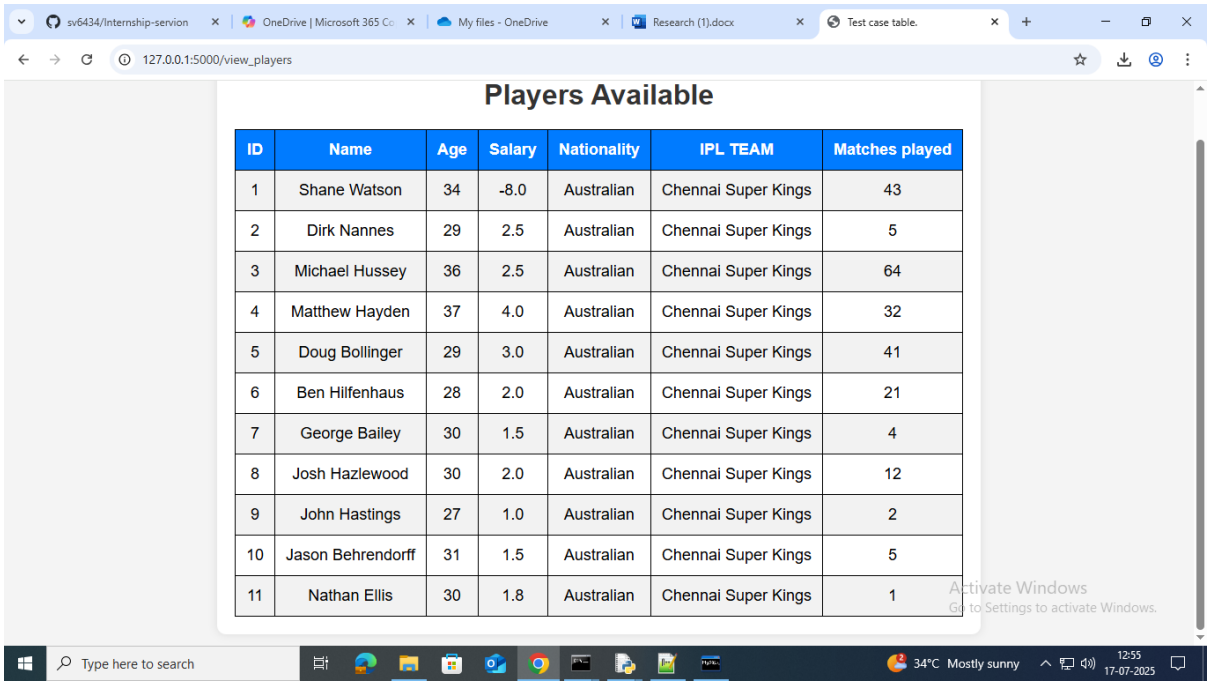


Fig4: The UI part which shows that there is an invalid entry in the salary column.

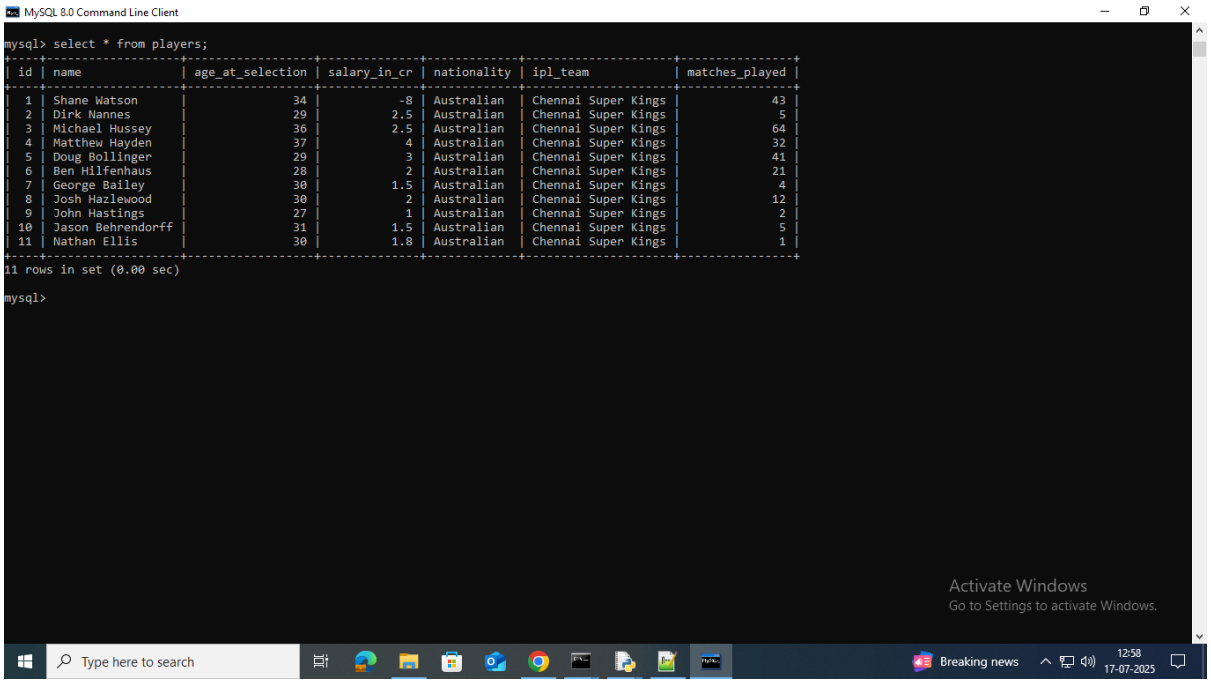


Fig5: The SQL viewer which shows that there is an invalid entry.

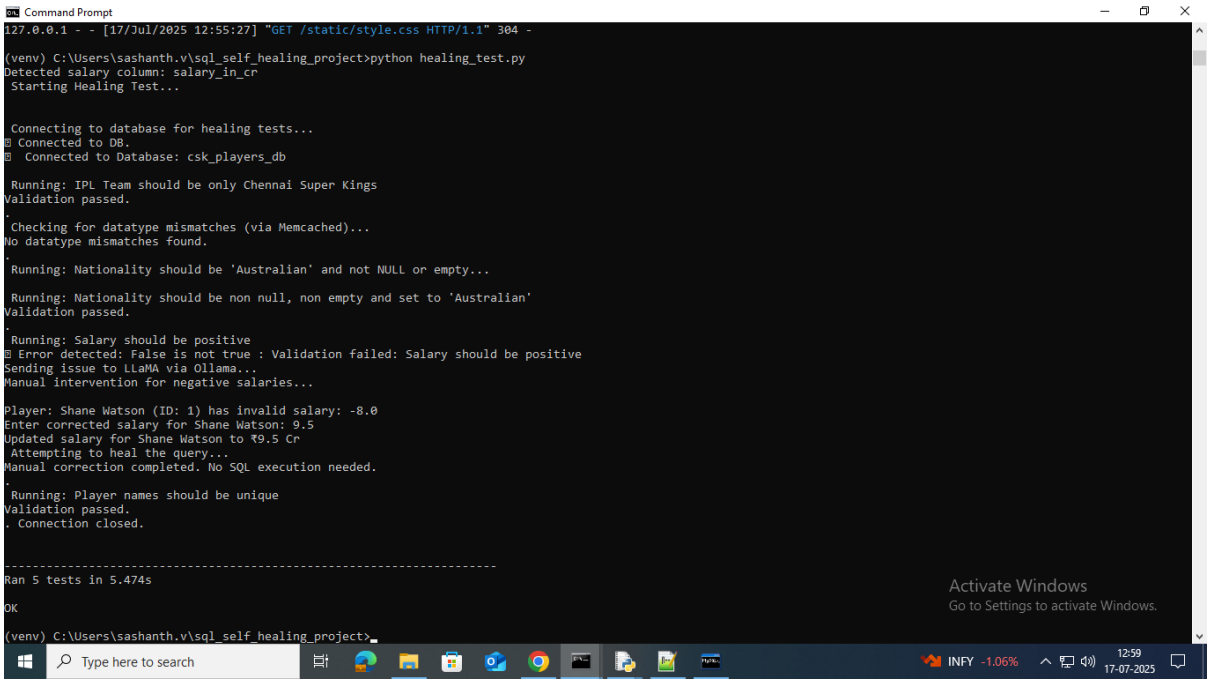


Fig6: Running the self healing test in the cmd.

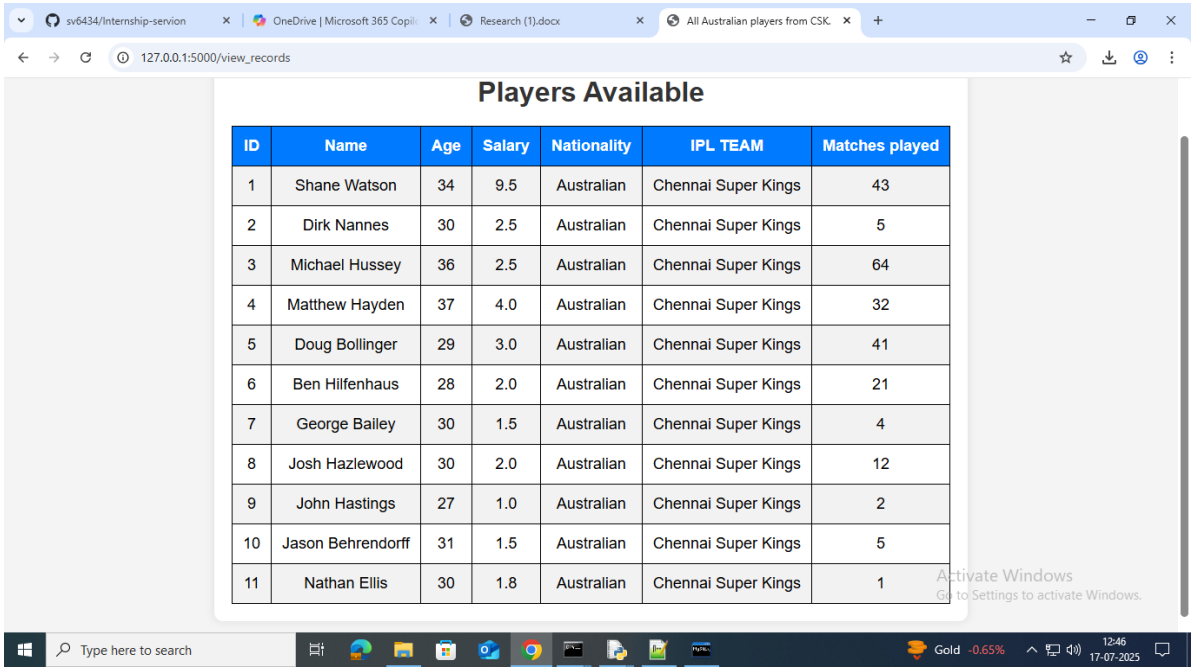


Fig7: The self healed DB.

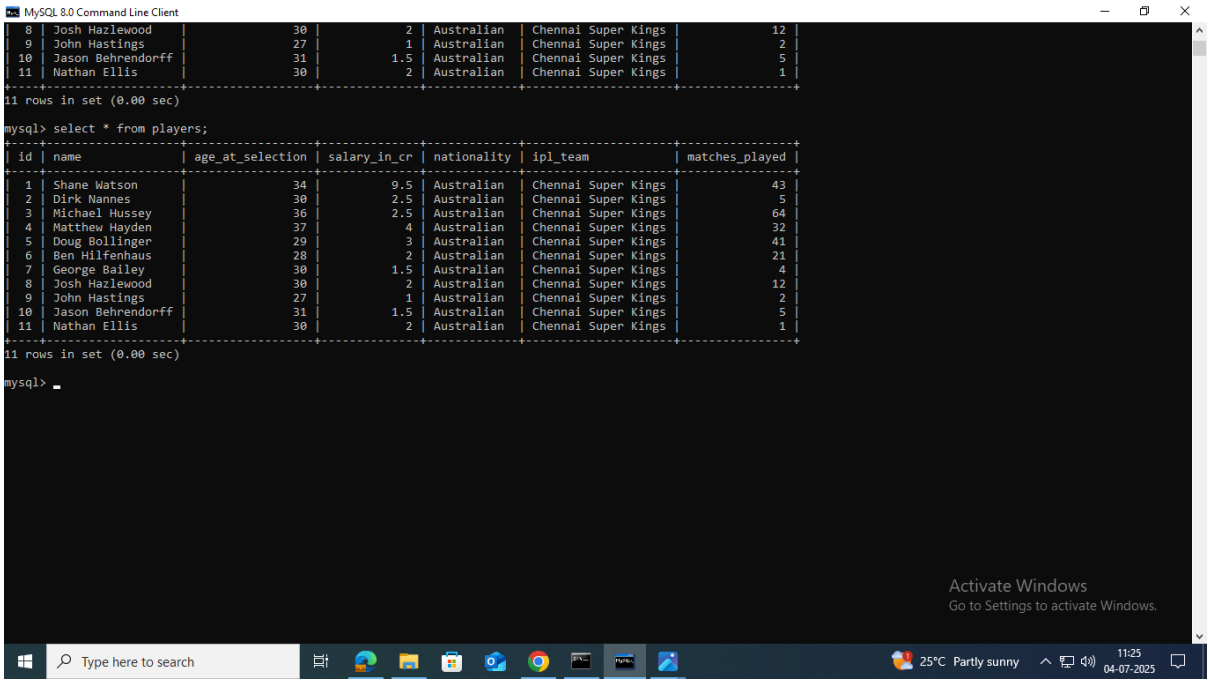
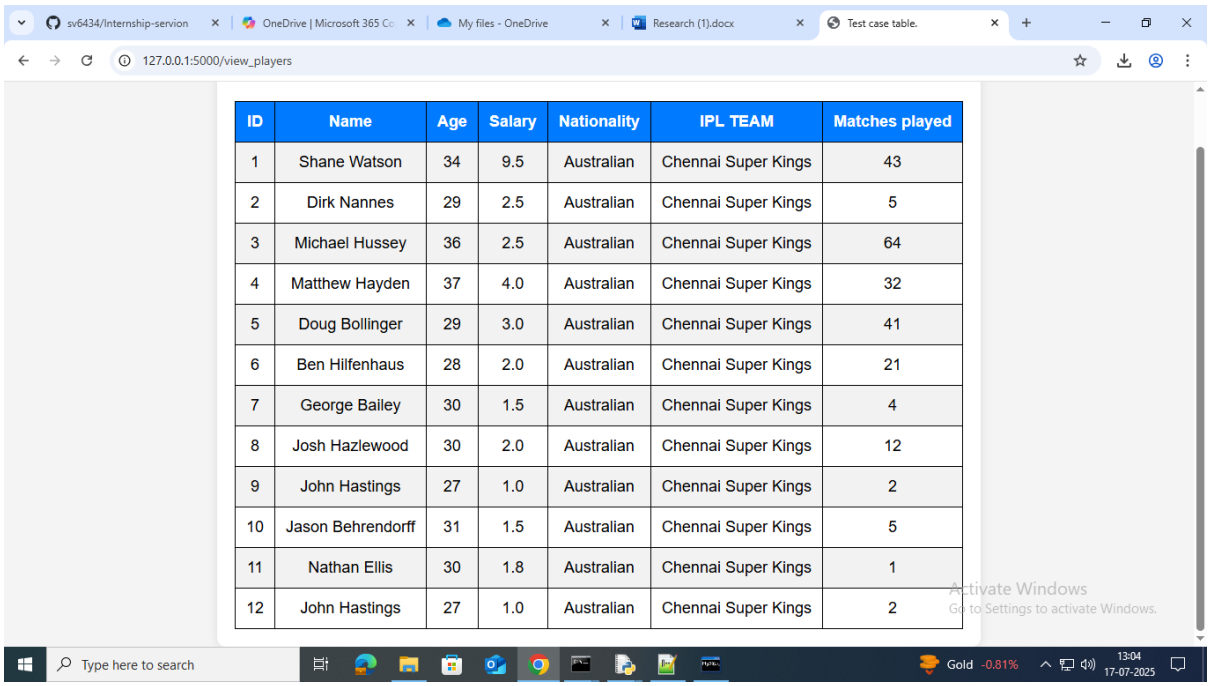


Fig8:Viewing it in MySQL.



ID	Name	Age	Salary	Nationality	IPL TEAM	Matches played
1	Shane Watson	34	9.5	Australian	Chennai Super Kings	43
2	Dirk Nannes	29	2.5	Australian	Chennai Super Kings	5
3	Michael Hussey	36	2.5	Australian	Chennai Super Kings	64
4	Matthew Hayden	37	4.0	Australian	Chennai Super Kings	32
5	Doug Bollinger	29	3.0	Australian	Chennai Super Kings	41
6	Ben Hilfenhaus	28	2.0	Australian	Chennai Super Kings	21
7	George Bailey	30	1.5	Australian	Chennai Super Kings	4
8	Josh Hazlewood	30	2.0	Australian	Chennai Super Kings	12
9	John Hastings	27	1.0	Australian	Chennai Super Kings	2
10	Jason Behrendorff	31	1.5	Australian	Chennai Super Kings	5
11	Nathan Ellis	30	1.8	Australian	Chennai Super Kings	1
12	John Hastings	27	1.0	Australian	Chennai Super Kings	2

Fig9: Seeing the database which has a duplicate entry in the 12th row.

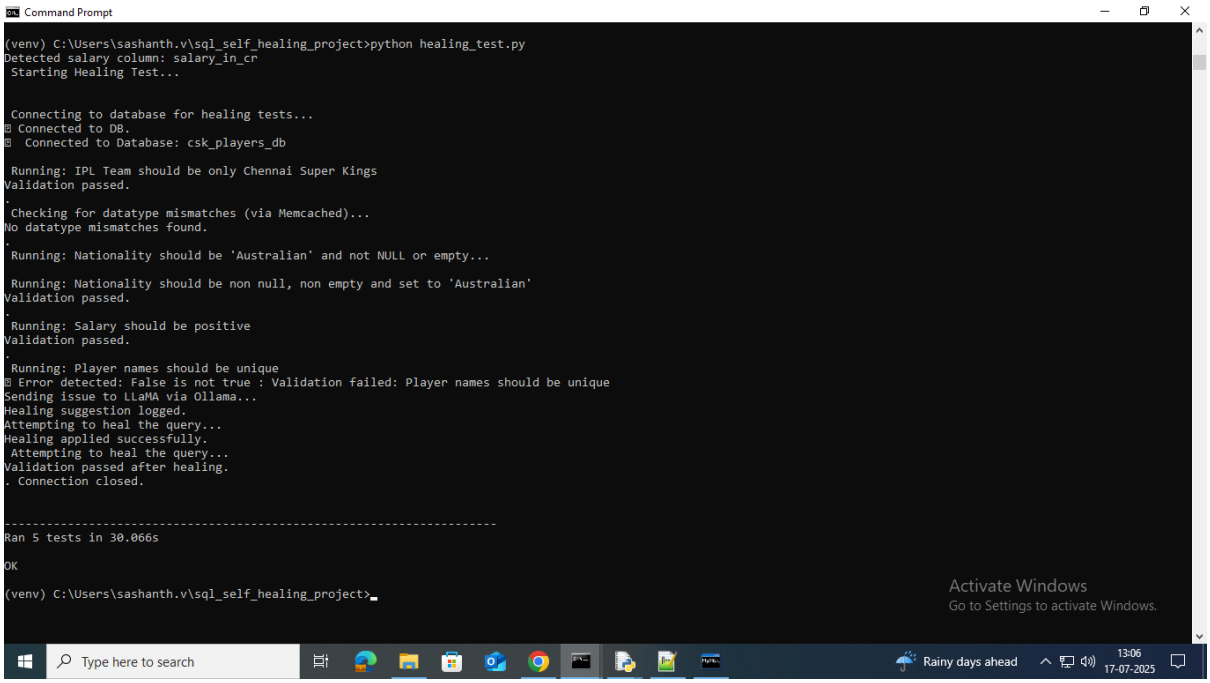
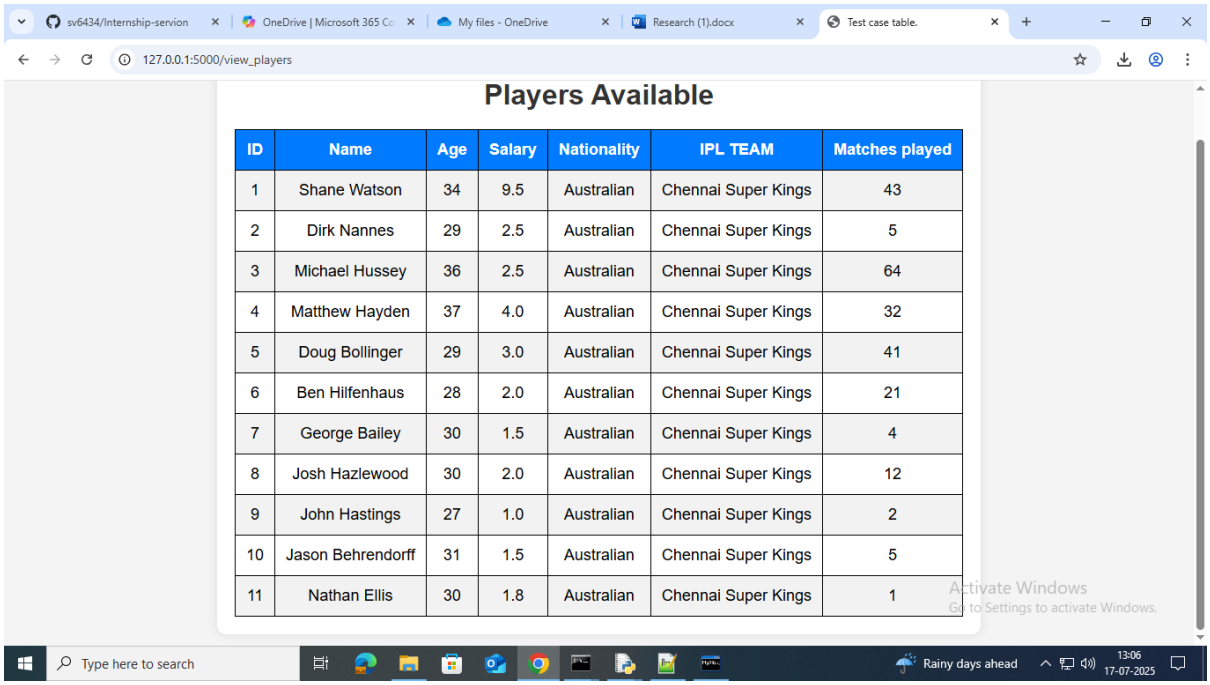


Fig10: Running the test in cmd.



ID	Name	Age	Salary	Nationality	IPL TEAM	Matches played
1	Shane Watson	34	9.5	Australian	Chennai Super Kings	43
2	Dirk Nannes	29	2.5	Australian	Chennai Super Kings	5
3	Michael Hussey	36	2.5	Australian	Chennai Super Kings	64
4	Matthew Hayden	37	4.0	Australian	Chennai Super Kings	32
5	Doug Bollinger	29	3.0	Australian	Chennai Super Kings	41
6	Ben Hilfenhaus	28	2.0	Australian	Chennai Super Kings	21
7	George Bailey	30	1.5	Australian	Chennai Super Kings	4
8	Josh Hazlewood	30	2.0	Australian	Chennai Super Kings	12
9	John Hastings	27	1.0	Australian	Chennai Super Kings	2
10	Jason Behrendorff	31	1.5	Australian	Chennai Super Kings	5
11	Nathan Ellis	30	1.8	Australian	Chennai Super Kings	1

Fig11; Viewing the self healed DB.

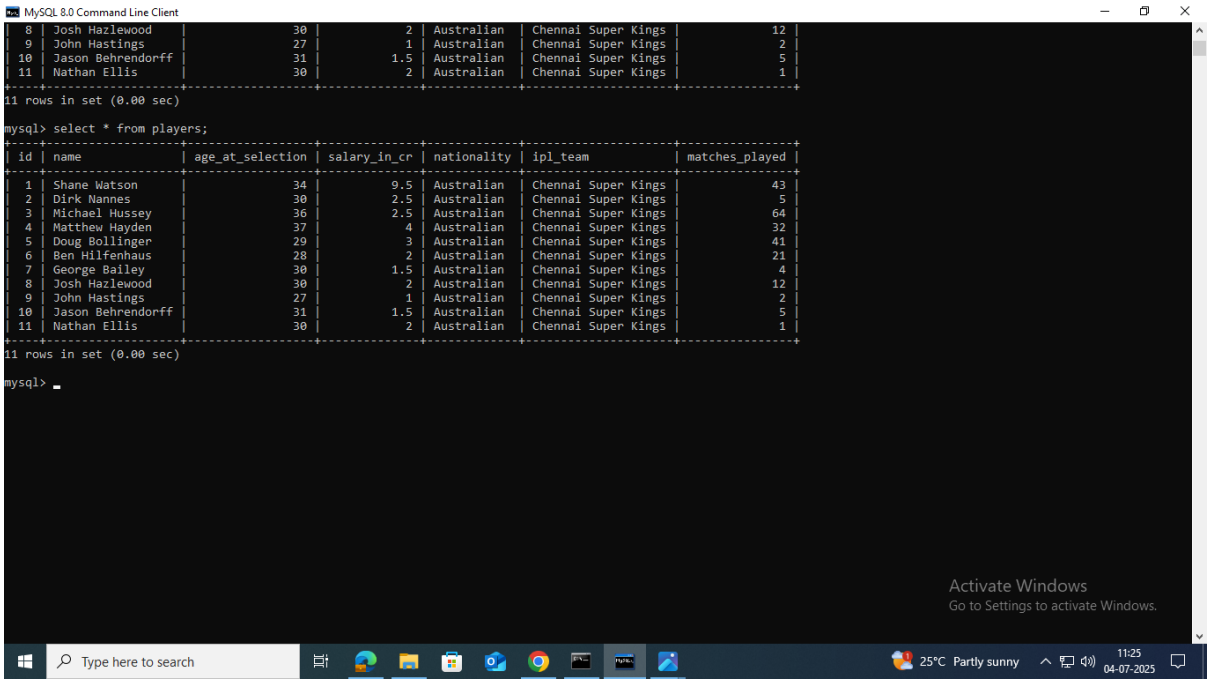


Fig12: Viewing in MySQL.

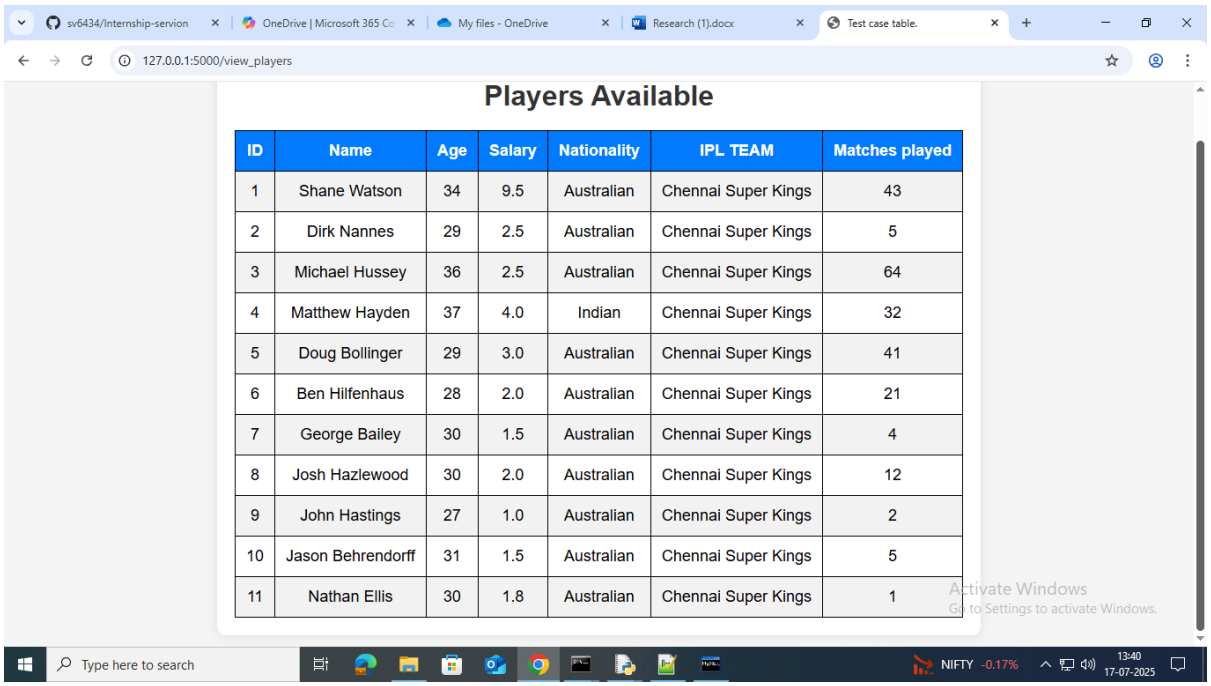


Fig13: Viewing the DB which has an invalid entry in the nationality column for person with id=4.

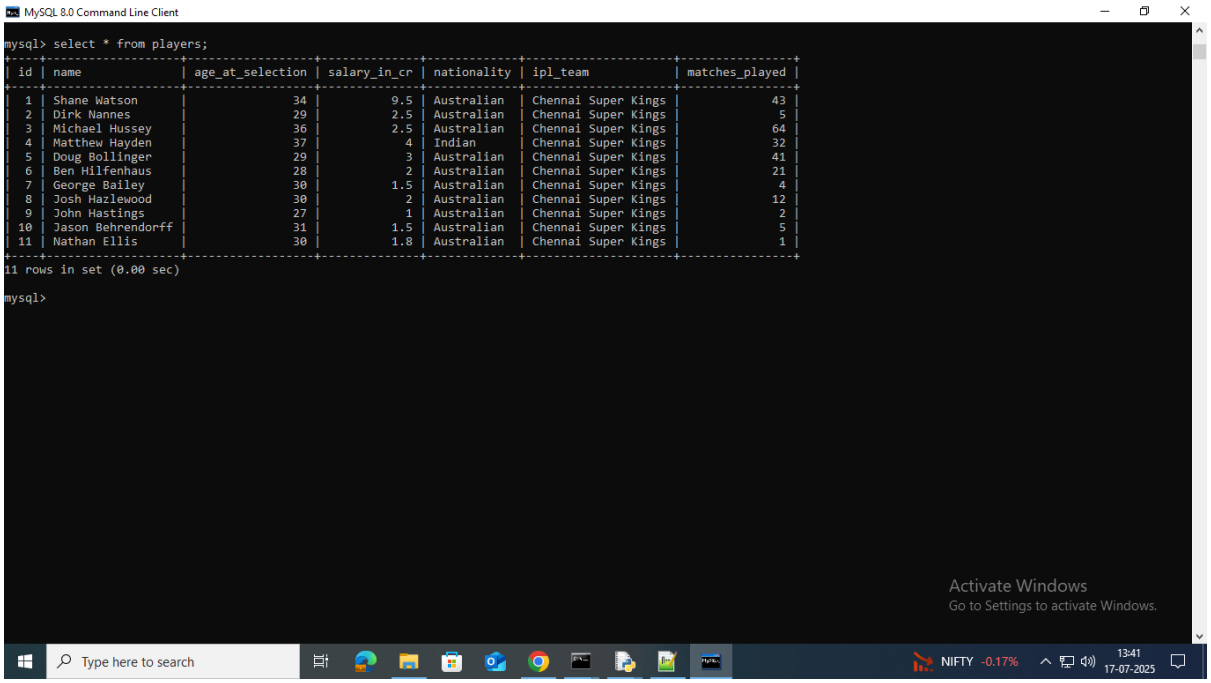


Fig14: The SQL viewer which shows that there is an invalid entry in the nationality column.

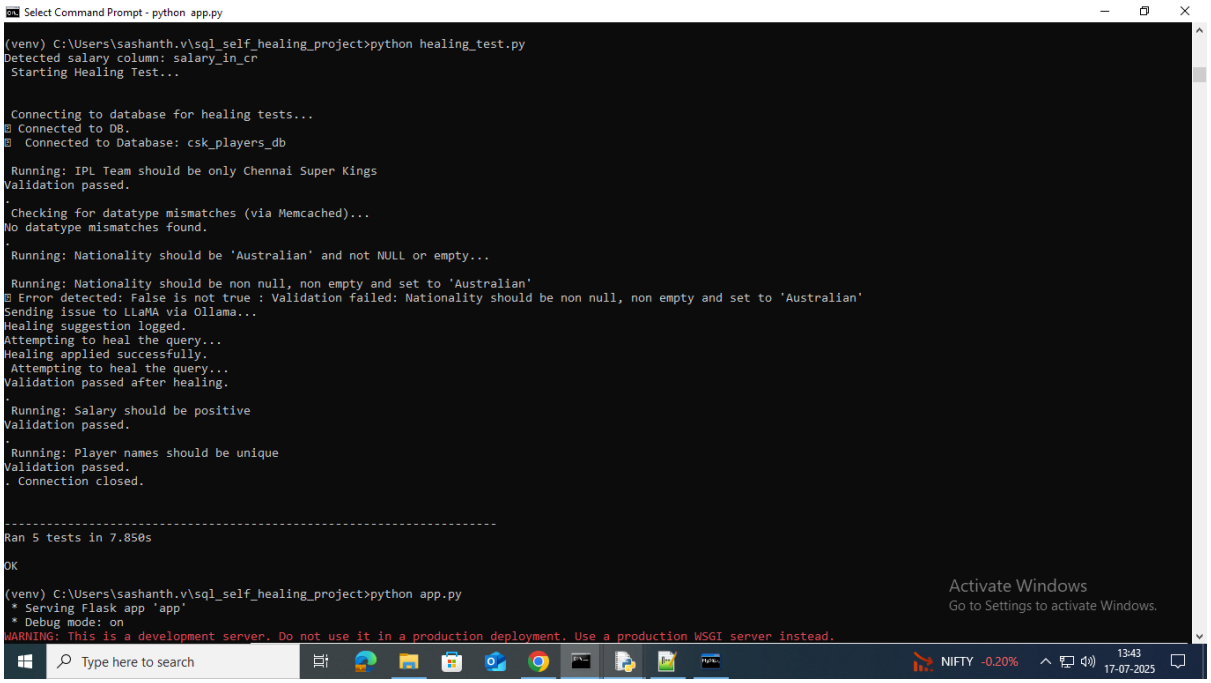


Fig15: Running the test via cmd

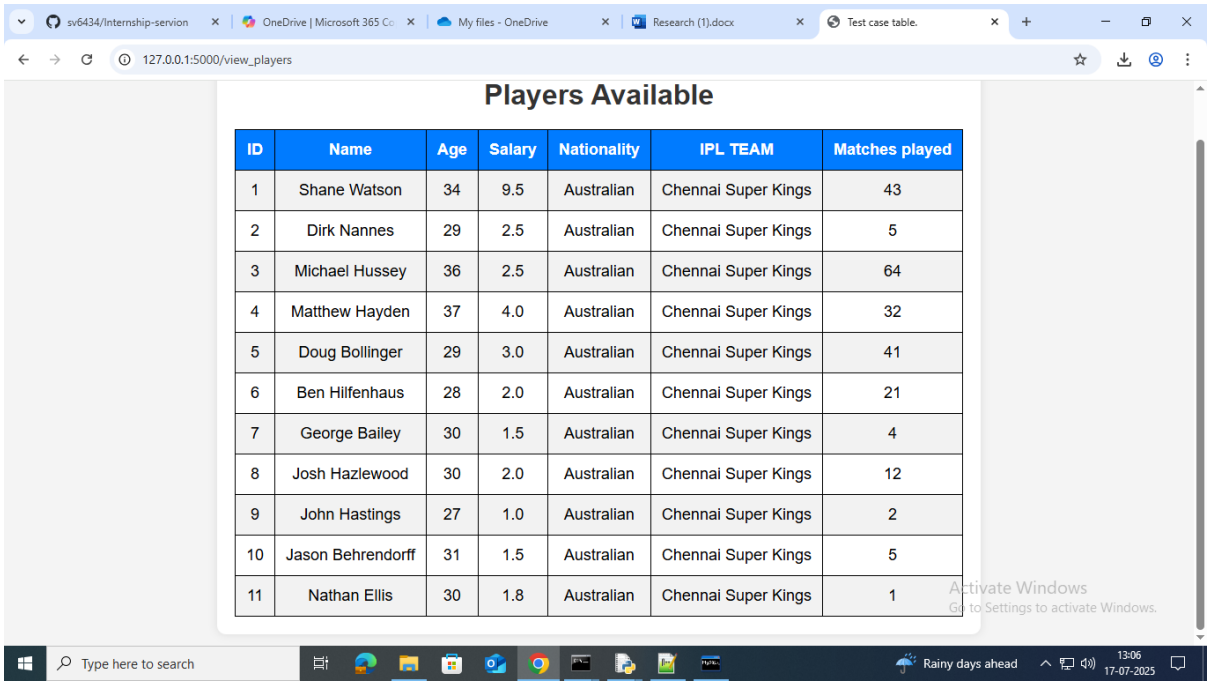


Fig16: The self-healed DB.

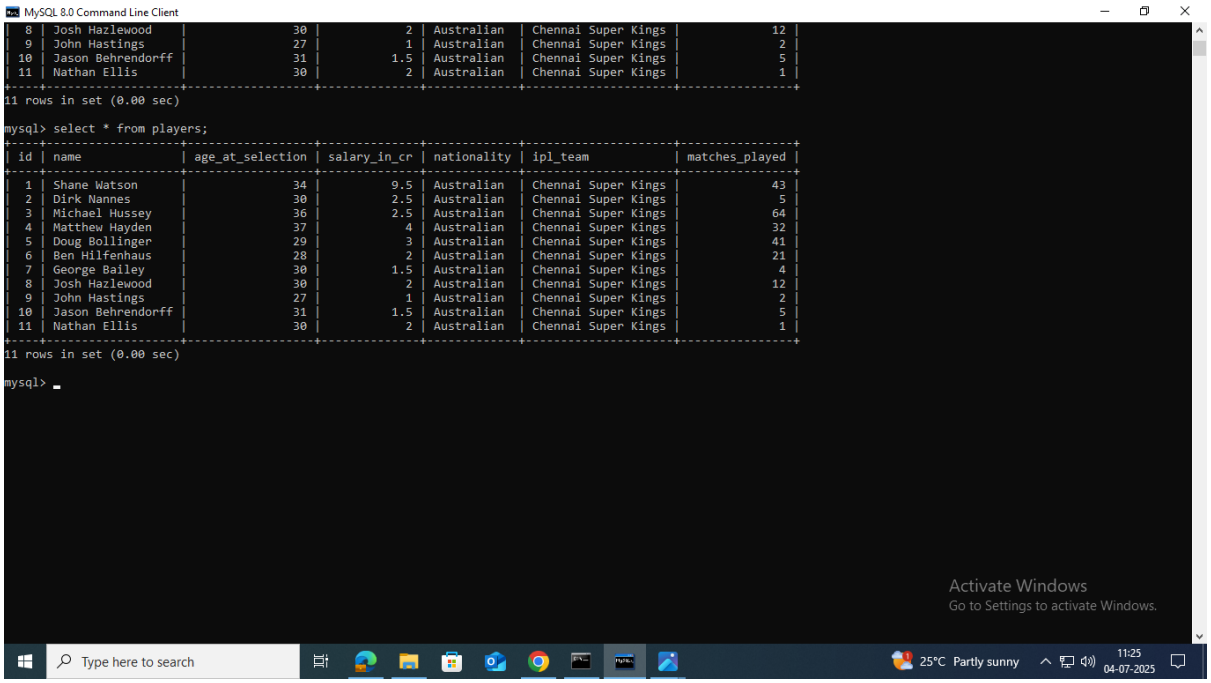


Fig17: Viewing it via MySQL.

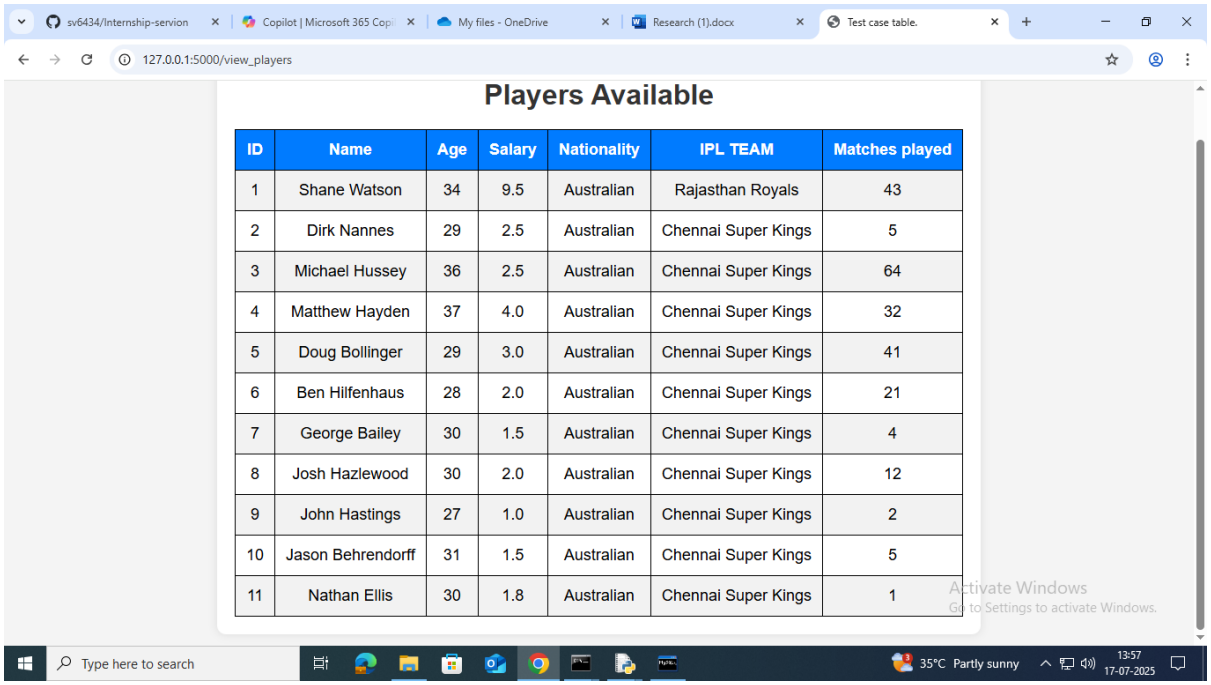


Fig18: The database with an invalid entry in the IPL TEAM column.

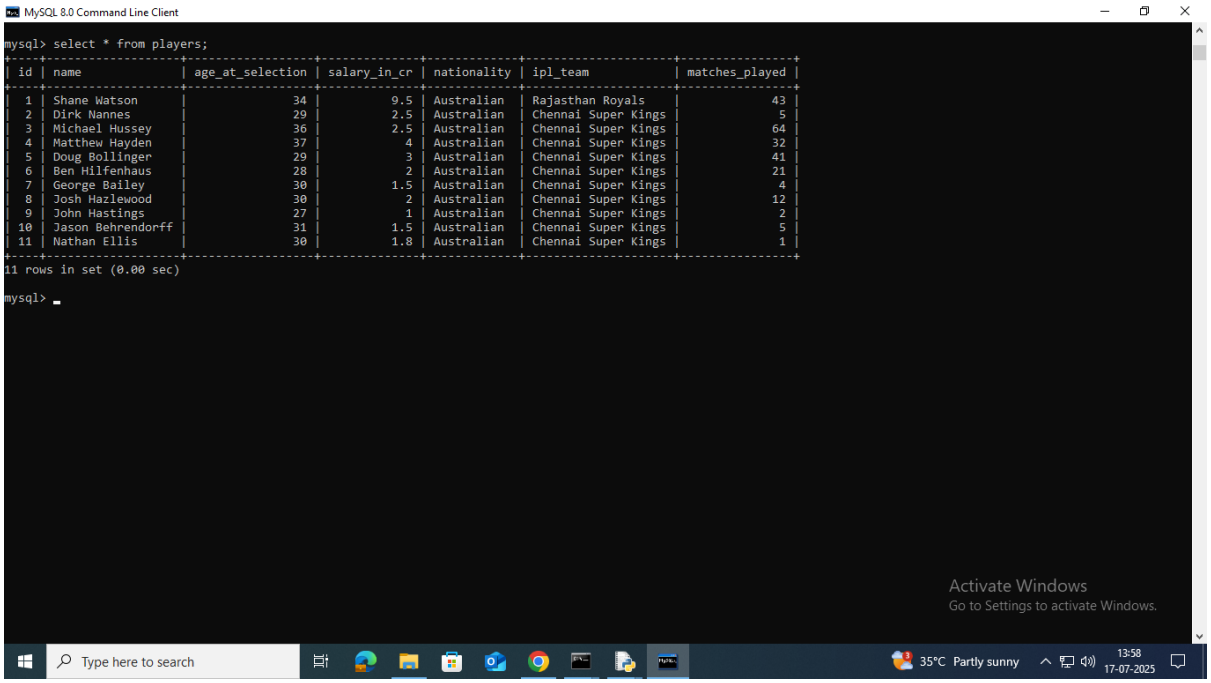


Fig19: Viewing the database in MySQL and seeing that there is an invalid entry in the ipl_team column.

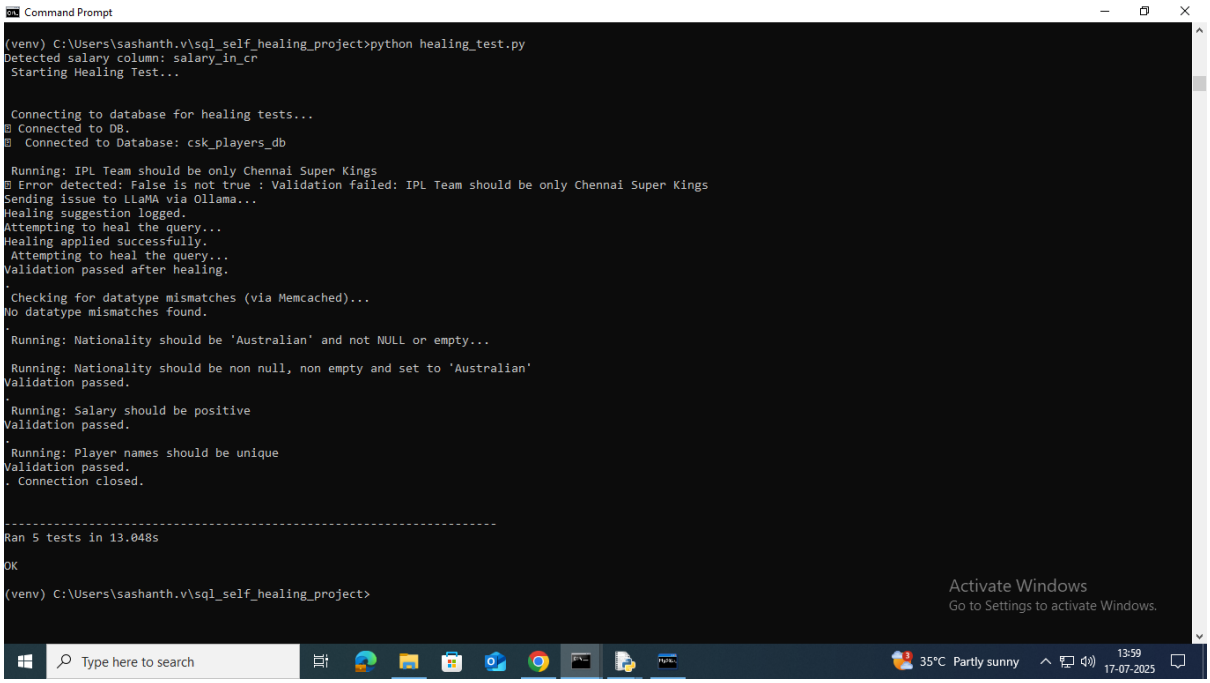
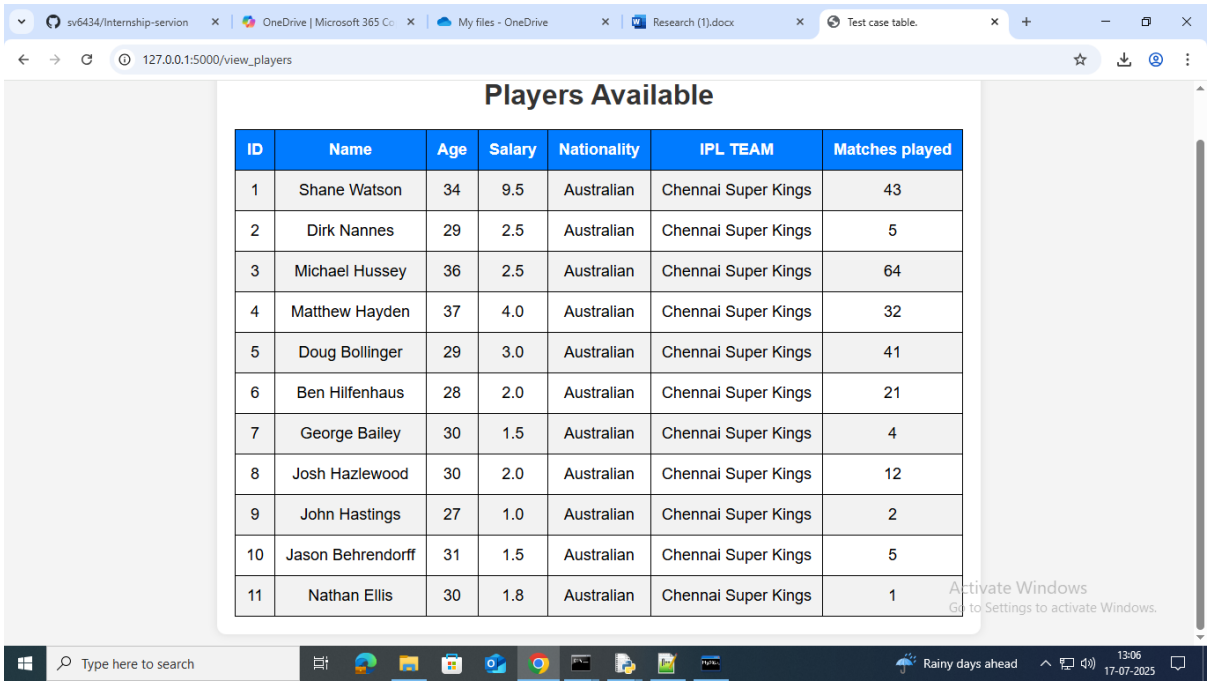


Fig20: Running the test via cmd.



ID	Name	Age	Salary	Nationality	IPL TEAM	Matches played
1	Shane Watson	34	9.5	Australian	Chennai Super Kings	43
2	Dirk Nannes	29	2.5	Australian	Chennai Super Kings	5
3	Michael Hussey	36	2.5	Australian	Chennai Super Kings	64
4	Matthew Hayden	37	4.0	Australian	Chennai Super Kings	32
5	Doug Bollinger	29	3.0	Australian	Chennai Super Kings	41
6	Ben Hilfenhaus	28	2.0	Australian	Chennai Super Kings	21
7	George Bailey	30	1.5	Australian	Chennai Super Kings	4
8	Josh Hazlewood	30	2.0	Australian	Chennai Super Kings	12
9	John Hastings	27	1.0	Australian	Chennai Super Kings	2
10	Jason Behrendorff	31	1.5	Australian	Chennai Super Kings	5
11	Nathan Ellis	30	1.8	Australian	Chennai Super Kings	1

Fig21: Viewing the self-healed DB.

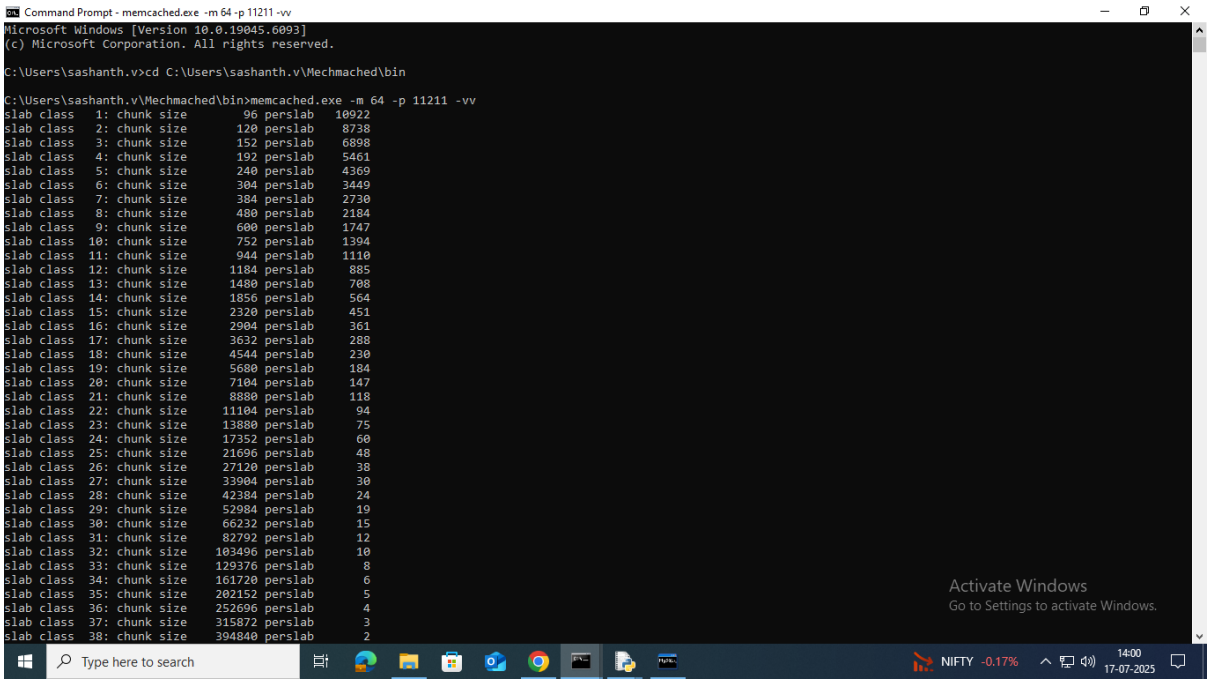


Fig22: Running the memcached.exe for testing the datatypes of our columns.

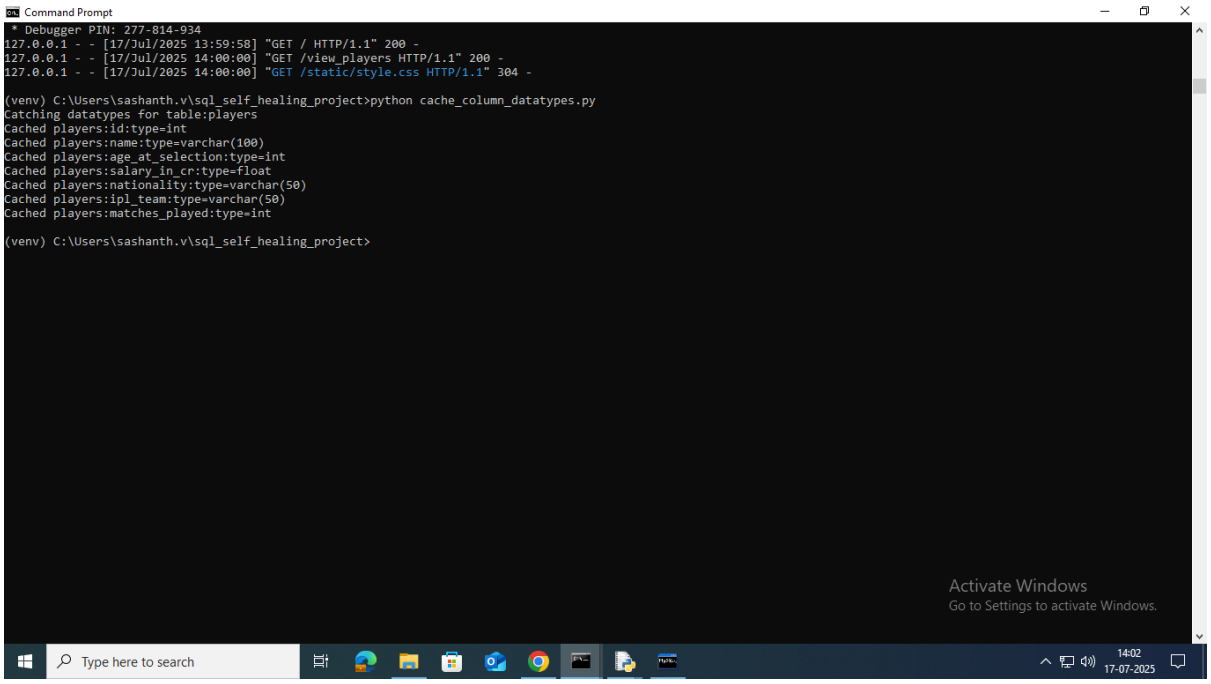


Fig23: Running the code cache_column_datatypes.py in cmd for caching the original datatypes oif each column.

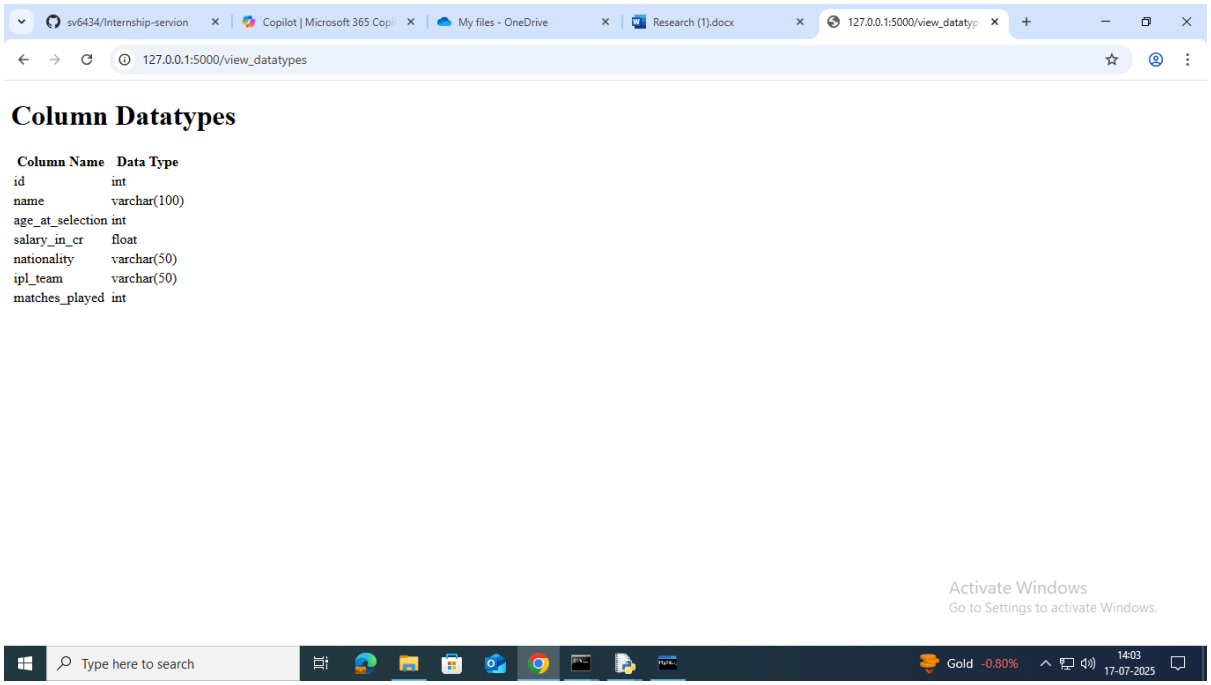


Fig24: Viewing the datatypes of each column.

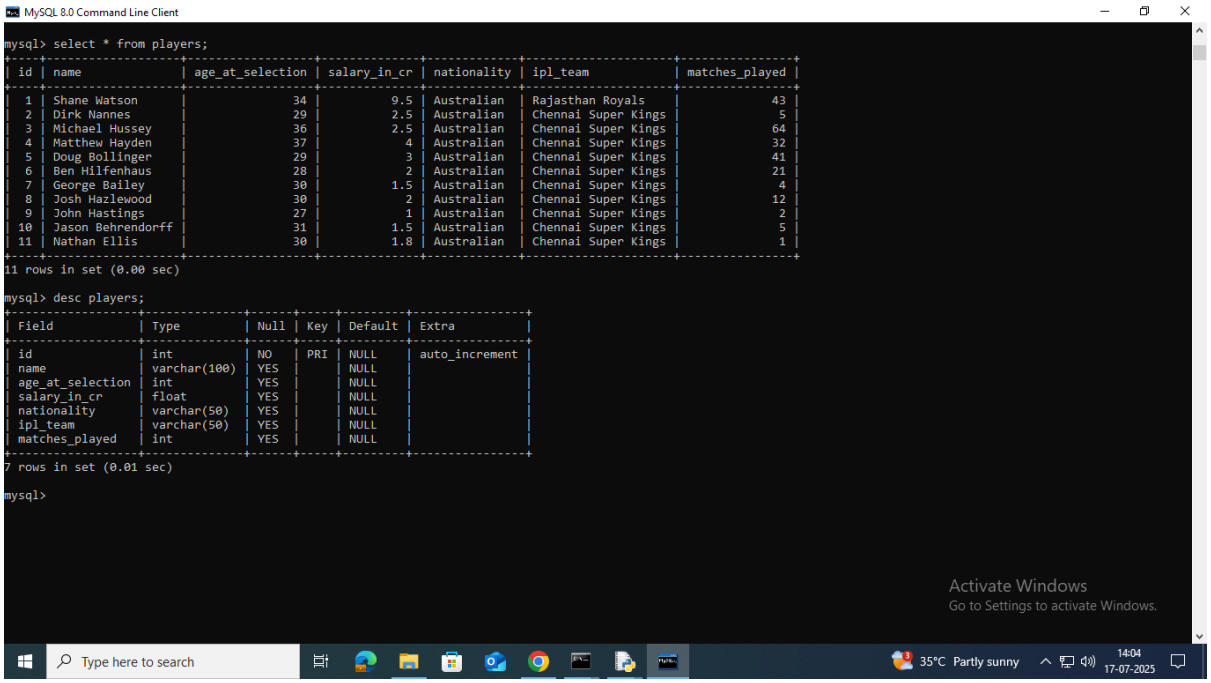


Fig25: Viewing the datatypes of each column at MySQL.

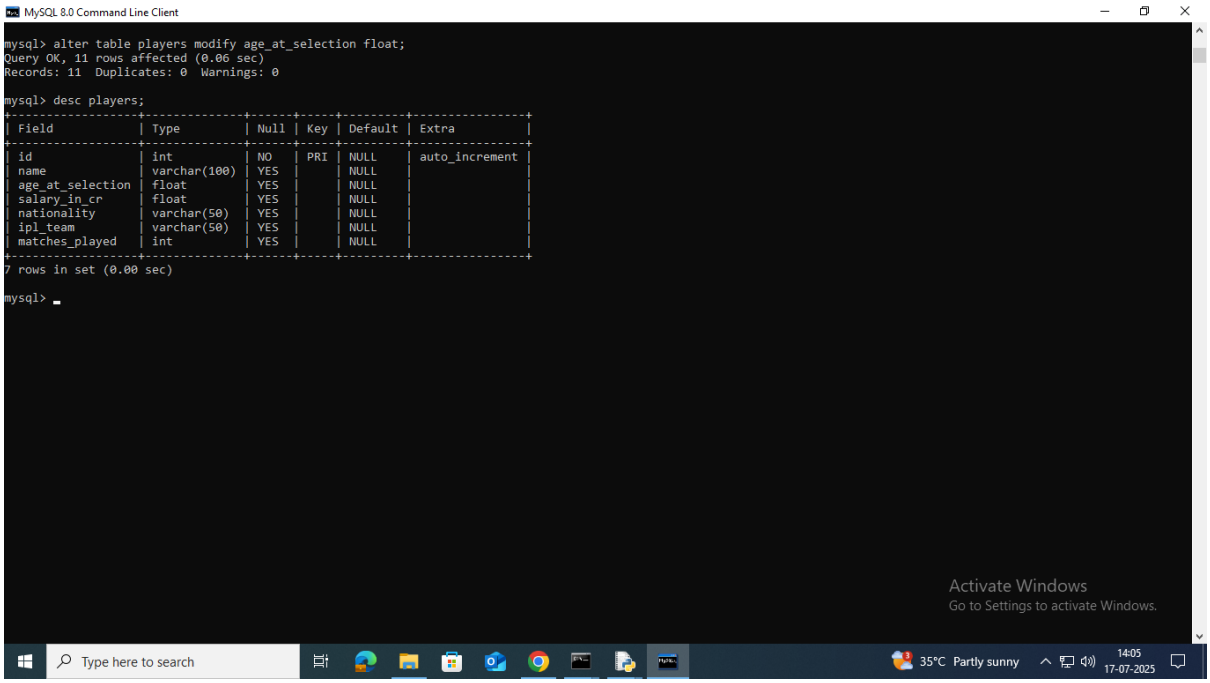


Fig26: Changing the datatype of the age column from int to float.

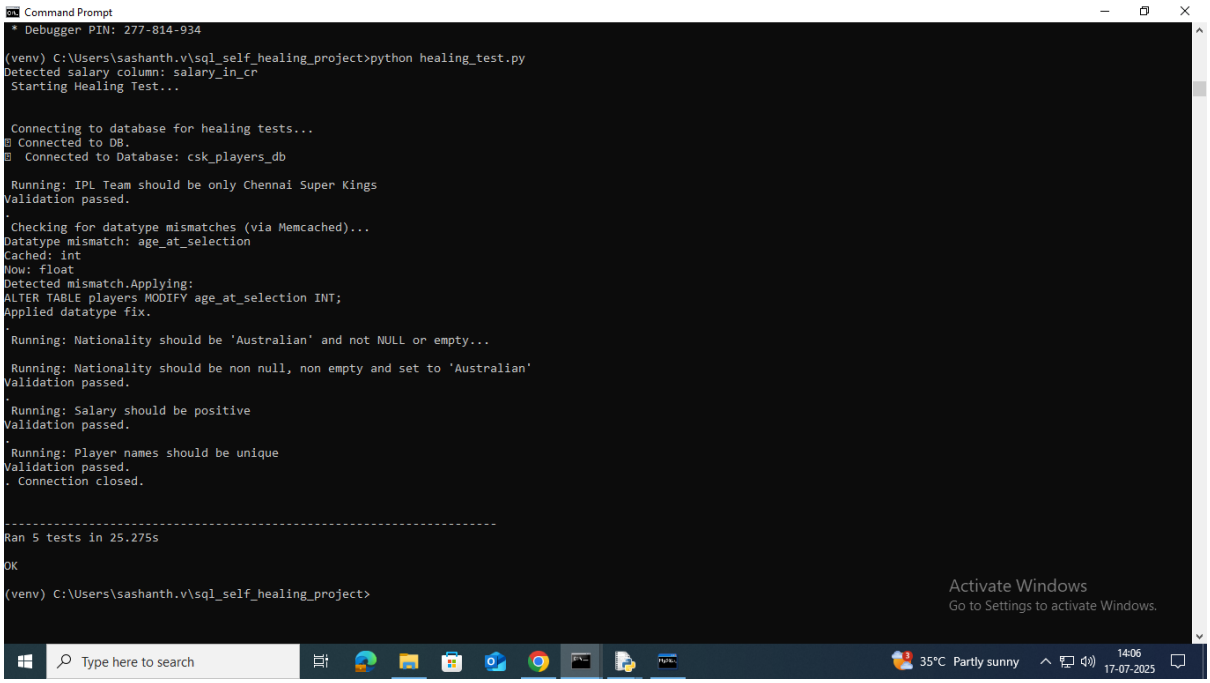


Fig27: Running the tests via cmd. Getting healed.

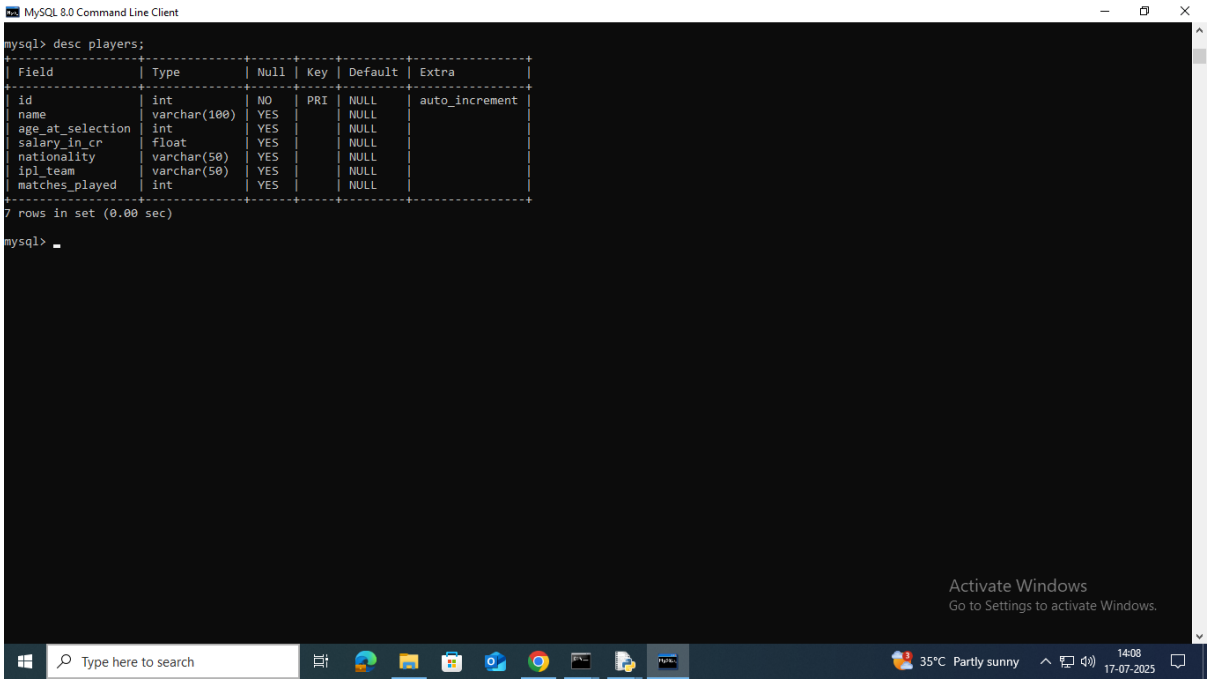


Fig28: Viewing the self-healed in the datatypes of columns.

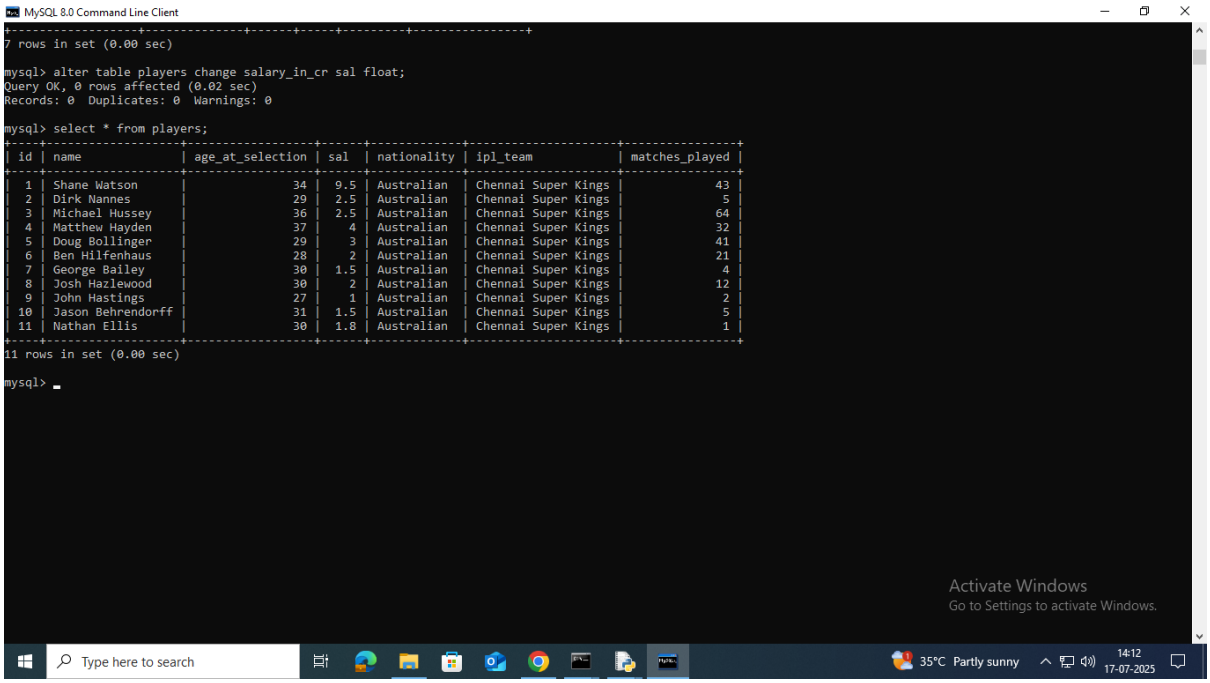


Fig29: Changing the column header name from salary_in_cr to sal.

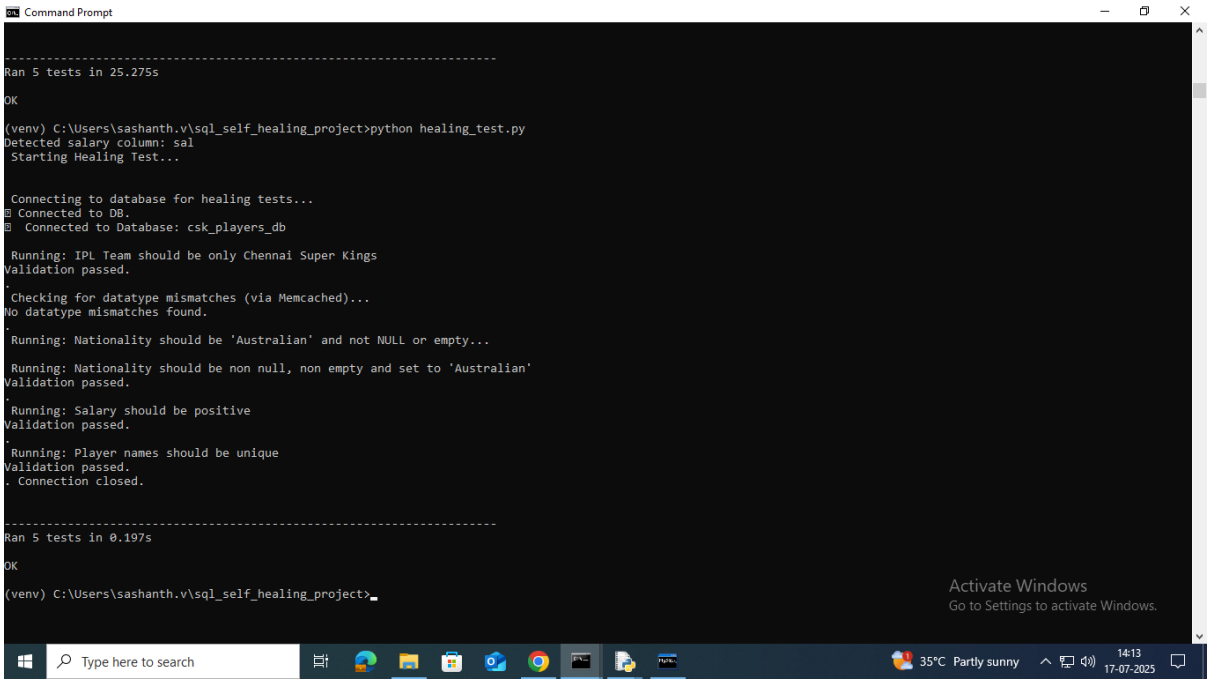


Fig30: The testing via cmd. The testing works even when the salary column name changes as we have used dynamic detection for identifying the salary column. We can observe here during testing that the self-healing agent Ollama identifies the column related to salary via certain keywords which we have provided and then proceeds to the testing of database. All tests are passed means that all the test-cases are valid and have been completed successfully.

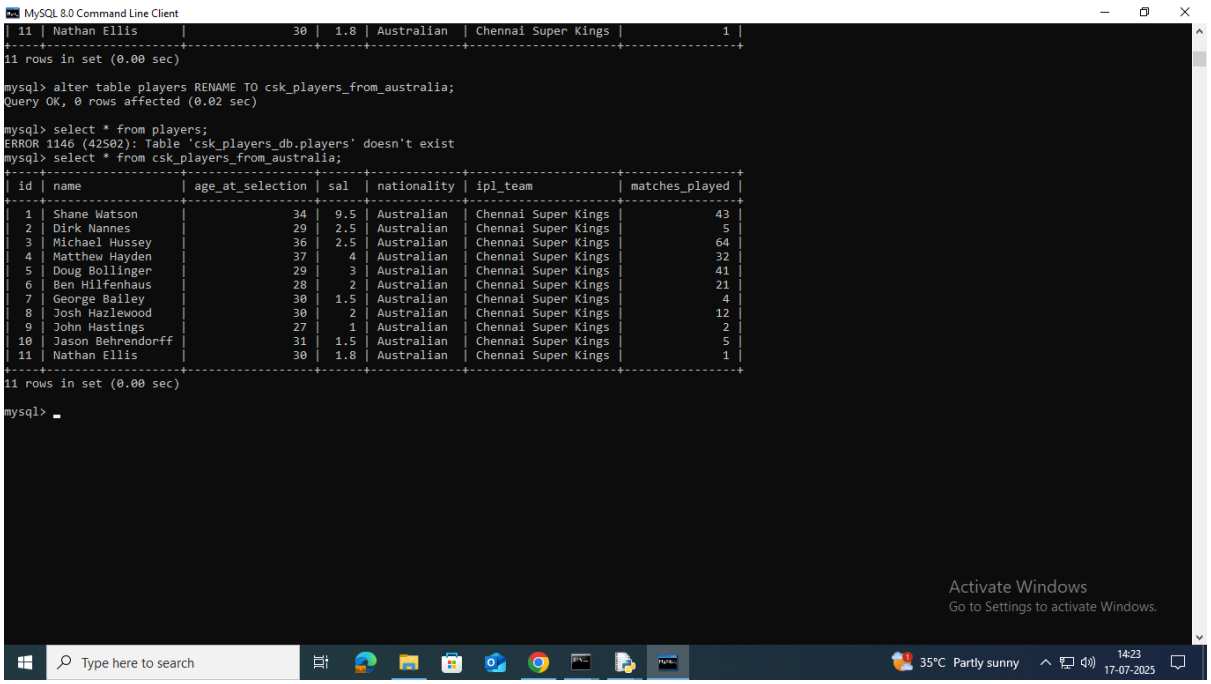


Fig31: Changing the name of the table from players to csk_players_from_australia.

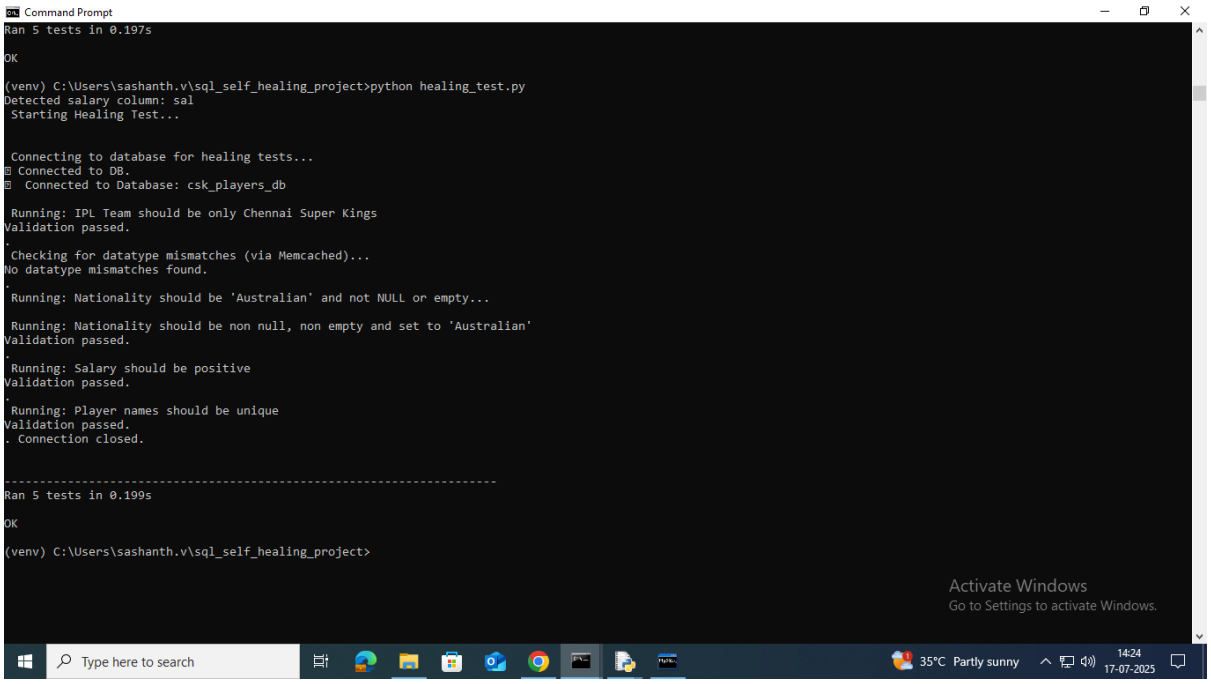


Fig32: Running the test via cmd. The testing works as we have also used the concept of dynamic detection for identifying the table name. We can observe here during testing that the self-healing agent Ollama identifies the table name and then proceeds to the testing of database. All tests are passed means that all the test-cases are valid and have been completed successfully.

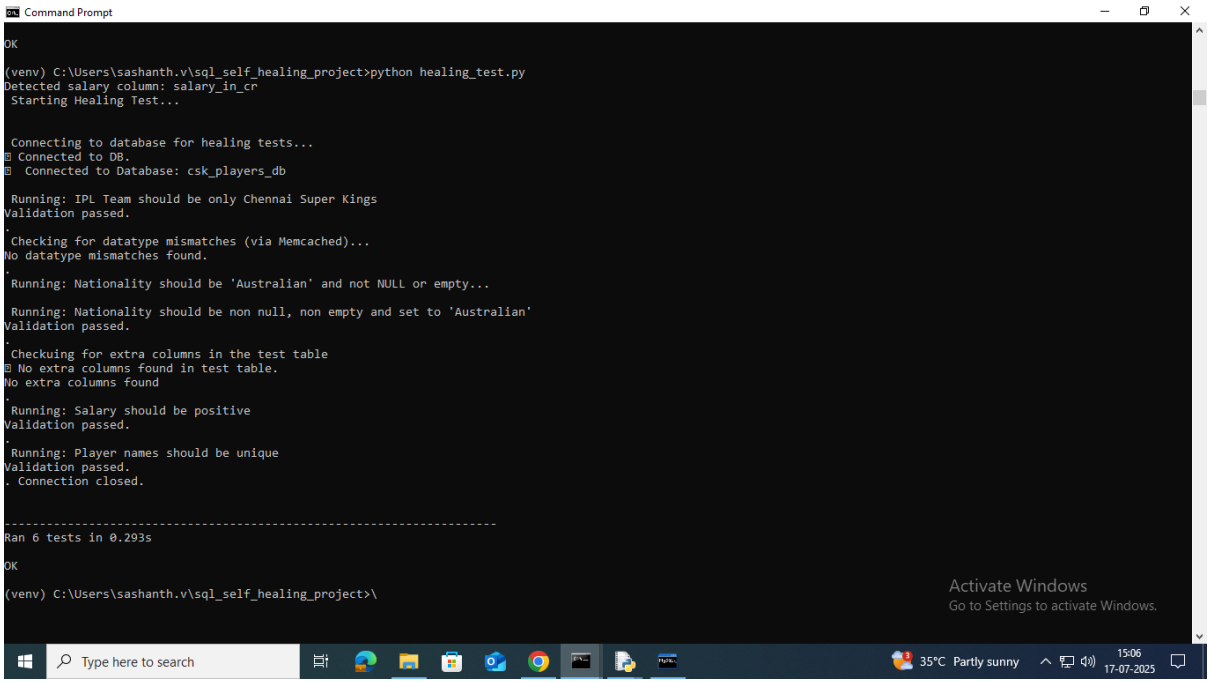


Fig33: Running the test for identifying any extra columns present in the database. If identified, it asks user permission for removing it and if yes it removes else it retains.

Result: Thus, self-healing in a database has been observed successfully and we are able to identify that the test-cases have been validated.

GitHub Link: https://github.com/sv6434/Internship-servion