Data Security

The intention of this activity will be to convey a fair and impartial perspective concerning the data security of submitted assignments over the duration of the course. These assignments encompass collaborated group submissions to individual assignments, all of which are subjected to an evaluation of their adherence to data security principles and practices.

Data security is the practice and implementation of safeguards to ensure digital data is free from any unauthorized access, misuse of data, disclosure, alteration, or destruction. There are generally accepted principle which should be adhered to in ensuring data security. These include confidentiality, integrity, transparency, lawfulness, data minimization and accountability.

Mapping the Health System

This assignment was a group submission on assessing the health ecosystem within South Africa.
Possible data security issues that could have arisen during the project lifecycle:

- A possible data security measure which could have risen was ensuring transparency of gathered information. Ensuring accurate referencing and validity check on the legitimacy of information should have been discussed more thoroughly during the assignment.

- Discussions among group members concerning the sharing and distribution of content of the assignment. Ensuring that content isn't shared outside of the allocated group.

Possible solutions in addressing the risk:

- Using a secure collaboration platform. Could have been achieved by implementing control measures within the collaboration platform to track any alterations made to the shared document.

Lab 1 Assignment

This lab was a group submission in which we implemented and applied a linear regression model to our chosen dataset. The chosen dataset related to life expectancy. The features within the dataset were "Country", "Year", "Developed/Developing", "Population_size", vaccines for: "Hepatitis B", "Measles", "Polio", "Diphtheria", prevalence of "HIV/AIDS", "infant deaths", "under-five deaths", "Total expenditure", "GDP", "BMI", "thinness 1-19 years", "Alcohol", and "Schooling".

- A possible data security measure which should have been implemented is data minimization. This method of ensuring data security is through assessment of datasets to ensure they only contain essential feature points required to fulfill the model's requirements.

Possible solutions in addressing the risk:

- To remove any data that is not vital for building and validating the linear regression model, especially data which contains sensitive information able to skew data results.

Activity 1 - Case study per activity requirement

The Importance of Data in the Health Sector.
District Health Information Software 2 (DHIS2) is an open-source health management information system. The platform helps the national department of health to monitor and evaluate health programs, track outbreaks, and ensure efficient use of resources.

- A data security measure which could have been implemented would been applying data quality management. This method of data security would be achieved by ensuring that the source of data information is accurate, reliable, and relevant through implementing quality assurance processes and data validation techniques.

Possible solutions in addressing the risk:

- In choosing the District Health Information Software 2 (DHIS2) for a case study, additional verification checks regarding the legitimacy of its software use should have been more thoroughly explored. Given that DHIS2 is open-source, deeper consideration of its practical application could have been undertaken.

- Third-Party Reviews and Use-Cases. This mitigation action would be to investigate third-party reviews and use-cases related to DHIS2 to gather insights regarding its practical applicability, reliability, and any associated risks in real-world scenarios.

Activity 2 – Johns Hopkins application of health analytics

Johns Hopkins effective use of health analytics to achieve competitive advantages.

- A data security issue identified was in compiling evidence for assessing competitive advantages was the possible risk of exposing sensitive patient data, which could lead to breaches of privacy and confidentiality.

- There could be unintentional bias or misrepresentation if the data used is not adequately represented in the presentation.

Possible solutions in addressing the risk:

- Achieved through ensuring that data utilized for the analysis is collected from diverse and representative sources to minimize data bias.

Lab 2 Assignment

Applying a logistic regression model to a chosen dataset.

Risk of Inadvertent Data Alteration

- Given the substantial size of the dataset, it is susceptible to unintentional modifications, deleted features, or corruption, which could detrimentally affect the precision and dependability of the model's output.

Possible solutions in addressing the risk:

- Make use of checksum algorithms to verify data integrity after introducing the dataset to the model application to ensure that any inadvertent modifications are immediately identified.

Throughout the submission phase of this module, various issues, extending beyond the domain of data security, were identified. Upon contemplation of data security considerations across all submissions, it became evident that the identified data security issues serve as additional considerations for applying health analytics skills in practical settings.

Subsequent lessons highlighted the criticalness and pertinence of data security, demonstrating how diverse factors can indirectly pose data security challenges.