Appendix

1 More Descriptive Analysis

In Tables 1 and 2 we provide descriptive statistics for other features in our prediction models. Specifically, Table 1 includes the categorical variables in the data set. These variables are for each individual: (i) Risk: their recidivism risk scaled into 3 levels. (ii) AdOffense: name of the offense committed that lead to them being admitted into the program. (iii) OffenseClass: class of the offense. (iv) Pdrug: Primary drug that the individual stated at admission to consume. (v) ReferralReason: reason for referral to the program. (vi) WhoReferred: Role of the person who referred the individual to the program. (vii) Gender. (viii) EmploymentS: employment type at admission . (ix) MaritalS: marital status at admission. (x) HousingS: states the living situation at admission. (xi) MedicaidSt: Medicaid enrollment status at admission. (xii) UniqueAgents: Number of unique agents that saw the person during his LOS. (xiii) FinalProgramPhase: Final program level that the person achieved. (xiv) RewardedBehavior: indicator to whether the persons received any reward during his LOS in the program, and (xv) Sanctions: indicator to whether the person received any sanction during his LOS in the program. Table 2: provides mean and standard deviation for continuous variables in the dataset. These variables are (i) AgeAtEnroll: Age of the individual at enrollment. (ii) CriminalHistScore: ranges from 0 to 30 and its the estimated score of the individual's past criminal history. (iii) and (iv) AvgMVistis (AvgReqMVisits) Average number of actual (required) monthly visits to the individual during their LOS. (v) and (vi) TotalMVistis (TotalReqMVisits): Total number of actual (required) visits to the individual during their LOS.

1.1 Details of the Outcome Prediction

Pre-processing. We apply the following pre-processing and feature selection process: (*i*) Data transformation: log transformation and square root transformation, were applied to enable the model to better understand the relative relationships between data points. (*ii*) Features grouping and outlier elimination: To enhance model performance, we visualized the distribution of unique values for each feature. Features with less than 3 unique values were dropped. For categorical features, we eliminated unique values that represented less than 10 percent of all unique values. For numerical features, we removed outliers using the Interquartile Range (IQR) method. (*iii*) Feature selection: we filtered features based on feature importance score for GBT model. We develop an automated data pre-processing pipeline to perform these three steps.

1.2 Figures for Arrivals and LOS

Table 1: Categorical Covariates Summary Statistics (N/A or Other Categories are Omitted)

Variable	Categories	County			
		DuPage	Cook	Will	Peoria
Risk	Highest	24.3	32.0	2.3	1.0
	High	60.7	26.2	35.1	24.7
	Medium	11.0	15.6	42.1	47.0
AdOffense	Drugs	43.0	67.8	31.7	37.0
	Property	31.1	17.6	52.5	46.3
	DUI	11.1	2.3	3.8	1.0
OffenseClass	Class 4	42.5	–	11.5	20.6
	Class 3	13.5	–	5.7	5.7
	Class 2	16.0	–	5.7	5.1
Pdrug	Heroin	27.0	43.6	32.3	9.5
	THC	18.6	18.5	17.5	21.6
	Coc.Crack	7.8	10.9	21.0	11.6
ReferralReason	Tech Violation	31.2	0.0	12.8	0.0
	3/4 Felon	20.5	70.5	59.2	80.0
	1/2 Felon	9.8	16.5	23.7	14.7
WhoReferred	Prob Officer	64.7	97.3	1.8	0.0
	Judge	32.0	1.3	0.7	91.3
	Pub. Defender	0.6	0.0	75.3	2.8
Gender	Female	25.2	21.3	21.7	19.8
	Male	74.8	77.5	78.2	80.0
EmplymntS	Full Time	49.7	85.7	38.2	6.7
	None	32.3	4.8	59.2	92.0
	Part Time	18.0	9.4	2.7	1.3
MaritalS	Single	86.4	85.6	15.0	22.9
	Married	5.9	7.1	1.8	5.7
	Divorced	4.7	2.3	0.2	1.8
EducationS	HighSchool	40.3	37.2	34.3	13.6
	No HighSchool	32.6	52.4	10.8	12.3
	Some College or Graduated	19.4	3.5	11.8	4.4
HousingS	Friend or Family	62.3	27.9	6.2	17.7
	Own/Rent	29.0	15.5	2.7	11.1
	No Home Reported	5.9	23.9	16.5	70.2
MedicaidS	Yes	23.8	48.4	8.3	3.3
UniqueAgents	4	11.6	2.2	8.6	–
	3	27.9	31.9	22.3	2.3
	2	60.6	65.9	69.1	97.7
FinalProgPhase	Level 3/4	11.1	15.7	32.3	0.3
	Level 1/2	56.5	14.4	22.7	3.1
	Level 0	2.9	35.5	7.0	27.0
RewardedBehv	Yes	4.0	29.1	2.5	1.5
Sanctions	Yes	91.8	99.3	89.8	41.1

Table 2: Continuous Covariates Summary Statistics

Variable	County			
	DuPage $\mu(\sigma)$	Cook $\mu(\sigma)$	Will $\mu(\sigma)$	Peoria $\mu(\sigma)$
AgeAtEnroll	30.5 (9.9)	41.8 (13.1)	33.6 (9.2)	35.5 (11.0)
CriminalHistScore	5.1 (1.6)	17.9 (11.3)	5.2 (1.7)	5.6 (1.4)
AvgMVisits	1.7 (1.0)	0.5 (0.5)	1.5 (0.8)	4.3 (1.9)
TotalMVisits	39.8 (30.5)	11.2 (11.0)	29.6 (20.0)	93.7 (56.2)
AvgReqMVisits	1.6 (1.2)	–	0.0 (0.0)	–
TotalReqMVisits	38.2 (35.3)	–	0.3 (0.6)	–

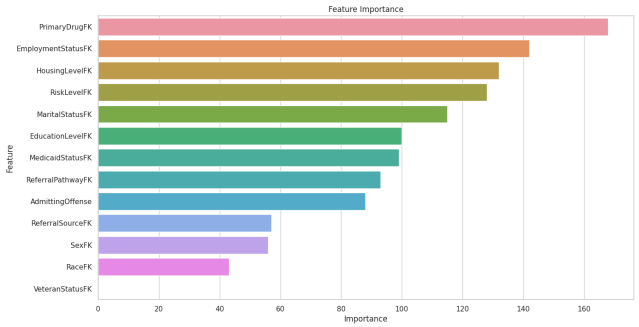
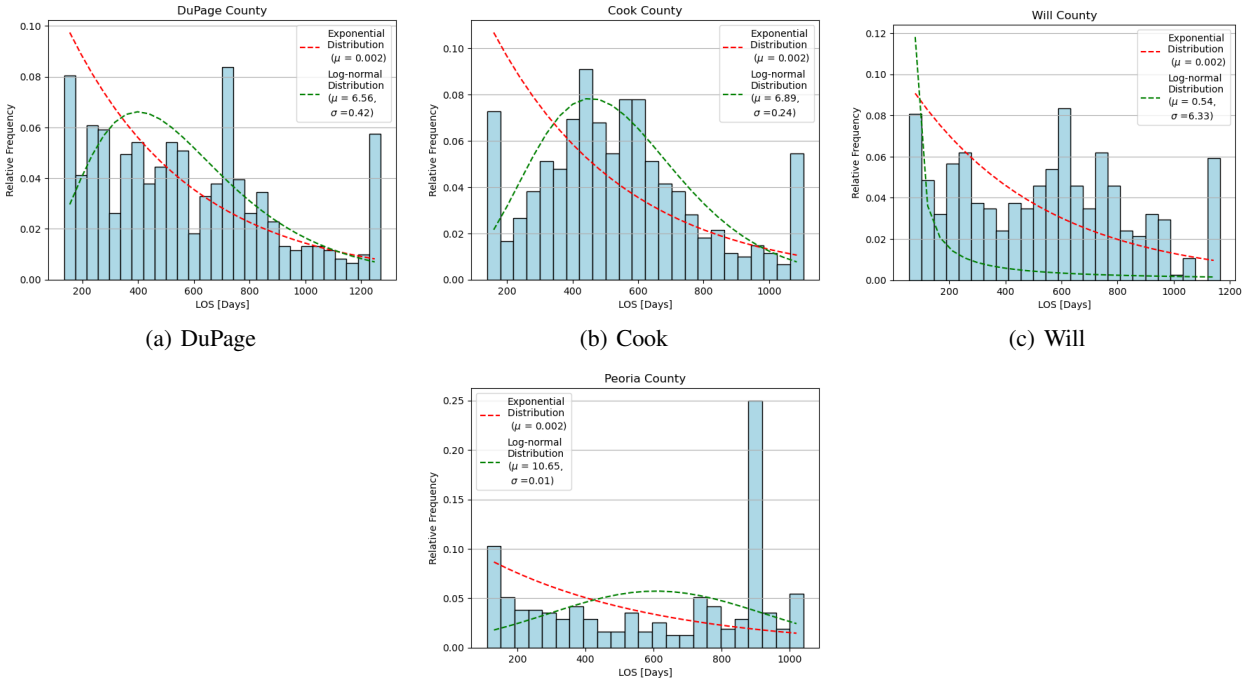


Figure 1: feature importance plot

Table 3: LOS by Outcome Summary Statistics

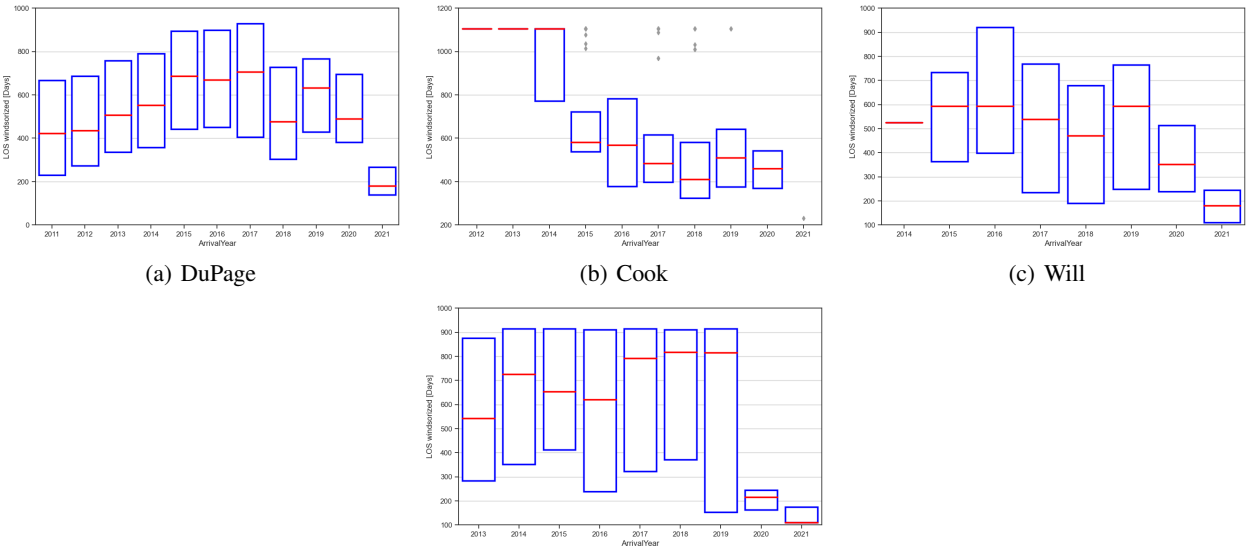
County	Race grp.	LOS yrs. by Outcome Mean (SD)		
		Com.	Not Com.	Revok.
DuPage (2011-2022)	All	1.6 (0.7)	1.4 (0.9)	1.7 (1.0)
	W	1.7 (0.7)	1.3 (0.9)	1.7 (0.9)
	AA	1.6 (0.8)	1.6 (1.0)	1.8 (1.1)
	H	1.5 (0.7)	1.6 (1.0)	1.7 (1.0)
	OT	1.3 (0.6)	0.5 (0.2)	1.9 (1.1)
Cook (2012-2022)	All	1.5 (0.5)	1.3 (0.8)	1.5 (0.8)
	W	1.8 (0.7)	1.3 (0.4)	1.1 (0.8)
	AA	1.5 (0.5)	1.3 (1.0)	1.5 (0.8)
	H	1.6 (0.4)	1.1 (1.0)	1.6 (0.8)
	OT	1.6 (0.7)	–	0.6 (0.3)
Will (2014-2022)	All	1.9 (0.6)	1.0 (0.9)	1.1 (0.8)
	W	1.9 (0.7)	0.9 (0.8)	1.1 (0.7)
	AA	2.0 (0.9)	1.3 (0.9)	1.2 (0.9)
	H	2.1 (0.7)	0.6(0.7)	1.3 (0.7)
	OT	2.2 (0.8)	1.0(1.4)	2.4 (–)
Peoria (2013-2022)	All	2.3 (0.5)	1.1 (0.7)	1.2 (1.0)
	W	2.2 (0.6)	1.3 (0.8)	1.1 (1.0)
	AA	2.3 (0.5)	1.1 (0.7)	1.3 (1.1)
	OT	2.3 (0.5)	0.8 (0.3)	–



(a) DuPage (b) Cook (c) Will

(d) Peoria

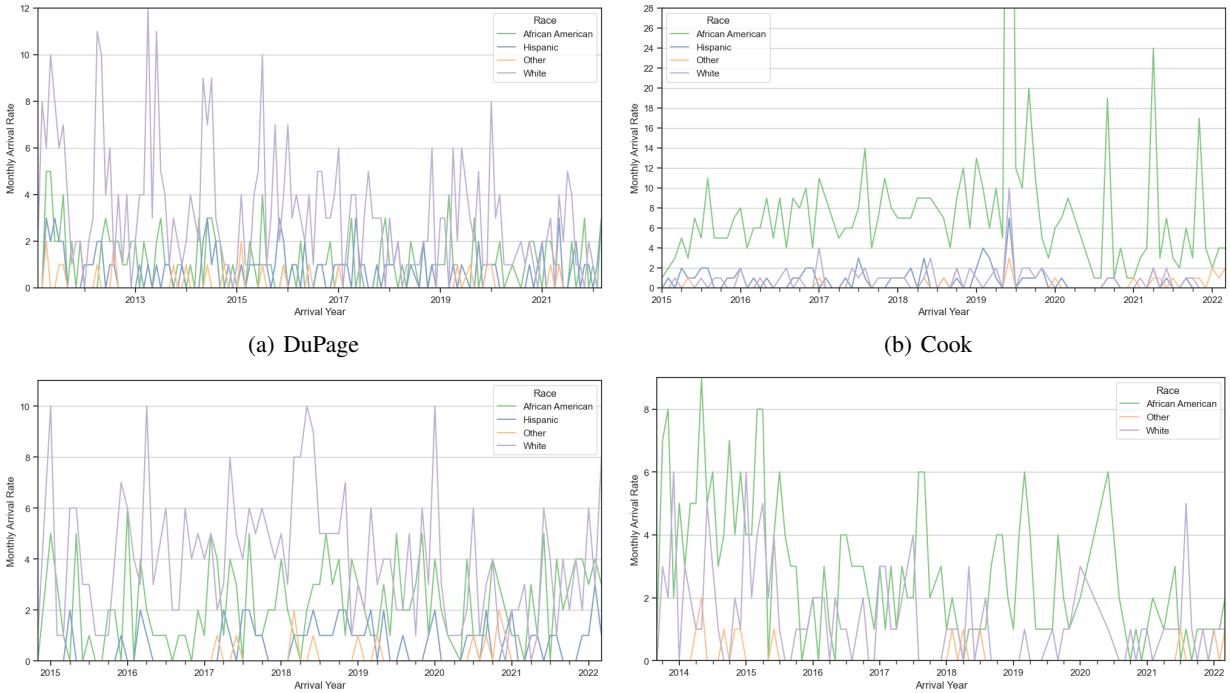
Figure 2: Length of Stay Distributions (Complete Data, Windsorized)



(a) DuPage (b) Cook (c) Will

(d) Peoria

Figure 3: Length of Stay per Year (Complete Data, Windsorized)



(a) DuPage (b) Cook

(c) Will (d) Peoria

Figure 4: Monthly Arrival Rate by Race

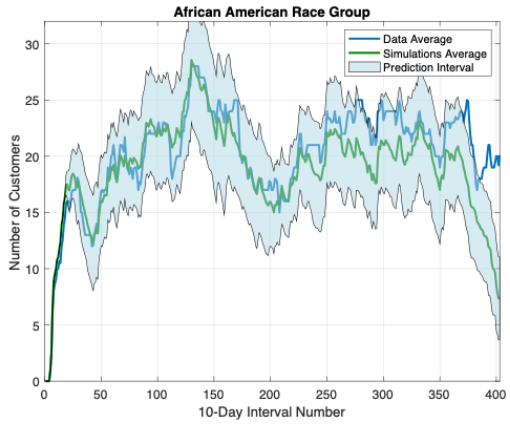


Figure 5: Simulation Calibration *without* solving the censoring issue: DuPage County