

## **Capstone Project Session 1**

Derrick Mills

Student

Michigan Engineering Professional Education AIML – Cohort 3

### **Aura: Intelligent Prediction and Recommendation Engine for Digital Health Analytics**

#### **Introduction**

The development of Aura, an intelligent prediction and recommendation engine for healthcare-focused digital marketing, requires a robust and reliable data pipeline capable of handling complex, heterogeneous datasets. The NSMES1988 dataset serves as the initial input for establishing the fundamental data ingestion, preprocessing, and optimization procedures that will support Aura's analytical, visualization, and machine-learning components.

The present report details the activities completed during Session 1, including data importation, structural evaluation, preliminary preprocessing, type refinement, memory footprint analysis, and exportation of optimized data artifacts. These steps establish a clean, validated, and computationally efficient foundation for subsequent modeling and system integration.

#### **Methods:**

##### **Data Importation Environment**

All tasks were conducted within Google Colab, using Python's scientific computing stack (pandas, NumPy). The dataset was imported from the file path /NSMES1988.csv. Colab provides a cloud-hosted execution environment that supports reproducible, scalable computational workflows suitable for healthcare analytics.

##### **Initial Structural Inspection**

Upon loading, the dataset was examined to characterize its dimensionality, variable composition, data types, and overall integrity. The inspection process involved:

- Evaluating dataset shape
- Reviewing variable data types
- Displaying summary statistics for numerical variables
- Assessing missingness patterns
- Identifying anomalous or non-standard values
- Characterizing categorical vs. continuous variables

##### **Data Cleanliness Assessment**

A complete analysis was performed using `df.isna().sum()`. Logical consistency checks were conducted to identify possible encoding abnormalities or out-of-range values. Special focus was given to socioeconomic variables such as income and clinical utilization variables that influence predictive modeling.

### **Type Optimization and Memory Reduction**

Given Aura's requirement to scale to larger datasets in future phases, memory efficiency was prioritized. The following transformations were applied:

- 1. Removal of redundant variables**

- Unnamed: 0, an auto-generated index column, was removed.

- 2. Categorical encoding**

Variables representing discrete states (e.g., health, adl, region, gender, married, employment, insurance, medicaid) were cast to pandas category types.

- 3. Numeric downcasting**

- Integer columns were downcast from `int64` to smaller integer types (`int16` or `int8`), where appropriate.
- Floating-point columns (age, income) were downcast from `float64` to `float32`.

Memory usage was measured before and after optimization using `df.info(memory_usage="deep")`.

### **Data Exportation**

To support downstream processing pipelines, two export formats were generated:

- The original dataset was exported as JSON (`NSMES1988.json`), enabling API-friendly ingestion.
- The cleaned, optimized dataset was exported as CSV (`NSMES1988new.csv`) for analytical and modeling tasks.

### **Results:**

#### **Dataset Structure and Composition**

The dataset contains 4,406 observations and 19 variables, encompassing:

- **Healthcare utilization** - visits, non-physician visits, physician visits, emergency services, hospital admissions
- **Health status indicators** - chronic conditions, functional limitations (ADL), self-reported health

- **Demographic factors** - age (expressed in decades), gender, marital status
- **Socioeconomic indicators** - education, income (in units of \$10,000), employment status
- **Insurance variables** - private insurance, Medicaid coverage
- **Geographic region** - urban, rural, or regional classifications

The dataset exhibits no missing values, indicating complete records across all variables.

## **Observed Anomalies and Considerations**

Two key observations emerged from the structural assessment:

### **1. Age Encoding**

The age variable ranges from approximately 6.6 to 10.9, indicating that age is measured in decades rather than years. This transformation will require reconversion or careful interpretative handling in later modeling stages.

### **2. Negative Income Values**

Income ranges from 1.0 to 54.8 (i.e., −\$10,000 to \$548,000). Negative values may reflect special coding, reporting errors, or debt and should be evaluated through business rules in subsequent preprocessing sessions.

## **Memory Utilization Before and After Optimization**

- Original dataset memory footprint - 2.4 MB
- Optimized dataset memory footprint - 123 KB

This represents a 95% reduction in memory usage, achieved through categorical encoding and numeric downcasting. Although this dataset is relatively small, the optimization approach establishes a scalable pattern suitable for large-volume healthcare datasets expected in later stages of Aura's development.

## **Exported Artifacts**

Two export files were successfully generated:

- NSMES1988.json - original dataset
- NSMES1988new.csv - cleaned and optimized dataset

These files are now available within the Colab working directory for downstream analysis.

## **Discussion**

The preprocessing performed in Session 1 establishes a rigorous and efficient foundation for integrating the NSMES1988 dataset into Aura's analytical framework. The absence of missing

data simplifies early modeling tasks; the presence of encoded or transformed values (e.g., age in decades, negative income values) warrants domain-specific treatment in later sessions.

The dramatic memory footprint reduction demonstrates the importance of dtype optimization in scalable analytics architectures. As Aura evolves to accommodate larger, multi-source, and potentially real-time healthcare datasets, such optimizations will play a pivotal role in supporting low-latency processing, cost-efficient cloud workloads, and smooth visualization rendering.

Furthermore, the structured export of both original and cleaned datasets ensures compatibility with diverse components of the Aura pipeline, including RESTful ingestion endpoints, visualization engines, and machine-learning model trainers.

## Conclusion

Session 1 accomplished the foundational tasks of data importation, structural assessment, data type optimization, memory analysis, and exportation of both original and cleaned datasets. The dataset is now prepared for more advanced phases of processing, including exploratory data analysis (Session 2), feature engineering, and predictive model development within the Aura engine.

The dataset's completeness, coupled with the systematic preprocessing conducted, provides a strong basis for subsequent analytical and modeling activities central to Aura's mission of delivering intelligent, data-driven insights for healthcare marketers.

## Recommendations for Future Work

To prepare for predictive modeling and visualization in subsequent sessions, the following steps are recommended:

1. **Address anomalous income values** through imputation, capping, or transformation based on domain policy.
2. **Normalize key continuous variables**, particularly income and age (if converted back to years).
3. **Create a correlation matrix and univariate distributions** to identify predictive signal strength.
4. **Explore visit utilization clusters** (e.g., high-utilizers vs. low-utilizers) as potential segmentation insights.
5. **Establish metadata documentation** for variable semantics and business rules to support auditability and reproducibility.

aura\_session1.ipynb

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text ▶ Run all

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

```
import numpy as np
import pandas as pd

# --- CONFIG ---
CSV_PATH = "/NOMES1988.csv" # update if your file is elsewhere, e.g. "NOMES1988.csv"

def load_data(csv_path: str) -> pd.DataFrame:
    """Load the NOMES1988 dataset from CSV."""
    df = pd.read_csv(csv_path)
    return df

def basic_inspection(df: pd.DataFrame) -> None:
    """Print basic info and cleanliness checks."""
    print("===== BASIC INSPECTION =====")
    print(f"Shape (rows, columns): {df.shape}\n")

    print("Column dtypes:")
    print(df.dtypes, "\n")

    print("First 5 rows:")
    print(df.head(), "\n")

    print("Summary statistics (numeric):")
    print(df.describe(), "\n")

    print("Missing values per column:")
    print(df.isna().sum(), "\n")

    if "age" in df.columns:
        print("Age stats:")
        print(df["age"].describe(), "\n")

    if "income" in df.columns:
        print("Income stats:")
        print(df["income"].describe(), "\n")

def memory_report(df: pd.DataFrame, title: str) -> None:
    """Print DataFrame info with deep memory usage."""
    print(f"===== MEMORY REPORT: {title} =====")
    df.info(memory_usage="deep")

def optimize_dtypes(df: pd.DataFrame) -> pd.DataFrame:
    """
    Optimize dtypes:
    - drop artificial index column if present
    - convert object columns to category
    - downcast integer and float columns
    """
    df_opt = df.copy()

    # Drop the 'Unnamed: 0' index column if present
    if "Unnamed: 0" in df_opt.columns:
        df_opt = df_opt.drop(columns=["Unnamed: 0"])

    # Categorical columns from the data dictionary
    cat_cols = [
        "health",
        "sex",
        "region",
        "gender",
        "married",
        "employed",
        "insurance",
        "medicaid",
    ]

    for col in cat_cols:
        if col in df_opt.columns:
            df_opt[col] = df_opt[col].astype("category")

    # Downcast integer columns
    int_cols = df_opt.select_dtypes(include=["int64"]).columns
    df_opt[int_cols] = df_opt[int_cols].apply(
        pd.to_numeric, downcast="integer"
    )

    # Downcast float columns
    float_cols = df_opt.select_dtypes(include=["float64"]).columns
    df_opt[float_cols] = df_opt[float_cols].apply(
        pd.to_numeric, downcast="float"
    )

    return df_opt

def export_files(df_original: pd.DataFrame, df_clean: pd.DataFrame) -> None:
    """
    Export:
    - original data to JSON (NOMES1988.json)
    - cleaned / optimized data to CSV (NOMES1988new.csv)
    """
    print("===== EXPORTING FILES =====")

    # JSON export of original dataframe
    json_path = "NOMES1988.json"
    df_original.to_json(json_path, orient="records")
    print(f"Exported original DataFrame to JSON: {json_path}")

    # CSV export of cleaned / optimized dataframe
    csv_new_path = "NOMES1988new.csv"
    df_clean.to_csv(csv_new_path, index=False)
    print(f"Exported cleaned DataFrame to CSV: {csv_new_path}")

# --- MAIN EXECUTION IN COLAB ---
```

How can I install Python libraries?

Load data from Google Drive

Show an example of training

What can I help you build?

+

Variables

Terminal

11:46 AM

Python 3

aura\_session1.ipynb

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text ▶ Run all

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

4

aura\_session1.ipynb

File Edit View Insert Runtime Tools Help

Q Commands + Code + Text + Run all

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

1319

1320

1321

1322

1323

1324

1325

1326

1327

1328

1329

1330

1331

1332

1333

1334

1335

1336

1337

1338

1339

1340

1341

1342

1343

1344

1345

1346

1347

1348

1349

1350

1351

1352

1353

1354

1355

1356

1357

1358

1359

1360

1361

1362

1363

1364

1365

1366

1367

1368

1369

1370

1371

1372

1373

1374

1375

1376

1377

1378

1379

1380

1381

1382

1383

1384

1385

1386

1387

1388

1389

1390

1391

1392

1393

1394

1395

1396

1397

1398

1399

1400

1401

1402

1403

1404

1405

1406

1407

1408

1409

1410

1411

1412

1413

1414

1415

1416

1417

1418

1419

1420

1421

1422

1423

1424

1425

1426

1427

1428

1429

1430

1431

1432

1433

1434

1435

1436

1437

1438

1439

1440

1441

1442

1443

1444

1445

1446

1447

1448

1449

1450

1451

1452

1453

1454

1455

1456

1457

1458

1459

1460

1461

1462

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474

1475

1476

1477

1478

1479

1480

1481

1482

1483

1484

1485

1486

1487

1488

1489

1490

1491

1492

1493

1494

1495

1496

1497

1498

<

```
aura_session1.pyb
File Edit View Insert Runtime Tools Help
nmands + Code + Text + Run all

    emergency    hospital    chronic    age    school \
count  4486.000000  4486.000000  4486.000000  4486.000000  4486.000000
mean    0.261104    0.295058    1.541088    7.402486    10.290285
std     0.783659    0.746398    1.349632    0.633485    3.738736
min     0.000000    0.000000    0.000000    6.000000    8.000000
25%     0.000000    0.000000    1.000000    6.500000    8.000000
50%     0.000000    0.000000    1.000000    7.300000    11.000000
75%     0.000000    0.000000    2.000000    7.800000    12.000000
max     12.000000    8.000000    8.000000    10.500000    18.000000

    income
count  4486.000000
mean    2.527132
std     2.924648
min    -1.812580
25%     0.912150
50%     1.698150
75%     3.172850
max     54.835180

Missing values per column:
Unnamed: 0  0
visits      0
nvisits     0
ovisits     0
novisits    0
emergency   0
hospital    0
health      0
chronic     0
adl         0
region      0
age         0
gender      0
married     0
school      0
income      0
employed    0
insurance   0
medicaid   0
dtype: int64

Age stats:
count  4486.000000
mean    7.402486
std     0.633485
min     6.000000
25%     6.500000
50%     7.300000
75%     7.800000
max     10.500000
Name: age, dtype: float64

Income stats:
count  4486.000000
mean    2.527132
std     2.924648
min    -1.812580

How can I install Python libraries? Load data from Google Drive Show an example of training:
```

```
aura_session1.pyb
File Edit View Insert Runtime Tools Help
nmands + Code + Text + Run all

Income stats:
count  4486.000000
mean    2.527132
std     2.924648
min    -1.812580
25%     0.912150
50%     1.698150
75%     3.172850
max     54.835180
Name: income, dtype: float64

--- MEMORY REPORT: Original DataFrame ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4486 entries, 0 to 4485
Data columns (total 19 columns):
# Column Non-Null Count Dtype
---  ---
0 Unnamed: 0 4486 non-null int64
1 visits 4486 non-null int64
2 nvisits 4486 non-null int64
3 ovisits 4486 non-null int64
4 novisits 4486 non-null int64
5 emergency 4486 non-null int64
6 hospital 4486 non-null int64
7 health 4486 non-null object
8 chronic 4486 non-null int64
9 adl 4486 non-null object
10 region 4486 non-null object
11 age 4486 non-null float64
12 gender 4486 non-null object
13 married 4486 non-null object
14 school 4486 non-null int64
15 income 4486 non-null float64
16 employed 4486 non-null object
17 insurance 4486 non-null object
18 medicaid 4486 non-null object
dtypes: float64(2), int64(9), object(8)
memory usage: 2.2 MB

--- MEMORY REPORT: Optimized DataFrame ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4486 entries, 0 to 4485
Data columns (total 18 columns):
# Column Non-Null Count Dtype
---  ---
0 visits 4486 non-null int8
1 nvisits 4486 non-null int8
2 ovisits 4486 non-null int16
3 novisits 4486 non-null int16
4 emergency 4486 non-null int8
5 hospital 4486 non-null int8
6 health 4486 non-null category
7 chronic 4486 non-null int8
8 adl 4486 non-null category
9 region 4486 non-null category
10 age 4486 non-null float32
11 gender 4486 non-null category

How can I install Python libraries? Load data from Google Drive Show an example of training:
```

```
1 visits 4486 non-null int64
2 nvisits 4486 non-null int64
3 ovisits 4486 non-null int64
4 novisits 4486 non-null int64
5 emergency 4486 non-null int64
6 hospital 4486 non-null int64
7 health 4486 non-null object
8 chronic 4486 non-null int64
9 aei 4486 non-null object
10 region 4486 non-null object
11 age 4486 non-null float64
12 gender 4486 non-null object
13 married 4486 non-null object
14 school 4486 non-null int64
15 income 4486 non-null float64
16 employed 4486 non-null object
17 insurance 4486 non-null object
18 medicaid 4486 non-null object
dtypes: float64(2), int64(9), object(8)
memory usage: 2.2 MB

=== MEMORY REPORT: Optimized DataFrame ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4486 entries, 0 to 4485
Data columns (total 18 columns):
# Column Non-Null Count Dtype
---
0 visits 4486 non-null int8
1 nvisits 4486 non-null int8
2 ovisits 4486 non-null int16
3 novisits 4486 non-null int16
4 emergency 4486 non-null int8
5 hospital 4486 non-null int8
6 health 4486 non-null category
7 chronic 4486 non-null int8
8 aei 4486 non-null category
9 region 4486 non-null category
10 age 4486 non-null float32
11 gender 4486 non-null category
12 married 4486 non-null category
13 school 4486 non-null int8
14 income 4486 non-null float32
15 employed 4486 non-null category
16 insurance 4486 non-null category
17 medicaid 4486 non-null category
dtypes: category(8), float32(2), int16(2), int8(6)
memory usage: 113.9 KB

=== EXPORTING FILES ===
Exported original DataFrame to JSON: NSHES1988.json
Exported cleaned DataFrame to CSV: NSHES1988new.csv

Done. Files now in the current working directory:
total 1.5M
dnuw----- 5 root root 4.0K Nov 29 16:41 drive
-rw-r--r-- 1 root root 1.2M Nov 29 16:46 NSHES1988.json
-rw-r--r-- 1 root root 259K Nov 29 16:46 NSHES1988new.csv
dnuw-r-x 1 root root 4.0K Nov 20 14:30 sample_data
```

[How can I install Python libraries?](#)

[Load data from Google Drive](#)

[Show an example of training](#)