

Data Cleaning

Key Words

- **Qualitative Data:** Non-numerical information that describes experiences, perceptions, and behavior.
- **Transcripts:** A written record of detailed content of an interview, focus group, or observation, usually produced from an audio or video recording of the event.
- **Redact:** To remove, alter or hide sensitive information in a document disclosure or publication.

What is Data Cleaning?

Careful review of transcripts before starting data analysis to correct errors or inconsistencies and redact the data to ensure participants personal information is not included.



Why is Data Cleaning Important?

Steps of Data Cleaning

1. Person 1 - Reviews transcript while listening to corresponding audio
2. Person 1 - Correct for errors and inaccuracies
 - This may require multiple reads, listens, etc.
3. Person 1 - Review again to redact personal information
4. Person 2 - Review transcript for accuracy and no personal information included

What do we Redact?

- Names of people
- Neighborhoods
- Locations (addresses, street names)
- Schools and libraries
- Parks and recreation centers
- Businesses
- Train stations or lines
- Dates
- Crimes

Example of Data Cleaning

one. I wo, if it's that bad, why don't you just like - you know what I'm saying? - find an alternative that won't get you in trouble?" You know what I'm saying? That's how I feel because where I live at, it's a lot of that stuff happening. It's a lot of crimes going on. It's Congress Heights. Y'all know what Congress Heights look like. Congress heights looks terrible. Especially Alabama Avenue, that y'all look terrible. So it's sort of like-- you know what I'm saying? Because I go to the bus stop, I see a bunch of people-- I don't want to say crackheads because you know what I'm saying? I see a bunch of crackheads.

S2: 17:13

[inaudible] too but.

S0: 17:16

A bunch of junkies.

find an alternative that won't get you in trouble?" You know what I'm saying? That's how I feel because where I live at, it's a lot of that stuff happening. It's a lot of crimes going on. It's ~~Congress Heights~~[neighborhood MASKED]. Y'all know what [neighborhood MASKED]~~Congress Heights~~ look ~~like~~like [neighborhood MASKED]. ~~Congress heights~~ looks terrible. Especially ~~Alabama~~[street MASKED]~~Avenue~~, that ~~y'all~~joint look terrible. So it's sort of like-- you know what I'm saying? Because I go to the bus stop, I see a bunch of people-- I don't want to say crackheads because you know what I'm saying? I see a bunch of crackheads.

S2: 17:13

~~[inaudible]~~I heard they call them cluckers now too ~~but~~.

S0: 17:16

A bunch of junkies.

Example of Data Cleaning

S2: 16:05

do you think police response times play a role in a number of homicides in our commended?

S3: 16:14

Yes, I do. Again, let me bring up the female Naomi when she passed away around here. My friend Arnique, she actually was in the hallway with the female, and my uncle was also. And they were like-- the ambulance. They were helping the little girl. They didn't have anything but t-shirts and garments. And you're not supposed to put dirty tissue or anything on a wound. So yeah, I do.

S2: 16:05

~~do~~Do you think police response times play a role in a number of homicides in our ~~commended~~community?

S3: 16:14

Yes, I do. Again, let me bring up the female, [name MASKED]~~Naomi~~ when she passed away around here. My friend, [name MASKED]~~Arnique~~, she actually was in the hallway with the female, and my uncle was also. And they were like, they were-- the ambulance. They were helping the little girl. They didn't have anything but t-shirts and garments. And you're not supposed to put dirty tissue or anything on a wound. So yeah, I do.

Example of Data Cleaning

supposed to run. So I'm watching security because they've put me in an apartment that's pretty much like a perch. I see everything through my three windows, bedroom-- my bedroom is right by the lobby door. So I hear all the noise coming in or going out.

S1: 09:55 Are you on the first floor?

S2: 09:57 I'm on the lobby level.

S1: 09:58 The lobby level.

S2: 09:59 My building is weird. Martin Luther King is a main [inaudible], and that's my address, 2652, right, Martin Luther King. The back of the building is actually where the lobby level is. You got to go through the-- but if you don't have a [file?] to get to the parking lot, you're not going to get back there to the lot. That's weird, so.

supposed to run. So I'm watching security because they've put me in an apartment that's pretty much like a perch. I see everything through my three windows, bedroom-- my bedroom is right by the lobby door. So I hear all the noise coming in or going out.

S1: 09:55 Are you on the first floor?

S2: 09:57 I'm on the lobby level.

S1: 09:58 The lobby level.

S2: 09:59 My building is weird. ~~Martin Luther King~~[street MASKED] is a main [throughway?inaudible], and that's my address, ~~2652, right, Martin Luther King~~[address MASKED]. The back of the building is actually where the lobby level is. You got to go through the-- but if you don't have a [file?]fob to get to the parking lot, you're not going to get back there to the lot. That's weird, so.

Activity

Activity Steps

1. Grab a highlighter and marker
2. Review the provided transcript
3. Mask (redact) any personal/sensitive information and replace with [Blank MASKED]
 - Examples: [street MASKED], [location MASKED], [train line MASKED]
4. Note any errors you come across in the text



Go Forth and Be Great!

This project was supported by Award No. 15PNIJ-22-GG-01420-RESS, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

Learn more about Safety 7 at <https://www.childtrends.org/project/centering-youth-experiences-youth-participatory-action>.