Bias In Data

Ethics CSCI-77800

Alicia Wade and Latoya Boland

As humans, we have many biases, both implicit and explicit. Biases are systematic errors in thinking influenced by cultural and personal experiences. Biases distort our perception and cause us to make incorrect decisions.

Computers, data, and algorithms are not actually completely objective. It is true that data analysis can help us make better decisions, but it is not immune to bias. Humans create technologies and algorithms. As a result, they often have human biases encoded into them.

Types of Biases

Selection Bias - Before we can analyze data or use machine learning algorithms, we need to collect data. Data collection is subject to selection bias (also called sample bias). Selection bias occurs when study subjects (i.e., the sample) are not representative of the population. Selection bias can be due to poor study design if the sample is too small or is not randomized. Selection bias can also crop up when the only data available is influenced by historical bias — systematic influence based on historic social and cultural beliefs.

Selection bias takes place when data is chosen in a way that is not reflective of real-world data distribution. This happens because proper randomization is not achieved when collecting data.

Types of selection Bias

1. Sampling bias: occurs when randomization is not properly achieved during data collection.
2. Convergence bias: occurs when data is not selected in a representative manner. e.g., when you collect data by only surveying customers who purchased your product and not another half, your dataset does not represent the group of people who did not purchase your product.
3. Participation bias: occurs when the data is unrepresentative due to participations gaps in the data collection process.

Let's say Apple launched a new iPhone and on the same day Samsung launched a new Galaxy Note. You send out surveys to 1000 people to collect their reviews. Now instead of randomly selecting the responses for analysis, you decide to choose the first 100 customers that responded to your survey. This will lead to sampling bias since those first 100 customers are more likely to be enthusiastic about the product and are likely to provide good reviews.

A Reuters article from 2018 highlights how the company Amazon produced a machine-learning algorithm that suffered from such a selection bias. The company designed the algorithm to help recruiters hire top talent. The model was trained on thousands of resumes from people that were or were not hired by Amazon. It learned 50,000 phrases associated with resumes and began to ignore common phrases, such as the names of programming languages. However, the algorithm also learned to downgrade resumes that contained the word "women's." This included resumes that referenced women's colleges, teams, or committees.

This is an example of selection bias because the data used to train the algorithm were not representative of the modern applicant pool. The majority of Amazon's past applicants and employees were male. This means a larger proportion of the successful resumes in the training

data came from male applicants. Amazon did not explicitly train the algorithm to use gender. Yet, the algorithm still found and used gender-associated terms to weed out women candidates.

We can do our best to avoid selection bias by doing everything possible to have a representative sample, not just a convenient one. For example, it's a good idea to include data inputs from multiple sources to diversify data. This is easier said than done, however, and we need to acknowledge and address historical bias in data sources and work towards building frameworks to increase inclusivity.

Bias in building and optimizing algorithms.

- **Algorithmic Bias**

  Algorithmic bias arises when an algorithm produces systematic and repeatable errors that lead to unfair outcomes, such as privileging one group over another. Algorithmic bias can be initiated through selection bias and then reinforced and perpetuated by other bias types.

  Facial recognition software is an area where algorithmic bias can do a lot of harm. This software is sold to police departments and used to recognize criminals in surveillance footage. If the software systematically makes more mistakes depending on race or gender, people in some groups will be incorrectly pursued more often, which has serious, negative outcomes for individuals.

  The Gender Shades project tested commercial facial recognition software for these kinds of biases. IBM, Microsoft, and Face++ are three companies that offer facial recognition software with a binary gender classifier feature. Researchers assessed the accuracy of these algorithms and discovered that they suffered from algorithmic bias.

The algorithms were good at identifying lighter males, okay at identifying darker males and lighter females, and very bad at identifying darker females.

Another key point when it comes to algorithmic bias in facial recognition software is that the algorithms are proprietary, making them “black boxes”. In addition to not knowing what data were used to train and test the algorithm, we can’t know how it was designed or how it works. As a result, it’s impossible to evaluate the algorithms themselves.

Avoiding algorithmic bias relies on transparency, especially concerning data used for training and testing an algorithm. In response to the poor performance of facial recognition with darker females, a new benchmarking dataset was developed (PPB) that is more representative of the full spectrum of humanity. This is a big step forward, as long as the new dataset is actually used by companies making and selling facial recognition software.

- **Evaluation Bias**

  Testing an algorithm with a non-representative dataset leads to evaluation bias. Testing with a non-representative benchmarking dataset would give high overall accuracy scores, even if the algorithms were inaccurate for certain groups.

Bias In interpreting results & Drawing Conclusions

Bias also influences the final stages of data analysis: interpreting results and drawing conclusions. The following bias types are ones we should watch out for when evaluating or generating data reports:

- **Confirmation Bias**

  Confirmation bias is our tendency to seek out information that supports our views. Confirmation bias influences data analysis when we consciously or unconsciously interpret results in a way that supports our original hypothesis. To limit confirmation bias, clearly state hypotheses, and goals before starting an analysis, and then honestly evaluate how they influenced our interpretation and reporting of results.

- **Overgeneralization Bias**

  Overgeneralization bias is inappropriately extending observations made with one dataset to other datasets, leading to overinterpreting results and unjustified extrapolation. To limit overgeneralization bias, be thoughtful when interpreting data, only extend results beyond the dataset used to generate them when it is justified, and only extend results to the proper population.

- **Reporting Bias**

  Reporting bias is the human tendency to only report or share results that affirm our beliefs or hypotheses, also known as “positive” results. Editors, publishers, and readers are also subject to reporting bias as positive results are published, read, and cited more often. To limit reporting bias, report negative results and cite others who do, too.

Coding Component - Impact of Biases in Data

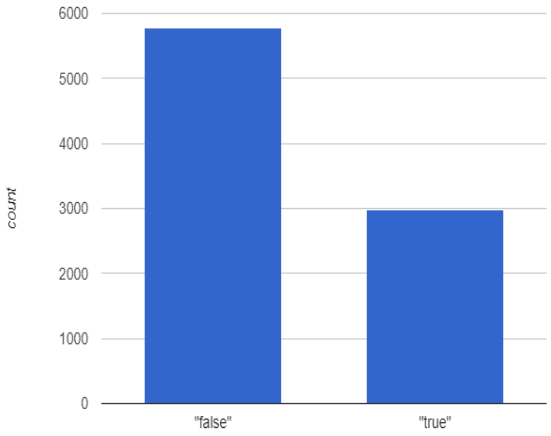
We analyzed the 2019 Stop, Question and Frisk Data Set to see what information we could gather. When analyzing Data, we aim to extract and catalogue data, to pinpoint and evaluate relationships, patterns, and trends so we can glean insights and draw conclusions based on the data and use these to make informed decisions.

Sometimes we approach data analyzation with the best of intentions, but many times we approach it from a confirmation biased point of view, where we extract data just to prove what we suspect is probably true. When we created our data visualizations, we did just that. We wanted to prove that stop and frisks were racially motivated, and that Law Enforcement were using it target young minorities. We sorted the Data and chose to create visualizations that proved this theory.

We used Reporting bias to prove our theory. We only gave visualizations that aimed to prove that stop and frisks were not only racially motivated, but that most of the stops were harassment by NYPD Cops.

What we show is that anyone can take a dataset and although the information they portray is accurate, they can skew it to only show what they want to represent. To avoid this type of bias, this is usually where ethics comes in. In order to prove how bias can be present; our visualizations show how some can create, interpret and report visualizations to convey the message that they would like to portray.

Stopped Vs Arrested



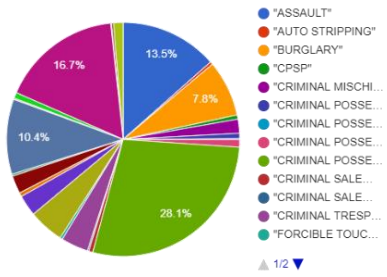
What suspected stop and frisk crimes lead to arrest?

Out of 8764 stop and frisk records for 2019, only 2986 lead to an actual arrest.

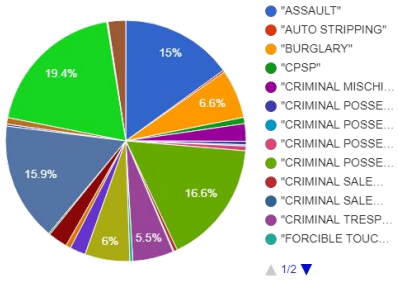
This number does not tell whether the number of arrests were convicted of the suspected crime.

One can conclude that about 65% of stop and frisk detainments are unnecessary and harassment by the cops. (Would this be accurate based on the chart?)

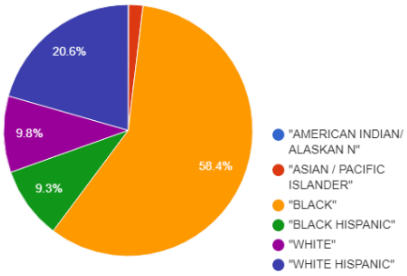
Percent of Persons Suspected by Crime

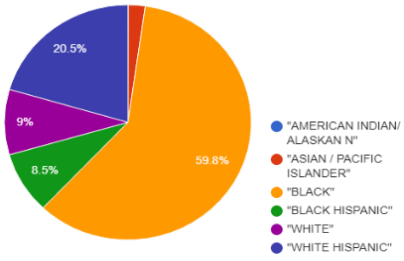


Percent of Arrested Persons by Crime



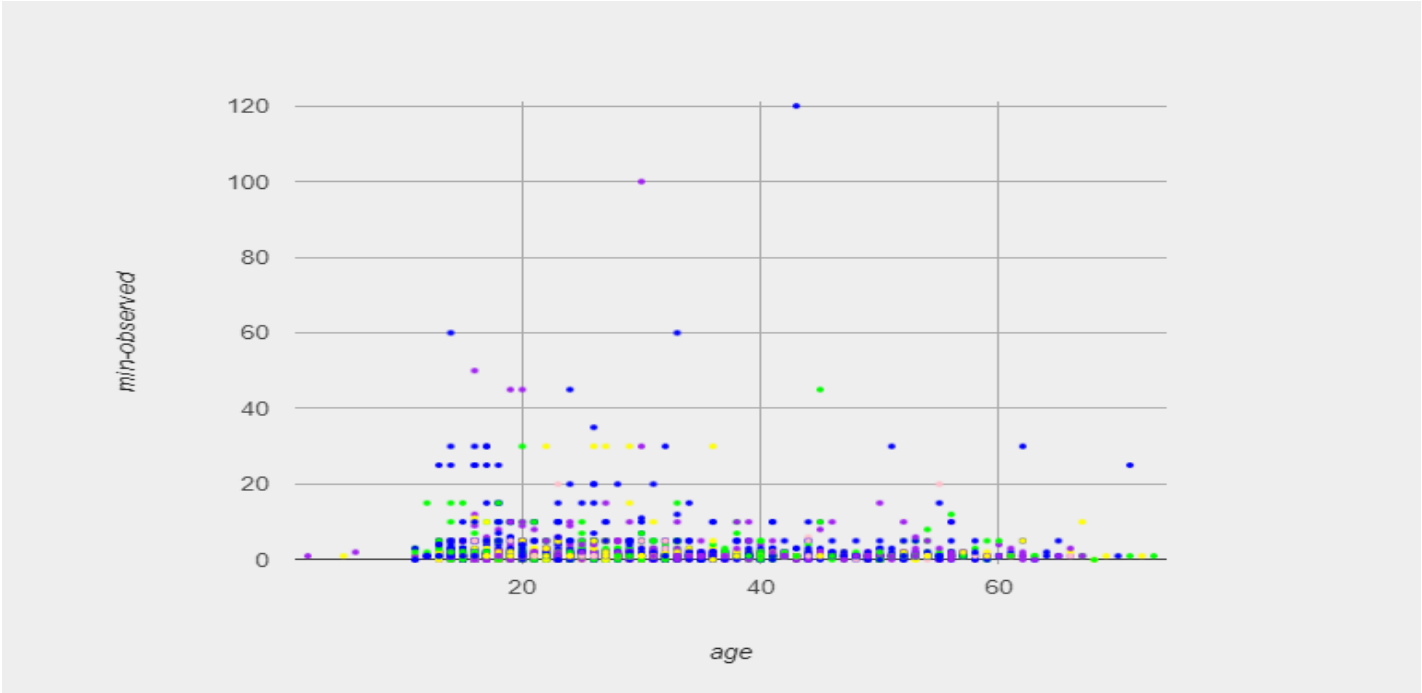
Percent of Arrested vs Suspected by Race





Charts represent the suspects who were stopped and frisked and those who were actually arrested for their suspected crimes. Compared to other races Blacks are twice times more likely to either be suspected or arrested during these stop and frisks. Does this mean that Blacks commit the most crimes? Or are they just twice times more likely to be stopped by a police officer and suspected of a crime. Is that a fair assumption?

Boroughs of Arrested Suspects

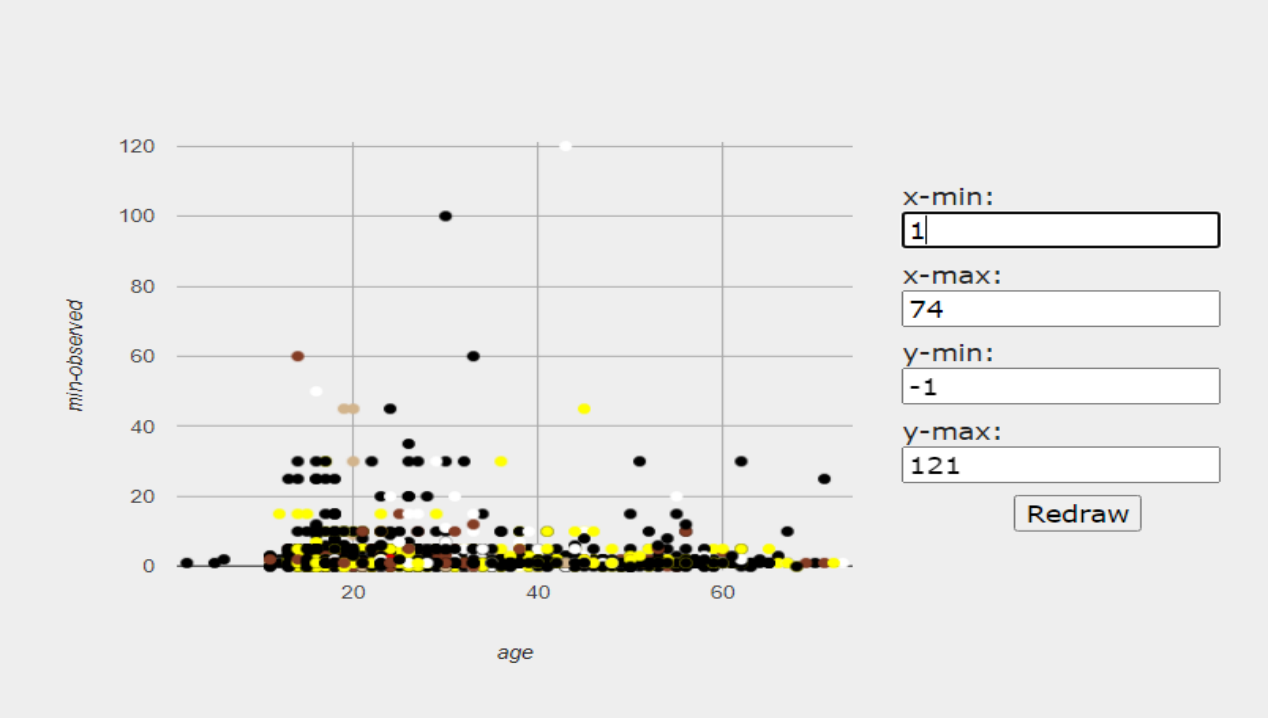


Scatter plot was made to show the different boroughs arrests were made in the stop and frisks, based on age and minutes observed for the suspected crimes. The Blue dots are

Surprisingly, the main Borough that arrests were made in was Manhattan.

Is it a fair assumption to make, that if you are a black teenager or young adult, and you visit the Manhattan area that you are likely to be suspected of and arrested for a crime?

Race of Arrested Suspects

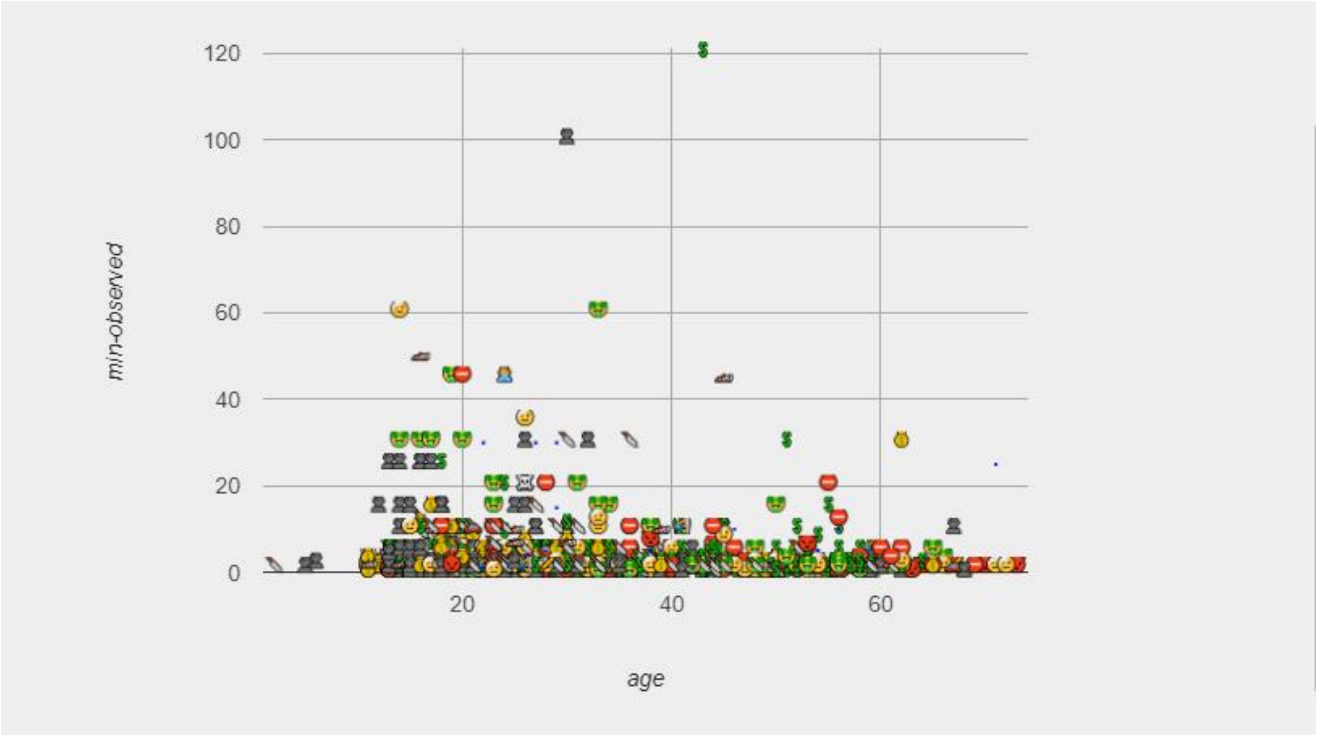


Scatter plot was made to show the different races of suspects arrested.

Here we see that blacks no matter the age are predominantly the ones that are arrested for suspected crimes.

Does this mean that Blacks are the main targets of the stop and frisk and law enforcement?

Arrested by Crime based on Age and Time observed

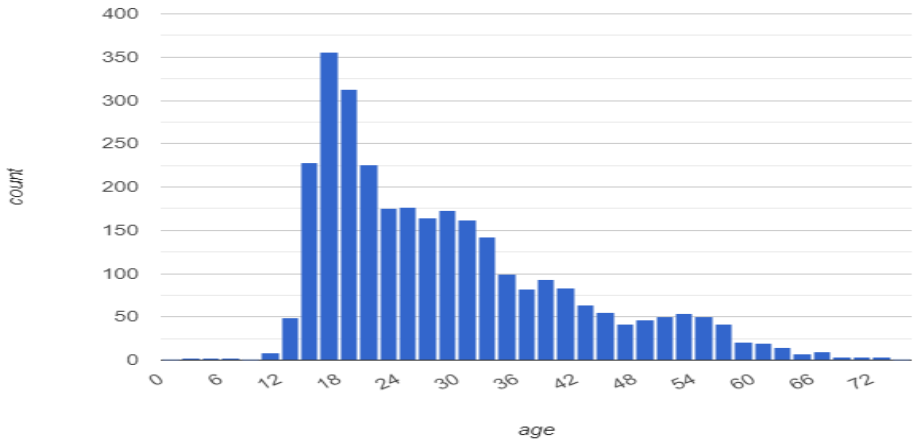


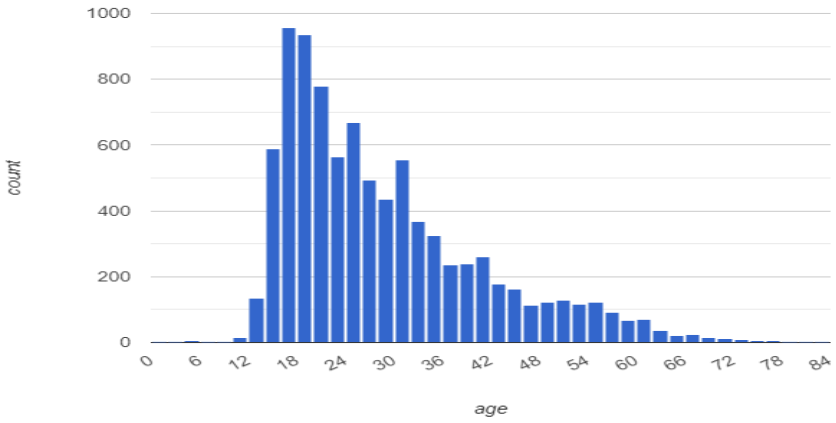
We analyzed whether the types of crimes had a correlation to age and whether that in itself had an impact on the time taken to observe the crime. We filtered to only show the individuals who were actually arrested.

Most crimes committed by teenagers to young adults appeared to be robbery. Most of the suspects were arrested within minutes of being suspected.

Is it fair to say that law enforcement is targeting this population or that this population prefers to steal from others and does not like to work for the things they want?

Age range of arrested vs suspected





Both charts represent the age range for both arrested and suspected stop and frisks. We can interpret that the general age range for both arrests and suspected cases is between the ages of 16 - 24.

Can we conclude that this is the reason why many minority teenagers and young adults have an aversion to the police?

Data Science Ethics

Today, data science has a significant impact on how businesses are conducted in disciplines as diverse as medical sciences, smart cities, and transportation. Whether it's the protection of personally identifiable data, implicit bias in automated decision-making, the illusion of free choice in psychographics, the social impacts of automation, or the apparent divorce of truth and trust in virtual communication, the dangers of data science without ethical considerations are as clear as ever. The need for a focus on data science ethics extends beyond a balance sheet of these potential problems because data science practices challenge our understanding of what it means to be human.

Algorithms, when implemented correctly, offer enormous potential for good in the world. When we employ them to perform jobs that previously required a person, the benefits may be enormous: cost savings, scalability, speed, accuracy, and consistency, to name a few. And because the system is more precise and reliable than a human, the outcomes are more balanced and less prone to social prejudice.

Ethical Practices

Making Decisions

Data scientists should never make judgments without contacting a client, even if the decision is for the interest of the project. The aims and objectives of projects must be understood by both data scientists and clients.

Let’s say a data scientist wishes to take action on behalf of a customer on a certain ongoing project. Even if the action is advantageous to the client and the project, it must be explained to the client, and no choice should be made on their behalf. Data scientists should only make decisions when it is expressly stated in the contract or when their authority allows them to.

Good Intentions with Data

Intentions of data collection and analyzing data must be good. Data professionals must be clear about how and why they use the data.

If a team is collecting data regarding users’ spending habits, to make an app to manage expenses, then the intention is good.

Data ownership

One of the important concepts of ethics in Data Science is that the individual has data ownership. Collecting someone’s personal data without their agreement is illegal and immoral. As a result, consent is required to acquire someone’s data.

Signed written agreements, digital privacy policies that require users to accept a company’s terms and conditions, and pop-ups with checkboxes that allow websites to track users’ online behavior using cookies are all typical approaches to get consent. To prevent ethical and legal issues, never assume a consumer agrees to you gathering their data; always ask for permission.

Transparency

Data subjects have a right to know how you plan to acquire, keep, and utilize their personal information, in addition to owning it. Transparency should be used when acquiring data. You should create a policy that explains how cookies are used to track user's activity and how the information gathered is kept in a secure database, as well as train an algorithm that gives a tailored online experience. It is a user's right to have access to this information so that they may choose whether or not to accept your site's cookies.

Privacy & Confidentiality of Data

Data scientists are continually involved in producing, developing, and receiving information. Data concerning client affiliates, customers, workers, or other parties with whom the clients have a confidentiality agreement is often included in this category. Then, regardless of the sort of sensitive information, it is the data scientist's responsibility to protect it. Only when the customer provides permission for data scientists to share or talk about this type of information should it be disclosed or spoken about. Complete privacy of clients' or customers' data must be maintained.

Even if a consumer consents to your organization collecting, storing, and analyzing their personally identifiable information (PII), that doesn't mean they want it made public.

Personally, identifiable information includes:

Phone Number, Address, Full Name, PAN card number, and so on.

To preserve people's privacy, make sure you're keeping the information in a secure database so it doesn't get into the wrong hands. Dual-authentication password protection and file encryption are two data security solutions that assist safeguard privacy.

Resources

- **Intro**
    - Intro - Resource 1

- **Types of Biases**
    - Types of Biases - Resource 1
    - Types of Biases - Resource 2

- **Data Science - Coding**
    - Bootstrap Data Science
    - Stop and Frisk Spreadsheet
    - Code | DOC
- **Data Science & Ethics**
    - Ethics Resource 1
    - Ethics Resource 2