Table 2 – Guidelines for the responsible application of data analytics.

1. General

   DO's

   1.1 **Governance**
   Ensure that a comprehensive governance framework is in place prior to, during, and for the relevant period after data acquisition, analysis and use activities, that it is commensurate with the activities' potential impacts, and that it encompasses:
   a. risk assessment and risk management from the perspectives of all affected parties
   b. express assignments of accountability, at an appropriate level of granularity

   1.2 **Expertise**
   Ensure that all individuals participating in the activities have education, training, and experience in relation to the real-world systems about which inferences are to be drawn, appropriate to the roles that they play

   1.3 **Compliance**
   Ensure that all activities are compliant with all relevant laws and established public policy positions within relevant jurisdictions, and with public standards of behaviour

2. Data Acquisition

   DO's

   2.1 **The Problem Domain**
   Understand the real-world systems about which inferences are to be drawn and to which data analytics are to be applied

   2.2 **The Data Sources**
   Understand each source of data, including:
   a. the data's provenance
   b. the purposes for which the data was created
   c. the meaning of each data-item at the time of creation
   d. the data quality at the time of creation
   e. the data quality and information quality at the time of use

   2.3 **Data Merger**
   If data is to be merged from multiple sources, assess the compatibility of the various collections, records and items of data, taking into account the data's provenance, purposes, meaning and quality, and the potential impact of mis-matching and mistaken assumptions

   2.4 **Data Scrubbing**
   If data is to be scrubbed, cleaned or cleansed, assess the reliability of the processes for the intended purpose and the potential impacts of mistaken assumptions and erroneous changes

   2.5 **Identity Protection**
   If the association of data with an entity is sensitive, apply techniques to the data whose effectiveness is commensurate with the risks to those entities, in order to ensure pseudonymisation (if the purpose is to draw inferences about individual entities), or de-identification (if the purpose is other than to draw inferences about individual entities)

   2.6 **Data Security**
   Minimise the risks arising from data acquisition, storage, access, distribution and retention, and manage the unavoidable risks

   DON'Ts

   2.7 **Identifier Compatibility**
   Don't merge data-sets unless the identifiers in each data-set are compatible with one another at a level of reliability commensurate with the potential impact of the inferences drawn

   2.8 **Content Compatibility**
   Don't merge data-sets unless the reliability of comparisons among the data-items in the sources reaches a threshold commensurate with the potential impact of the inferences drawn

3. Data Analysis

   DO's

   3.1 **Expertise**
   Ensure that all staff and contractors involved in the analysis have:
   a. appropriate professional qualifications
   b. training in the specific tools and processes
   c. sufficient familiarity with the real-world system to which the data relates and with the manner in which the data purports to represent that real-world system
   d. accountability for their analyses

   3.2 **The Nature of the Tools**
   Understand the origins, nature and limitations of data analytic tools that are considered for use

Table 2 – (*continued*)

3.3 The Nature of the Data Processed by the Tools
Understand the assumptions that data analytic tools make about the data that they process, and the extent to which the data to be processed is consistent with those assumptions. Important areas in which assumptions may exist include:
a. the presence of values in relevant data-items
b. the presence of only specific, pre-defined values in relevant data-items
c. the scales against which relevant data-items have been measured
d. the precision with which relevant data-items have been expressed

3.4 The Suitability of the Tool and the Data
Demonstrate the applicability of each particular data analytic tool to the particular data that it is proposed be processed using it

DON'Ts

3.5 Inappropriate Data
Don't apply data analytics unless the data satisfies threshold tests commensurate with the potential impact of the inferences drawn, in relation to data quality, internal consistency, and reliable correspondence with the real-world systems about which inferences are to be drawn

3.6 Humanly-Understandable Rationale
Don't apply an analytical tool that lacks transparency, by which is meant that the rationale for inferences that it draws is expressible in humanly-understandable terms

4. Use of the Inferences

DO's

4.1 The Impacts
Understand the potential negative impacts on stakeholders of reliance on the inferences drawn, taking into account the quality of the data and the data analysis process

4.2 Evaluation
Where decisions based on inferences from data analytics may have material negative impacts, evaluate the advantages and disadvantages of proceeding, by conducting cost-benefit analysis and risk assessment from an organisational perspective, and impact assessments from the perspectives of other internal and external stakeholders

4.3 Reality Testing
Test a sufficient sample of the results of the analysis against the real world, in order to gain insight into the reliability of the data as a representation of relevant real-world entities and their attributes

4.4 Safeguards
Design, implement and maintain safeguards and mitigation measures, together with controls that ensure the safeguards and mitigation measures are functioning as intended, commensurate with the potential impacts of the inferences drawn

4.5 Proportionality
Where specific decisions based on inferences from data analytics may have material negative impacts on individuals, consider the reasonableness of the decisions prior to committing to them

4.6 Contestability
Where actions are taken based on inferences drawn from data analytics, ensure that the rationale for the decisions is transparent to people affected by them, and that mechanisms exist whereby stakeholders can access information about, and if appropriate complain about and dispute interpretations, inferences, decisions and actions

4.7 Breathing Space
Provide stakeholders who perceive that they will be negatively impacted by the action with the opportunity to understand and to contest the proposed action

4.8 Post-Implementation Review
Ensure that actions and their outcomes are audited, and that adjustments are made to reflect the findings

DON'Ts

4.9 Humanly-Understandable Rationale
Don't take actions based on inferences drawn from an analytical tool in any context that may have a material negative impact on any stakeholder unless the rationale for each inference is readily available to those stakeholders in humanly-understandable terms

4.10 Precipitate Actions
Don't take actions based on inferences drawn from data analytics until stakeholders who perceive that they may be materially negatively impacted by the action have had a reasonable opportunity to understand and to contest the proposed action. Denial of a reasonable opportunity is only justifiable on the basis of emergency, as distinct from urgency or mere expediency or efficiency. Where a reasonable opportunity is not provided, ensure that stringent safeguards, mitigation measures and controls are designed, implemented and maintained in relation to justification, reporting, review, and recourse in the case of unjustified or disproportionate actions

4.11 Automated Decision-Making
Don't delegate to a device any decision that has potentially harmful effects without ensuring that it is subject to specific human approval prior to implementation, by a person who is acting as an agent for the accountable organisation