
Ethical Data Handling

DATA
POLICY

HAPPY DATA™

Data handling (how to procure, store, manage, use, and dispose of data) must align with ethical, as well as legal, principles in order to protect an organization's reputation, employees, and customers. Ethical handling often requires understanding of technical and statistical concepts. Please reach out to the Data Management team if you are not familiar with these concepts.

Introduction

[OUR COMPANY] is committed to handling data, including personal data, in an ethical way. This includes considering the impact personal data handling has on individuals and guarding against misuse. The ethical principles in this policy aim to protect the firm and its people from inaccurate or unjust inferences as technology advances and data becomes more available than ever. We also take our legal and regulatory obligations seriously. You can read about our compliance practices in our privacy policies: [LINK TO YOUR PRIVACY POLICY].

Avoiding Misrepresentation

1. **Data fishing:** users of company data shall not selectively gather or report data in a way that distorts or obscures reality. Prohibited practices include, but are not limited to, cherry-picking data, omitting some data points, and strategically timing data capture at opportune times.
2. **Statistical smoothing:** users of company data shall not engage in misleading statistical smoothing, nor engage in “data mining snooping” or any other practice involving the performance of mass correlations on a dataset in search of statistically significant findings.
3. **Misleading visualizations:** users of company data shall not use charts and graphs to present data in a misleading manner. Prohibited practices include, but are not limited to, manipulating a chart’s scale in a way that distorts or obscures reality (e.g., making trends look better or worse), omitting data points, ignoring accepted visual conventions, and comparing facts or groups without clarifying their relationship.
4. **Unclear definitions:** users of company data must clarify which definitions they are using for data in their analyses, especially where there might be multiple accepted definitions.

DO: Define the question you are trying to answer, and data sets and points you will need to answer that question, before gathering data.

DO: Familiarize yourself with data visualization best practices.

DO: Fully explain your visualizations. A best practice is to include a separate document explaining that variables have the same scale, any relationships between compared groups, and any definitions or assumptions relied on.

DO: Utilize the company's data visualization team!

Minimizing Bias

1. **Biased use of collected data:** users of company data shall draw conclusions that naturally and logically flow from the collected data. Users shall not manipulate data or its interpretation to make it support a pre-determined conclusion.
2. **“Hunch and search”:** users of company data shall not use only the data that satisfies a hunch while failing to account for other possibilities the data may surface. Before drawing conclusions from a correlation, users should attempt to identify explainer variables that may account for the correlation (i.e., ensure you are not declaring causation when there is only correlation).
3. **Biased sampling:** users of company data shall, to the best of their ability, limit bias in data sampling by using statistical tools to select sample sets and by selecting adequate sample sizes.
4. **Cultural or contextual bias:** users of company data, especially where data is used in a decision-making capacity, shall make all due effort to understand where cultural bias may affect outcomes. Users shall further make all due effort to add context to data-based decisions such that data is presented and actioned in the most neutral way possible.

DO: Define the question you aim to answer and your data collection plan before you start gathering data.

DO: Account for other causes of a correlation beyond your variable of interest.

Ensuring Data Quality

1. **Quality dimensions:** [THE COMPANY] (through the Data Operations team) shall develop procedures to measure and improve accuracy, consistency, integrity, timeliness, and validity of critical data.
2. **Lineage:** [THE COMPANY] (through the Data Operations team) shall develop procedures to certify datasets and establish the lineage of those datasets.
3. **Metadata:** [THE COMPANY] (through the Data Operations team) shall develop reliable metadata for critical data, including consistent definitions of individual data elements, lineage, steward information, and protection level.

Types of Projects at Higher Risk for Ethical Issues

1. **Identification projects:** thoughtfully consider the data selection method and the demographic data needed.

2. **Behavior capture projects:** consider content, capture method, and any legal review that may be necessary.
3. **BI/Analytics/Data Science:** including profiling prospects and forecasting activities.
4. **Decision-making:** including granting or denial of permissions, status, or relationship changes (Note: personal information may not be used in automated decision-making without consultation from Data Governance and Privacy Counsel).