

Analyse de la performance des stations de distribution de pétrole en Tunisie

Projet SID

Année universitaire 2024-2025

Table des matières

1	Introduction	2
2	Étapes du projet	2
2.1	Analyse des sources de données	2
2.2	Préparation et transformation des données	3
2.3	Implémentation ETL sous Talend	5
2.4	Identification des KPIs	7
2.5	Visualisation avec Power BI	8
3	Conclusion	9

1 Introduction

Dans un environnement économique de plus en plus concurrentiel, les entreprises opérant dans le secteur pétrolier doivent s'appuyer sur des outils décisionnels efficaces pour piloter leur activité et prendre des décisions stratégiques éclairées. Face à la diversité des branches commerciales — Carburants, Gaz, Lubrifiants et Aviation — il devient essentiel de disposer d'un système capable de centraliser, structurer et valoriser les données issues de ces différentes activités.

Ce projet s'inscrit dans cette logique. Il a pour objectif de concevoir une application d'aide à la décision dédiée au suivi et à l'analyse des performances commerciales. À travers une chaîne décisionnelle complète — allant de l'analyse des sources de données à la visualisation des indicateurs via Power BI — ce projet vise à mettre en place un datamart optimisé, alimenté par des flux ETL conçus sous Talend, permettant de restituer des tableaux de bord dynamiques et interactifs.

Ce rapport retrace l'ensemble des étapes de réalisation du projet, depuis l'étude des données sources jusqu'à la création des visualisations finales, en passant par la modélisation, le choix des indicateurs clés (KPIs) et le développement des processus ETL.

2 Étapes du projet

2.1 Analyse des sources de données

La première étape du projet a consisté à explorer les données mises à disposition, fournies sous forme de fichiers plats (CSV/Excel). Ces fichiers contiennent des informations de vente détaillées par transaction.

Les attributs clés de ces fichiers incluent :

- Code produit et libellé produit
- Quantité vendue
- Montant total
- Client et segment client
- Région et point de vente
- Canal de distribution

Voici quelques aperçus des fichiers sources.

Voici un extrait du fichier de codification des clients :

SERVICE	CODE	STATION	GOUVERNORAT
RESEAU	150720	ARIANA	ARIANA
RESEAU	152510	MINIHLA	ARIANA
RESEAU	153360	Rte de RAOUED	ARIANA
RESEAU	153650	C.B. Ammar	ARIANA
RESEAU	153990	ENNASSI ARIANA	ARIANA
RESEAU	154060	S.A. BOUKHTIOUA	ARIANA
RESEAU	155380	Rte de RAOUED	ARIANA
RESEAU	155500	S.A. BOUKHTIOUA	ARIANA
RESEAU	155850	MINIHLA 2	ARIANA
RESEAU	156020	Rte de RAOUED	ARIANA
RESEAU	154130	IDRHC	BEJA
RESEAU	154380	MEDIEZ EL BAB	BEJA
RESEAU	154400	SIDI SMAIL	BEJA
RESEAU	154530	SLUGUJIA	BEJA
RESEAU	154770	AMDOUJ	BEJA
RESEAU	154900	GUEBELLAT	BEJA
RESEAU	155240	OUCHTATA	BEJA
RESEAU	155250	GRIAT N. PK29	BEJA
RESEAU	155260	GRIAT S. PK29	BEJA
RESEAU	155300	BEJA	BEJA
RESEAU	155390	GRIAT N. PK29	BEJA
RESEAU	155400	GRIAT S. PK29	BEJA
RESEAU	155480	GRIAT N. PK29	BEJA
RESEAU	155530	GRIAT S. PK29	BEJA
RESEAU	155660	GRIAT N. PK29	BEJA
RESEAU	155670	GRIAT S. PK29	BEJA
RESEAU	155780	SLUGUJIA	BEJA

FIGURE 1 – Fichier Excel — Codification des clients

Voici le fichier contenant les informations sur les stations de distribution :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	CODIVR	CODLOC	LIBLOC	A definir													
2	1	1	TUNIS	EL ATTAR													
3	1	10	TUNIS	LA MARSA													
4	1	11	TUNIS	SEDJOURNI													
5	1	12	TUNIS	LA GOULETTE													
6	1	13	TUNIS	TUNIS CARTHAGE													
7	1	20	TUNIS	A definir													
8	1	21	TUNIS	CARTHAGE													
9	1	34	TUNIS	SIDI BOUSAID													
10	1	40	TUNIS	BAB SAADOUN													
11	1	41	TUNIS	BAB ALIOUA													
12	1	42	TUNIS	TUNIS VILLE													
13	1	43	TUNIS	LA KASSRA													
14	1	44	TUNIS	SAIDA													
15	1	45	TUNIS	EL GUARDIA													
16	1	46	TUNIS	KAGARRA													
17	1	47	TUNIS	MONTLEURY													
18	1	48	TUNIS	DUBOSVILLE													
19	1	49	TUNIS	JEBEL JELLOUD													
20	1	50	TUNIS	SIDI FATHALLAH													
21	1	51	TUNIS	SAINT GOBIN ZA													
22	1	52	TUNIS	JEBEL LAHMAN													
23	1	55	TUNIS	MAJELLE VILLE													
24	1	56	TUNIS	BELVEDERE													
25	1	57	TUNIS	CITE ELKADRA													
26	1	58	TUNIS	EL MENSAH													
27	1	59	TUNIS	A definir													
28	1	60	TUNIS	KHEIREDDINE													
29	1	61	TUNIS	LA CAGRA													
30	1	62	TUNIS	BERGES DU LAC													
31	1	63	TUNIS	AN MACRUM													

FIGURE 2 – Fichier Excel — Informations sur les stations

Voici le fichier regroupant les données de vente par transaction :

Account

Project_SID_MI_DSSD.pdf

Configuration client/ux

Statistiques

vente1015_2016new/ux

Mediate & Niveau maintenance

Menu

Accueil

Insertion

Mise en page

Formules

Données

Revision

Affichage

Outils

Boîte à outils intelligente

Portager

Format Printer

Color & Font

B

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

Plage horaire

Dans la soirée

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

R

FIGURE 3 – Fichier Excel — Détail des ventes

2.2 Préparation et transformation des données

Afin de garantir la qualité des données, une phase de nettoyage a été réalisée à l'aide d'un script Python. Cette étape correspond à la zone de *staging* dans un processus décisionnel.

Les principales opérations appliquées sont :

- Suppression des lignes et colonnes entièrement vides
- Élimination des doublons
- Standardisation des noms de colonnes (minuscules, sans espaces)
- Suppression manuelle des colonnes non pertinentes

Voici un aperçu du script Python utilisé pour ces transformations :

```

transformation.py X
C:\Users\Lenovo> OneDrive > Bureau > projetSID > transformation.py > ...
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns # type: ignore
4
5 # Lecture des Feuilles Excel
6 vente_2015 = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/vente2015_2016new.xlsx", sheet_name="2015")
7 vente_2016 = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/vente2015_2016new.xlsx", sheet_name="2016")
8 vente_SQL_statement = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/vente2015_2016new.xlsx", sheet_name="SQL Statement")
9
10 stations_categorie = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/Stations.xlsx", sheet_name="categorie")
11 stations_localite = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/Stations.xlsx", sheet_name="localite_gouvernorat")
12
13 clients_LISTESTATION = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/codification_clients.xlsx", sheet_name="LISTESTATION")
14 clients_LISTECLIENTS = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/codification_clients.xlsx", sheet_name="LISTECLIENTS")
15 clients_LISTECLIENTSGAZ = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/codification_clients.xlsx", sheet_name="LISTECLIENTSGAZ")
16 clients_LISTECLIENTSJET = pd.read_excel("C:/Users/Lenovo/OneDrive/Bureau/projetSID/data/codification_clients.xlsx", sheet_name="LISTECLIENTSJET")
17
18 dfs_raw = {
19     "Vente 2015": vente_2015,
20     "Vente 2016": vente_2016,
21     "Stations Categorie": stations_categorie,
22     "Stations localite": stations_localite,
23     "Clients Station": clients_LISTESTATION,
24     "Clients": clients_LISTECLIENTS,
25     "Clients Gaz": clients_LISTECLIENTSGAZ,
26     "Clients Jet": clients_LISTECLIENTSJET
27 }
28
29 print("\n----- VISUALISATION EN MODE STAGING (AVANT NETTOYAGE) -----")
30 for nom, df in dfs_raw.items():
31     print(f"\n * Aperçu de (nom)")
32     print("Shape:", df.shape)
33     print("Colonnes:", list(df.columns))

```

FIGURE 4 – Extrait du script Python — Partie 1

```

transformation.py X
C:\Users\Lenovo> OneDrive > Bureau > projetSID > transformation.py > ...
69 plt.suptitle("Aperçu des dimensions des tables avant nettoyage", fontsize=16)
70 plt.tight_layout(rect=[0, 0, 1, 0.95])
71 plt.show()
72
73 # Fonction de nettoyage
74 def nettoyer_dataframe(df):
75     df.columns = df.columns.str.strip().str.lower()
76     df = df.dropna(axis=1, how="all")
77     df = df.dropna(axis=0, how="all")
78     df.columns = df.columns.str.strip()
79     df = df.drop_duplicates()
80     return df
81
82 # Application du nettoyage
83 del vente_SQL_statement
84 clients_LISTECLIENTS_clean = clients_LISTECLIENTS.drop(columns=["GOUVERNORAT"])
85 vente_2015_clean = nettoyer_dataframe(vente_2015)
86 vente_2016_clean = nettoyer_dataframe(vente_2016)
87 stations_categorie_clean = nettoyer_dataframe(stations_categorie)
88 stations_localite_clean = nettoyer_dataframe(stations_localite)
89 clients_LISTESTATION_clean = nettoyer_dataframe(clients_LISTESTATION)
90 clients_LISTECLIENTS_clean = nettoyer_dataframe(clients_LISTECLIENTS)
91 clients_LISTECLIENTSGAZ_clean = nettoyer_dataframe(clients_LISTECLIENTSGAZ)
92 clients_LISTECLIENTSJET_clean = nettoyer_dataframe(clients_LISTECLIENTSJET)
93
94 # Affichage des dimensions
95 print("Vente 2015 shape:", vente_2015_clean.shape)
96 print("Vente 2016 shape:", vente_2016_clean.shape)
97 print("Stations categorie shape:", stations_categorie_clean.shape)
98 print("Stations localite shape:", stations_localite_clean.shape)
99 print("Clients station shape:", clients_LISTESTATION_clean.shape)
100 print("Clients Jet shape:", clients_LISTECLIENTSJET_clean.shape)

```

FIGURE 5 – Extrait du script Python — Partie 2

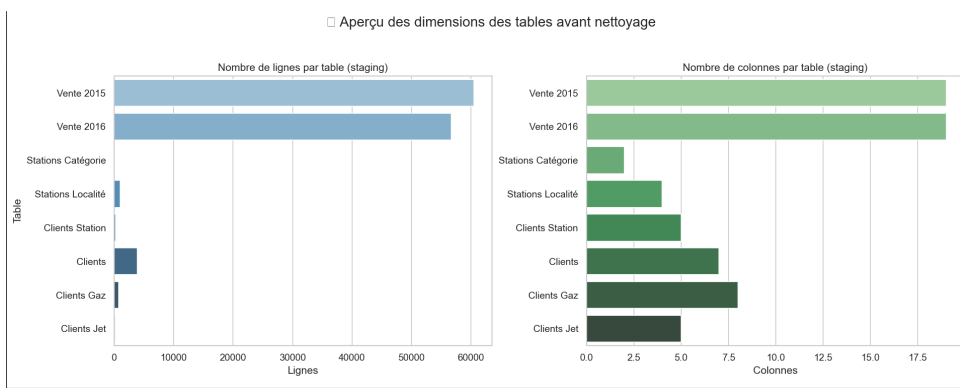


FIGURE 6 – Extrait du script Python — Partie 3

Voici les fichiers générés après nettoyage :

clients_clean	✓	22/04/2025 18:43	Feuille de calcul X...	220 Ko
clients_station_clean	✓	19/04/2025 19:47	Feuille de calcul X...	18 Ko
clients_gaz_clean	✓	19/04/2025 19:37	Feuille de calcul X...	53 Ko
clients_jet_clean	✓	19/04/2025 19:37	Feuille de calcul X...	13 Ko
station_categorie_clean	✓	19/04/2025 19:37	Feuille de calcul X...	7 Ko
station_localite_clean	✓	19/04/2025 19:37	Feuille de calcul X...	28 Ko
vente_2016_clean	✓	19/04/2025 19:37	Feuille de calcul X...	5513 Ko
vente_2015_clean	✓	19/04/2025 19:36	Feuille de calcul X...	5834 Ko

FIGURE 7 – Fichiers CSV générés après nettoyage

2.3 Implémentation ETL sous Talend

La phase suivante a consisté à concevoir les flux ETL (Extract, Transform, Load) avec l'outil Talend Open Studio. Ces flux permettent de charger les données nettoyées dans un schéma en étoile comportant une table de faits et plusieurs dimensions.

Voici un aperçu de l'implémentation des différentes dimensions :

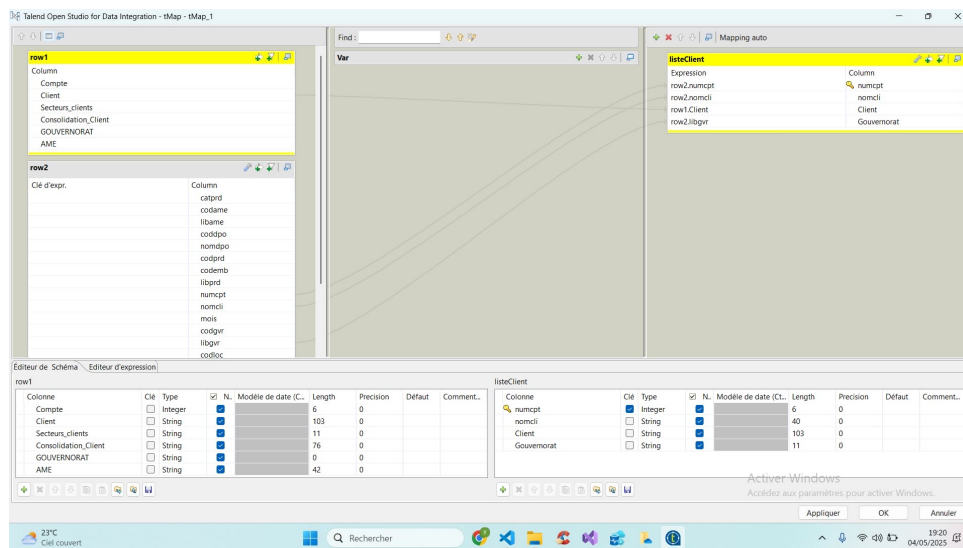


FIGURE 8 – Composant tMap — Transformation des données client

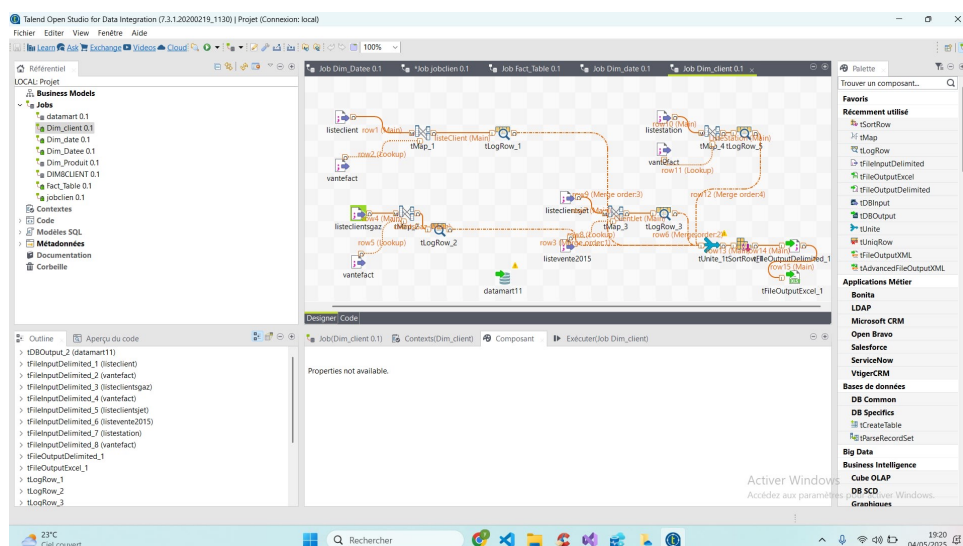


FIGURE 9 – Dimension Client — Structure cible

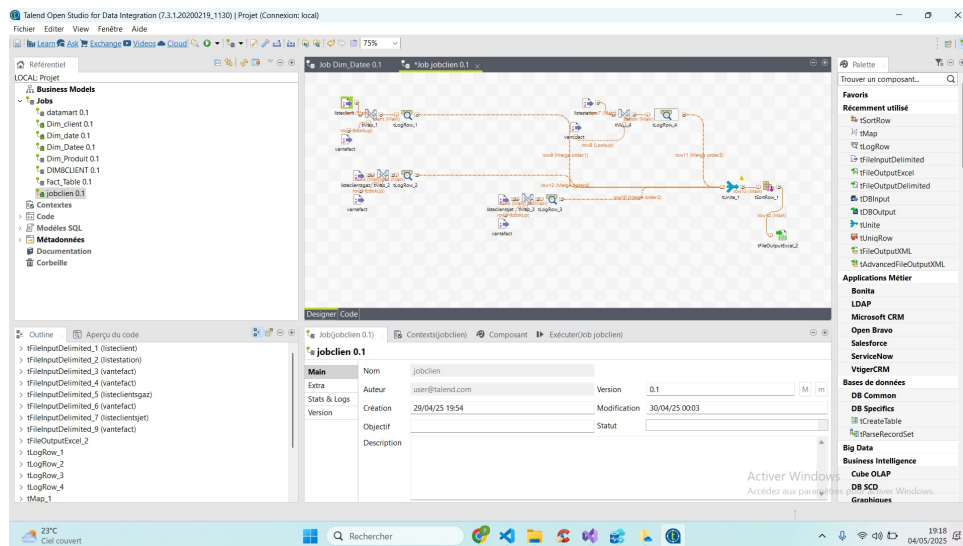


FIGURE 10 – Job Talend — Chargement de la dimension Date

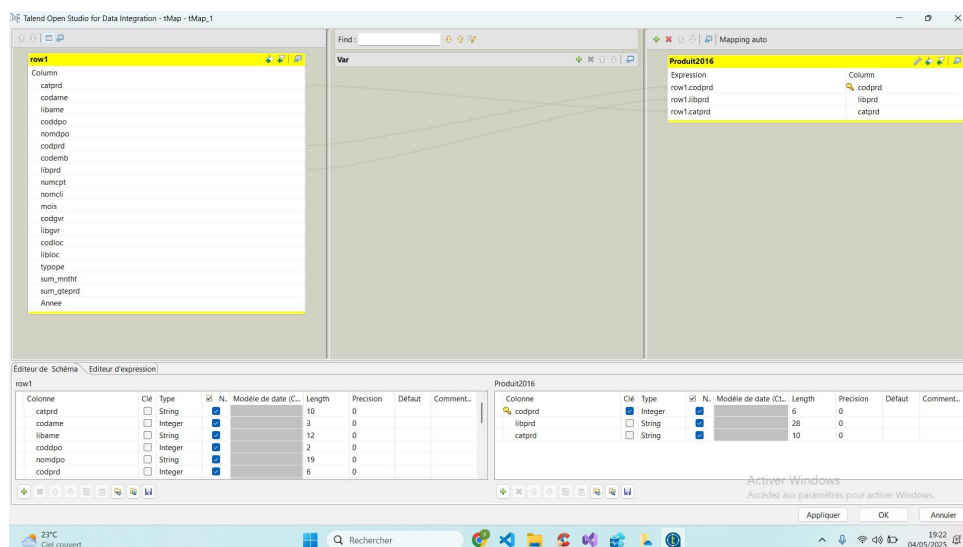


FIGURE 11 – Composant tMap — Transformation des données produit

Voici la structure de la table de faits *vente* :

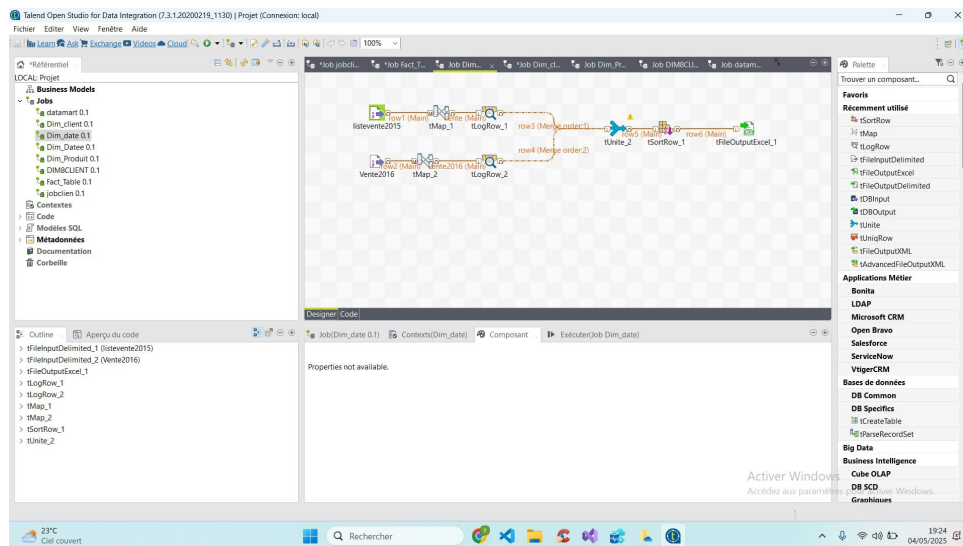


FIGURE 12 – Table de faits — Vente

Voici le flux Talend alimentant cette table de faits :

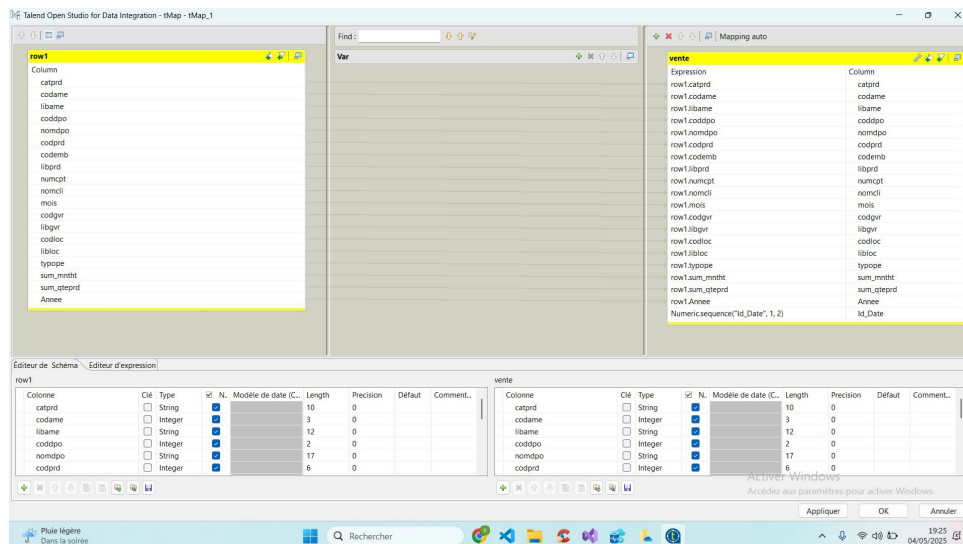


FIGURE 13 – Composant tMap — Alimentation de la table de faits vente

2.4 Identification des KPIs

Pour orienter la visualisation et l'analyse, plusieurs indicateurs clés de performance (KPIs) ont été identifiés en collaboration avec les parties prenantes métiers. Ces indicateurs sont les suivants :

- Chiffre d'affaires par région, station et produit
- Volume de vente par canal et type de client
- Top produits en termes de ventes
- Répartition géographique et temporelle des ventes
- Évolution mensuelle du chiffre d'affaires
- Nombre total de transactions
- Nombre de clients par année
- Nombre de ventes
- Nombre de clients par mois
- Nombre de clients par gouvernorat
- Répartition géographique des clients (carte)
- Proportion de clients par type de produit
- Montant HT total
- Quantité vendue totale

2.5 Visualisation avec Power BI

Une fois les données chargées et modélisées, nous avons conçu un tableau de bord interactif dans Power BI. Il est structuré en plusieurs pages thématiques :

- **Page d'Accueil** : Vue synthétique de la composition du dashboard
- **Vue d'ensemble** : Analyse globale des ventes par période, région, canal
- **Analyse des mesures** : Exploration par station, produit et client des mesures (montant hors taxe et de la quantité totale vendue)

Voici les différentes captures d'écran des pages Power BI créées :

Voici la page d'accueil du tableau de bord Power BI :

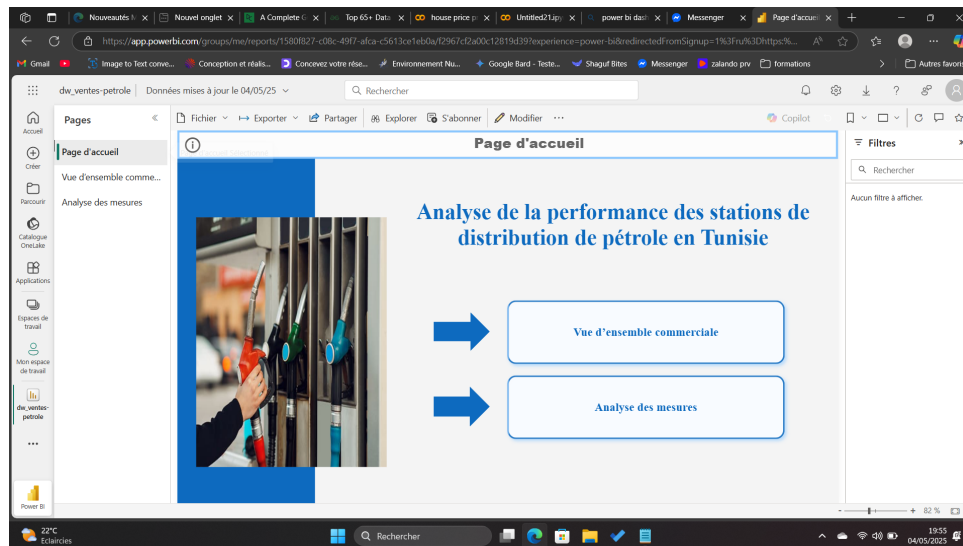


FIGURE 14 – Page d'accueil — Tableau de bord Power BI

Voici la vue d'ensemble :

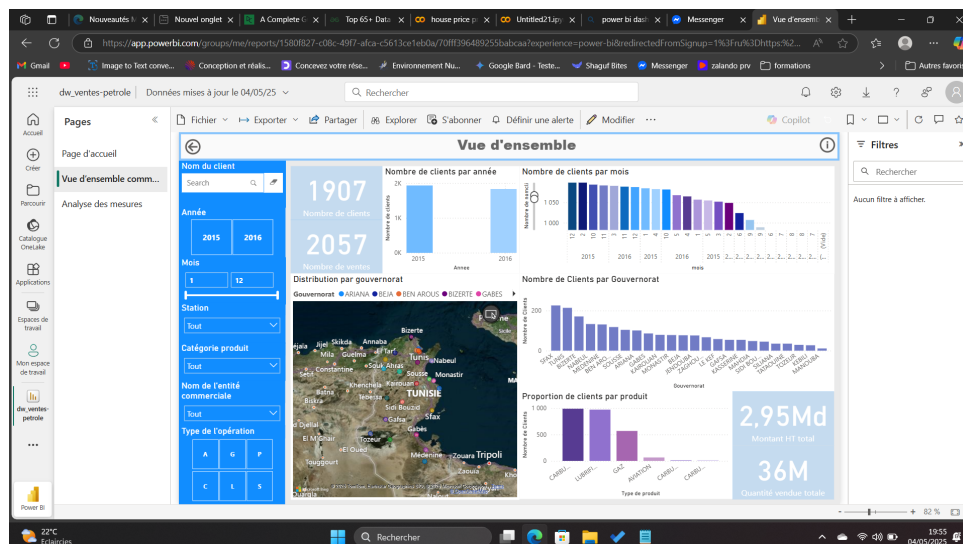


FIGURE 15 – Vue d'ensemble — Analyse globale

Voici une analyse détaillée :

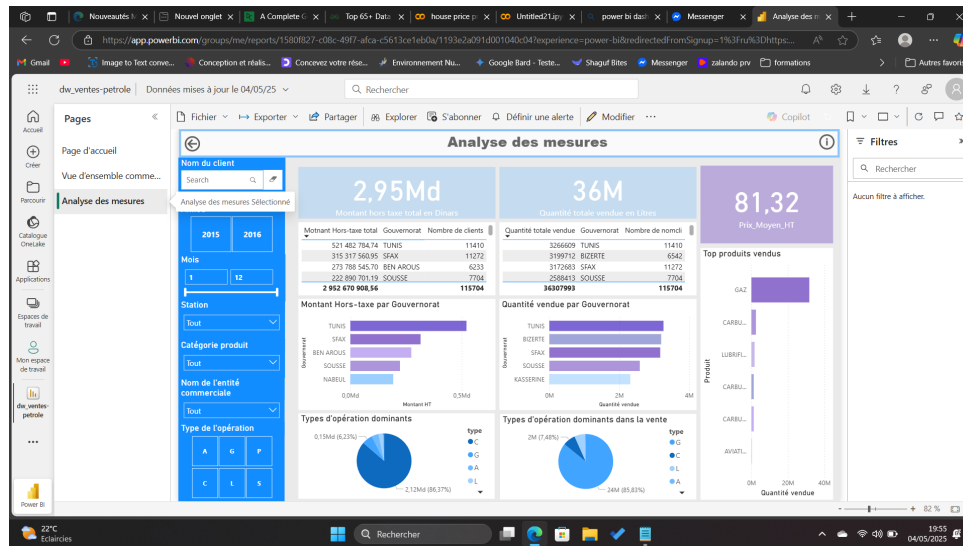


FIGURE 16 – Analyse des mesures

3 Conclusion

Ce projet a permis la mise en œuvre d'une solution complète de Business Intelligence dédiée au secteur pétrolier. Grâce à une architecture décisionnelle bien pensée, l'intégration automatisée via Talend et des visualisations pertinentes sous Power BI, l'entreprise peut désormais suivre et analyser efficacement la performance de ses stations de distribution.

Cette démarche permet non seulement de centraliser les données mais aussi d'améliorer la prise de décision à tous les niveaux de l'organisation. Les indicateurs dynamiques offrent une meilleure compréhension de l'activité commerciale et facilitent le suivi des objectifs.