
 <p>CEFSa Centro Educacional da Fundação Salvador Arena</p>	<p>Código: EC E – 1060 Disciplina: Gestão Estratégica da Informação Curso: Engenharia da Computação Turma: EC10 Prof.: Daniel Henrique Coord.: Michele B. De Souza</p>	 <p>FACULDADE DE TECNOLOGIA TERMOPRECISA FUNDACÃO SALVADOR ARENA</p>
<p>Aluno (a): Caio William Favoretto RA: 082170041 Aluno (a): Kaique Cirto Mafrá RA: 082160029 Aluno (a): Matheus Boareto RA: 082160025 Aluno (a): William Rocha de Oliveira RA: 082160033</p>		<p>Nota:</p>
<p>A N2 do primeiro bimestre consistirá no desenvolvimento de um projeto completo de <i>Business Intelligence</i>, versando a contextualização do problema a ser analisado, estruturação da(s) base(s), construção de dashboard e análise de implementação da solução.</p> <p>Vocês deverão buscar alguma(s) base(s), entender quais são os principais dados a serem analisados nessa base, explorar a mesma e construir um dashboard interativo que possibilite que pessoas que não conheçam o assunto possam entender de forma simples o comportamento desses dados.</p> <p>Assim, deverão ser executados os seguintes passos:</p> <ul style="list-style-type: none">• Busca de base (contendo pelo menos 1000 linhas)• Contextualização e descrição da base, e do problema a ser analisado (Por que estou utilizando essa base? Qual problema pretendo resolver? Quais são as variáveis da base?)• Tratamento dos dados, análise descritiva (análise estatística da base) e construção de Dashboard que possibilite analisar o comportamento dos dados• Proposta de Uso e implantação (Onde o dashboard seria utilizado? Como ele seria implementado? Com que frequência os dados serão atualizados?) <p>Cada grupo deverá apresentar o projeto realizado no dia 30/03/2021, versando os principais pontos analisados. O entregável é um arquivo Power BI contendo o dashboard, e um arquivo word ou pdf detalhando o contexto do problema a ser analisado, descrição das variáveis, proposta de uso e implementação do dashboard.</p> <p>Todos os arquivos relacionados à atividade deverão ser postados no moodle, na atividade "N2 – Primeiro Bimestre".</p> <p>Orientações:</p> <ul style="list-style-type: none">• Caso tenham atividades iguais, ambas terão nota zero;• A nota da atividade será a nota de N2 do 1º bimestre;• A atividade deverá ser postada no Moodle, na disciplina Gestão Estratégica da Informação, apenas por um integrante do grupo e deverá conter o nome e RA de todos os componentes;• A atividade a ser entregue deverá ser postada, IMPRETERÍVELMENTE, até, no máximo, dia 30/03, às 22h45min. Após este horário e data, o Moodle estará travado para receber as atividades e a nota será zero;• A não realização da atividade fará com que o aluno tenha nota zero na N2 do 1º bimestre;• Abaixo critérios de correção: <p>A atividade será corrigida por meio dos seguintes critérios:</p> <p>a. Descrição e contextualização do problema (20%) b. Clareza e interpretabilidade das análises realizadas (40%) c. Criatividade e Dataviz (30%) c. Proposta de Uso e Implantação da Solução (10%)</p>		

Sumário

1. Contextualização do problema e descrição da base	2
2. Análise e Tratamento	2
2.1. Análise Descritiva/Estatística da base	2
2.2. Tratamento dos Dados	7
3. Proposta de implementação	7

1. Contextualização do problema e descrição da base

O Dataset utilizado, é um compilado extraído de quatro bases de dados vinculadas e organizadas por tempo e local, montada com foco em encontrar sinais correlacionados ao aumento das taxas de suicídio entre diferentes pontos de maneira global, em todo o espectro socioeconômico, além de considerar variáveis como idade, sexo, Índice de Desenvolvimento Humano da região, país e entre outros que serão demonstrados abaixo.

As bases utilizadas para montar a atual base analisada foram:

1. Programa das Nações Unidas para o Desenvolvimento. (2018). Índice de desenvolvimento humano (IDH).
 - Obtido em <http://hdr.undp.org/en/indicators/137506>
2. Banco Mundial. (2018). Indicadores de desenvolvimento mundial: PIB (US \$ corrente) por país: 1985 a 2016.
 - Obtido em <http://databank.worldbank.org/data/source/world-development-indicators#>
3. Szamil. (2017). Suicídio no Século XXI [conjunto de dados].
 - Obtido em <https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook>
4. Organização Mundial da Saúde. (2018). Prevenção do suicídio.
 - Obtido em https://www.who.int/mental_health/suicide-prevention/en/

A base foi montada com o intuito de mapear possíveis estatísticas relacionadas ao suicídio e viabilizar de alguma maneira um auxílio para quem trabalha com a prevenção deste problema social.

2. Análise e Tratamento

2.1. Análise Descritiva/Estatística da base

Utilizando-se da biblioteca python [Pandas Profiling](#), realizamos inicialmente uma análise estatística da base, com o intuito de compreender, de forma macro, a necessidade de tratamento dos dados. O código abaixo mostra como foi realizado:

```
In [4]: import pandas as pd
import pandas_profiling

base = pd.read_csv("Suicide Rates Overview 1985 to 2016.csv")

pandas_profiling.ProfileReport(df=base, explorative=True)
```

Como resultado a biblioteca fornece um report rápido com estatísticas gerais da base e individuais para cada variável.

Em **Overview** podemos observar que possuímos uma base com 12 variáveis, sendo 6 categóricas e 6 numéricas, mais de 27 mil registros e, talvez o mais importante, com apenas 5,8% de células missing, ou seja, temos uma base bem preenchida.

Overview

Warnings10

Reproduction

Dataset statistics

Number of variables	12
Number of observations	27820
Missing cells	19456
Missing cells (%)	5.8%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	2.5 MiB
Average record size in memory	96.0 B

Variable types

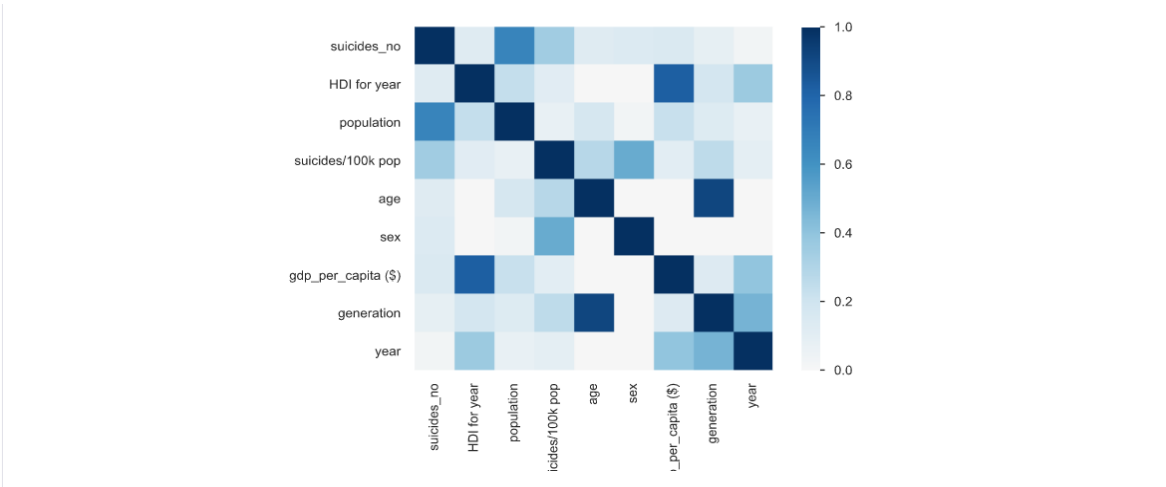
Categorical	6
Numeric	6

Porém, em **Warnings** temos mais algumas informações importantes. Possuímos 3 colunas com uma alta cardinalidade, o que significa que possuem muitos valores repetidos e que poderiam quebrados em dimensões específicas com o intuito de reduzir redundância, economia de espaço em disco e memória (o que permite melhor performance no processamento dos relatórios), além de ser uma boa prática de organização e modelagem de dados.

E outros dois pontos importantes de atenção, onde as variáveis **“HDI for year”**, **“suicides_no”** e **“suicides/100k pop”** possuem uma taxa considerável de missing e zero, o que pode impactar ou até enviesar as análises.

Overview	Warnings 10	Reproduction
Warnings		
country	has a high cardinality: 101 distinct values	High cardinality
country-year	has a high cardinality: 2321 distinct values	High cardinality
gdp_for_year (\$)	has a high cardinality: 2321 distinct values	High cardinality
HDI for year	has 19456 (69.9%) missing values	Missing
sex	is uniformly distributed	Uniform
age	is uniformly distributed	Uniform
country-year	is uniformly distributed	Uniform
gdp_for_year (\$)	is uniformly distributed	Uniform
suicides_no	has 4281 (15.4%) zeros	Zeros
suicides/100k pop	has 4281 (15.4%) zeros	Zeros

Ainda em uma visão macro da base, temos uma análise rápida de correlação entre as variáveis, que pode nos dar uma ideia inicial interessante de como os dados se comportam nessa base e, possivelmente, nos dar uma direção inicial para a análise.



E passando agora para uma visão mais individual de cada variável, podemos ver estatísticas específicas de cada uma e entender como cada variável pode nos ajudar a obter informações interessantes dessa base.

Variables

country

Categorical

HIGH CARDINALITY

Distinct	101
Distinct (%)	0.4%
Missing	0
Missing (%)	0.0%
Memory size	217.5 KiB

Austria382

Iceland382

Mauritius382

Netherlands382

Mexico372

Other values (96)25920

Toggle details

year

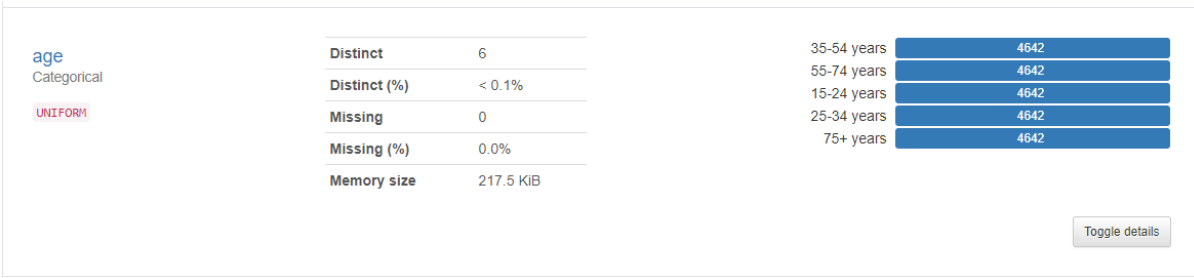
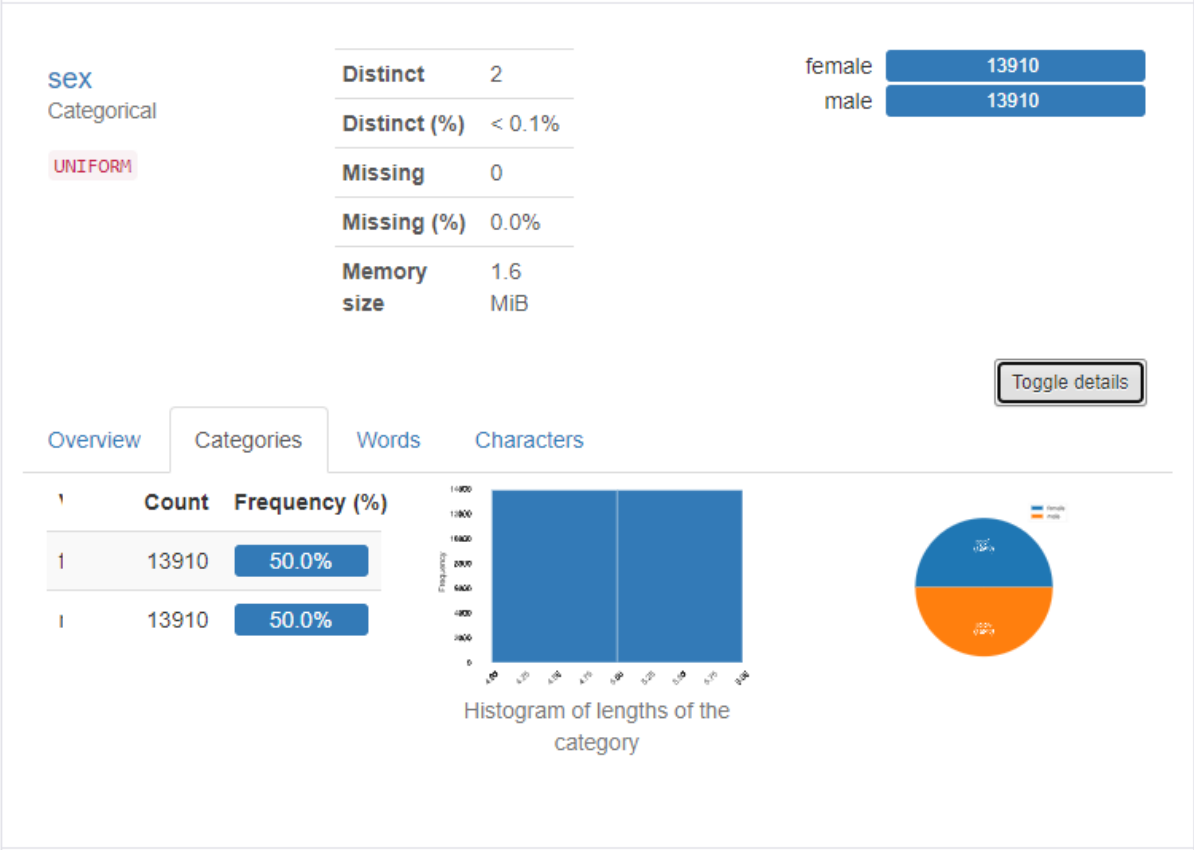
Real number (ℝ₃₂)

Distinct	32
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

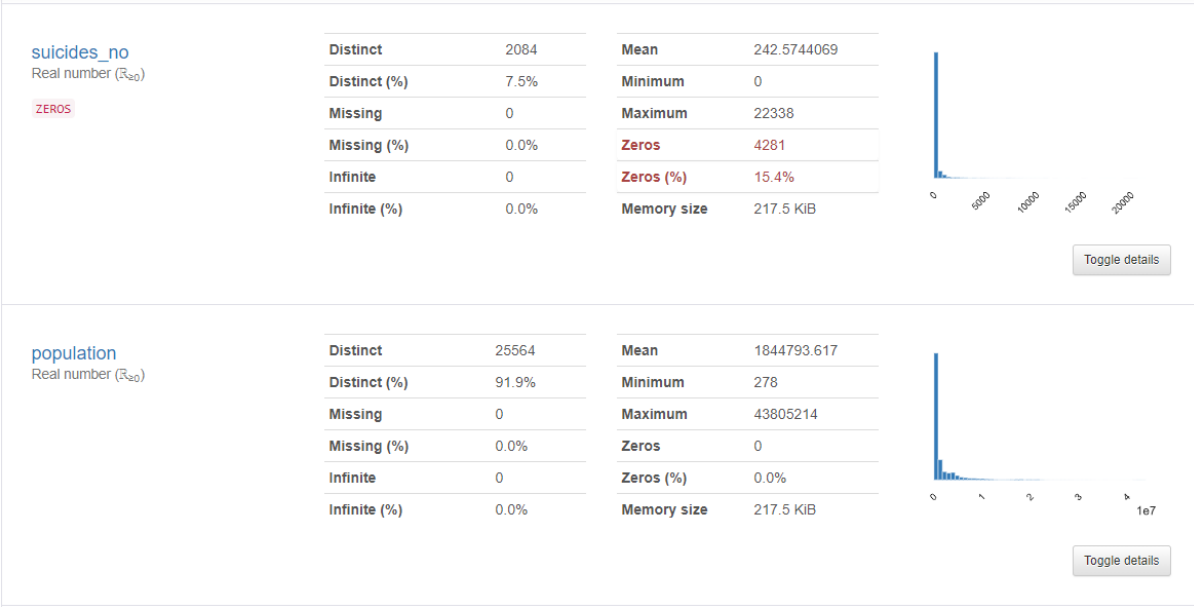
Mean	2001.258375
Minimum	1985
Maximum	2016
Zeros	0
Zeros (%)	0.0%
Memory size	217.5 KiB

A bar chart showing the frequency of each year from 1985 to 2016. The x-axis is labeled with years 1980, 2000, and 2010. The bars show a relatively uniform distribution of data points across the years, with a slight increase in frequency towards the end of the period.

Toggle details



** É importante observar que a distribuição das variáveis “sex” e “age” são iguais entre suas categorias na base, o que não deve influenciar no risco de enviesar as análises.



population

Real number ($\mathbb{R}_{\geq 0}$)

Distinct

25564

Distinct (%)

91.9%

Missing

0

Missing (%)

0.0%

Infinite

0

Infinite (%)

0.0%

Mean

1844793.617

Minimum

278

Maximum

43805214

Zeros

0

Zeros (%)

0.0%

Memory size

217.5 KiB

Toggle details

0

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

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

1001

1002

1003

1004

1005

1006

1007

1008

1009

1010

1011

1012

1013

1014

1015

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

1066

1067

1068

1069

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

1080

1081

1082

1083

1084

1085

1086

1087

1088

1089

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

1109

1110

1111

1112

1113

1114

1115

1116

1117

1118

1119

1120

1121

1122

1123

1124

1125

1126

1127

1128

1129

1130

1131

1132

1133

1134

1135

1136

1137

1138

1139

1140

1141

1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187

1188

1189

1190

1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207

1208

1209

1210

1211

1212

1213

1214

1215

1216

1217

1218

1219

1220

1221

1222

1223

1224

1225

1226

1227

1228

1229

1230

1231

1232

1233

1234

1235

1236

1237

1238

1239

1240

1241

1242

1243

1244

1245

1246

1247

1248

1249

1250

1251

1252

1253

1254

1255

1256

1257

1258

1259

1260

1261

1262

1263

1264

1265

1266

1267

1268

1269

1270

1271

1272

1273

1274

1275

1276

1277

1278

1279

1280

1281

1282

1283

1284

1285

1286

1287

1288

1289

1290

1291

1292

1293

1294

1295

1296

1297

1298

1299

1300

1301

1302

1303

1304

1305

1306

1307

1308

1309

1310

1311

1312

1313

1314

1315

1316

1317

1318

131

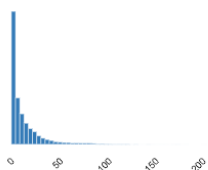
suicides/100k pop

Real number ($\mathbb{R}_{\geq 0}$)

ZEROS

Distinct	5298
Distinct (%)	19.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	12.81609741
Minimum	0
Maximum	224.97
Zeros	4281
Zeros (%)	15.4%
Memory size	217.5 KiB



Toggle details


country-year

Categorical

HIGH CARDINALITY
UNIFORM

Distinct	2321
Distinct (%)	8.3%
Missing	0
Missing (%)	0.0%
Memory size	217.5 KiB

Lithuania2003	12
Malta2003	12
Russian Federation2000	12
Kyrgyzstan1993	12
Belize2014	12
Other values (2316)	27760



Toggle details

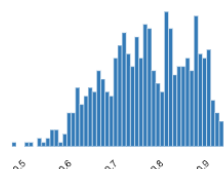
HDI for year

Real number ($\mathbb{R}_{\geq 0}$)

MISSING

Distinct	305
Distinct (%)	3.6%
Missing	19456
Missing (%)	69.9%
Infinite	0
Infinite (%)	0.0%

Mean	0.7766011478
Minimum	0.483
Maximum	0.944
Zeros	0
Zeros (%)	0.0%
Memory size	217.5 KiB



Toggle details


gdp_for_year (\$)

Categorical

HIGH CARDINALITY
UNIFORM

Distinct	2321
Distinct (%)	8.3%
Missing	0
Missing (%)	0.0%
Memory size	217.5 KiB

431,120,310,089	12
342,617,007,104	12
57,008,425,296	12
314,267,667,675	12
262,709,948,091	12
Other values (2316)	27760



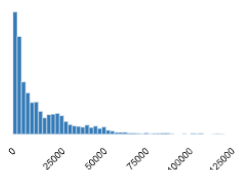
Toggle details

gdp_per_capita (\$)

Real number ($\mathbb{R}_{\geq 0}$)

Distinct	2233
Distinct (%)	8.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	16866.46441
Minimum	251
Maximum	126352
Zeros	0
Zeros (%)	0.0%
Memory size	217.5 KiB



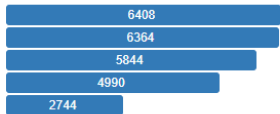
Toggle details

generation

Categorical

Distinct	6
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	217.5 KiB

Generation X	6408
Silent	6364
Millenials	5844
Boomers	4990
G.I. Generation	2744

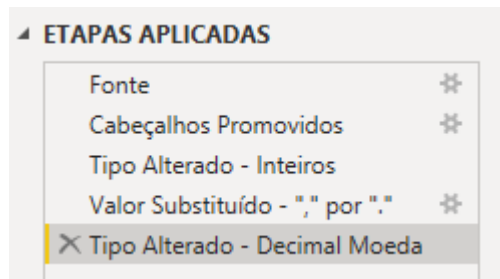


Toggle details

2.2. Tratamento dos Dados

Para a construção do Dashboard, foi necessário antes realizar alguns tratamentos nos dados. Para isto, foi utilizando as próprias ferramentas que o PowerBI fornece, onde foi construída uma “pipeline” de tratamentos que será executada todas as vezes que a base for carregada ou atualizada na ferramenta.

As etapas de tratamento aplicadas são mostradas na imagem abaixo, onde:



- **Fonte:** etapa de load da base;
- **Cabeçalhos Promovidos:** etapa em que a primeira linha do arquivo CSV é definida como os nomes das colunas;
- **Tipo Alterado - Inteiros:** etapa de conversão de colunas numéricas para tipo “Número Inteiro”;
- **Valor Substituído - “,” por “.”:** etapa onde o caractere “,” é substituído por “.” em colunas com números decimais;
- **Tipo Alterado - Decimal Moeda:** etapa de conversão de colunas numéricas para tipo “Número decimal fixo”.

2.3. Proposta de implementação

Como proposta de uso e implementação do Dashboard, é sugerido que o relatório seja publicado utilizando-se uma conta corporativa da Microsoft, de forma que seja possível compartilhar o relatório na web adicionando os usuários que teriam permissão de acesso ao dashboard.

Da mesma forma, a base de dados, por ser um simples arquivo CSV, poderia ser armazenada na própria ferramenta da Microsoft, o OneDrive, pois existe uma boa integração entre os produtos do Office 365.