

Anomaly Detection for Behavior Data



Justin Holladay, 2021

The Problems

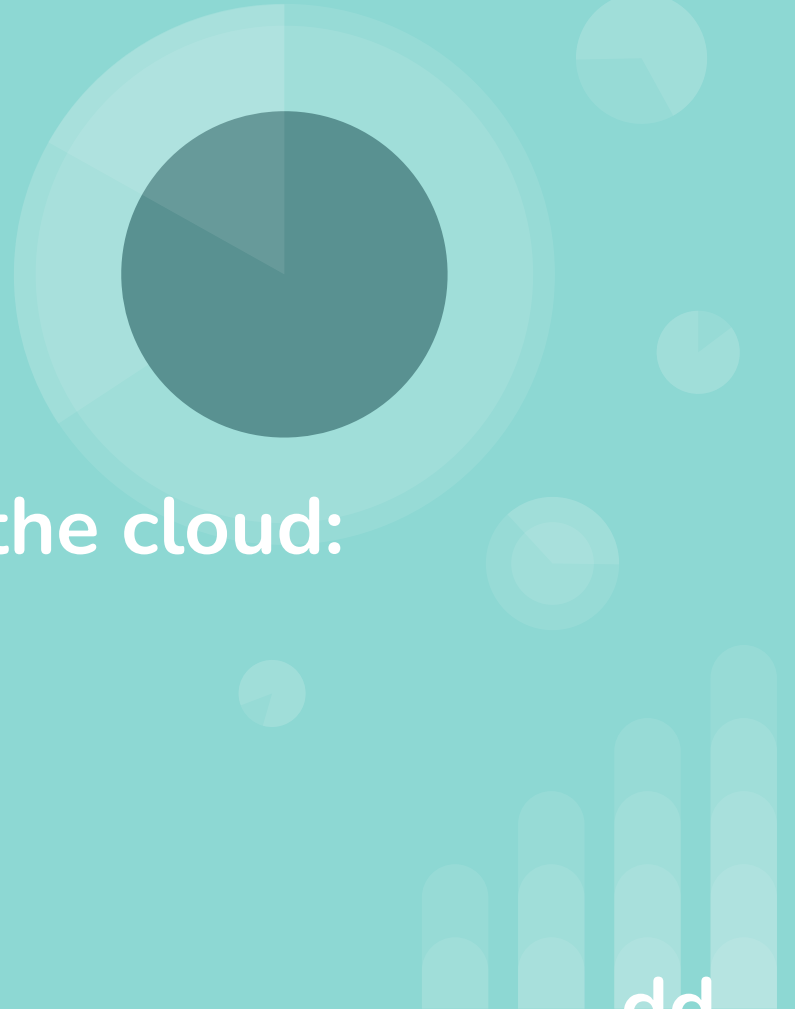
- Data in many formats
- Behavioral data is complicated
- Quick action is essential



Data in Many Formats

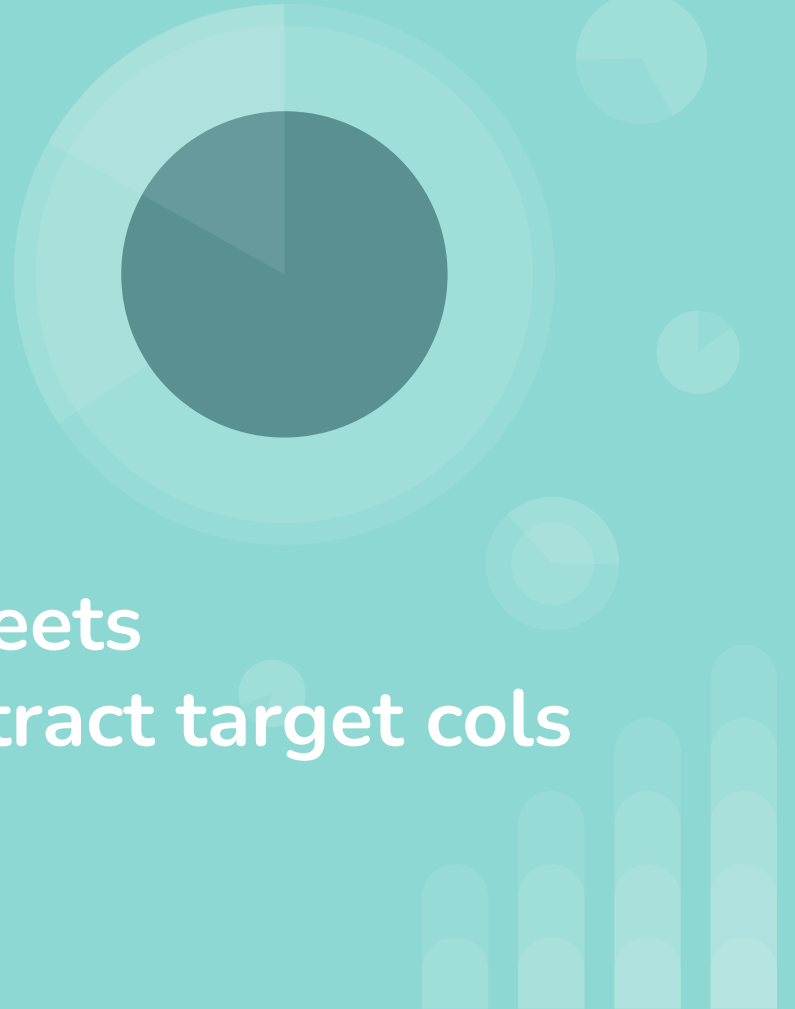
Moving data from paper to the cloud:

- Different People
- Different Companies
- Different Methods



Standardizing formats

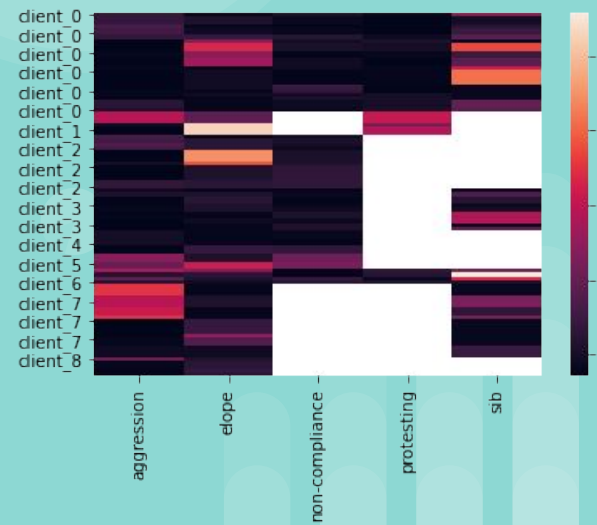
- Merge and sort spreadsheets
- Use FuzzyMatching to extract target cols differently named



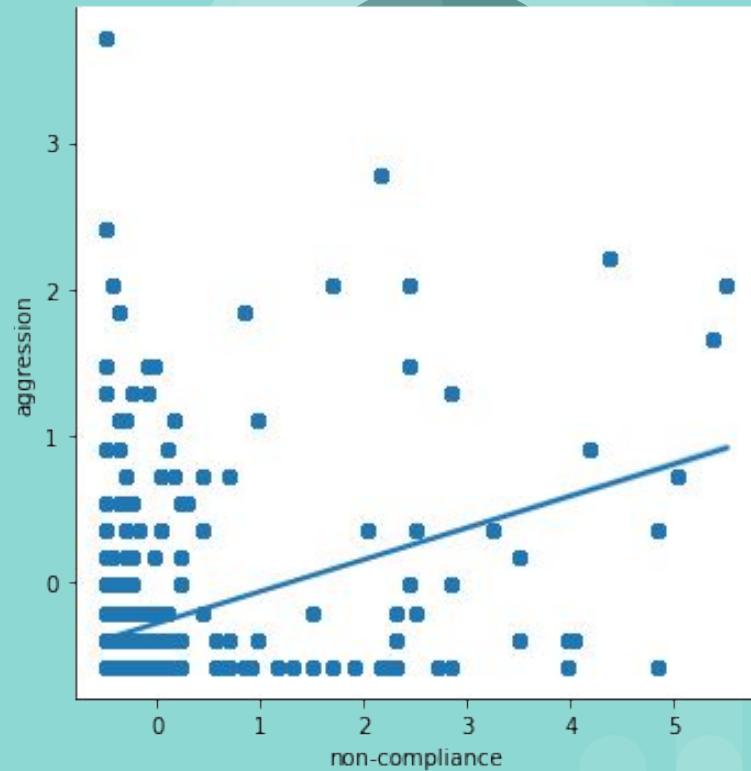
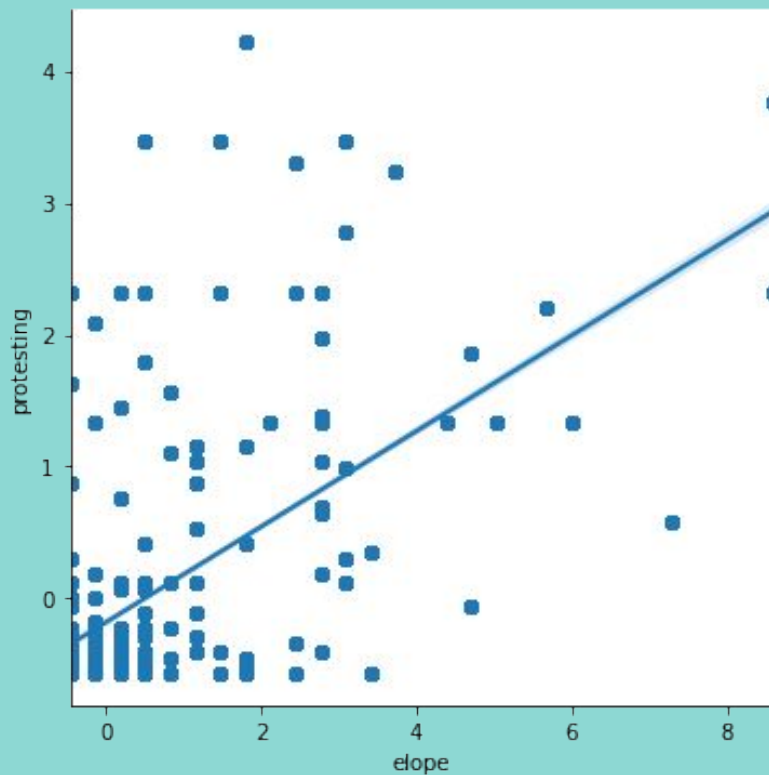
Exploring this data

Relationships:

- All clients and a bx
- Between different bx
- Individuals bx and intensity



Exploring (cont.)



Standardizing for Modeling

- Used StandardScaler()
- New featured generated
- NaN kept
- One-hot for categories



	Problem	Date	Day	Staff Member	Duration	aggression	elope	non-compliance	protesting	sib	client
client_3	elope	2021-06-17	3	staff_17	1.50	-0.585974	0.508188	-0.495307	NaN	0.405570	client_3
client_2	aggression	2021-04-05	0	staff_14	4.33	1.096018	0.508188	0.172368	NaN	NaN	client_2
client_6	na	2021-01-22	4	staff_7	7.00	0.348466	0.508188	-0.495307	0.411591	9.176212	client_6
client_3	na	2021-03-08	0	staff_21	4.00	-0.585974	-0.137430	-0.295005	NaN	3.860672	client_3
client_3	elope	2021-06-17	3	staff_19	4.00	-0.585974	0.508188	-0.495307	NaN	0.405570	client_3

Modeling

Manually filtered bx using deviation from mean to serve as true values

- Used DBSCAN
- Used KMeans



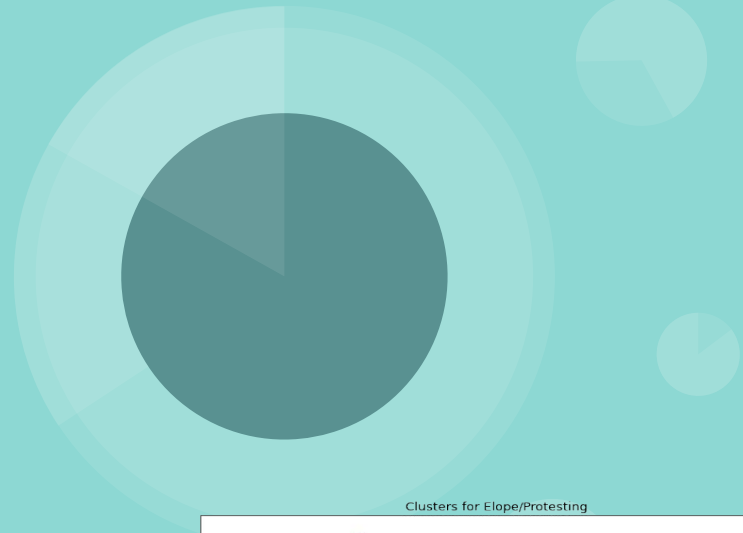
Modeling True Values

- Began with 1000 observations
- Filtered with deviation of mean
- 95 observations found 'extreme'



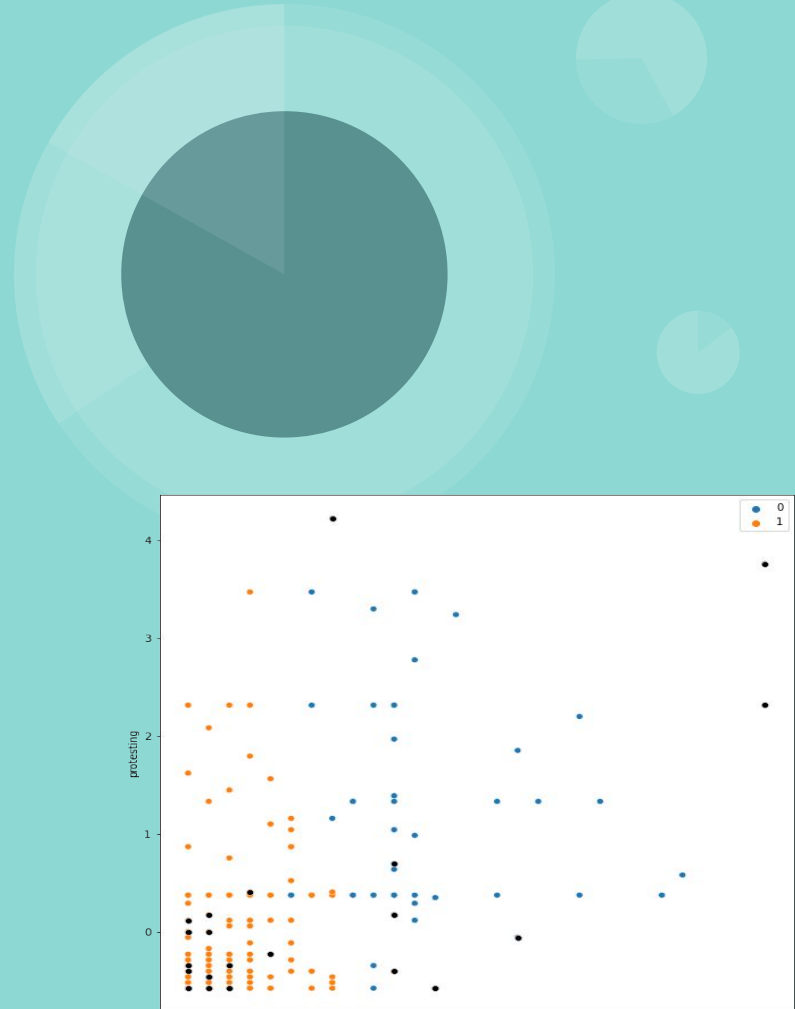
Modeling DBSCAN

- After preprocessing, 405 observations
- 15 observations found 'extreme'

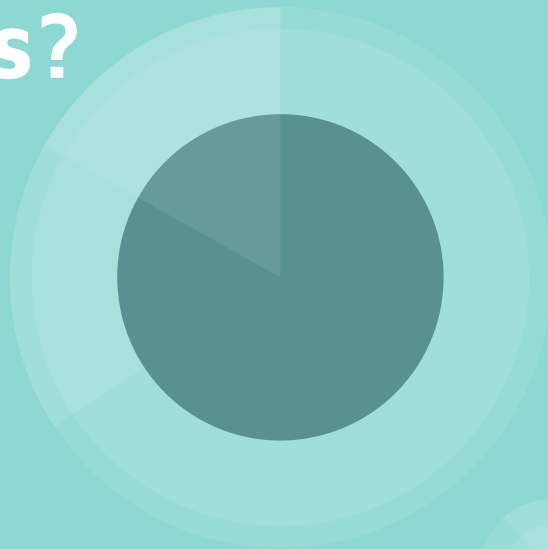


Modeling KMeans

- After preprocessing, 405 observations
- 394 observations found 'extreme'



What's the Point of This?



- Flag IOA issues
- Identify relationships in client bx
- Detect anomalies between clients/staff
- Discover novel relationships

Conclusion

- Many relationships worth exploring
- Manual method best
- DBSCAN > KMeans here



Furture work:

- Extend to skills targets
- Collect location, time of day and other vars
- Identify more features