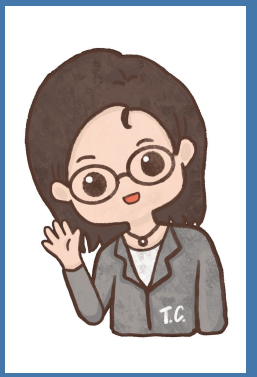
MTH 208 Exploratory Data Analysis

Lesson 10: Ethical Considerations in EDA

Ying-Ju Tessa Chen, PhD

Associate Professor
Department of Mathematics
University of Dayton

@ying-ju
ying-ju
ychen4@udayton.edu



Learning Objectives

- Introduction to Data Ethics
- Key Ethical Concerns in EDA
- Group Discussion on Case Studies
- Best Practices

Introduction to Data Ethics

Definition

Data ethics encompasses the moral obligations involved in gathering, protecting, and using personally identifiable information, particularly how these actions affect individuals and society.

Importance

Ethical considerations are paramount as they guide professionals in making responsible decisions that respect privacy and ensure fairness. Highlighting the impact of unethical data handling through recent cases can emphasize the real-world significance of ethical practices.

The Core Questions of Data Ethics

- `Is this the right thing to do?`
    - Example: Using data from users without their explicit consent for research.
- `Can we do better?`
    - Hypothetical: Could anonymization techniques be enhanced to protect individual privacy more effectively?

Insight from Academia: Professor Dustin Tingley describes data ethics as a dynamic field that extends beyond legal compliance to encompass continuous improvement and questioning of data practices.

Five Principles Of Data Ethics

- `Transparency:` Clarity about data usage, collection, and sharing methods.
- `Accountability:` Taking responsibility for ethical data management and readiness to address and rectify any arising issues.
- `Integrity:` Honest and fair data collection and use, supporting data accuracy and validity.
- `Protection:` Implementing robust security measures and respecting data privacy.
- `Fairness:` Avoiding discrimination and actively mitigating biases in data practices.

These principles form the backbone of trustworthy data handling, ensuring ethical compliance and fostering public confidence in data-driven technologies.

Ensuring Data Privacy in EDA

Definition of Data Privacy

Data privacy involves guidelines and practices that ensure sensitive information is accessed and shared only by authorized parties. It protects personally identifiable information (PII) and personal health information (PHI).

Examples of Sensitive Data

- Personal identifiers like names, social security numbers, birthdates, and contact details.
- Sensitive records including financial information, medical records, and employment details.

Significance in Business

Sensitive data is crucial for operations and development in businesses and involves customers, employees, and shareholders. Proper data handling maintains confidentiality and builds trust, supporting ethical business operations.

Discussion Questions

- What types of data do you consider sensitive in your own life?
- Why is it important to protect such data?

Protecting Data Privacy: Laws and Real-World Implications

Regulatory Frameworks

Data privacy laws are crucial in setting standards for how data should be handled to protect individual privacy and ensure transparency. Here are some key examples of data privacy regulations from around the world:

- `General Data Protection Regulation (GDPR):` Enforced across the European Union, GDPR sets stringent rules for data handling and grants significant rights to individuals regarding their personal data.
- `California Consumer Privacy Act (CCPA):` This law mandates protections against unauthorized data access and ensures transparency in data usage within California.
- `Brazil's General Data Protection Law (LGPD):` Similar to GDPR, the LGPD regulates the processing of personal data of individuals in Brazil, emphasizing consent, rights, and transparency.
- `Personal Information Protection and Electronic Documents Act (PIPEDA):` This Canadian law governs how private sector organizations collect, use, and disclose personal information in the course of commercial business.

These frameworks not only protect data but also empower consumers by providing them control over their personal information.

Protecting Data Privacy: Laws and Real-World Implications (Continued)

Case Study

- `Scenario:` A healthcare provider fails to secure patient records, resulting in unauthorized access.
- `Consequences:` Legal repercussions, financial penalties, and a loss of trust from the public.
- `Discussion Questions:`
  - How can organizations ensure they comply with data privacy regulations?
  - Discuss potential strategies to prevent data breaches in a healthcare setting.

Case Studies

Case 1: Improper Data Sharing

A company shares customer data with a third party without explicit consent, leading to a data breach.

- Questions for Discussion:
  - What were the ethical lapses in this scenario?
  - How could the company have prevented this situation?
  - What steps should be taken moving forward to restore trust?

Case 2: Biased Algorithm Deployment

A job screening tool disproportionately filters out candidates from certain demographic backgrounds.

- `Questions for Discussion:`
  - Identify the biases present in the data or algorithm.
  - Discuss the consequences for affected candidates.
  - Propose measures to correct and prevent such biases in future tools.

Case Studies (Continues)

Case 3: Security Negligence

A hospital’s inadequate security measures lead to a significant data breach of patient records.

- Questions for Discussion:
  - What are the ethical violations in this case?
  - What could have been done differently to protect the data?
  - Suggest long-term strategies to improve data security in healthcare settings.

Best Practices for Data Privacy and Protection

Regular Audits and Updates to Security Measures

- **Importance:** Ensures ongoing protection against emerging threats and vulnerabilities.
- **Examples:** Patching software, updating encryption methods, revising access controls.

Clear and Accessible Privacy Policies

- **Purpose:** Allows users to understand how their data is used and what controls they have.
- **Tips:** Use plain language, be concise, provide clear consent and opt-out options.

Strategies to Mitigate Bias in Data Analysis

Techniques for Detecting and Correcting Bias

- **Methods:** Use statistical tools and algorithms designed to identify and adjust biases.
- **Examples:** Regression analysis to adjust for known biases, anomaly detection for outlier analysis.

Inclusion of Diverse Datasets and Perspectives

- **Value:** Avoids echo chambers and improves the robustness of findings.
- **Strategies:** Engage with stakeholders from varied demographics, incorporate multiple data sources.

Accountability and Transparency

Documentation of Data Sources and Methodology

- **Importance:** Enables traceability and reproducibility of findings.
- **Examples:** Maintaining detailed data dictionaries, version-controlled code repositories, comprehensive project logs.

Openness in Sharing Results and Admitting Errors

- **Benefits:** Builds trust and supports a culture of continuous improvement.
- **Practices:** Share both successes and lessons from failures, engage in transparent dialogue with stakeholders.

References

The lectures of this course are based on the ideas from the following references.

- Exploratory Data Analysis by John W. Tukey
- A Course in Exploratory Data Analysis by Jim Albert
- The Visual Display of Quantitative Information by Edward R. Tufte
- Data Science for Business: what you need to know about data mining and data-analytic thinking by Foster Provost and Tom Fawcett
- Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic