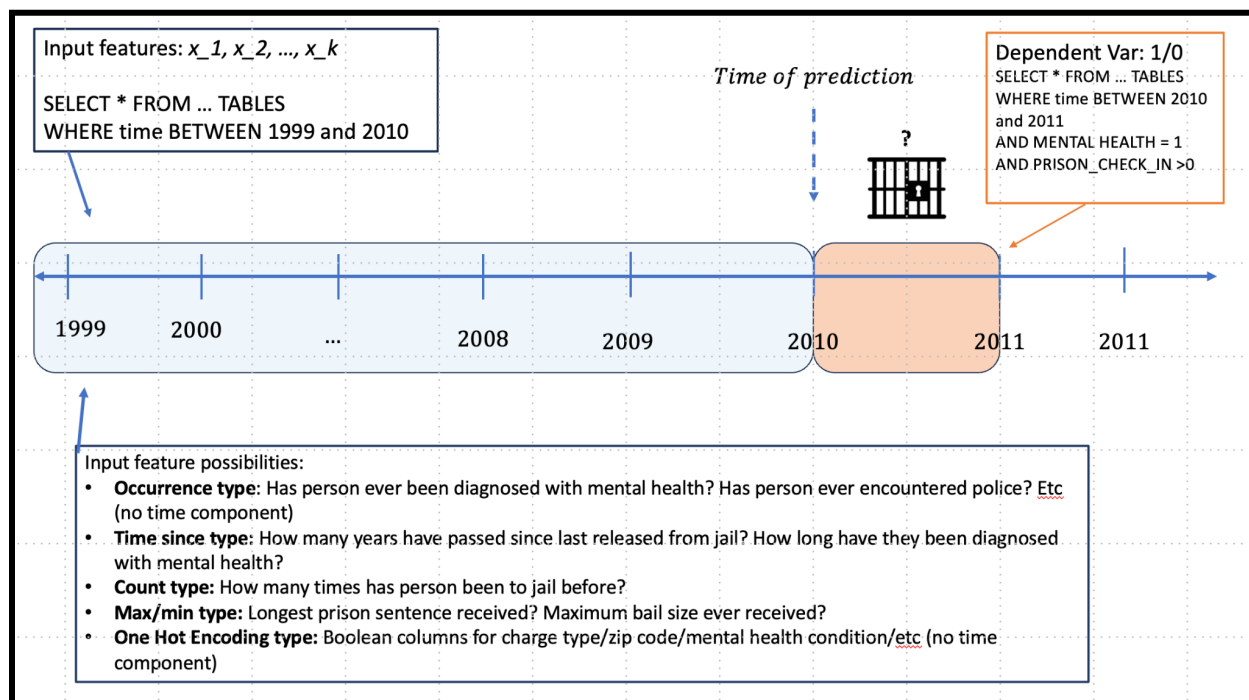


Appendix 2: Preliminary Analysis of Outcome Variable & Row Counts of Groups

Overview:

Eventually, we will need to build a dataset for our prediction task which takes time into consideration. If we wanted to predict whether or not some re-offends between years 2010 and 2011, for example, we would construct a feature vector for that individual that only looked at data before 2010. That feature vector would join together data across all our datasets, aggregating information to the person level. We will have to carefully consider exactly how to best engineer our dataset to be represented on the individual level. A schematic of that process is shown below:



Due to time constraints, however, we have not constructed this dataset yet. Instead, we have created a static dataset across all time on the individual level.

Below are row counts by different characteristics:

Group:	Number of Rows:	% of Total:
Total JO ID's:	919165	100%
People who are incarcerated 1 time or more	173376	18.9%

People who are incarcerated 2 times or more	74862	8.1%
People who show any sign of mental health issues (any screen result, medact incident with MH, MHC client, services)	131205	14.3%
BJMHS: referred to JCMHC	6958	0.8%
LSIR: MH need identified	15743	1.7%
PTA: MH_flag	5559	0.6%
Medical incident related with MH	8505	0.9%
Admitted to JCMHC (patient)	96091	10.5%
Called JCMHC with MH related issues	60681	6.6%
Received JCMHC service(s)	80828	8.8%
JC Department of Health encounters with MH related issues	21	0.002%

We then define our outcome variable:

If:

- Incarcerated 2 or more times = TRUE
AND
- Have mental health problems = TRUE

Then:

Outcome = 1

Else:

Outcome = 0

Below are cross tabulations of features w.r.t. our outcome variable:

Feature	Outcome = 0		Outcome = 1	
	# NULL:	Mean/Mode	# NULL:	Mean/Mode
Inmate incarcerated two times or more		Total 60601 inmates		Total 16045 inmates
Race		A 1490		A 2

		B 50833 I 1154 O 3617 P 3075 W 213911		B 3220 W 12823
Gender		FEMALE 205410 MALE 270959		FEMALE 4168 MALE 11877
Age		31.293237		31.260328
Mental Health needs identified		No 1266 Yes 3023		No 39 Yes 12820
Marital Status		D 10767 M 30195 OTHER 19975 S 101729 W 782		D 528 M 1556 OTHER 558 S 13210 W 193
Resident		False 323236 True 818784		False 9311 True 12094