Google

Reading: The practices and principles of good data stewardship

As you have been learning, all data professionals are responsible for ensuring the quality, integrity, accessibility, and security of data. Data stewardship is the practice of ensuring that data is accessible, usable, and safe. Making data stewardship a normal part of your work habits will benefit everyone who relies on your analysis, both inside and outside of your organization. In this reading, you will learn more about data stewardship and receive some best practices that can assist in guiding your career in data analytics.

Respect privacy

Earlier in this course, you learned about Information that permits the identity of an individual to be inferred by either direct or indirect means. This kind of information is commonly referred to as personally identifiable information or PII. When users share personal information, they are putting a high level of trust into an organization. It is the responsibility of all who have access within the organization to help protect the privacy of their users. As a data analytics professional, it is important to be thoughtful about any personal data and exhibit great care to protect it. In different parts of the world, laws are in place to guide best practices for data privacy. Laws provide a foundation for best practices as you grow in knowledge and experience on how to support and sustain privacy. One of your responsibilities as a data professional will be to stay up to date with any change in data laws and regulations that govern data. Depending on your organization's location or industry considerations, there may be additional regulations and policies in place. Here are a couple of regional examples:

- General Data Protection Regulation or GDPR (European Union law):
    - The GDPR is described on their website as the toughest privacy and security law in the world. It imposes obligations onto organizations anywhere, so long as they target or collect data related to people in the European Union.
- Lei Geral de Proteção de Dados Pessoais or LGPD (Brazil's general law for the protection of personal data):
    - The LGPD is a data protection law that governs how companies collect, use, disclose, and process personal data belonging to people in Brazil. LGPD applies to companies that process data about individuals in Brazil.
- The California Consumers Privacy Act or CCPA (Privacy rights for California consumers):
    - The CCPA gives consumers more control over the personal information that businesses collect about them. These regulations provide guidance on how to implement the law.
    - Additionally, states like Virginia, Colorado, New York, Utah, and Connecticut have enacted similar legislation to protect consumer privacy in their states.

Be cautious of unintentional harm

Data analytics is expanding its influence across an increasing range of industries. Companies are using the results of data analysis to make informed decisions. Many of these decisions have the potential to impact people across a broad range of social and economic factors. It is good practice to continually strive to produce information that is accurate, while respecting cultural and social norms.

Due to the global marketplace, decisions play out differently in different cultures. Taking these issues and considerations into account is very important for the executive team of an organization. Also, companies are known to take a position on particular politicized social and cultural issues, and these can be reflected in their policies. As a data analytics professional, you must be cognizant of your company's policies. When presented with challenges, it is best to seek guidance from leadership within your organization on how to navigate.

Avoid creating or reinforcing bias

You have learned about bias within data and how it can have an impact on your analysis. Identifying bias is not always simple. A good practice when working with data is to keep in mind that data gathering is a task managed by humans–and that process is informed by people from different backgrounds, experiences, beliefs, and worldviews. These and other types of biases can affect the data and the results, which in turn can have an impact on business decisions. You will learn more about bias within data as you progress through the program.

Consider inclusivity

Often in your role as a data analytics professional, you will have access to data collected in a variety of ways. You will need to consider whether the methods of data collection have excluded information from particular populations. Inclusionary approaches can expand how any organization collects and analyzes data. Building diverse research teams, communicating clearly with user communities, and engaging in careful and critical analysis that considers equity and inclusion benefits all stakeholders.

Uphold high standards of scientific excellence

The processes and technology that you will interact with as a data analytics professional are deeply rooted in the scientific method. As you continue in your data professional journey, embrace inquiry, intellectual discussion, and collaboration. Invite feedback and assess feedback. Remember, artificial intelligence still depends heavily on the instructions provided by data professionals. The more time and consideration that goes into the process of data analytics, the better the results.

Different industries have different standards. In your role as a data analytical professional, you will need to be aware of the standards for the industries you are working in. Each industry will have its own standards based on industry conventions.

Conventions that work well in the transportation industry may not necessarily be as high of a priority for the healthcare industry. For example, in transportation, data is collected to create predictive

analytics models to analyze the best route based on traffic patterns. In the healthcare industry, data is analyzed in medical imaging, predicting genetic factors, and speeding up the development of treatments.

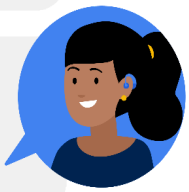
Data stewardship and ethics conversations

As you progress through your career as a data analytics professional, you will need to consider issues of ethical concern. For example, you may encounter situations where you address questions of bias or need to protect user data and personally identifiable information (PII). When these types of questions arise, many seek guidance and support from online communities of data professionals who have dealt with similar issues. The following graphics present scenarios involving these kinds of issues. You can also find text alternative versions of these conversations in the Data stewardship and ethics conversations transcript.

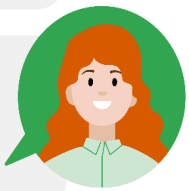
Data_Analyst_Luca My manager has asked me to do an analysis of our customers. How can I make sure that my analysis is representative and inclusive?

Reply from user 'Chidi_is_calculating'
My advice is to include as many contextual points within your analysis of the data as possible. Let me offer you an example. Let's say you have a data column for retail price and it's populated with numbers. Do those numbers reflect the local currency? If they are converted, do they reflect the correct conversion rate? It's those kinds of details that contain the contextual information...it can make a big difference.

Reply from user 'Analyzing_Aria'
Yeah...I experienced something like this while working with real estate data. The price paid for a property has changed over the years! If a family purchased a home decades ago, that could change the context and impact your understanding of the results.

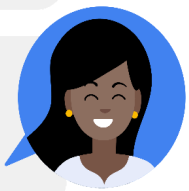


Reply from user 'Meryl_cool'
I'll add on to that point, if there is a column of data that isn't labeled clearly, be very careful not to make an assumption to what that data represents. Always reach out with questions.



Reply from user 'Estrella_EComm'
I've found that it's a good idea that whenever I encounter areas where there is some doubt about my interpretation, I'll reach out to other teammates or even stakeholders in different areas.

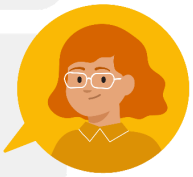
Reply from user 'Shareefah'
Are you dealing with the entire database or are you working with a sampling from the available data? Do you know if the data sample that you have been asked to work with truly represents all customers?



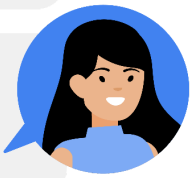
Reply from user 'Zoe_knows'
Good analysis begins with even better questions. In my experience, people see what they expect to see within data, unless there is an effort made to identify bias. You can begin the process by asking for the opinions of teammates and stakeholders.

Hiroki_DA I've heard about human-centered approaches to data science... what does that mean? Should I really be intervening in the data analysis process?

Reply from user 'Xavier_the_prof'
If I had to describe the idea of a human-centered approach to data analysis to someone outside of our field, I'd probably break it down to the interaction between humans, computers, and technology.



Reply from user 'Sofia_is_knowledge'
It's also important to point out that as a data analytics professional, you have the ability to influence any data science process.

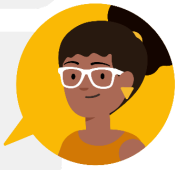


Reply from user 'Ning_Zhao_DA'
There is plenty of discussion about the tech side of what we do, but the truth is that data science is not the computers running data analytics on their own!

Reply from user 'Martina_Rojas'
So true!!! Analytics projects use both the power of machines and the expertise and judgment of the people that run them. It's important to realize the role that you play and the ways your decisions can impact the use of technology to get insights.

Data_driven_Dimitri What are some of the ways you have handled data and analytics in sensitive industries like healthcare and finance?

Reply from user 'DS_Megan'
I've always advised those who I've mentored to start with a high-level of respect for sensitive information–especially data that contains any personally identifiable info. If you include them into your normal operations, they become ingrained into the way you approach each project you work on.



Reply from user 'Logical_Luana'
I work within the healthcare industry and I can tell you that respecting privacy is one of the more important aspects of my job. Every data professional in healthcare must protect the privacy of patients. No results of tests or other medical records are shared. In fact, all of the identifying data is encrypted and it can't be accessed.

Reply from user 'Trending_Theo'
Your experience is quite similar to what I have encountered in my work with financial institutions. There is a very high level of data stewardship that is maintained at all times. Same as in healthcare–any information that could identify an individual's financial records are encrypted. If you offer financial services, you also need to be aware of laws that protect the privacy of personal information. In the United States, companies are required to explain how they share customer information and safeguard sensitive data.

Key takeaways

Data stewardship is the responsibility of every data professional. This responsibility goes beyond interactions with the data. By conducting your work in ways that are socially beneficial and inclusive, you will increase your ability to identify human bias. Guide your efforts through scientific and ethical principles and stay aware of possible bias throughout the data analysis process.
