

## VIDEO DEMO

### DESAFIO 2

#### INCONSISTÊNCIAS DE DADOS NOS SISTEMAS DOS TRIBUNAIS

Como podemos identificar e corrigir na base do Datajud as inconsistências nos metadados dos processos em tramitação nos sistemas dos tribunais?

#saneamento

## Desafio 2 - GRUPO 11

### JAMILA FONSECA

Advogada 4.0 – Dir. Tributário, com experiência em Tecnologia, Ciência de Dados, Inteligência Artificial e Robotização.



ADV

### LUIS ASSUNÇÃO

Arquiteto de Soluções e Cientista de Dados, Especialista em ferramentas de Gestão, Analytics, AI, DW e Big Data. Desenvolvedor Python.



DEV

### LUCIANA WEILER

Analista de sistemas com MBA em Governança de TI. Experiência como Gestora de TI e em Desenvolvimento de Sistemas.



UX

### PROBLEMA

- Diversidade de Sistemas
- Inconsistência Informação
- Falta de Uniformidade
- Lacunas de Informação
- Falta de Visibilidade

### REQUISITOS

- Gestão de Metadados
- Identificar Inconsistências
- Corrigir e Sanear Dados
- Interoperabilidade
- Integração com DATAJUD

### SOLUÇÃO

- Ciência de Dados
- Algoritmo Inconsistências
- Algoritmo de Saneamento
- Machine Learning
- Painel de Gestão



	TRIBUNAL	(All)	Campos Válidos, sem nenhum erro ou inconsistência	Campos sem nenhum preenchimento (vazios)	Campos preenchidos com ZEROS (Ex: "0", "0.0", "00000" ...)	Campos de DATA inválida; ou Num Processo com Digito Verificador Errado, ou valores fora do domínio.	Consistência verificada através do cruzamento com as Tabelas: Classe, Assunto e Serventia.
	TRIBUNAL_JUSTICA	(All)					
	MUNICIPIO	(All)					
	UF	(All)					
	NOME_ORGAO	(All)					
Row Labels	TOT PROCESSOS	QDE VALIDOS	% VALIDOS	QDE VAZIOS	QDE com ZEROS	QDE ERROS	QDE INCONSISTENCIA
11-01-DATA_AUJIZAMENTO	2.994.290	2.589.652	86,49%	0	0	404.638	0
12-02-NUM_PROCESSO	2.994.290	2.976.168	99,39%	0	0	18.122	0
13-03-GRAU_JURISDICAO	2.994.290	2.994.290	100,00%	0	0	0	0
14-04-SIGLA_TRIBUNAL	2.994.290	2.994.290	100,00%	0	0	0	0
15-05-ORGAO_NOME	2.994.290	2.981.082	99,56%	13.208	0	0	0
16-06-ORGAO_CODIGO	2.994.290	2.954.440	98,67%	129	3.588	39.850	0
17-07-ORGAO_MUN_IRGE	2.994.290	2.568.765	85,79%	0	173.882	251.643	0
18-08-COD_LOCALIDADE	2.994.290	1.998.561	66,75%	577	40.317	954.835	0
19-09-CLASSE_PROCESSUAL	2.994.290	2.736.788	91,40%	0	0	0	257.502
20-10-TIPO_PROCESSO	2.994.290	1.915.035	63,96%	1.045.208	30.849	3.198	0
21-11-SIST_ELETRONICO	2.994.290	1.819.976	60,78%	1.174.314	0	0	0
22-12-VALOR_CAUSA	2.994.290	1.145.692	38,26%	1.464.270	326.557	57.771	0
23-13-ASSUNTO01_COD_LOCAL	2.994.290	165.070	5,51%	2.822.945	6.275	0	0
24-14-ASSUNTO01_COD_PAI	2.994.290	163.389	5,46%	2.822.945	7.956	0	0
25-15-ASSUNTO01_DESCRICAO	2.994.290	136.843	4,57%	2.857.447	0	0	0
26-16-ASSUNTO01_COD_NACIONAL	2.994.290	2.394.287	79,96%	394.094	45	0	205.864
27-17-MOV_026_DISTRIBUICAO	2.994.290	1.973.964	65,92%	1.020.326	0	0	0
28-18-MOV_193_SENTENCA	2.994.290	1.203.205	40,18%	1.791.085	0	0	0
29-19-MOV_848_TRANSITO_JULGADO	2.994.290	486.534	16,25%	2.507.756	0	0	0
30-20-MOV_022_BAIXA	2.994.290	404.436	13,51%	2.589.854	0	0	0
31-21-MOV_246_ARQUIVADO	2.994.290	31.374	1,05%	2.044.906	0	0	918.010
32-22-MOV_NACIONAL_TODOS	2.994.290	2.882.949	96,28%	111.341	0	0	0



Campos Válidos,  
sem nenhum erro ou  
inconsistência

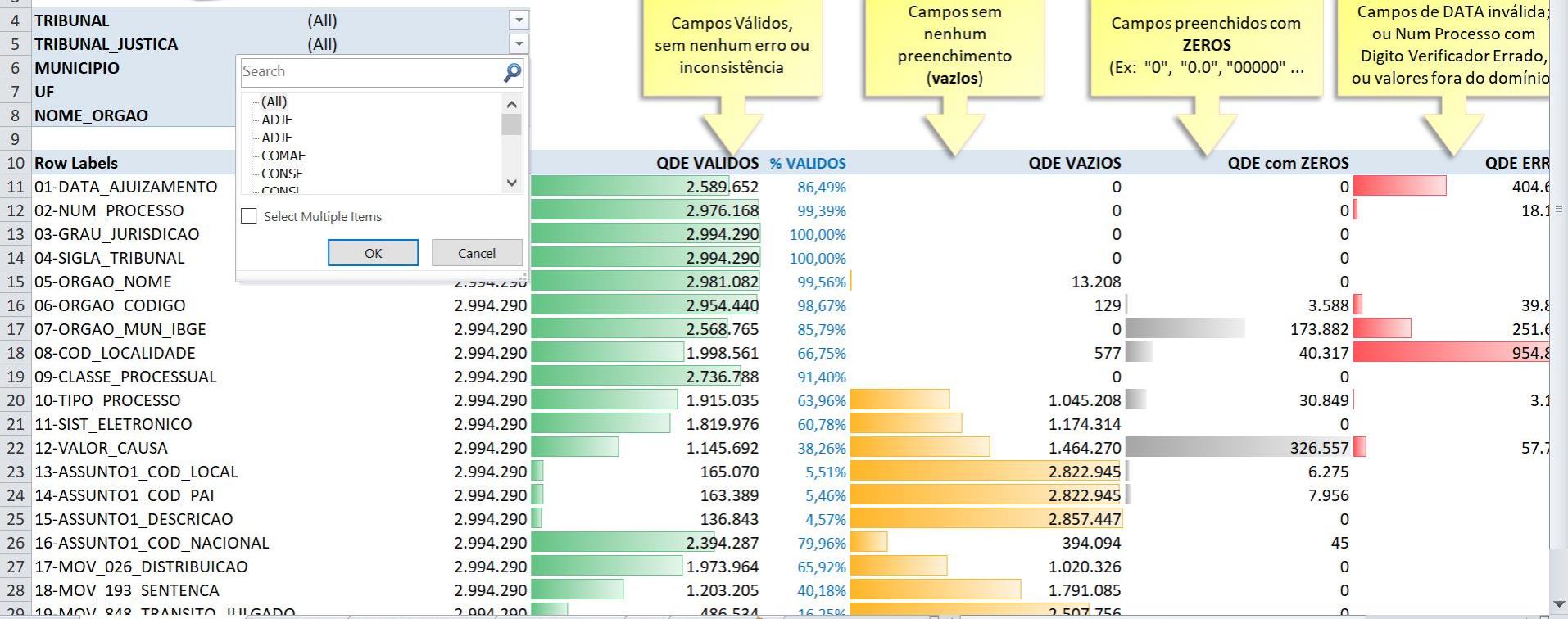
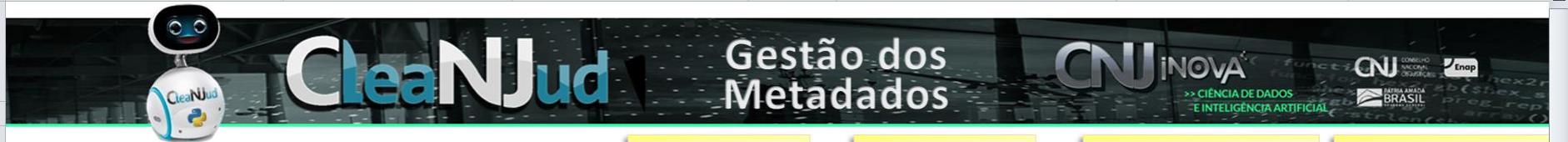
Campos sem  
nenhum  
preenchimento  
(vazios)

Campos preenchidos com  
**ZEROS**  
(Ex: "0", "0.0", "00000" ...)

Campos de DATA inválida;  
ou Num Processo com  
Digito Verificador Errado,  
ou valores fora do domínio.

Consisténcia verificada  
através do cruzamento com  
as Tabelas: Classe, Assunto  
e Serventia.

TOT PROCESSOS	QDE VALIDOS	% VALIDOS	QDE VAZIOS	QDE com ZEROS	QDE ERROS	QDE INCONSISTENCIA
2.994.290	2.589.652	86,49%	0	0	404.638	0
2.994.290	2.976.168	99,39%	0	0	18.122	0
2.994.290	2.994.290	100,00%	0	0	0	0
2.994.290	2.994.290	100,00%	0	0	0	0
2.994.290	2.981.082	99,56%	13.208	0	0	0
2.994.290	2.954.440	98,67%	129	3.588	39.850	0
2.994.290	2.568.765	85,79%	0	173.882	251.643	0
2.994.290	1.998.561	66,75%	577	40.317	954.835	0
2.994.290	2.736.788	91,40%	0	0	0	257.502
2.994.290	1.915.035	63,96%	1.045.208	30.849	3.198	0
2.994.290	1.819.976	60,78%	1.174.314	0	0	0
2.994.290	1.145.692	38,26%	1.464.270	326.557	57.771	0
2.994.290	165.070	5,51%	2.822.945	6.275	0	0
2.994.290	163.389	5,46%	2.822.945	7.956	0	0
2.994.290	136.843	4,57%	2.857.447	0	0	0
2.994.290	2.394.287	79,96%	394.094	45	0	205.864
2.994.290	1.973.964	65,92%	1.020.326	0	0	0
2.994.290	1.203.205	40,18%	1.791.085	0	0	0
2.994.290	496.524	16,75%	2.507.756	0	0	0





## RANKING POR ESTADO

RANKING DA QUANTIDADE TOTAL DE INCONSISTÊNCIAS (SOMA DOS INDICADORES: VAZIOS + ZERADOS + ERROS + INCONSISTÊNCIAS)

TRIBUNAL\_JUSTICA

(All)

OBJETIVO / AÇÃO: Identificar os Principais Ofensores, para ações de correção na ORIGEM do problema

Sum of TOTAL\_INCONSISTENCIAS Column Labels

		01-DATA_AUJIZAMENTO	02-NUM_PROCESSO	06-ORGÃO_CODIGO	07-ORGÃO_MUN_BGE	08-COD_LOCALIDADE	09-CLASSE PROCESSUAL	12-VALOR_CUSA	13-ASSUNTO1_COD_LOCAL	21-MOV_246_ARQUIVADO	Grand Total
Row Labels											
10	SP	39.615	59	0	5.439	58.946	20.778	127.666	238.653	241.217	732.373
11	SC	25.035	7.119	0	42.697	56.952	8.380	116.661	164.533	163.552	584.929
12	PR	28.822	567	0	3.144	85.890	36.919	96.966	125.620	162.682	540.610
13	BA	10.943	67	0	25.593	61.671	6.041	123.700	118.940	122.263	469.218
14	MG	15.548	1	0	35.318	61.840	15.108	63.983	132.071	142.815	466.684
15	RS	19.985	7.682	0	290	36.105	13.592	95.572	133.161	139.603	445.990
16	RN	22.401	1.080	0	63.754	73.976	5.538	44.158	107.722	113.940	432.569
17	GO	15.367	85	0	55.865	59.577	7.569	66.178	107.071	108.516	420.228
18	DF	21.116	2	0	12.631	78.540	5.559	60.726	108.190	115.482	402.246
19	RJ	1.039	0	0	1.087	52.307	6.143	98.240	112.443	113.118	384.377
20	MS	11.902	197	0	59.842	62.813	9.498	34.757	99.157	104.846	383.012
21	PA	3.046	292	0	3.032	59.241	6.353	84.997	109.130	112.344	378.435
22	PE	4.032	1	0	5.539	14.316	5.018	97.086	119.625	122.641	368.258
23	PB	11.707	2	0	27.483	33.285	5.590	63.440	98.364	98.521	338.392
24	AM	12.620	6	0	1.418	31.617	5.006	95.169	93.706	94.998	334.540
25	CE	3.757	0	0	4.579	11.602	4.685	43.968	119.848	121.557	309.996
26	MA	22.494	8	0	2.441	50.358	5.718	43.088	88.694	91.288	304.089
27	SE	297	0	0	1.640	2.817	2.859	69.398	93.702	89.325	260.038
28	(blank)	18.658	372	0	0	0	58.831	59.311	57.033	59.292	253.497
29	TO	312	0	0	50.422	50.499	1.736	20.467	61.239	61.179	245.854
30	RO	27.574	7	0	5.417	22.426	3.303	22.361	79.687	83.730	244.505
31		11.000	0	0	1.100	1.000	1.000	1.000	22.000	22.000	22.000



1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387



#### RANKING POR NOME DO ORGÃO

RANKING DA QUANTIDADE TOTAL DE INCONSISTÊNCIAS (SOMA DOS INDICADORES: VAZIOS + ZERADOS + ERROS +)

UF	(All)	OBJETIVO / AÇÃO: Identificar os Principais Ofensores, para ações de correção na ORIGEM do problema									
Sum of TOTAL_INCONSISTENCIAS	Column Labels	01-DATA_AJUAMENTO	02-NUM_PROCESSO	06-ORGAO_CODIGO	07-ORGAO_MUN_BGE	08-COD_LOCALIDADE	09-CLASSE_PROCESUAL	12-VALOR_CAUSA	13-ASSUNTO1_COD_LOCAL	21-MOV_246_ARQUIVADO	
Row Labels											
VARAE		136.965	7.908	0	263.176	573.838	38.436	808.432	1.246.575	1.231.853	
GRA1T		132.937	19	0	0	46.081	107.419	332.881	751.110	815.264	
VARAF		29.160	7.037	0	89.660	234.754	13.607	216.549	222.050	265.065	
ZONA		19.327	1.930	0	4.839	8.637	3.012	130.731	130.050	120.623	
VARAM		10.507	839	0	25.805	54.748	21.288	33.896	109.003	109.974	
(INATIVO) 2º JUIZADO ESPECIAL CRIMINAL CE		0	0	0	1	1	0	1	1	1	
(INATIVO) JUIZADO ESPECIAL CRIMINAL DO D		0	0	0	1	1	0	1	1	1	
º JUIZADO ESPECIAL CÍVEL E CRIMINAL DE SAN		0	0	0	0	83	3	44	83	83	
1º Juizado Especial		0	0	0	209	209	0	9	209	209	
ª AUDITORIA DA JUSTIÇA MILITAR DE MG		150	1	0	151	151	122	2.543	3.272	3.272	
ª AUDITORIA MILITAR		0	0	0	460	460	12	753	1.183	1.183	
ª AUDITORIA MILITAR DE PASSO FUNDO		240	1	0	0	296	4	552	552	552	
ª AUDITORIA MILITAR DE PORTO ALEGRE		780	705	0	0	1.097	22	1.937	1.951	1.951	
ª AUDITORIA MILITAR DE SANTA MARIA		492	2	0	0	376	13	906	908	908	
ª CÂMARA		3	0	0	3	3	252	252	252	252	
ª JUIZADO ESPECIAL DE FAZENDA PÚBLICA		0	0	0	0	0	724	3	2.655	2.655	
ª RELATORIA DA 1º TURMA RECURSAL DOS JE		14	0	0	0	15	0	15	15	15	
ª RELATORIA DA 1º TURMA RECURSAL DOS JE		109	0	0	0	260	0	260	260	260	
ª RELATORIA DA 2º TURMA RECURSAL DOS JE		27	0	0	0	45	0	45	45	45	
ª RELATORIA DA 2º TURMA RECURSAL DOS JE		181	0	0	0	452	0	451	452	452	
ª RELATORIA DA 3º TURMA RECURSAL DOS JE		4	0	0	0	19	0	19	19	19	
ª RELATORIA DA 3º TURMA RECURSAL DOS JE		0	0	0	0	0	105	105	105	105	

	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5								
6								
7								
8	NOME CAMPO CLEANJUD	cod_atribut	JSON_nivel_1	JSON_nivel_2	JSON_nivel_3	JSON_nivel_4	descricao_campo	
9								
10			1 millisInsercao					
11	03-GRAU_JURISDICO	2 grau						
12	04-SIGLA_TRIBUNAL	3 siglaTribunal						
13		4 dadosBasicos	assunto	principal				
14	16-ASSUNTO1_COD_NACIONAL	5 dadosBasicos	assunto	codigoNacional				
15	13-ASSUNTO1_COD_LOCAL	6 dadosBasicos	assunto	assuntoLocal	codigoAssunto			
16	14-ASSUNTO1_COD_PAI	7 dadosBasicos	assunto	assuntoLocal	codigoPaiNacional			
17	15-ASSUNTO1_DESCRICAO	8 dadosBasicos	assunto	assuntoLocal	descricao			
18	05-ORGAO_NOME	9 dadosBasicos	orgaoJulgador	nomeOrgao				
19	07-ORGAO_MUN_IBGE	10 dadosBasicos	orgaoJulgador	codigoMunicipioIBGE				
20	06-ORGAO_CODIGO	11 dadosBasicos	orgaoJulgador	codigoOrgao				
21		12 dadosBasicos	orgaoJulgador	instancia				
22		13 dadosBasicos	outrosNumeros					
23	02-NUM_PROCESSO	14 dadosBasicos	numero					
24		15 dadosBasicos	competencia					
25	09-CLASSE_PROCESSUAL	16 dadosBasicos	classeProcessual					
26	08-COD_LOCALIDADE	17 dadosBasicos	codigoLocalidade					
27		18 dadosBasicos	nivelSigilo					
28		19 dadosBasicos	intervencaoMP					
29		20 dadosBasicos	tamanhoProcesso					
30	01-DATA_AUJIZAMENTO	21 dadosBasicos	dataAjuizamento					
31	10-TIPO_PROCESSO	22 dadosBasicos	procEl					
32	11-SIST_ELETRONICO	23 dadosBasicos	dscSistema					
33		24 dadosBasicos	relacaoIncidental					
34		25 dadosBasicos	prioridade					
35	12-VALOR_CAUSA	26 dadosBasicos	valorCausa					
36		27 dadosBasicos	complemento					
37		28 dadosBasicos	dataProcesso					
38		29 dadosBasicos	dataRecebimento					
39		30 dadosBasicos	dataProcessoFinal					
40		31 dadosBasicos	dataRecebimentoFinal					
41		32 dadosBasicos	dataProcessoFinal					
42		33 dadosBasicos	dataRecebimentoFinal					
43		34 dadosBasicos	dataProcessoFinal					
44		35 dadosBasicos	dataRecebimentoFinal					
45		36 dadosBasicos	dataProcessoFinal					
46		37 dadosBasicos	dataRecebimentoFinal					
47		38 dadosBasicos	dataProcessoFinal					
48		39 dadosBasicos	dataRecebimentoFinal					
49		40 dadosBasicos	dataProcessoFinal					
50		41 dadosBasicos	dataRecebimentoFinal					
51		42 dadosBasicos	dataProcessoFinal					
52		43 dadosBasicos	dataRecebimentoFinal					
53		44 dadosBasicos	dataProcessoFinal					
54		45 dadosBasicos	dataRecebimentoFinal					
55		46 dadosBasicos	dataProcessoFinal					
56		47 dadosBasicos	dataRecebimentoFinal					
57		48 dadosBasicos	dataProcessoFinal					
58		49 dadosBasicos	dataRecebimentoFinal					
59		50 dadosBasicos	dataProcessoFinal					
60		51 dadosBasicos	dataRecebimentoFinal					
61		52 dadosBasicos	dataProcessoFinal					
62		53 dadosBasicos	dataRecebimentoFinal					
63		54 dadosBasicos	dataProcessoFinal					
64		55 dadosBasicos	dataRecebimentoFinal					
65		56 dadosBasicos	dataProcessoFinal					
66		57 dadosBasicos	dataRecebimentoFinal					
67		58 dadosBasicos	dataProcessoFinal					
68		59 dadosBasicos	dataRecebimentoFinal					
69		60 dadosBasicos	dataProcessoFinal					
70		61 dadosBasicos	dataRecebimentoFinal					
71		62 dadosBasicos	dataProcessoFinal					
72		63 dadosBasicos	dataRecebimentoFinal					
73		64 dadosBasicos	dataProcessoFinal					
74		65 dadosBasicos	dataRecebimentoFinal					
75		66 dadosBasicos	dataProcessoFinal					
76		67 dadosBasicos	dataRecebimentoFinal					
77		68 dadosBasicos	dataProcessoFinal					
78		69 dadosBasicos	dataRecebimentoFinal					
79		70 dadosBasicos	dataProcessoFinal					
80		71 dadosBasicos	dataRecebimentoFinal					
81		72 dadosBasicos	dataProcessoFinal					
82		73 dadosBasicos	dataRecebimentoFinal					
83		74 dadosBasicos	dataProcessoFinal					
84		75 dadosBasicos	dataRecebimentoFinal					
85		76 dadosBasicos	dataProcessoFinal					
86		77 dadosBasicos	dataRecebimentoFinal					
87		78 dadosBasicos	dataProcessoFinal					
88		79 dadosBasicos	dataRecebimentoFinal					
89		80 dadosBasicos	dataProcessoFinal					
90		81 dadosBasicos	dataRecebimentoFinal					
91		82 dadosBasicos	dataProcessoFinal					
92		83 dadosBasicos	dataRecebimentoFinal					
93		84 dadosBasicos	dataProcessoFinal					
94		85 dadosBasicos	dataRecebimentoFinal					
95		86 dadosBasicos	dataProcessoFinal					
96		87 dadosBasicos	dataRecebimentoFinal					
97		88 dadosBasicos	dataProcessoFinal					
98		89 dadosBasicos	dataRecebimentoFinal					
99		90 dadosBasicos	dataProcessoFinal					
100		91 dadosBasicos	dataRecebimentoFinal					
101		92 dadosBasicos	dataProcessoFinal					
102		93 dadosBasicos	dataRecebimentoFinal					
103		94 dadosBasicos	dataProcessoFinal					
104		95 dadosBasicos	dataRecebimentoFinal					
105		96 dadosBasicos	dataProcessoFinal					
106		97 dadosBasicos	dataRecebimentoFinal					
107		98 dadosBasicos	dataProcessoFinal					
108		99 dadosBasicos	dataRecebimentoFinal					
109		100 dadosBasicos	dataProcessoFinal					
110		101 dadosBasicos	dataRecebimentoFinal					
111		102 dadosBasicos	dataProcessoFinal					
112		103 dadosBasicos	dataRecebimentoFinal					
113		104 dadosBasicos	dataProcessoFinal					
114		105 dadosBasicos	dataRecebimentoFinal					
115		106 dadosBasicos	dataProcessoFinal					
116		107 dadosBasicos	dataRecebimentoFinal					
117		108 dadosBasicos	dataProcessoFinal					
118		109 dadosBasicos	dataRecebimentoFinal					
119		110 dadosBasicos	dataProcessoFinal					
120		111 dadosBasicos	dataRecebimentoFinal					
121		112 dadosBasicos	dataProcessoFinal					
122		113 dadosBasicos	dataRecebimentoFinal					
123		114 dadosBasicos	dataProcessoFinal					
124		115 dadosBasicos	dataRecebimentoFinal					
125		116 dadosBasicos	dataProcessoFinal					
126		117 dadosBasicos	dataRecebimentoFinal					
127		118 dadosBasicos	dataProcessoFinal					
128		119 dadosBasicos	dataRecebimentoFinal					
129		120 dadosBasicos	dataProcessoFinal					
130		121 dadosBasicos	dataRecebimentoFinal					
131		122 dadosBasicos	dataProcessoFinal					
132		123 dadosBasicos	dataRecebimentoFinal					
133		124 dadosBasicos	dataProcessoFinal					
134		125 dadosBasicos	dataRecebimentoFinal					
135		126 dadosBasicos	dataProcessoFinal					
136		127 dadosBasicos	dataRecebimentoFinal					
137		128 dadosBasicos	dataProcessoFinal					
138		129 dadosBasicos	dataRecebimentoFinal					
139		130 dadosBasicos	dataProcessoFinal					
140		131 dadosBasicos	dataRecebimentoFinal					
141		132 dadosBasicos	dataProcessoFinal					
142		133 dadosBasicos	dataRecebimentoFinal					
143		134 dadosBasicos	dataProcessoFinal					
144		135 dadosBasicos	dataRecebimentoFinal					
145		136 dadosBasicos	dataProcessoFinal					
146		137 dadosBasicos	dataRecebimentoFinal					
147		138 dadosBasicos	dataProcessoFinal					
148		139 dadosBasicos	dataRecebimentoFinal					
149		140 dadosBasicos	dataProcessoFinal					
150		141 dadosBasicos	dataRecebimentoFinal					
151		142 dadosBasicos	dataProcessoFinal					
152		143 dadosBasicos	dataRecebimentoFinal					
153		144 dadosBasicos	dataProcessoFinal					
154		145 dadosBasicos	dataRecebimentoFinal					
155		146 dadosBasicos	dataProcessoFinal					
156		147 dadosBasicos	dataRecebimentoFinal					
157		148 dadosBasicos	dataProcessoFinal					
158		149 dadosBasicos	dataRecebimentoFinal					
159		150 dadosBasicos	dataProcessoFinal					
160		151 dadosBasicos	dataRecebimentoFinal					
161		152 dadosBasicos	dataProcessoFinal					
162		153 dadosBasicos	dataRecebimentoFinal					
163		154 dadosBasicos	dataProcessoFinal					
164		155 dadosBasicos	dataRecebimentoFinal					
165		156 dadosBasicos	dataProcessoFinal					
166		157 dadosBasicos	dataRecebimentoFinal					
167		158 dadosBasicos	dataProcessoFinal					
168		159 dadosBasicos	dataRecebimentoFinal					
169		160 dadosBasicos	dataProcessoFinal					
170		161 dadosBasicos	dataRecebimentoFinal					
171		162 dadosBasicos	dataProcessoFinal					
172		163 dadosBasicos	dataRecebimentoFinal					
173		164 dadosBasicos	dataProcessoFinal					
174		165 dadosBasicos	dataRecebimentoFinal					
175		166 dadosBasicos	dataProcessoFinal					
176		167 dadosBasicos	dataRecebimentoFinal					
177		168 dadosBasicos	dataProcessoFinal					
178		169 dadosBasicos	dataRecebimentoFinal					
179		170 dadosBasicos	dataProcessoFinal					
180		171 dadosBasicos	dataRecebimentoFinal					
181		172 dadosBasicos	dataProcessoFinal					
182		173 dadosBasicos	dataRecebimentoFinal					
183		174 dadosBasicos	dataProcessoFinal					
184		175 dadosBasicos	dataRecebimentoFinal					
185		176 dadosBasicos	dataProcessoFinal					
186		177 dadosBasicos	dataRecebimentoFinal					
187		178 dadosBasicos	dataProcessoFinal					
188		179 dadosBasicos	dataRecebimentoFinal					
189		180 dadosBasicos	dataProcessoFinal					
190		181 dadosBasicos	dataRecebimentoFinal					
191		182 dadosBasicos	dataProcessoFinal					
192		183 dadosBasicos	dataRecebimentoFinal					
193		184 dadosBasicos	dataProcessoFinal					
194		185 dadosBasicos	dataRecebimentoFinal					
195		186 dadosBasicos	dataProcessoFinal					
196		187 dadosBasicos	dataRecebimentoFinal					
197		188 dadosBasicos	dataProcessoFinal					
198		189 dadosBasicos	dataRecebimentoFinal					
199		190 dadosBasicos	dataProcessoFinal					
200		191 dadosBasicos	dataRecebimentoFinal					
201		192 dadosBasicos	dataProcessoFinal					
202		193 dadosBasicos	dataRecebimentoFinal					
203		194 dadosBasicos	dataProcessoFinal					
204		195 dadosBasicos	dataRecebimentoFinal					
205		196 dadosBasicos	dataProcessoFinal					
206		197 dadosBasicos	dataRecebimentoFinal					
207		198 dadosBasicos	dataProcessoFinal					
208		199 dadosBasicos	dataRecebimentoFinal					
209		200 dadosBasicos	dataProcessoFinal					
210		201 dadosBasicos	dataRecebimentoFinal					
211		202 dadosBasicos	dataProcessoFinal					
212		203 dadosBasicos	dataRecebimentoFinal					
213		204 dadosBasicos	dataProcessoFinal					
214		205 dadosBasicos	dataRecebimentoFinal					
215		206 dadosBasicos	dataProcessoFinal					
216		207 dadosBasicos	dataRecebimentoFinal					
2								



TRIBUNAL	(All)	<input type="button" value="▼"/>
TRIBUNAL_JUSTICA	(All)	<input type="button" value="▼"/>
MUNICIPIO	(All)	<input type="button" value="▼"/>
UF	(All)	<input type="button" value="▼"/>
NOME_ORGAO	(All)	<input type="button" value="▼"/>

Campos Válidos,  
sem nenhum erro ou  
inconsistência

Campos sem  
nenhum  
preenchimento  
**(vazios)**

Campos preenchidos com  
**ZEROS**  
(Ex: "0", "0.", "00000" ...)

Campos de DATA inválida;  
ou Num Processo com  
Digito Verificador Errado,  
ou valores fora do domínio.

Consistência verificada  
através do cruzamento com  
as Tabelas: Classe, Assunto  
e Serventia.

Row Labels	TOT PROCESSOS	QDE VALIDOS	% VALIDOS	QDE VAZIOS	QDE com ZEROS	QDE ERROS	QDE INCONSISTENCIA
01-DATA_AJUIZAMENTO	2.994.290	2.766.679	92,40%	0	0	227.611	0
02-NUM_PROCESSO	2.994.290	2.976.168	99,39%	0	0	18.122	0
03-GRAU_JURISDICO	2.994.290	2.994.290	100,00%	0	0	0	0
04-SIGLA_TRIBUNAL	2.994.290	2.994.290	100,00%	0	0	0	0
05-ORGAO_NOME	2.994.290	2.994.290	100,00%	0	0	0	0
06-ORGAO_CODIGO	2.994.290	2.981.082	99,56%	13.208	0	0	0

### ANTES

DO SANEAMENTO

- 01-DATA\_AJUIZAMENTO
- 02-NUM\_PROCESSO
- 06-ORGAO\_CODIGO
- 07-ORGAO\_MUN\_IBGE
- 08-COD\_LOCALIDADE
- 11-SIST\_ELETRONICO

### Processos com Dados VÁLIDOS

2.589.652	86,49%
2.976.168	99,39%
2.954.440	98,67%
2.568.765	85,79%
1.998.561	66,75%
1.819.976	60,78%

### DEPOIS

DO SANEAMENTO

### Processos com Dados VÁLIDOS

2.766.679	92,40%
2.976.168	99,39%
2.982.026	99,59%
2.994.147	100,00%
2.978.850	99,48%
2.395.639	80,01%

**MELHORIA**  
em p.p.

**6%**

**1%**

**14%**

**33%**

**19%**

257.502

0

0

0

0

0

0

205.864

0

0

0

0

0

0

0

0

0

0

## ARQUITETURA



CONSELHO  
NACIONAL  
DE JUSTIÇA

Staging  
Area



Tratamento



Amazon S3



DATA LAKE



Amazon ES  
Elastic Search



Pesquisas



BIG DATA



HADOOP  
HDFS  
HIVE  
IMPALA

Data  
Warehouse



SGBD relacional  
Star Schema

Analytics



QlikView



Microsoft  
Power BI



eproc

@-SAJ

PJe



VM

Máquina Virtual  
( Linux ou Win )

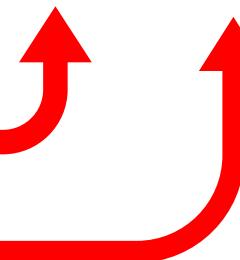
python  
Versão 3.6+



STORAGE

Base Dados  
Saneada

Inconsistências  
e Anomalias



Segurança

Integração

LGPD

Escalabilidade

Interoperabilidade

## FLUXO DE DADOS



LOG de execução

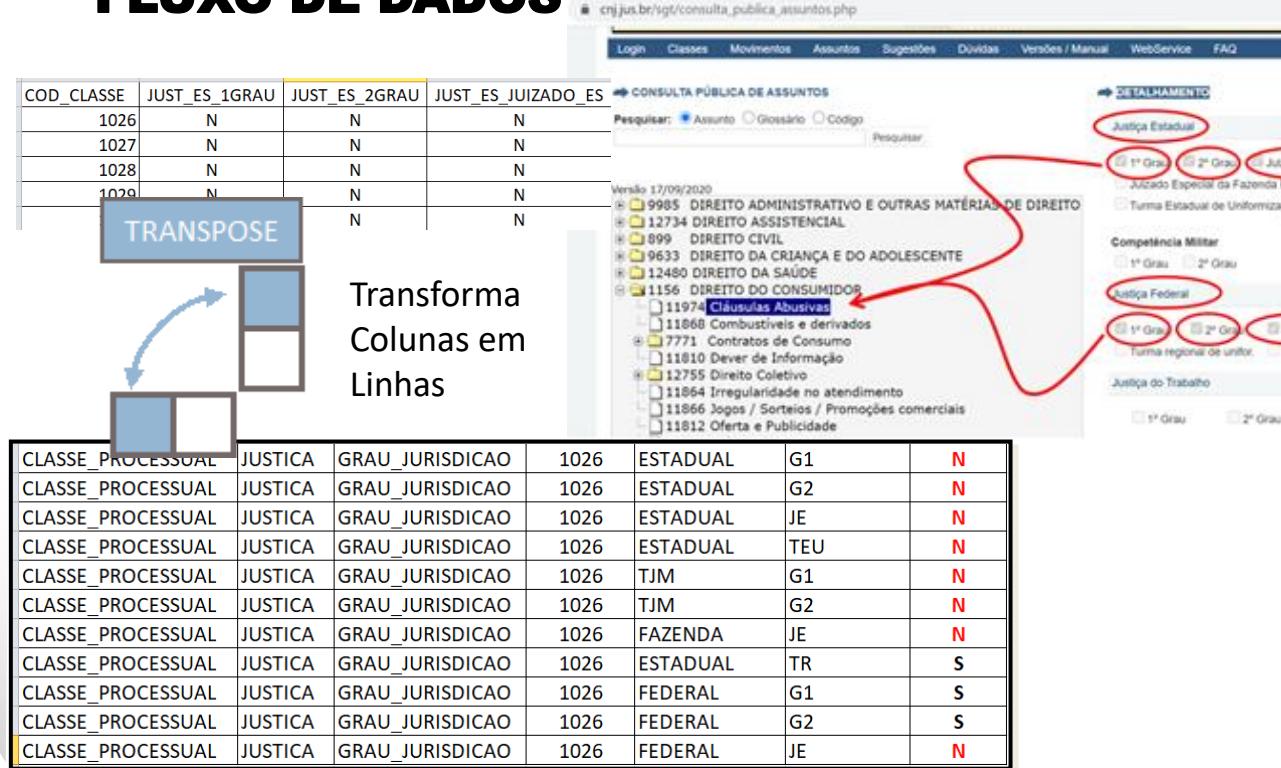
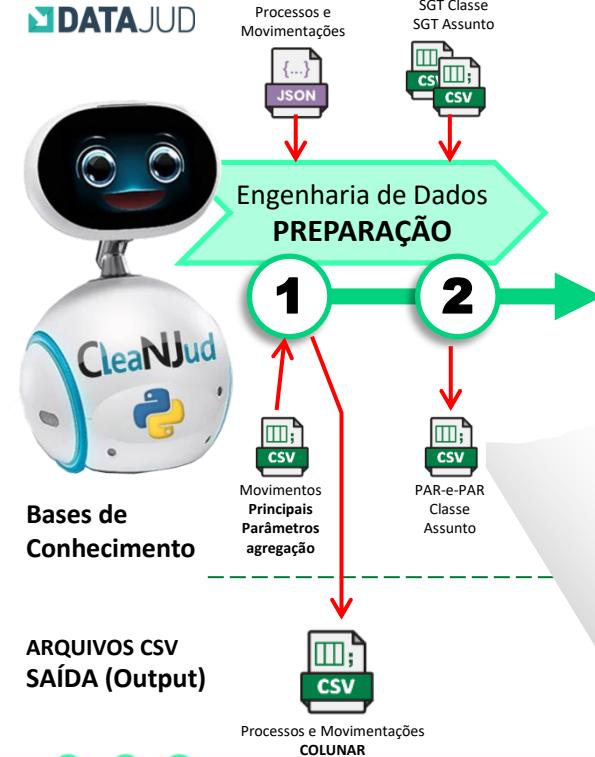
```
log_cleanjud_1_parse_2020_10_17_034053.txt - Notepad
File Edit Format View Help
LOG REPORT - 2020-10-17 03:40:53
CNJ Inova - HACKATHON - Desafio 2
CleaNJud - Saneamento do DATAJUD
Version: 17/10/2020 - v 1.0.0
-----
Start at : 2020-10-17 03:40:53
Input Directory: C:\bases_CNJ
Output Directory: D:\CNJ\output\
Arq JSON Processados= 613
Qde Processos = 2994290
Qde Movimentacoes = 132520434
Completo: 2020-10-17 05:28:22
```

**613** arquivos JSON  
**2.994.290** processos  
**132.520.434** movimentações

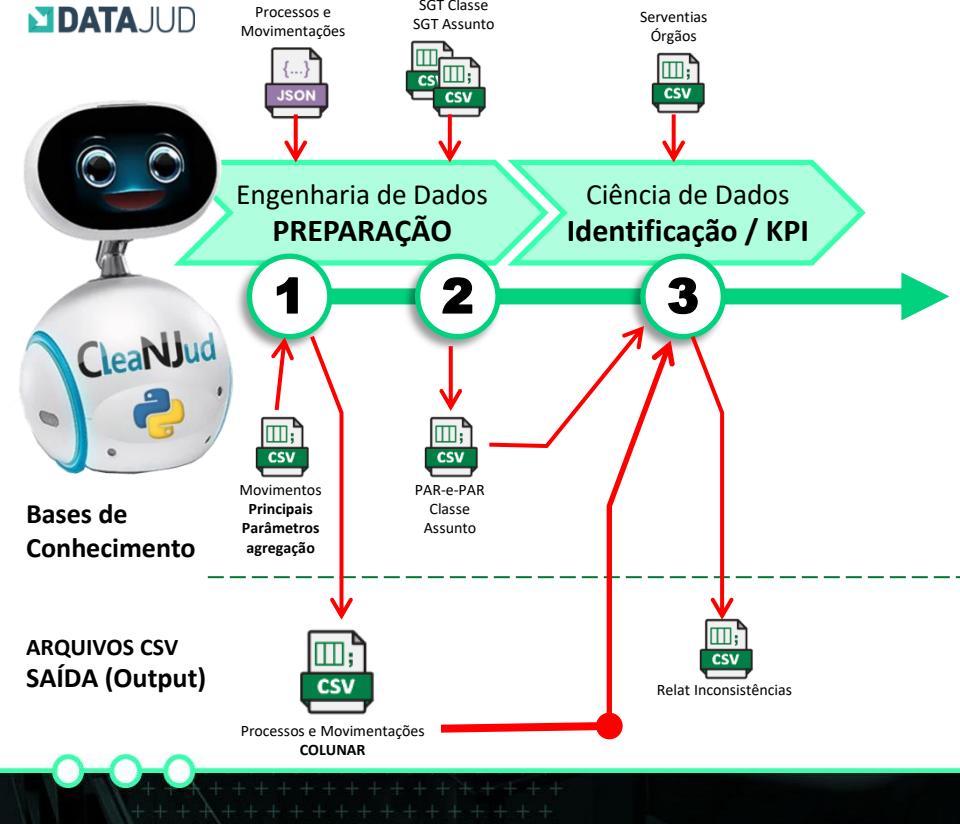
456	MOV_193_SENTENCA	Extinção
12324	MOV_193_SENTENCA	Desfavorável
12323	MOV_193_SENTENCA	Favorável em parte
12322	MOV_193_SENTENCA	Favorável
26	MOV_026_DISTRIB	Distribuição
848	MOV_848_JULGADO	Trânsito em Julgado
22	MOV_022_BAXA	Baixa Definitiva
246	MOV_246_ARQUIVA	Arquivamento Definitivo
861	MOV_246_ARQUIVA	Arquivamento

CLASSE PROCESSUAL	:	002994290	perc:(100.00%)
TIPO PROCESSO	:	002002236	perc:(66.87%)
SIST_ELETRONICO	:	001943116	perc:(64.89%)
VALOR CAUSA	:	001763988	perc:(58.91%)
ASSUNTO1_COD_LOCAL	:	000171345	perc:(5.72%)
ASSUNTO1_COD_PAIS	:	000171345	perc:(5.72%)
ASSUNTO1_DESCRICAO	:	000138657	perc:(4.63%)
ASSUNTO1_PRINCIPAL	:	002828178	perc:(94.45%)
ASSUNTO1_COD_NACIONAL	:	002644957	perc:(88.33%)
ASSUNTO2_COD_LOCAL	:	000019815	perc:(0.66%)
ASSUNTO2_COD_PAIS	:	000019815	perc:(0.66%)
ASSUNTO2_DESCRICAO	:	000016719	perc:(0.54%)

## FLUXO DE DADOS



## FLUXO DE DADOS

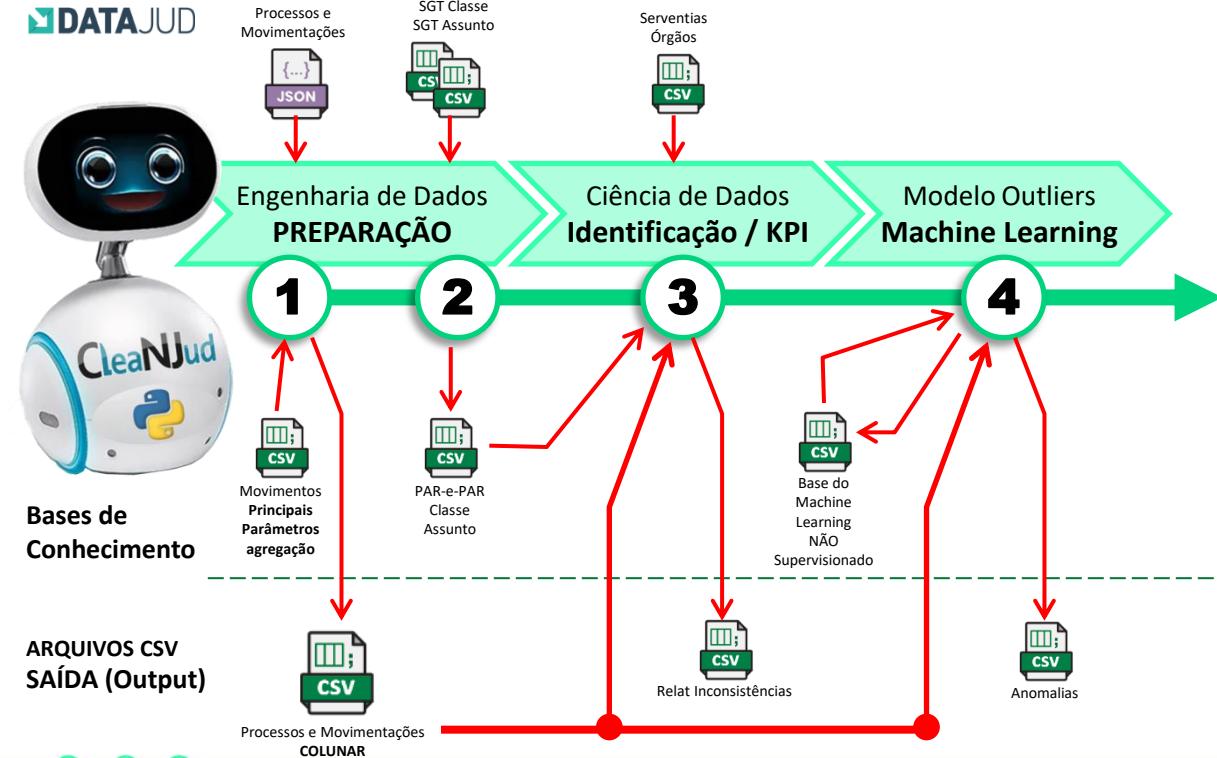


- Realiza contagens: Campos Vazios e Zerados
- Valida os Campos Data
- Valida dígito verificador do Num Processo
- Faz o batimento com tabelas de Domínio
- Movimentações Inconsistentes (fora da sequência)
- Inovação: Algoritmo de Batimento Parametrizado:

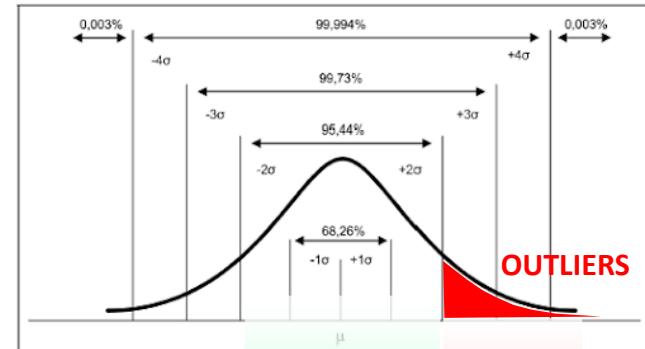
Comparação de Pares de Valores,  
Baseado no Conhecimento Jurídico:

CLASSE x Órgão  
Assunto x Órgão

## FLUXO DE DADOS



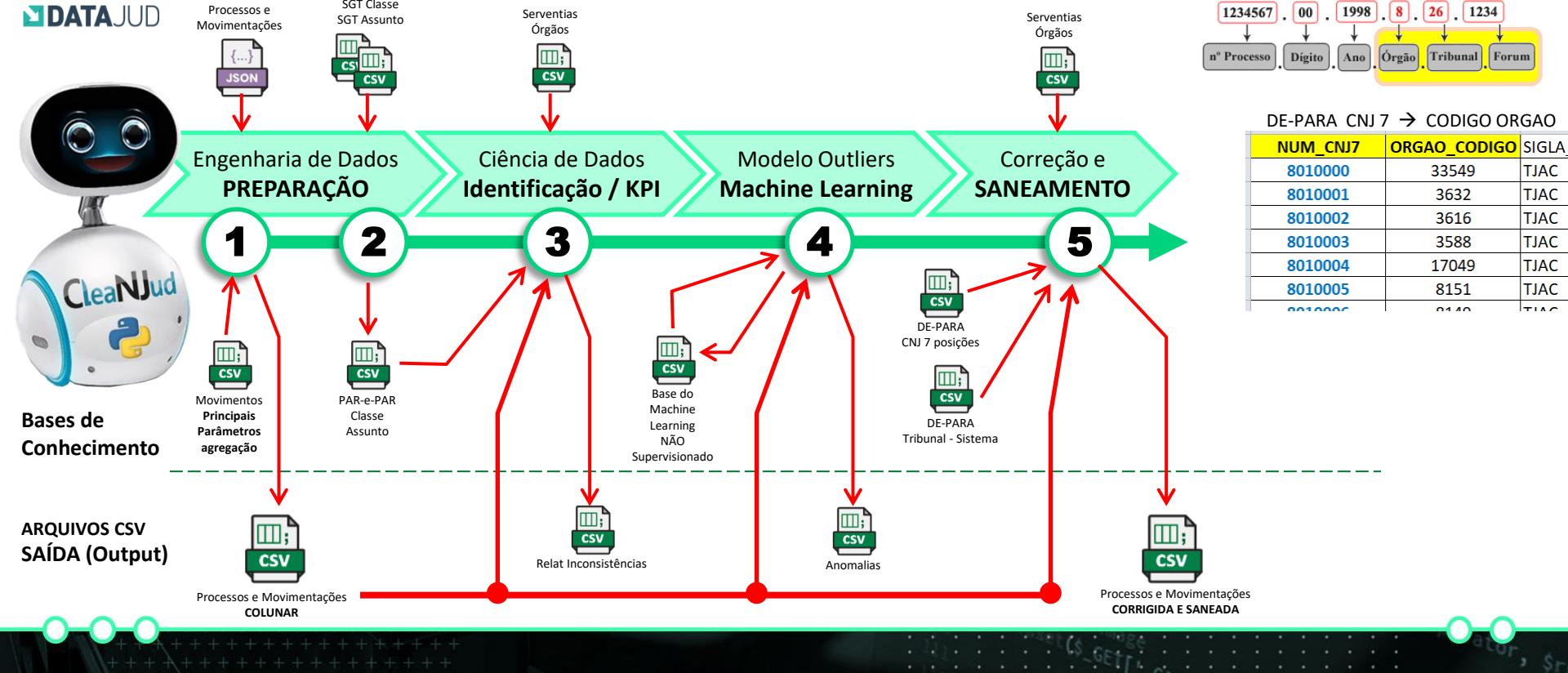
**Modelo Estatístico de Distribuição de Probabilidades – Curva Normal**



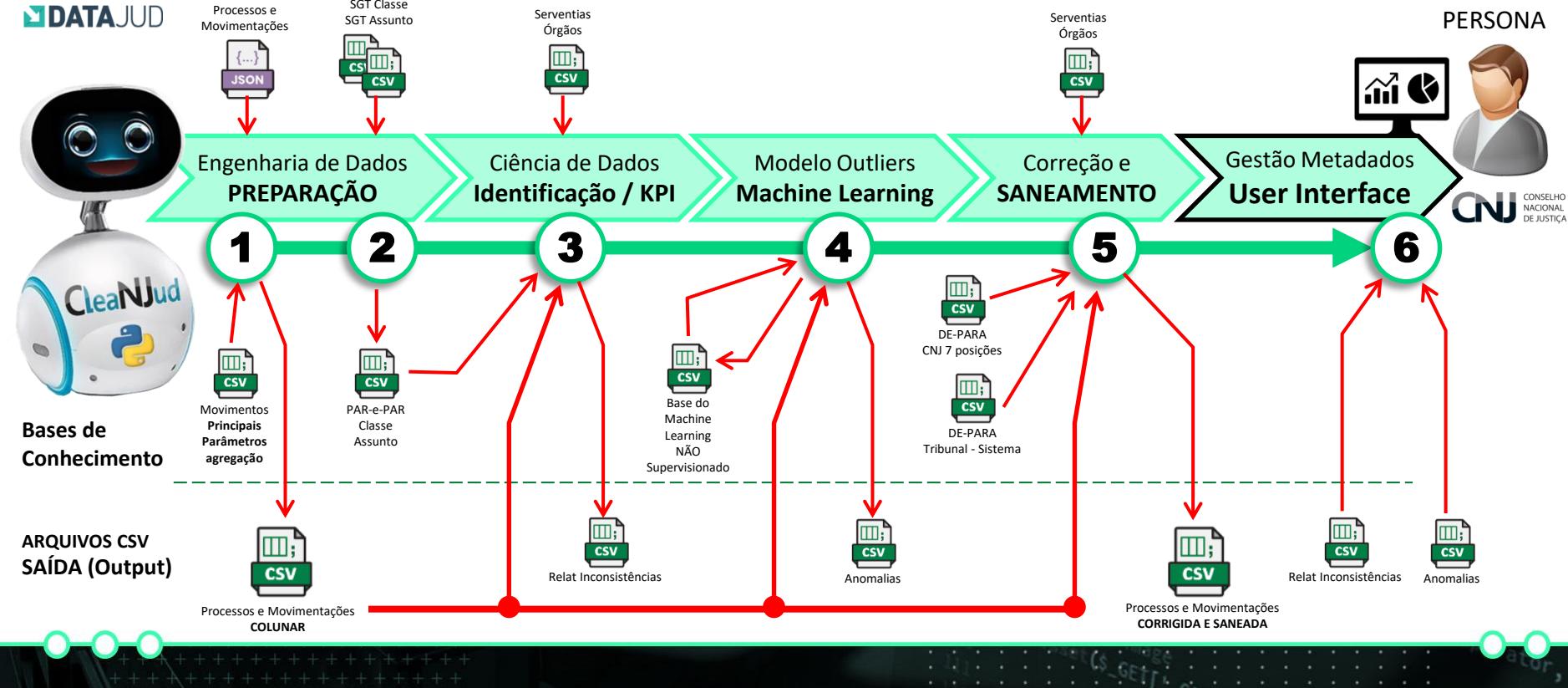
**Base do Machine Learning NÃO Supervisionado**

**ANOMALIAS (suspeitas)**

## FLUXO DE DADOS



## FLUXO DE DADOS



### Efetividade na resolução do problema

- Identificação Inconsistências
- **MELHORIA DE 1% até 33% (pontos percentuais) em algumas informações, após o SANEAMENTO**
- Gestão de Metadados

### Qualidade da Solução

- **100% em PYTHON**
- **100% Software Livre**
- **100% suportado mercado**
- **Totalmente Transparente**
- **Uso no AMBIENTE CNJ**
- **Não trafega dados**
- **Segurança e LGPD**

### Completude da Solução

- UI – Painel Gestão Dados
- Simplicidade
- Facilidade de Uso
- Informações Objetivas
- Documentado

### Viabilidade de Implantação

- **Aproveita toda a estrutura existente no CNJ DATAJUD**
- **Não requer HW específico**
- **Implantado em qq VM**
- **Escalabilidade Flexível**
- **Já em uso pelas equipes do CNJ (Python)**
- **Tecnicamente Replicável**

## AGRADECIMENTOS ESPECIAIS AOS MENTORES:



**Gabriela Moreira Soares**  
Diretora Executiva  
Dep Pesquisas Jud/CNJ  
Brasília



**Isabela Fontana da Mota**  
Advogada  
Dep Pesquisas Jud/CNJ  
Brasília



**Alexander Monteiro**  
TI / Cientista de Dados  
Dep Pesquisas Jud/CNJ  
Brasília



**Thales Falek**  
Advogado e Sócio da  
Carrilho Donas Advocacia  
Brasília



**Rosfran Borges**  
SW Architect / Data Scientist  
CNJ  
Brasília



**Amauri Barros**  
CEO  
AB8 Marketing  
Rio de Janeiro



**Renata Ramos de Castro**  
Consultora jurídica  
TRweb  
Belo Horizonte

EQUIPE:

SHQWEE





# CleaNJud

**DECISÕES MAIS EFICAZES COM  
A QUALIDADE DO DATAJUD**