

Cardiovascular Health Analysis

Notebooks as Business Reports

Due Friday 20th October 2023 18:00

Cardiovascular health is a key determinant of overall health and well-being. Cardiovascular diseases, including heart disease and stroke, are among the leading causes of death worldwide. By understanding and optimizing cardiovascular health through lifestyle choices and medical intervention when necessary, individuals can significantly reduce their risk of cardiovascular diseases and improve their overall quality of life.

This assignment seeks to uncover valuable insights regarding the relationship between demographics, health, and lifestyle factors in relation to cardiovascular disease by analysing the cardiac health dataset. The analysis will be conducted by leveraging Python programming in Google Colab notebooks, exploring SQLite databases, data visualization, and applying industry best practices in programming.

Assignment Objectives

Analyse the provided data and SQLite databases, data exploration, and SQLite databases in programming.

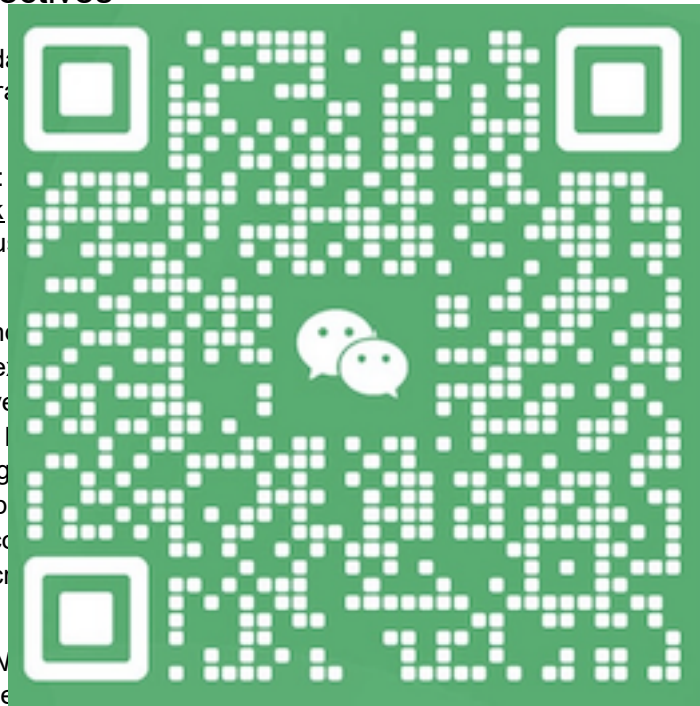
Deliver two notebooks: a development notebook (for development) and a production notebook (for production) and a business report (for business) and a best practice not (for best practice not).

Learning Objectives

- Understand and
 - Perform data e
 - Create effective
 - Apply industry b
 - design, reusing
 - Select and app
 - Interpret and co
 - Demonstrate cr
- creating a modular

The *Business Report* must be a *business report* and not a *technical report*. It should contain no additional steps other than running code.

Note: You can have code-cells in the notebook set up the Colab instance, for example, copy data, python scripts, or other notebooks. But other than running a code cell your notebook should require no further interaction from the user/reader of the notebook.



Tasks:

- Set up the environment:
 - Create a new Google Colab notebook.
 - Connect the notebook to your GitHub account.
 - Import the necessary libraries (SQLite3, Pandas, Matplotlib, and ipywidgets).
- Access the database:
 - Connect to the cardiohealth SQLite database using the SQLite3 library.
 - Examine the schema of the database and understand the structure of the tables.
- Data extraction and manipulation:
 - Write SQL queries to extract relevant information from the tables
 - Use pandas to load the query results into data frames and perform data manipulation tasks such as filtering, grouping, and aggregation.
 - Clean and pre-process the data, addressing any missing or inconsistent values.
- Interpretation and conclusion:
 - Summarise the main insights you have gained from the data analysis.
 - Discuss any limitations of your analysis and suggest possible improvements.
 - Reflect on the process of data analysis.

Analysing and Visualising Data

Note that you may need to use the following:

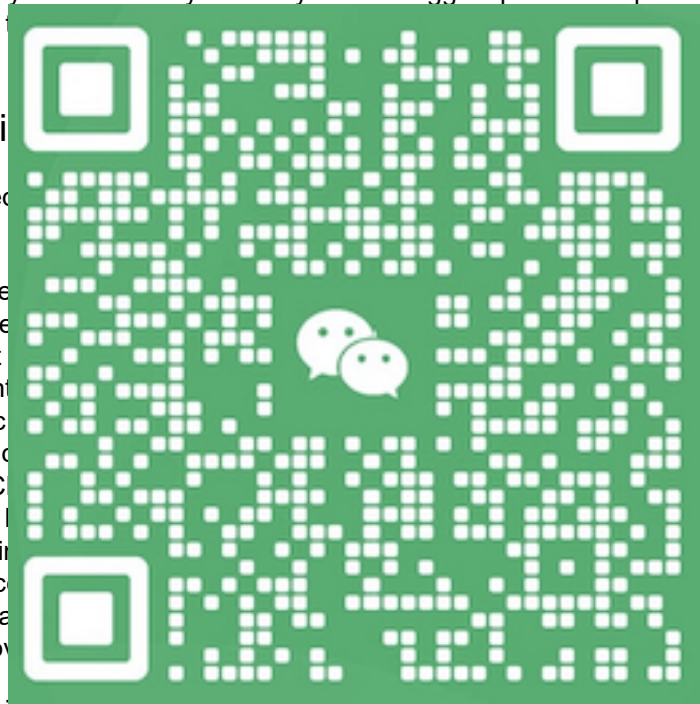
Dataset

- **age:** Age of the patient
- **gender:** Gender of the patient
- **height:** Height of the patient
- **weight:** Weight of the patient
- **ap_hi:** Systolic blood pressure
- **ap_lo:** Diastolic blood pressure
- **cholesterol:** Cholesterol level (normal, high, or very high)
- **gluc:** Glucose level (normal, high, or very high)
- **smoke:** Smoking status (yes or no)
- **alco:** Alcohol consumption (yes or no)
- **active:** Physical activity level (yes or no)
- **cardio:** Cardiovascular disease status (yes or no)

Examine the relationship between the occurrence of cardiovascular disease and the following factors within the provided cardio health dataset:

1. **Choose one:** Investigate either **Age groups** or **Gender**.
2. **Choose one:** Explore either **BMI** (Body Mass Index) or **Blood pressure** (Systolic and Diastolic).
3. **Choose one:** Analyze either **BMI and Cholesterol**, **Glucose and Blood pressure**, or **Cholesterol and Blood pressure**.
4. **Choose one:** Study either the connection between **Smoking and physical activity** or **Alcohol and physical activity**.

Please select one option from each group and assess how it impacts the presence or absence of cardiovascular disease.



Visualising Data

Its.

ormal).

GitHub

Version control is an industry best practice technique for monitoring changes to a file or group of files over time and reverting to a previous version. For this assignment, you are required to create a **new PRIVATE GitHub repository** to store the notebook and any support files. The assignment GitHub repository will contain:

- README
- Non-Conformance Report (if applicable)
- Notebooks required for the assignment
- Python scripts required for the assignment
- Any other relevant documents

Evaluation

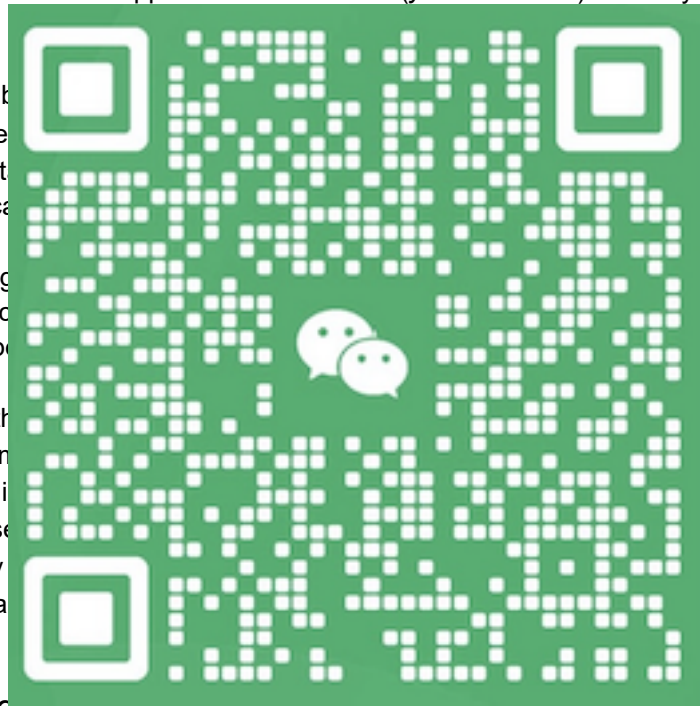
As an IS Professional, you are expected to meet the specification to the best of your ability. This specification is to be treated as the output of a meeting between yourself and a client. Your instructor will take on the role of the client. If you want to implement any functionality or behaviour not described in this specification, please seek approval from the client (*your instructor*) **before** you begin writing your program.

Your submission will be evaluated based on the criteria mentioned in this document. This problem is designed to test your ability to analyse a problem, provide clear and detailed solutions, and demonstrate your problem-solving, and communication skills.

You may only use programming concepts covered in the textbook. If you plan to use any advanced concepts, you must seek approval from your instructor **before** you begin writing your program.

The code must follow the guidelines outlined in the assignment brief, which include:

- Meaningful naming
- Correct capitalisation
- Appropriate use of comments
- Reference any external resources
- Use of space and indentation



Submission Guidelines

Save your Google Colab notebook(s) as an .ipynb file and push it to your GitHub repository. Write a brief README.md file describing the assignment and the purpose of the repository. Your GitHub repository should be private and contain all documents relevant to this assignment.

Submit the link and zip file to your GitHub repository containing the notebook and README.md file.

This assignment is to be completed individually. The assignment is **due 18:00 Friday 20th October 2023**. The entire assignment GitHub project folder must be submitted as a single compressed archive file to the unit's BlackBoard site submission link.

Non-Conformance Report (NCR)

A non-conformance report (NCR) is a document that addresses issues where there has been a deviation from the project specification or where work fails to meet agreed quality standards. If you cannot implement some functionality or have difficulty meeting any of the requirements, you will need to provide a NCR. An example might be unable to produce the plots, or deviation from the style guide. For each non-conformance issue, you need to document:

- The problem
- Severity and impact
- How it occurred
- How to prevent it from happening again
- Plan or time estimate to fix

Grading Criteria

Your assignment will be graded based on the following criteria:

- Clarity and organization of your code (comments, modular design, code reuse).
- Proper use of version control with GitHub.
- Quality and content of your reports (clear explanations, and visualizations).
- Effectiveness of your code (correctness, efficiency, and readability).
- Overall data analysis (correctness, completeness, and accuracy).
- Critical thinking (ability to identify and solve problems).

Academic Integrity

Curtin's Academic Integrity Policy is a commitment to the highest standards of academic integrity. The Academic Integrity tab in the course page provides more details, go to the Academic Integrity tab in the course page. All assignments must adhere to the Copyright Act of 1969.

