

2hack2furious



Team Goal: Data
De-Sensitizer/Anonymizer

Tasks:

- Develop data-cleaning framework
- Develop system to evaluate data sensitivity (internally comparable and objectively)
- Working proof-of-concept

Possible Concerns/Issues:

- Every dataset is different. It may not be easy to account for these differences

2data2anonymous

```
graph LR; A[Upload CSV] --> B[Clean data]; B --> C[Create score for potential sensitivity]; C --> D[Recommend if we should delete or substitute]; D --> E[Perform deletions & substitutions]; E --> F[Return data to user];
```

2data2anonymous

Upload CSV

Francesco	18	UNC	0.72	118 Green Dr, ...
Sean	27	SFU	1.45	243 Blue Rd, ...
Kai	64	UNC	2.32	19 Blue Rd, ...
Rax	53	UTexas	0.18	8 Orange Ln, ...
Michelle	25	UVic	1.74	117 Red Rd, ...
Je'aimé	19	SFU	4.39	97 Blue Rd, ...
Ricardo	24	UVic	5.25	18 Orange Ln, ...
Sasmita	72	UNC	1.34	13 Green Dr, ...
Ziggy	84	UTexas	1.27	4 Red Rd, ...
Charlie	39	UTexas	0.29	6 Red Rd, ...

*	<21	UNC	0.72	Green Dr, ...
*	25-30	SFU	1.45	Blue Rd, ...
*	60-65	UNC	2.32	Blue Rd, ...
*	50-55	UTexas	0.18	Orange Ln, ...
*	25-30	UVic	1.74	Red Rd, ...
*	<21	SFU	4.39	Blue Rd, ...
*	20-25	UVic	5.25	Orange Ln, ...
*	70-75	UNC	1.34	Green Dr, ...
*	80-85	UTexas	1.27	Red Rd, ...
*	35-40	UTexas	0.29	Red Rd, ...

Download