

— Data Science Ethics

Matt Brems, Noelle Brown, Adi Bronshtein

Plagiarism

Simply put, plagiarism involves both:

- stealing someone else's work or idea
- and lying about it afterward.

Plagiarized work includes turning in code that you didn't write or building someone else's idea without giving credit.

Using Others' Work or Ideas

We do this all the time!

- We see how others have solved a problem.
- We search for examples of code online.
- We reference solution code.

THIS IS OK! (with *some* limitations)



Lying About It Afterward

This is where things get messy.

By putting your name (or username) on a piece of work, you are accepting credit for the work!

- Blog Post
- Lab
- Assessments
- Projects

[How do I safely write code in my own words and not plagiarize?](#)

General Assembly Policy

Open source code is inherently open and usually available for free use to modify and implement into any non-commercial project. However, please review their license and usage guidelines, give credit to the project and/or author, and do not try to hide or disguise the usage of such code.

Example

2. Plot the rolling_mean for Weekly_Sales. What general trends do you observe?

```
#This function here will
role1 = store1_sales.rolling(window = 1).mean()

role4 = store1_sales.rolling(window = 4).mean()

role13 = store1_sales.rolling(window = 13).mean()

#plot the figure here to compare the different rolling means.

fig = plt.figure(figsize=(12, 8))

mean = plt.plot(role1, color='red', label='Rolling Mean (1 week)')

mean = plt.plot(role4, color='blue', label='Rolling Mean(4 week)')

mean = plt.plot(role13, color='green', label='Rolling Mean(13 week)')

plt.legend(loc='best')
plt.title('Rolling Mean for 1,4 and 13 Months')
plt.show()
```

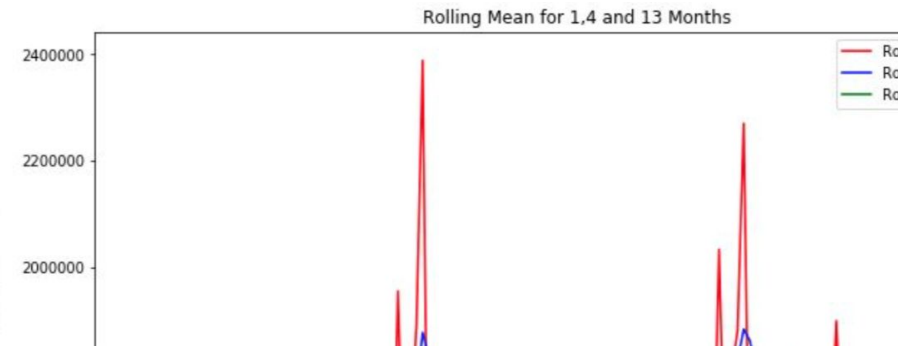


2. Plot the rolling_mean for Weekly_Sales. What general trends do you observe?

```
rolmean1 = store1_sales.rolling(window = 1).mean()
rolmean4 = store1_sales.rolling(window = 4).mean()
rolmean13 = store1_sales.rolling(window = 13).mean()

fig = plt.figure(figsize=(12, 8))
mean = plt.plot(rolmean1, color='red', label='Rolling Mean (1 week)')
mean = plt.plot(rolmean4, color='blue', label='Rolling Mean(4 week)')
mean = plt.plot(rolmean13, color='green', label='Rolling Mean(13 week)')

plt.legend(loc='best')
plt.title('Rolling Mean for 1,4 and 13 Months')
plt.show()
```



How do we avoid lying about who deserves credit?

By providing credit!

```
In [2]: # This code has been slightly adapted from code  
  
def plot_continuous_pdf(low, high, dist_name =  
w = 5):
```

My classmate Riley Dallas shared this line of code.

<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>



Why is plagiarism important?

Let's say there's a student who plagiarized work.

- This isn't fair to students who did their work "the right way."
- This isn't fair to the career coaches who are evaluated on their ability to help students get jobs.
- This isn't fair to an instructor who recommends/vouched for that student.
- This isn't fair to all of the other GA job-seekers.
- That student is getting credit for work they didn't do!

It can also be very embarrassing for the person who plagiarizes...



https://twitter.com/math_rachel/status/1210301328792203264?s=20

YouTube thinkfluencer Siraj Raval admits he plagiarized boffins' neural qubit papers – as ESA axes his workshop

Oops I did it again. And by it, we mean, ripped people off

https://www.theregister.co.uk/2019/10/14/ravel_ai_youtube/



<https://twitter.com/stephaniehicks/status/1212077611125338113?s=20>

General Assembly Policy

All of the following are considered plagiarism or cheating:

- Turning in work that is not your own.
- Turning in someone else's work as your own.
- Hiring, or paying someone to do your work for you.
- Copying words or code without giving credit.
- Building or copying someone else's idea without their knowledge or giving credit.
- Giving incorrect information about a source.
- Changing words, variable names, etc. but copying the code or files of a source without giving credit.
- Copying so many ideas or code blocks from a source that it makes up the majority of your work, whether you give credit or not.
- Failing to put a quotation in quotation marks.

How can I avoid plagiarism?

1. Avoid copying or “copying” where possible.
2. Provide credit to any source you copy or “copy” from.
3. If you’re unsure of whether or not you should provide credit, err on the side of providing credit!

General Assembly Policy

General Assembly has a zero tolerance policy towards plagiarism and cheating. It is destructive to classroom culture, and exhibits a clear lack of respect for classmates, instructors, the company, and the greater community. Any work considered to have been plagiarised will not be accepted and will not count towards graduation requirements. If a project exhibits evidence of plagiarism or cheating, the student will not be able to display the project at a GA-sponsored class “science fair” or “meet & greet.” Any student found plagiarising or attempting to plagiarise will be disciplined accordingly (including but not limited to removal from class).

That being said, I need to give credit for this presentation...

- Matt Brems
- Noelle Brown
- Data Science Ethics (University of Michigan) - available on [Coursera](#)
- Dylan Brown (Noelle's brother!)



— Let's talk about ethics



Learning Objectives

By the end of the lesson, you should be able to...

- Identify how data science is implemented every day in obvious and non-obvious ways.
- Understand some of the ethical responsibilities associated with gathering and using data.
- Understand the implications of data science decisions from various perspectives.



Disclaimer

Though I try to be ethical, I am neither a lawyer nor a professional ethicist.

This lesson is not about my ethical code or even your ethical code. The purpose of this lesson is to get you to think critically and empathetically about decisions that you make in data science.



What is Ethics?

“The discipline dealing with what is good and bad and with moral duty and obligation” - [Merriam-Webster](#)

Ethical Obligations

You don't have to be a data scientist to get value out of this.

Ethical obligations aren't a new thing.

- Doctors
- Lawyers
- Businesses
- Government
- Parents

— Norms

Case Study 1: Crowd-Sourced Pothole Identification



<https://www.wired.com/insights/2014/03/potholes-big-data-crowdsourcing-way-better-government/>

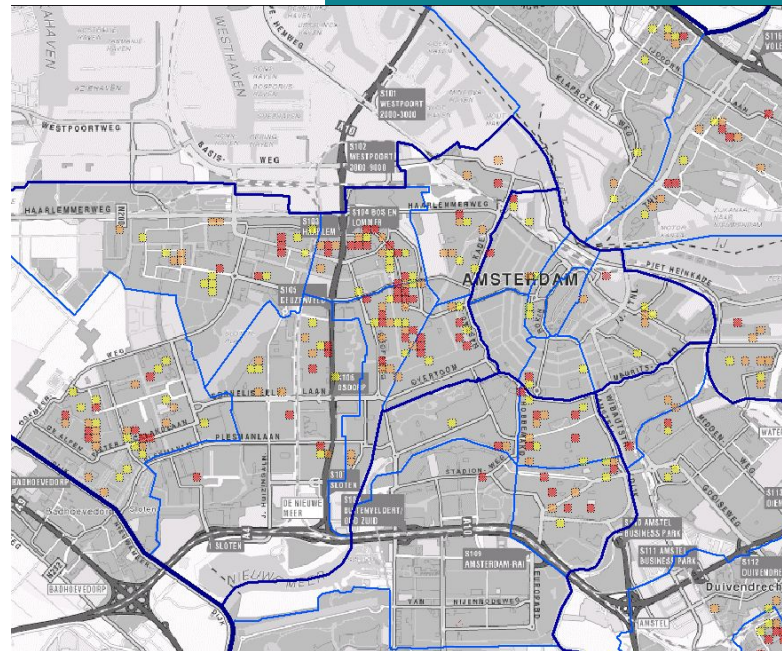
Case Study 2:

Target Ads



<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#2c1c3a666686>

Case Study 3: Predictive Policing



<https://nij.ojp.gov/topics/articles/overview-predictive-policing>

How can we make sure we consider different perspectives?

Having diverse data teams is crucial to ensuring you understand the consequences of any decisions that you make.

<https://teachdatascience.com/diversity/>

<https://www.spsnational.org/the-sps-observer/fall/2014/why-diversity-important-science>

Bias

Code is still written by humans.

For example - [Coded Bias](#)





Discussion:

Food for Thought



Something to think about:

1. What are some applications of data science that can be unintentionally biased?
2. What are some ways that you can think of that could help reduce bias in technology/data science?

Data Privacy

If I take a picture of you, who owns that picture?

Data Collection: Worst-Case Scenario

- Think: “If this data were to fall into the hands of the individual who could do the most harm with it, what would happen?”
- While we don’t always want to focus on the worst thing that can happen, it’s important to keep this in mind!
 - This mindset can help us identify what data we don’t actually need to collect.



Conclusion

1. **Be aware** of what your algorithm is prioritizing and what implications that has.
2. Encourage **diversity in data science** teams.
3. **Don't collect data** you don't need to collect.



