Classification Results

		LR	RF	XGB
adult	**Categorical Split Encoding**	75.7876	77.2816	75.0394
	Frequency Encoding	72.8844	78.2556	75.0896
	GLM Encoding	75.6678	78.417	75.3024
	Integer Encoding	66.8544	71.0506	69.9004
	Leaf Encoding	75.3048	77.1084	75.2244
	OHE	76.3114	78.1406	75.4004
	Target Encoding	75.6398	78.5878	74.637
amazon_employee_access	**Categorical Split Encoding**	62.665	65.4658	56.8594
	Frequency Encoding	50	63.8804	51.0432
	GLM Encoding	55.5932	63.6458	59.718
	Integer Encoding	50	63.2416	51.4284
	Leaf Encoding	60.585	60.212	59.7148
	OHE			
	Target Encoding	61.6614	62.0066	60.7562
churn	**Categorical Split Encoding**	59.7844	85.5696	84.0796
	Frequency Encoding	58.7066	87.3506	84.3664
	GLM Encoding	58.7856	87.2356	84.3004
	Integer Encoding	58.0268	86.2814	83.0782
	Leaf Encoding	59.938	87.0632	83.9848
	OHE	61.0804	83.3652	81.349
	Target Encoding	60.4074	87.0278	84.3976
click_prediction_small	**Categorical Split Encoding**	50.777	50.852	50.9246
	Frequency Encoding	50.328	53.6684	51.9058
	GLM Encoding	52.312	54.3484	53.57
	Integer Encoding	50.3156	52.8	51.929
	Leaf Encoding	50.323	50.1434	51.6358
	OHE			
	Target Encoding	53.1274	55.707	53.9782
eucalyptus	**Categorical Split Encoding**	88.7426	89.0624	88.2952
	Frequency Encoding	86.6796	89.3372	88.8494
	GLM Encoding	90.5234	89.5718	88.1198
	Integer Encoding	84.1916	89.9348	88.5352
	Leaf Encoding	86.3172	90.044	88.0694
	OHE	89.2364	88.686	88.8358
	Target Encoding	90.2708	89.7484	88.33
hpc_job_scheduling	**Categorical Split Encoding**	84.1206	82.8062	87.3852
	Frequency Encoding	76.1892	90.5872	83.1726
	GLM Encoding	83.2148	90.4398	86.9328
	Integer Encoding	75.936	90.9826	83.4086
	Leaf Encoding	76.0164	83.9058	81.5136
	OHE	84.454	88.9904	85.9996
	Target Encoding	82.6828	91.1302	87.2656
kdd_internet_usage	**Categorical Split Encoding**	85.4754	85.2066	84.7562
	Frequency Encoding	78.8512	85.1718	85.78
	GLM Encoding	83.2014	85.3326	85.4676
	Integer Encoding	80.4084	84.8994	85.181
	Leaf Encoding	73.8374	77.8546	73.7662
	OHE	81.5186	84.7984	85.9678
	Target Encoding	84.9166	85.2332	85.1912
kick	**Categorical Split Encoding**	50.9088	51.247	50.8568
	Frequency Encoding	50.0458	61.4496	61.2354
	GLM Encoding	50.069	61.1788	61.1258
	Integer Encoding	52.7592	61.8188	61.387
	Leaf Encoding	52.1912	50.7122	50.4344
	OHE			
	Target Encoding	61.777	62.05	61.575
open_payments	**Categorical Split Encoding**	62.83846154	70.93076923	68.24769231
	Frequency Encoding	50.01153846	54.85923077	54.77
	GLM Encoding	59.9384	62.0744	61.863
	Integer Encoding	50.05538462	62.52	61.09076923
	Leaf Encoding	57.88	59.18307692	58.76769231
	OHE			
	Target Encoding	68.60461538	72.13461538	72.76
road_safety_drivers_sex	**Categorical Split Encoding**	59.319	59.4	59.4632
	Frequency Encoding	50.0142	59.6208	59.3658
	GLM Encoding	59.2892	60.3954	60.2222
	Integer Encoding	50	56.8586	57.6362
	Leaf Encoding	50.9484	57.7102	58.21
	OHE			
	Target Encoding	59.9698	59.8932	59.813

Table 1: roc_auc_score mean results

	CSE	OHE	Integer	Freq	Leaf	Target	GLM
adult	7.241973	0.413774	0.000219	0.074521	1.235876	0.032395	215.5779
amazon_employee_access	21.87355		0.030021	0.110466	3.371026	0.055553	241.9505
churn	0.633244	0.04133	0.001277	0.017413	0.126782	0.005665	12.72808
click_prediction_small	25.76951		0.193475	0.508508	96.05683	0.29463	428.8492
eucalyptus	49.6976	0.008617	0.000778	0.002933	0.049808	0.037042	15.53983
hpc_job_scheduling	0.545946	0.006562	0.000459	0.002513	0.061058	0.033899	20.48228
kdd_internet_usage	4.876017	0.384013	0.024575	0.139149	1.676657	0.09818	305.8383
kick	21.31131		0.032055	0.156882	4.271318	0.074401	367.9059
open_payments	16.94361		0.021098	0.069052	2.150941	0.037592	153.0357
road_safety_drivers_sex	8.677993		7.98E-05	0.126583	14.16809	0.065366	104.8335

Table 2: Time Information (sec)

	CSE	OHE	Integer	Freq	Leaf	Target	GLM
adult	26.26	105	0	0	7	0	0
amazon_employee_access	201.4		0	0	1	0	0
churn	29	82	0	0	15	0	0
click_prediction_small	88.68		0	0	6	0	0
eucalyptus	64.34	91	0	0	15	20	0
hpc_job_scheduling	37.42	26	0	0	6	6	0
kdd_internet_usage	143.7	575	0	0	1	0	0
kick	98.52		0	0	15	0	0
open_payments	96.15385		0	0	1	0	0
road_safety_drivers_sex	18.24		0	0	3	0	0

Table 3: Number of columns generated

Regression Results

		LR	RF	XGB
ames_housing	**Categorical Split Encoding**	10.367	9.7226	9.6666
	Frequency Encoding	12.8268	9.7656	9.3334
	GLM Encoding	10.21143	9.235714	8.912857
	Integer Encoding	12.8186	9.7456	9.4088
	Leaf Encoding	17.2384	9.9884	10.6962
	OHE	9.6816	10.0098	9.2372
	Target Encoding	10.6484	9.3042	10.6962
avocado_sales	**Categorical Split Encoding**	14.6716	8.7218	8.9676
	Frequency Encoding	18.39	8.048	9.582
	GLM Encoding	14.6266	7.9206	9.4982
	Integer Encoding	18.0458	7.9038	9.2228
	Leaf Encoding	18.6394	10.859	13.1544
	OHE	14.6142	9.1422	9.286
	Target Encoding	14.6312	7.6074	13.1544
employee_salaries	**Categorical Split Encoding**	11.3848	9.3124	6.9938
	Frequency Encoding	24.1622	12.274	9.325
	GLM Encoding	11.349	6.8702	7.152
	Integer Encoding	24.2862	11.9354	8.6874
	Leaf Encoding	21.5322	10.9464	10.6812
	OHE	8.9626	11.986	10.6236
	Target Encoding	11.367	6.72	10.6812
medical_charges	**Categorical Split Encoding**	15.2776	10.2946	12.955
	Frequency Encoding	67.5216	35.2564	18.5574
	GLM Encoding	16.0376	10.8758	10.7184
	Integer Encoding	70.2962	37.7726	13.9874
	Leaf Encoding	77.9146	55.3988	58.2548
	OHE			
	Target Encoding	16.6842	8.894	58.2548

Table 4: MAPE scores

		LR	RF	XGB
ames_housing	**Categorical Split Encoding**	9.47	8.7794	8.787
	Frequency Encoding	10.9808	8.6136	8.5622
	Integer Encoding	11.3504	8.638	8.7448
	Leaf Encoding	15.135	8.9164	9.87
	Target Encoding	9.756	8.323	9.87
	OHE	8.5882	8.8878	8.5364
avocado_sales	**Categorical Split Encoding**	13.7862	8.3608	8.69
	Frequency Encoding	17.0888	7.7764	9.2586
	Integer Encoding	16.7862	7.6428	8.926
	Leaf Encoding	17.257	10.3992	12.5432
	Target Encoding	13.756	7.36	12.5432
	OHE	13.7426	8.7164	8.9766
employee_salaries	**Categorical Split Encoding**	10.1152	8.631	6.9506
	Frequency Encoding	21.671	11.8814	9.0316
	Integer Encoding	22.0816	11.6874	8.5214
	Leaf Encoding	19.8512	10.0532	10.3754
	Target Encoding	10.0228	6.6886	10.3754
	OHE	8.7492	10.5966	9.8338
medical_charges	**Categorical Split Encoding**	14.2128	10.8624	16.864
	Frequency Encoding	52.1756	28.9644	17.198
	Integer Encoding	54.2452	30.4676	13.9552
	Leaf Encoding	59.0788	39.2856	41.5734
	Target Encoding	14.9956	9.5808	41.5734
	OHE			

Table 5: Weighted MAPE

	CSE	OHE	Integer	Freq	Leaf	Target	GLM
ames_housing	1.464756	0.330576	0.008078	0.045479	0.599917	0.051563	11.92808
avocado_sales	2.575029	0.422287	0.003171	0.019508	0.184444	0.006582	2.391262
employee_salaries	4.026082	2.909022	0.002392	0.012925	0.350085	0.008398	2.5291
medical_charges	30.93197		0.014441	0.064269	7.849549	0.036902	12.94454

Table 6: Time Information (sec)

	CSE	OHE	Integer	Freq	Leaf	Target	GLM
ames_housing	86.1	353	0	0	35	0	0
avocado_sales	40.22	68	0	0	9	0	0
employee_salaries	162.88	1121	0	0	2	0	0
medical_charges	208.78		0	0	2	0	0

Table 7: Time Information (sec)

Random Missing Values Generation in Categorical Columns

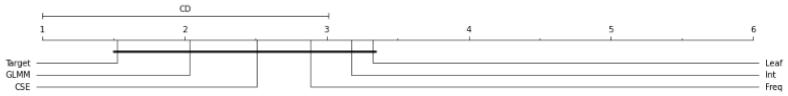
0.1					0.25				0.4			
		LR	RF	XGB		LR	RF	XGB		LR	RF	XGB
amazon_employee_access	Categorical Split Encoding	60.2025	61.7525	57.0225	Categorical Split Encoding	59.8775	60.1275	57.6025	Categorical Split Encoding	57.8225	58.7825	55.3675
	Leaf Encoding	55.795	59.94	59.2925	Leaf Encoding	57.84	57.875	53.9775	Leaf Encoding	51.2725	52.69	51.0175
	Target Encoding	59.6525	59.5675	59.43	Target Encoding	59.045	59.2725	58.415	Target Encoding	57.8275	58.3725	57.315
churn	Categorical Split Encoding	58.655	80.32	80.59	Categorical Split Encoding	59.005	76.8175	76.7475	Categorical Split Encoding	58.775	76.37	77.1975
	Leaf Encoding	58.1825	83.2375	80.745	Leaf Encoding	59.15	82.695	80.8675	Leaf Encoding	55.8075	75.625	72.5825
	Target Encoding	59.36	82.05	81.1075	Target Encoding	60.155	81.35	80.135	Target Encoding	58.97	77.93	77.51
click_prediction_small	Categorical Split Encoding	51.055	50.6975	50.1975	Categorical Split Encoding	51.265	50.4375	50.2875	Categorical Split Encoding	50.705	50.2875	50.0975
	Leaf Encoding	50.32	50.1775	51.405	Leaf Encoding	50.32	50.1775	51.405	Leaf Encoding	50.32	50.175	51.405
	Target Encoding	53.37	54.475	52.845	Target Encoding	53.46	54.4	52.66	Target Encoding	53.695	54.3875	52.745
eucalyptus	Categorical Split Encoding	86.6275	88.815	88.045	Categorical Split Encoding	87.085	87.77	86.7375	Categorical Split Encoding	86.56	87.9725	86.695
	Leaf Encoding	87.0925	88.9875	86.6225	Leaf Encoding	85.85	89.1075	87.37	Leaf Encoding	84.765	89.08	87.3
	Target Encoding	87.515	88.3025	86.9375	Target Encoding	86.0075	87.83	87.135	Target Encoding	85.9125	87.5275	86.89
hpc_job_scheduling	Categorical Split Encoding	82.045	81.7025	85.2275	Categorical Split Encoding	80.3825	79.6475	83.7625	Categorical Split Encoding	79.615	80.725	82.435
	Leaf Encoding	76.6075	77.8225	73.4275	Leaf Encoding	75.9925	84.6275	82.2525	Leaf Encoding	72.615	78.8475	75.9675
	Target Encoding	81.4675	87.325	85.75	Target Encoding	80.2975	85.7025	84.19	Target Encoding	79.7275	84.4025	82.745
kdd_internet_usage	Categorical Split Encoding	83.5325	83.6375	82.715	Categorical Split Encoding	80.3175	80.755	79.165	Categorical Split Encoding	76.47	77.395	75.365
	Leaf Encoding	72.6725	76.9725	77.1275	Leaf Encoding	68.6875	66.3175	64.33	Leaf Encoding	64.1575	63.905	59.15
	Target Encoding	83.575	83.695	83.0525	Target Encoding	81.1575	81.0525	79.9675	Target Encoding	78.105	77.095	74.9825
kick	Categorical Split Encoding	50.5375	50.9225	50.805	Categorical Split Encoding	50.86	50.8275	51.0775	Categorical Split Encoding	50.805	50.605	50.87
	Leaf Encoding	52.835	50.9525	50.5875	Leaf Encoding	53.14	50.5675	50.7975	Leaf Encoding	50.09	50.295	50.38
	Target Encoding	55.6375	59.7575	60.47	Target Encoding	52.6775	54.25	55.59	Target Encoding	51.4975	51.0125	50.6125
open_payments	Categorical Split Encoding	61.17	66.155	62.3325	Categorical Split Encoding	60.9275	62.73	59.72	Categorical Split Encoding	53.6925	58.4225	56.875
	Leaf Encoding	53.935	54.2	54.335	Leaf Encoding	54.91	57.585	58.535	Leaf Encoding	54.285	52.64	52.9175
	Target Encoding	62.715	63.245	62.1	Target Encoding	60.06	60.07	59.18	Target Encoding	57.9475	57.9375	57.3025
road_safety_drivers_sex	Categorical Split Encoding	58.625	59.07	58.615	Categorical Split Encoding	57.3225	57.79	57.13	Categorical Split Encoding	56.3025	56.81	56.13
	Leaf Encoding	53.095	56.035	56.785	Leaf Encoding	53.745	58.255	59.4025	Leaf Encoding	51.1925	55	58.155
	Target Encoding	58.7	59.4225	57.955	Target Encoding	57.465	59.555	57.54	Target Encoding	56.5575	59.2175	55.975

Table 8: Classification - roc-auc-scores at various null rates

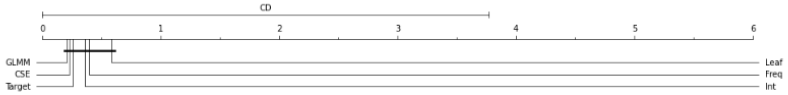
0.1					0.25				0.4			
		LR	**RF**	**XGB**		**LR**	**RF**	**XGB**		**LR**	**RF**	**XGB**
ames_housing	**Categorical Split Encoding**	0.1	0.1	1	**Categorical Split Encoding**	0.11	0.1	1	**Categorical Split Encoding**	0.12	0.1	1
	Leaf Encoding	0.165	0.095	0.5	**Leaf Encoding**	0.1625	0.1	0.4975	**Leaf Encoding**	0.1725	0.1	0.5
	Target Encoding	0.1075	0.09	1	**Target Encoding**	0.11	0.0975	1	**Target Encoding**	0.1175	0.1	1
avocado_sales	**Categorical Split Encoding**	0.28	0.26	0.33	**Categorical Split Encoding**	0.2825	0.25	0.33	**Categorical Split Encoding**	0.29	0.245	0.33
	Leaf Encoding	0.2725	0.2275	0.4225	**Leaf Encoding**	0.2875	0.2325	0.445	**Leaf Encoding**	0.29	0.2275	0.43
	Target Encoding	0.27	0.235	0.33	**Target Encoding**	0.275	0.23	0.33	**Target Encoding**	0.28	0.2325	0.33
employee_salaries	**Categorical Split Encoding**	0.18	0.1275	1	**Categorical Split Encoding**	0.2225	0.1475	1	**Categorical Split Encoding**	0.2475	0.17	1
	Leaf Encoding	0.185	0.165	0.51	**Leaf Encoding**	0.195	0.2075	0.49	**Leaf Encoding**	0.2175	0.2125	0.51
	Target Encoding	0.1325	0.11	1	**Target Encoding**	0.16	0.13	1	**Target Encoding**	0.1725	0.15	1
medical_charges	**Categorical Split Encoding**	0.31	0.37	1	**Categorical Split Encoding**	0.94	0.6125	1	**Categorical Split Encoding**	1.17	0.7125	1
	Leaf Encoding	0.56	0.655	0.5825	**Leaf Encoding**	0.675	0.87	0.725	**Leaf Encoding**	1.1925	0.525	0.5425
	Target Encoding	0.29	0.2975	1	**Target Encoding**	0.3775	0.3575	1	**Target Encoding**	0.4325	0.4125	1

Table 9: Regression - prmse scores at various missing rates

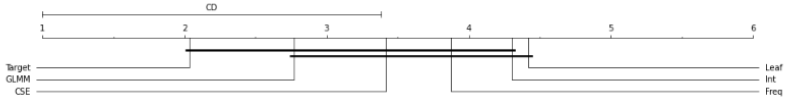
OVERALL



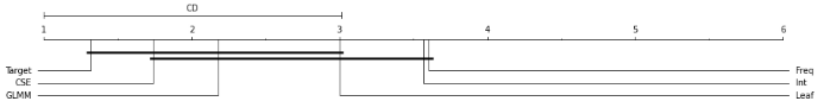
Regression Results



Classification Results



LR Results



RF Results

XGB Results