Ethics & Big Data

Modern databases document lots of data throughout our livesThis information explosion is especially useful because so much of this data can be combined using values like an address or government id number. For this activity, you will need to use the following information that Bureau of Governmental Unified Surveillance (BoGUS) received on a US citizen named Pat Smith with SS# 123-45-6789 to answer the questions on the next page.

Citibank Visa Account 4700345687562

4/1/19	ECMC Substance Abuse Clinic - Buffalo, NY	$	50.00
5/3/19	ECMC Substance Abuse Clinic - Buffalo, NY	$	50.00
5/29/19	ECMC Substance Abuse Clinic - Buffalo, NY	$	50.00
6/20/19	St. Joseph's Emergency Room – Buffalo, NY	$	100.00
6/24/19	AA.com round trip between BUF & Bogata, Colombia	$	1,200.00
6/30/19	Hilton Grand Hotel, Bogata, Colombia	$	500.00
7/5/19	ECMC Substance Abuse Clinic - Buffalo, NY	$	50.00
8/18/19	AA.com round trip between BUF & Bogata, Colombia	$	958.00
8/24/19	Hilton Grand Hotel, Bogata, Colombia	$	630.00

Department of Homeland Security Passport 444555666

6/24/19	Exit MIA flight to Bogata, Colombia
6/30/19	Enter MIA from Bogata, Colombia
8/18/19	Exit MIA flight to Bogata, Colombia
8/24/19	Enter MIA from Bogata, Colombia

HSBC Commercial Bank Business Account 987654

7/1/19	Deposit	$	1,000,000.00
8/26/19	Deposit	$	10,000.00

ATT Cellphone Pen Register Account 716-867-5309

6/23/19	Incoming call from Bogata, Colombia -- 20 minutes
6/24/19	Outgoing call to Bogata, Colombia -- 5 minutes
8/16/19	Incoming call from Bogata, Colombia -- 4 minutes
8/18/19	Outgoing call to Bogata, Colombia -- 23 minutes

Buffalo & Erie County Library Services Card ID 568877

6/1/19	445566 - "PreColombian Art"
6/1/19	324455 "FARC, Colombian Drug Cartels, & the Cocaine Trade"
6/1/19	45112 "Sightseeing in Bogata"

Buffalo Police Department Records matching "Pat Smith"

6/20/19	11-41 for cocaine overdose & 10-29a on "Pat Smith"

Using this information, what would you conclude Pat is doing on their trips to Colombia? Can you think of any other conclusions about Pat that could be consistent with the information?

United States regulations require banks report **all** cash deposits of $10,000 or more and **all** withdrawals over $5,000. It also makes it illegal to "structure" a series of withdrawals or deposits to evade those reports. These regulations only apply to banks which have operations inside the United States. So Pat's lack of withdrawals means that they either received a lot of cash from other people or maintain an account outside the United States. Does this change your earlier conclusions?

The US Government keeps a "Do Not Fly" list of suspected terrorists. This likely includes people linked to the cocaine cartels (whose crimes also include people smuggling, slavery, and political assassinations). The "Do Not Fly" list is not public, the standards to be added to this list are not public, and the "redress" process to remove your name from this list is not public. Given all of this, should Pat be added to the "Do Not Fly" list?

BoGUS is… bogus (I made them up), but the Federal Government really collects and mines all kinds of data. International travelers are now required to specify their social media accounts. Would you feel more or less comfortable adding Pat to the "Do Not Fly" list if the data analytics included social media contacts? What about if it could include the pictures posted to social media?

How would using people's social media interests and contacts bias who gets added to "Do Not Fly" list? Is it possible to mine social media data in a way that limits those biases? What unbiased data sets should be used?

Suppose your company was awarded the contract to build BoGUS's data mining system and automate adding the names of suspected terrorists to the federal "Do Not Fly" list. What will your team do when you jointly reassigned to the group developing this system?