**FUNDAÇÃO ESCOLA DE COMÉRCIO ÁLVARES PENTEADO**
**FECAP**
**ANÁLISE E DESENVOLVIMENTO DE SISTEMAS**

**ALEXANDRA CHRISTINE SILVA RAIMUNDO - 24026156**

**CARLOS AUGUSTO SANTOS DE ALMEIDA - 20010535**

**HEBERT DOS REIS ESTEVES - 24026079**

**JOSÉ BENTO ALMEIDA GAMA - 24026127**

# CANNOLI INTELLIGENCE

## PREPARAR OS DADOS

**São Paulo - SP**
**2025**

# SUMÁRIO

# 01. INTRODUÇÃO

Este relatório apresenta a segunda etapa do projeto Cannoli Intelligence, dedicada à preparação, enriquecimento e carga dos dados no banco de dados MySQL. O principal objetivo desta fase foi transformar as bases originais — inicialmente brutas, heterogêneas e inconsistentes — em conjuntos de dados limpos, padronizados e integrados, prontos para alimentar os dashboards analíticos e os módulos de inteligência da aplicação final.

Durante essa etapa, foram realizadas ações fundamentais para a consolidação do ambiente de dados, incluindo a criação de novas bases complementares, a integração entre diferentes fontes, o tratamento de valores nulos e inconsistentes, a padronização textual, e o enriquecimento automatizado com apoio de inteligência artificial. Por fim, os dados tratados foram importados e estruturados no MySQL, garantindo integridade relacional, consistência e compatibilidade com as futuras camadas de visualização e análise do sistema.

# 02. PREPARAÇÃO E ENRIQUECIMENTO DOS DADOS

## 02.1. Criação da base de estabelecimentos

Para viabilizar a autenticação individual de cada estabelecimento na aplicação, foi criada uma base denominada Estabelecimentos.csv. Essa base contém os campos establishment_id, store_name, segmento, cidade e UF. Foram gerados 12 estabelecimentos distintos, representando diferentes tipos de negócios e localidades. Essa estrutura serviu como referência principal para relacionar os dados de clientes, pedidos, campanhas e filas de mensagens com seus respectivos estabelecimentos.

| establishment_id | storeName | segmento | cidade | uf |
|---|---|---|---|---|
| EST001 | Pizzaria Bella Massa | Alimentacao | Sao Paulo | SP |
| EST002 | Cafe do Centro | Bebidas | Campinas | SP |
| EST003 | Hamburgueria Prime | Alimentacao | Rio de Janeiro | RJ |
| EST004 | Cantina Italia | Alimentacao | Curitiba | PR |
| EST005 | Doceria Sabor e Arte | Confeitaria | Belo Horizonte | MG |
| EST006 | Restaurante Vila Gourmet | Alimentacao | Recife | PE |
| EST007 | Churrascaria Gaucha | Alimentacao | Porto Alegre | RS |
| EST008 | Padaria Delicia | Panificacao | Salvador | BA |
| EST009 | Sorveteria Tropical | Confeitaria | Florianopolis | SC |
| EST010 | Lanchonete Express | Fast Food | Brasilia | DF |
| EST011 | Veg e Fit Saladas | Saudavel | Sao Paulo | SP |
| EST012 | Temakeria Nippon | Sushi | Santos | SP |

Figura 01 – Criação da base de estabelecimentos.

## 02.2. Integração dos estabelecimentos nas bases principais

Para que cada registro pudesse ser associado a um estabelecimento real, foi necessário realizar o mapeamento do campo establishment_id em todas as bases do projeto — Customer, Order, Campaign e CampaignQueue.

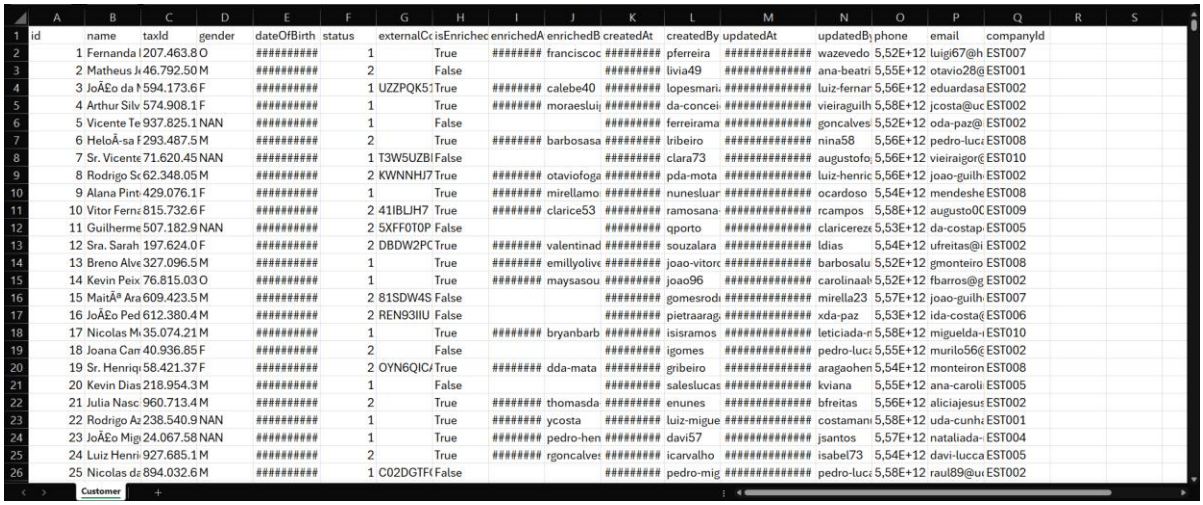
Esse processo foi conduzido com o apoio de IA generativa (ChatGPT), utilizada para automatizar o cruzamento e a distribuição coerente dos dados entre as 12 lojas criadas na base Estabelecimentos.csv.

A inteligência artificial analisou padrões de distribuição e criou a relação entre cada cliente, pedido, campanha e fila de mensagens com um storeId ou companyId específico.

Durante a execução, foram aplicadas regras de consistência para garantir que todos os relacionamentos entre tabelas fossem válidos, mantendo integridade

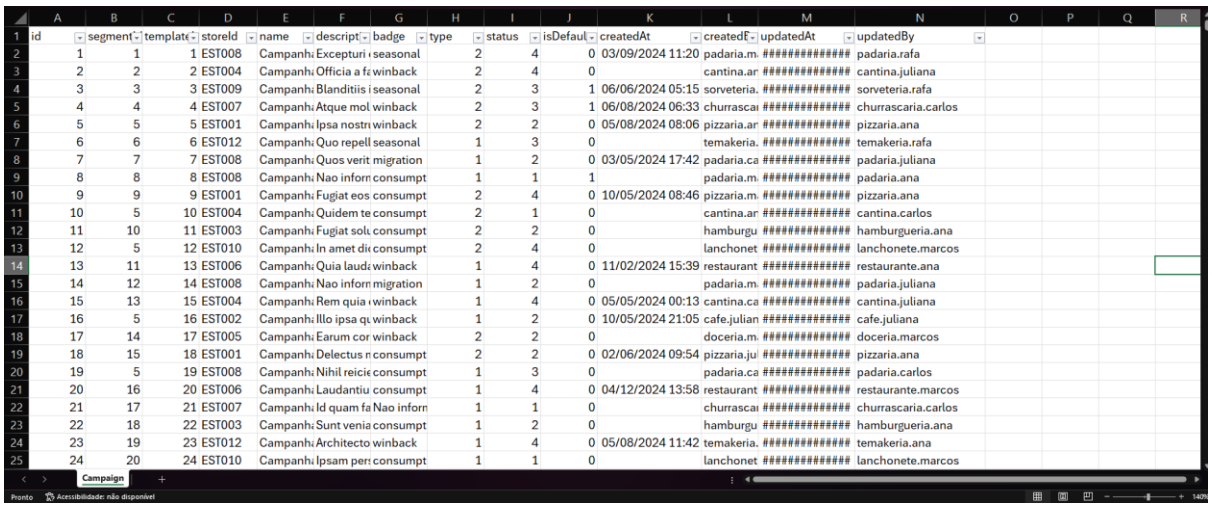
referencial entre os dados (por exemplo, cada campanha pertencer a uma loja existente e cada pedido estar vinculado a um cliente e a um estabelecimento real).

Com isso, foi possível preparar as bases para o funcionamento da autenticação no sistema web, permitindo que cada usuário (loja) visualize apenas seus próprios dados, enquanto o perfil de administrador mantém acesso consolidado a todos.



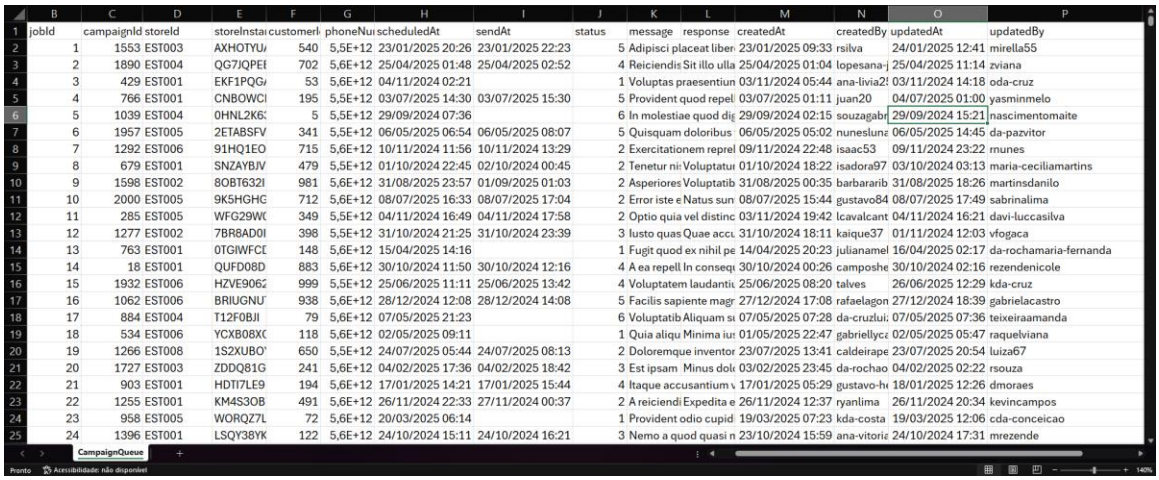
| id | name | taxId | gender | dateOfBirth | status | externalCo | isEnriched | enrichedA | enrichedB | createdAt | createdBy | updatedAt | updatedBy | phone | email | companyId |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Fernanda I | 207.463.8 | O | ########## | 1 | | True | ######## | franciscoc | ######### | pferreira | ########## | wazevedo | 5,52E+12 | luigi67@h | EST007 |
| 2 | Matheus Ja | 46.792.50 | M | ########## | 2 | | False | | | ######### | livia49 | ########## | ana-beatri | 5,55E+12 | otavio28@ | EST001 |
| 3 | JoÃ£o da M | 594.173.6 | F | ########## | 1 | UZZPQK51 | True | ######## | calebe40 | ######### | lopesmari | ########## | luiz-fernar | 5,56E+12 | eduardasa | EST002 |
| 4 | Arthur Silv | 574.908.1 | F | ########## | 1 | | True | ######## | moraesluiz | ######### | da-concei | ########## | vieiraguilh | 5,58E+12 | jcosta@uc | EST002 |
| 5 | Vicente Te | 937.825.1 | NAN | ########## | 1 | | False | | | ######### | ferreirama | ########## | goncalves | 5,52E+12 | oda-paz@ | EST002 |
| 6 | HeloÃ-sa F | 293.487.5 | M | ########## | 2 | | True | ######## | barbosasa | ######### | lribeiro | ########## | nina58 | 5,56E+12 | pedro-luc | EST008 |
| 7 | Sr. Vicente | 71.620.45 | NAN | ########## | 1 | T3W5UZBI | False | | | ######### | clara73 | ########## | augustofo | 5,56E+12 | vieiraigor | EST010 |
| 8 | Rodrigo Sc | 62.348.05 | M | ########## | 2 | KWNNHJ7 | True | ######## | otaviofoga | ######### | pda-mota | ########## | luiz-henric | 5,56E+12 | joao-guilh | EST002 |
| 9 | Alana Pint | 429.076.1 | F | ########## | 1 | | True | ######## | mirellamo | ######### | nunesluar | ########## | ocardoso | 5,54E+12 | mendeshe | EST008 |
| 10 | Vitor Ferna | 815.732.6 | F | ########## | 2 | 41IBLJH7 | True | ######## | clarice53 | ######### | ramosana | ########## | rcampos | 5,58E+12 | augusto0( | EST009 |
| 11 | Guilherme | 507.182.9 | NAN | ########## | 2 | 5XFF0T0P | False | | | ######### | qporto | ########## | claricereze | 5,53E+12 | da-costap | EST005 |
| 12 | Sra. Sarah | 197.624.0 | F | ########## | 2 | DBDW2PC | True | ######## | valentinad | ######### | souzalara | ########## | ldias | 5,54E+12 | ufreitas@i | EST002 |
| 13 | Breno Alve | 327.096.5 | M | ########## | 1 | | True | ######## | emillyolive | ######### | joao-vitorc | ########## | barbosalu | 5,52E+12 | gmonteiro | EST008 |
| 14 | Kevin Peix | 76.815.03 | O | ########## | 1 | | True | ######## | maysasou | ######### | joao96 | ########## | carolinaab | 5,52E+12 | fbarros@g | EST002 |
| 15 | MaitÃª Ara | 609.423.5 | M | ########## | 2 | 81SDW4S | False | | | ######### | gomesrodr | ########## | mirella23 | 5,57E+12 | joao-guilh | EST007 |
| 16 | JoÃ£o Ped | 612.380.4 | M | ########## | 2 | REN93IIU | False | | | ######### | pietraarag | ########## | xda-paz | 5,53E+12 | ida-costa( | EST006 |
| 17 | Nicolas Mi | 35.074.21 | M | ########## | 1 | | True | ######## | bryanbarb | ######### | isisramos | ########## | leticiada-r | 5,58E+12 | miguelda-r | EST010 |
| 18 | Joana Cam | 40.936.85 | F | ########## | 2 | | False | | | ######### | igomes | ########## | pedro-luc | 5,55E+12 | murilo56@ | EST002 |
| 19 | Sr. Henriq | 58.421.37 | F | ########## | 2 | OYN6QIC | True | ######## | dda-mata | ######### | gribeiro | ########## | aragaohen | 5,54E+12 | monteiron | EST008 |
| 20 | Kevin Dias | 218.954.3 | M | ########## | 1 | | False | | | ######### | saleslucas | ########## | kviana | 5,55E+12 | ana-caroli | EST005 |
| 21 | Julia Nasc | 960.713.4 | M | ########## | 2 | | True | ######## | thomasda | ######### | enunes | ########## | bfreitas | 5,56E+12 | aliciajesus | EST002 |
| 22 | Rodrigo Az | 238.540.9 | NAN | ########## | 1 | | True | ######## | ycosta | ######### | luiz-migue | ########## | costamani | 5,58E+12 | uda-cunh | EST001 |
| 23 | JoÃ£o Mig | 24.067.58 | NAN | ########## | 1 | | True | ######## | pedro-hen | ######### | davi57 | ########## | jsantos | 5,57E+12 | nataliada- | EST004 |
| 24 | Luiz Henri | 927.685.1 | M | ########## | 2 | | True | ######## | rgoncalves | ######### | icarvalho | ########## | isabel73 | 5,54E+12 | davi-lucca | EST005 |
| 25 | Nicolas da | 894.032.6 | M | ########## | 1 | C02DGTF( | False | | | ######### | pedro-mig | ########## | pedro-luc | 5,58E+12 | raul89@u | EST002 |

Figura 02 – Base Customer integrada aos estabelecimentos.



| id | segment | template | storeId | name | descript | badge | type | status | isDefaul | createdAt | createdB | updatedAt | updatedBy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | EST008 | Campanh | Excepturi | seasonal | 2 | 4 | 0 | 03/09/2024 11:20 | padaria.m | ########## | padaria.rafa |
| 2 | 2 | 2 | EST004 | Campanh | Officia a f | winback | 2 | 4 | 0 | | cantina.ar | ########## | cantina.juliana |
| 3 | 3 | 3 | EST009 | Campanh | Blanditiis | seasonal | 2 | 3 | 1 | 06/06/2024 05:15 | sorveteria. | ########## | sorveteria.rafa |
| 4 | 4 | 4 | EST007 | Campanh | Atque mol | winback | 2 | 3 | 1 | 06/08/2024 06:33 | churrasca | ########## | churrascaria.carlos |
| 5 | 5 | 5 | EST001 | Campanh | Ipsa nostr | winback | 2 | 2 | 0 | 05/08/2024 08:06 | pizzaria.ar | ########## | pizzaria.ana |
| 6 | 6 | 6 | EST012 | Campanh | Quo repell | seasonal | 1 | 3 | 0 | | temakeria. | ########## | temakeria.rafa |
| 7 | 7 | 7 | EST008 | Campanh | Quos verit | migration | 1 | 2 | 0 | 03/05/2024 17:42 | padaria.ca | ########## | padaria.juliana |
| 8 | 8 | 8 | EST008 | Campanh | Nao inforr | consumpt | 1 | 1 | 1 | | padaria.m | ########## | padaria.ana |
| 9 | 9 | 9 | EST001 | Campanh | Fugiat eos | consumpt | 2 | 4 | 0 | 10/05/2024 08:46 | pizzaria.m | ########## | pizzaria.ana |
| 10 | 5 | 10 | EST004 | Campanh | Quidem te | consumpt | 2 | 1 | 0 | | cantina.ar | ########## | cantina.carlos |
| 11 | 10 | 11 | EST003 | Campanh | Fugiat sol | consumpt | 2 | 2 | 0 | | hamburgu | ########## | hamburgueria.ana |
| 12 | 5 | 12 | EST010 | Campanh | In amet di | consumpt | 2 | 4 | 0 | | lanchonet | ########## | lanchonete.marcos |
| 13 | 11 | 13 | EST006 | Campanh | Quia laud | winback | 1 | 4 | 0 | 11/02/2024 15:39 | restaurant | ########## | restaurante.ana |
| 14 | 12 | 14 | EST008 | Campanh | Nao inforr | migration | 1 | 2 | 0 | | padaria.m | ########## | padaria.juliana |
| 15 | 13 | 15 | EST004 | Campanh | Rem quia | winback | 1 | 4 | 0 | 05/05/2024 00:13 | cantina.ca | ########## | cantina.juliana |
| 16 | 5 | 16 | EST002 | Campanh | Illo ipsa q | winback | 1 | 2 | 0 | 10/05/2024 21:05 | cafe.julian | ########## | cafe.juliana |
| 17 | 14 | 17 | EST005 | Campanh | Earum cor | winback | 2 | 2 | 0 | | doceria.m | ########## | doceria.marcos |
| 18 | 15 | 18 | EST001 | Campanh | Delectus r | consumpt | 2 | 2 | 0 | 02/06/2024 09:54 | pizzaria.ju | ########## | pizzaria.ana |
| 19 | 5 | 19 | EST008 | Campanh | Nihil reici | consumpt | 1 | 3 | 0 | | padaria.ca | ########## | padaria.carlos |
| 20 | 16 | 20 | EST006 | Campanh | Laudantiu | consumpt | 1 | 4 | 0 | 04/12/2024 13:58 | restaurant | ########## | restaurante.marcos |
| 21 | 17 | 21 | EST007 | Campanh | Id quam fa | Nao inforr | 1 | 1 | 0 | | churrasca | ########## | churrascaria.carlos |
| 22 | 18 | 22 | EST003 | Campanh | Sunt venia | consumpt | 1 | 2 | 0 | | hamburgu | ########## | hamburgueria.ana |
| 23 | 19 | 23 | EST012 | Campanh | Architecto | winback | 1 | 4 | 0 | 05/08/2024 11:42 | temakeria. | ########## | temakeria.ana |
| 24 | 20 | 24 | EST010 | Campanh | Ipsam per | consumpt | 1 | 1 | 0 | | lanchonet | ########## | lanchonete.marcos |

Figura 03 – Base Campaign integrada aos estabelecimentos



| jobId | campaignId | storeId | storeInstar | customerI | phoneNu | scheduledAt | sendAt | status | message | response | createdAt | createdBy | updatedAt | updatedBy |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1553 | EST003 | AXHOTYU | 540 | 5,5E+12 | 23/01/2025 20:26 | 23/01/2025 22:23 | 5 | Adipisci | placeat liber | 23/01/2025 09:33 | rsilva | 24/01/2025 12:41 | mirella55 |
| 2 | 1890 | EST004 | QG7JQPEI | 702 | 5,6E+12 | 25/04/2025 01:48 | 25/04/2025 02:52 | 4 | Reiciendis | Sit illo ulla | 25/04/2025 01:04 | lopesana-j | 25/04/2025 11:14 | zviana |
| 3 | 429 | EST001 | EKF1PQGA | 53 | 5,6E+12 | 04/11/2024 02:21 | | 1 | Voluptas | praesentiun | 03/11/2024 05:44 | ana-livia2! | 03/11/2024 14:18 | oda-cruz |
| 4 | 766 | EST001 | CNBOWCI | 195 | 5,5E+12 | 03/07/2025 14:30 | 03/07/2025 15:30 | 5 | Provident | quod repell | 03/07/2025 01:11 | juan20 | 04/07/2025 01:00 | yasminmelo |
| 5 | 1039 | EST004 | 0HNL2K6: | 5 | 5,5E+12 | 29/09/2024 07:36 | | 6 | In molestiae | quod dig | 29/09/2024 02:15 | souzagabr | 29/09/2024 15:21 | nascimentomaite |
| 6 | 1957 | EST005 | 2ETABSFV | 341 | 5,5E+12 | 06/05/2025 06:54 | 06/05/2025 08:07 | 5 | Quisquam | doloribus | 06/05/2025 05:02 | nunesluna | 06/05/2025 14:45 | da-pazvitor |
| 7 | 1292 | EST006 | 91HQ1EO | 715 | 5,6E+12 | 10/11/2024 11:56 | 10/11/2024 13:29 | 2 | Exercitationem | reprel | 09/11/2024 22:48 | isaac53 | 09/11/2024 23:22 | rnunes |
| 8 | 679 | EST001 | SNZAYBJV | 479 | 5,5E+12 | 01/10/2024 22:45 | 02/10/2024 00:45 | 2 | Tenetur ni | Voluptatur | 01/10/2024 18:22 | isadora97 | 03/10/2024 03:13 | maria-ceciliamartins |
| 9 | 1598 | EST002 | 8OBT632I | 981 | 5,6E+12 | 31/08/2025 23:57 | 01/09/2025 01:03 | 2 | Asperiores | Voluptatib | 31/08/2025 00:35 | barbararib | 31/08/2025 18:26 | martinsdanilo |
| 10 | 2000 | EST005 | 9K5HGHC | 712 | 5,6E+12 | 08/07/2025 16:33 | 08/07/2025 17:04 | 2 | Error iste e | Natus sun | 08/07/2025 15:44 | gustavo84 | 08/07/2025 17:49 | sabrinalima |
| 11 | 285 | EST005 | WFG29WC | 349 | 5,5E+12 | 04/11/2024 16:49 | 04/11/2024 17:58 | 2 | Optio quia | vel distinc | 03/11/2024 19:42 | lcavalcant | 04/11/2024 16:21 | davi-luccasilva |
| 12 | 1277 | EST002 | 7BR8AD0I | 398 | 5,5E+12 | 31/10/2024 21:25 | 31/10/2024 23:39 | 3 | Iusto quas | Quae accu | 31/10/2024 18:11 | kaique37 | 01/11/2024 12:03 | vfogaca |
| 13 | 763 | EST001 | 0TGIWFCE | 148 | 5,6E+12 | 15/04/2025 14:16 | | 1 | Fugit quod | ex nihil pe | 14/04/2025 20:23 | julianamel | 16/04/2025 02:17 | da-rochamaria-fernanda |
| 14 | 18 | EST001 | QUFD08D | 883 | 5,6E+12 | 30/10/2024 11:50 | 30/10/2024 12:16 | 4 | A ea repell | In consequ | 30/10/2024 00:26 | camposhe | 30/10/2024 02:16 | rezendenicole |
| 15 | 1932 | EST006 | HZVE906Z | 999 | 5,5E+12 | 25/06/2025 11:11 | 25/06/2025 13:42 | 4 | Voluptatem | laudantiu | 25/06/2025 08:20 | talves | 26/06/2025 12:29 | kda-cruz |
| 16 | 1062 | EST006 | BRIUGNU | 938 | 5,6E+12 | 28/12/2024 12:08 | 28/12/2024 14:08 | 5 | Facilis sapiente | magn | 27/12/2024 17:08 | rafaelagon | 27/12/2024 18:39 | gabrielacastro |
| 17 | 884 | EST004 | T12F0BJI | 79 | 5,6E+12 | 07/05/2025 21:23 | | 6 | Voluptatib | Aliquam si | 07/05/2025 07:28 | da-cruzlui | 07/05/2025 07:36 | teixeiraamanda |
| 18 | 534 | EST006 | YCXB08XC | 118 | 5,6E+12 | 02/05/2025 09:11 | | 1 | Quia aliqu | Minima iu | 01/05/2025 22:47 | gabriellyca | 02/05/2025 05:47 | raquelviana |
| 19 | 1266 | EST008 | 1S2XUBO | 650 | 5,5E+12 | 24/07/2025 05:44 | 24/07/2025 08:13 | 2 | Doloremque | inventor | 23/07/2025 13:41 | caldeirape | 23/07/2025 20:54 | luiza67 |
| 20 | 1727 | EST003 | ZDDQ81G | 241 | 5,6E+12 | 04/02/2025 17:36 | 04/02/2025 18:42 | 3 | Est ipsam | Minus dol | 03/02/2025 23:45 | da-rochao | 04/02/2025 02:22 | rsouza |
| 21 | 903 | EST001 | HDTI7LE9 | 194 | 5,6E+12 | 17/01/2025 14:21 | 17/01/2025 15:44 | 4 | Itaque accusantium | v | 17/01/2025 05:29 | gustavo-h | 18/01/2025 12:26 | dmoraes |
| 22 | 1255 | EST001 | KM4S3OB | 491 | 5,6E+12 | 26/11/2024 22:33 | 27/11/2024 00:37 | 2 | A reiciendi | Expedita e | 26/11/2024 12:37 | ryanlima | 26/11/2024 20:34 | kevincampos |
| 23 | 958 | EST005 | WORQZ7L | 72 | 5,6E+12 | 20/03/2025 06:14 | | 1 | Provident | odio cupid | 19/03/2025 07:23 | kda-costa | 19/03/2025 12:06 | cda-conceicao |
| 24 | 1396 | EST001 | LSQY38YK | 122 | 5,6E+12 | 24/10/2024 15:11 | 24/10/2024 16:21 | 3 | Nemo a quod | quasi n | 23/10/2024 15:59 | ana-vitoria | 24/10/2024 17:31 | mrezende |

Figura 04 – Base CampaignQueue integrada aos estabelecimentos

| id | companyId | containerI | createdAt | customer | displayId | engineId | engineNar | engineTyp | extraInfo | integrated | integratior | isTest | orderTimir | orderType | salesChan | scheduledAt | status | preparatio | takeOutTir | totalAmou | updatedAt | version |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | EST011 | AKVXT2FH | 08/01/2025 15:47 | 525 | O8CC98 | E3R037 | DirectOrdi | POS | Adipisci m | True | 2452 | False | IMMEDIAT | DELIVERY | ANOTAAI | | DISPATCH | 45 | 2131 | 90.91 | 09/01/2025 03:32 | v3.7.8 |
| 2 | EST004 | KKMY5TG5 | 03/11/2024 22:30 | 694 | HJ89VA | EJVOBV | DirectOrdi | APP | | False | 4688 | True | IMMEDIAT | DELIVERY | WHATSAPP | | CONCLUD | 33 | 374 | 99.69 | 04/11/2024 03:05 | v2.3.0 |
| 3 | EST004 | I87JLGTV | 09/09/2024 05:49 | 491 | XJJDH2 | UDX1ZF | KDSPro | POS | | False | 8501 | False | IMMEDIAT | INDOOR | EPADOCA | | CONCLUD | 22 | 247 | 45.97 | 09/09/2024 09:32 | v1.9.5 |
| 4 | EST004 | W3ZXOQH | 22/05/2025 06:22 | 846 | J8QDM3 | 30FLOC | KDSPro | WEB | Autem mo | True | 3560 | False | SCHEDULI | TAKEOUT | 99FOOD | 22/05/2025 14:22 | CANCELEI | 35 | 1766 | 104.31 | 22/05/2025 21:39 | v2.8.5 |
| 5 | EST006 | NA0C0FO: | 22/02/2025 01:23 | 77 | H4LV73 | QI3URN | KDSPro | POS | | True | 3196 | False | SCHEDULI | DELIVERY | 99FOOD | 24/02/2025 11:23 | DISPATCH | 20 | 1219 | 105.5 | 22/02/2025 13:53 | v2.1.3 |
| 6 | EST002 | NGNJYJIK | 17/12/2024 14:54 | 463 | WS9UVN | M1XWGY | KDSPro | APP | | True | 1550 | False | SCHEDULI | INDOOR | 99FOOD | 18/12/2024 07:54 | CONFIRMI | 68 | 1689 | 51.42 | 18/12/2024 09:12 | v2.8.3 |
| 7 | EST011 | F5Q4MYIX | 17/05/2025 16:51 | 625 | XJA3WX | 6ILJ42 | CannoliEn | POS | | False | 5252 | False | SCHEDULI | TAKEOUT | SITE | 18/05/2025 01:51 | PENDING | 43 | 2922 | 110.33 | 18/05/2025 12:12 | v2.8.8 |
| 8 | EST003 | QR7D6E5F | 02/06/2025 15:22 | 395 | 0H74IS | RCAXHC | CannoliEn | APP | Consectet | True | 7965 | False | SCHEDULI | INDOOR | SITE | 05/06/2025 13:22 | CANCELEI | 20 | 1797 | 21.41 | 03/06/2025 06:56 | v1.4.6 |
| 9 | EST001 | O97PBYD4 | 11/06/2025 02:18 | 633 | DJXMVM | TYGLF5 | CannoliEn | WEB | | False | 4098 | False | IMMEDIAT | INDOOR | SITE | | PENDING | 33 | 3243 | 28.36 | 11/06/2025 18:50 | v2.3.4 |
| 10 | EST004 | UDDVA3P( | 16/12/2024 09:27 | 350 | PYH5H2 | O5LTG6 | IfoodBridg | APP | | False | 6282 | False | SCHEDULI | DELIVERY | SITE | 17/12/2024 22:27 | PLACED | 51 | 1062 | 48.1 | 17/12/2024 01:56 | v1.2.1 |
| 11 | EST009 | QIZRMT0N | 11/10/2024 16:49 | 34 | HB62HL | N6GOQS | IfoodBridg | APP | | True | 9201 | False | SCHEDULI | INDOOR | EPADOCA | 13/10/2024 12:49 | DISPATCH | 85 | 3548 | 39.21 | 12/10/2024 03:13 | v2.4.4 |
| 12 | EST009 | J5UFVFCW | 15/04/2025 20:22 | 920 | GQ8X2M | PXTGU0 | DirectOrdi | POS | | False | 2944 | False | SCHEDULI | DELIVERY | EPADOCA | 16/04/2025 00:22 | PENDING | 32 | 2155 | 107.72 | 16/04/2025 10:18 | v1.9.9 |
| 13 | EST012 | 109S8B04 | 16/11/2024 07:53 | 518 | A9YZ5S | CZH36H | IfoodBridg | APP | | True | 1015 | False | SCHEDULI | INDOOR | ANOTAAI | 16/11/2024 21:53 | CONFIRMI | 30 | 820 | 14.15 | 16/11/2024 20:38 | v1.1.1 |
| 14 | EST006 | 952B5IKR | 14/01/2025 09:12 | 932 | O8UA4N | UTHO69 | CannoliEn | APP | Vitae eaq | False | 9517 | False | IMMEDIAT | DELIVERY | IFOOD | | PENDING | 48 | 3435 | 44.53 | 15/01/2025 04:58 | v3.3.6 |
| 15 | EST008 | BEHHW9B | 03/05/2025 03:19 | 19 | 2MPGL9 | U0U83T | KDSPro | WEB | | True | 9117 | False | IMMEDIAT | DELIVERY | 99FOOD | | DISPATCH | 40 | 2318 | 114.57 | 03/05/2025 13:26 | v3.5.5 |
| 16 | EST007 | CXHHDAS | 10/10/2024 13:12 | 365 | CO1RYW | YMCS3Y | DirectOrdi | WEB | | False | 1701 | False | IMMEDIAT | INDOOR | WHATSAPP | | PLACED | 82 | 1126 | 140.07 | 11/10/2024 08:49 | v1.9.0 |
| 17 | EST009 | 1FNTTUYS | 07/07/2025 16:58 | 34 | HCJQT2 | JAVYEJ | KDSPro | POS | Eius quo d | True | 2177 | False | SCHEDULI | TAKEOUT | IFOOD | 09/07/2025 10:58 | PLACED | 54 | 367 | 46.37 | 08/07/2025 11:33 | v2.0.0 |
| 18 | EST002 | KLL5ZNLN | 19/12/2024 00:30 | 39 | CJ7KZY | 0UQC4W | IfoodBridg | POS | | True | 6461 | False | IMMEDIAT | DELIVERY | DELIVERYVIP | | CONFIRMI | 18 | 2092 | 98.13 | 19/12/2024 11:03 | v2.1.1 |
| 19 | EST002 | E3OZKDZV | 12/08/2025 09:43 | 99 | AC2HZE | Q3MN9X | IfoodBridg | POS | Qui assum | True | 4627 | False | IMMEDIAT | TAKEOUT | WHATSAPP | | CANCELEI | 29 | 1650 | 37.15 | 13/08/2025 03:27 | v3.4.1 |
| 20 | EST002 | 4UXMWA1 | 05/12/2024 07:26 | 301 | 8FXPIF | 8RDZRQ | CannoliEn | POS | Nostrum i | True | 7106 | False | IMMEDIAT | TAKEOUT | IFOOD | | CONFIRMI | 63 | 2405 | 47.72 | 05/12/2024 17:12 | v3.1.5 |
| 21 | EST002 | KJS7YUCF | 11/01/2025 07:45 | 348 | 6WXQGL | DXZFLM | DirectOrdi | WEB | | True | 2822 | False | SCHEDULI | DELIVERY | IFOOD | 14/01/2025 06:45 | CONFIRMI | 51 | 3181 | 18.53 | 11/01/2025 18:33 | v2.2.5 |
| 22 | EST005 | XS5B4M6E | 26/09/2024 07:01 | 233 | ER7EB | LQWXW0 | DirectOrdi | WEB | Quaerat c | True | 4011 | False | SCHEDULI | INDOOR | ANOTAAI | 26/09/2024 15:01 | CANCELEI | 80 | 1846 | 33.69 | 26/09/2024 11:54 | v3.4.1 |
| 23 | EST001 | ABED5RBE | 23/11/2024 23:42 | 66 | MGBCS2 | R3AKEZ | IfoodBridg | POS | Voluptas q | False | 3265 | False | SCHEDULI | DELIVERY | SITE | | CONCLUD | 15 | 3370 | 40.94 | 24/11/2024 00:23 | v2.3.3 |
| 24 | EST003 | SZLCDNTC | 07/02/2025 13:25 | 115 | ICHEX7 | SAOEOG | KDSPro | APP | | True | 9347 | False | SCHEDULI | TAKEOUT | WHATSAPI | 09/02/2025 21:25 | CONFIRMI | 53 | 3094 | 70.15 | 08/02/2025 03:26 | v3.3.7 |
| 25 | EST004 | RJLBFRJS | 15/07/2025 07:07 | 71 | IJ7P3O | PBL120 | CannoliEn | POS | Ea numqu | False | 9823 | False | IMMEDIAT | TAKEOUT | IFOOD | | DISPATCH | 13 | 2214 | 16.35 | 15/07/2025 23:02 | v1.7.7 |
| 26 | EST001 | XX36DOLK | 27/11/2024 07:34 | 661 | 32J1WM | PQ30VZ | DirectOrdi | APP | | True | 1905 | False | IMMEDIAT | INDOOR | WHATSAPP | | CANCELEI | 52 | 1441 | 63.31 | 27/11/2024 22:56 | v1.0.0 |
| 27 | EST003 | BN26F4DC | 10/03/2025 02:30 | 73 | WML7O6 | 10HLSE | IfoodBridg | POS | Officia nec | True | 4028 | False | SCHEDULI | DELIVERY | IFOOD | 10/03/2025 16:30 | CONFIRMI | 14 | 74 | 24.69 | 10/03/2025 11:36 | v2.4.9 |
| 28 | EST001 | IGBU990M | 19/09/2024 01:43 | 416 | Z9HGNA | 0B6NWY | KDSPro | WEB | | False | 9517 | False | IMMEDIAT | INDOOR | 99FOOD | | CANCELEI | 49 | 578 | 30.37 | 19/09/2024 07:24 | v1.6.4 |
| 29 | EST003 | 8OIG4E8Q | 21/08/2025 00:09 | 595 | ENVGUA | SMCJ2I | IfoodBridg | POS | | True | 1537 | False | IMMEDIAT | INDOOR | SITE | | PENDING | 41 | 948 | 112.35 | 21/08/2025 02:46 | v1.0.7 |
| 30 | EST006 | OC19YQ5( | 07/11/2024 03:56 | 469 | HS5CD3 | BMYU8T | KDSPro | POS | | True | 3423 | False | SCHEDULI | DELIVERY | ANOTAAI | 07/11/2024 22:56 | DISPATCH | 47 | 241 | 38.35 | 07/11/2024 23:15 | v2.4.8 |
| 31 | EST007 | D0NWEVH | 14/09/2024 09:57 | 822 | QE5KN0 | W0BRNB | DirectOrdi | WEB | Officia eur | True | 7008 | False | IMMEDIAT | INDOOR | 99FOOD | | PENDING | 80 | 3124 | 21.58 | 14/09/2024 18:22 | v1.4.3 |

Figura 05 – Base Order integrada aos estabelecimentos

## 02.3. Geração de novos pedidos

Após o mapeamento, a base Order foi enriquecida por meio de IA generativa (ChatGPT), com o objetivo de aumentar o volume de registros e criar dados mais próximos de um cenário de negócio real.

A IA foi responsável por gerar 1.500 novos pedidos sintéticos, mantendo coerência nas variáveis originais — como salesChannel, status, totalAmount, createdAt e storeId.

O algoritmo foi configurado para simular tendências temporais e comportamentos comerciais típicos, criando pedidos mais recentes (de meses próximos a 2024 e 2025), o que enriquece as análises temporais nos dashboards.

Os registros foram distribuídos proporcionalmente entre os 12 estabelecimentos, garantindo diversidade e volume equilibrado.

Esse processo resultou em uma base mais robusta e representativa, ideal para visualização de métricas como ticket médio, receita por canal e volume mensal de vendas.

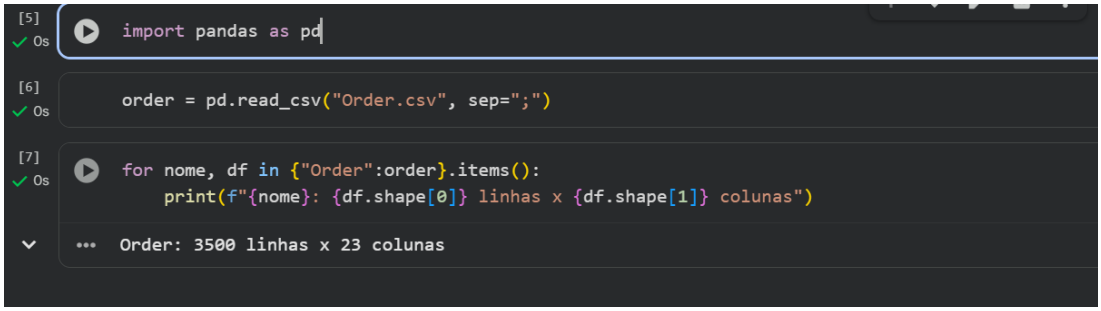



Figura 06 – Ampliação da base de pedidos.

## 03. TRATAMENTO E PREPARAÇÃO DOS DADOS

O processo de tratamento e preparação dos dados foi realizado no Google Colab, utilizando as bibliotecas pandas e numpy. O objetivo foi transformar as bases enriquecidas da plataforma Cannoli em conjuntos de dados limpos, consistentes e prontos para integração com o banco de dados MySQL e os dashboards interativos. O fluxo de trabalho foi estruturado em sete etapas principais: leitura dos arquivos, conversões de tipo, tratamento de valores nulos, padronização e criação de novos campos, inferência de gênero, validações finais e exportação dos arquivos tratados. Cada etapa foi cuidadosamente documentada e implementada para garantir a integridade e qualidade dos dados.

### 03.1. Leitura dos Arquivos CSV

Na primeira etapa, foram lidas as quatro bases principais do projeto: Order.csv, Campaign.csv, CampaignQueue.csv e Customer.csv. O carregamento foi feito utilizando o separador “;” e, após a leitura, foi realizada a verificação da estrutura de cada arquivo, exibindo o número de linhas e colunas no console. Essa verificação permitiu confirmar que todos os arquivos haviam sido corretamente importados após o enriquecimento com os identificadores dos estabelecimentos e a geração de novos pedidos.

A base Order, por exemplo, passou a conter 3.500 registros (2.000 originais mais 1.500 gerados artificialmente para simular novos pedidos recentes), totalizando 23 colunas. As demais bases — Campaign, CampaignQueue e Customer — mantiveram suas quantidades esperadas de linhas e colunas, já com os campos de relacionamento (companyId e storeId) devidamente populados e consistentes com a tabela de estabelecimentos.


Figura 07 – Leitura dos arquivos CSV e exibição das dimensões das bases (linhas e colunas).

### 03.2. Conversões Básicas de Tipo

A segunda etapa teve como foco garantir que todas as colunas possuíssem o tipo de dado adequado para análise.

Na base Order, as colunas de data (createdAt, updatedAt e scheduledAt) foram convertidas para o tipo datetime, e a coluna totalAmount — originalmente armazenada como texto com separadores inconsistentes — foi tratada e convertida em formato numérico em uma nova coluna chamada totalAmount_num. As variáveis textuais (salesChannel, engineName, engineType, orderType, orderTiming e version) foram padronizadas para texto simples, com remoção de espaços em branco e caracteres indesejados.

Na base Campaign, as colunas de data (createdAt, updatedAt, _createdAt) também foram convertidas para datetime, e os campos textuais (badge, description, type e storeId) foram padronizados e limpos.

Na base CampaignQueue, as datas (scheduledAt, sendAt, createdAt, updatedAt) foram tratadas, e as colunas textuais (message, response, storeId, storeInstanceId) foram normalizadas.

Já na base Customer, campos de data (dateOfBirth, createdAt, updatedAt, enrichedAt) foram convertidos para datetime, e campos de identificação (name, gender, externalCode, email, phone) foram tratados e convertidos para texto padronizado.

Essas conversões foram essenciais para viabilizar cálculos temporais, agrupamentos mensais e filtragens por status em etapas posteriores.


Figura 8 – Conversão de colunas de texto, datas e valores numéricos nas bases de Customer e Campaign.

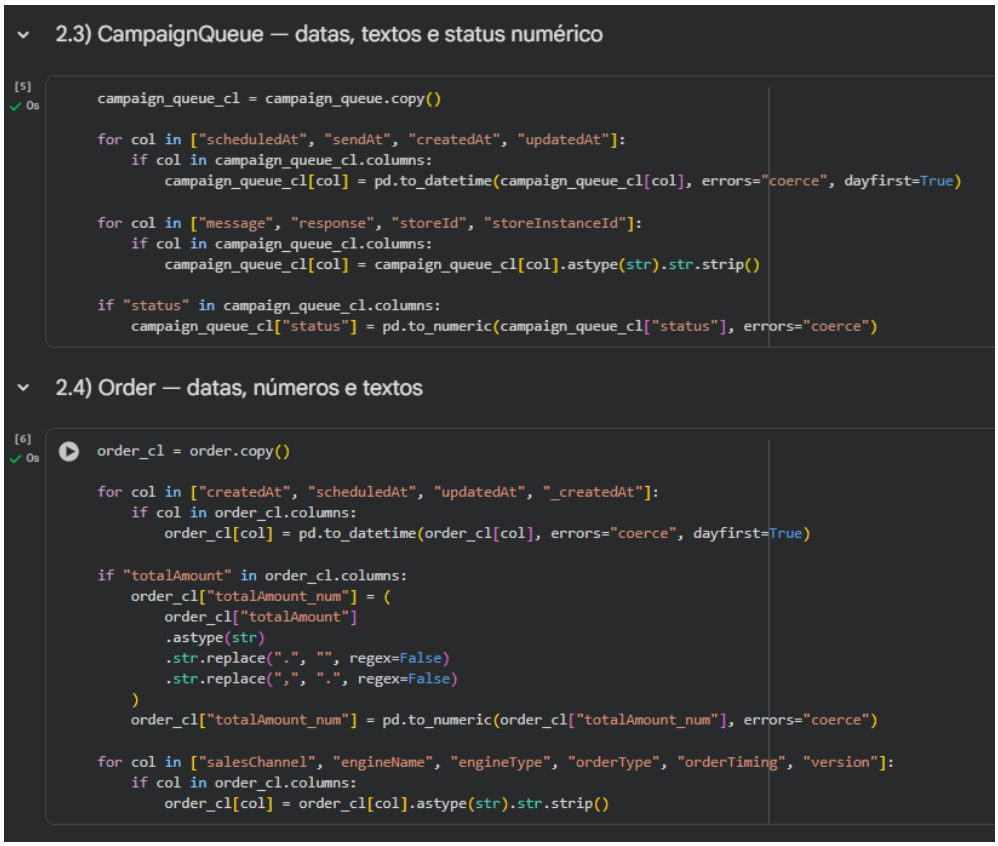



Figura 9 – Conversão de colunas de texto, datas e valores numéricos nas bases de CampaignQueue e Order.

### 03.3. Tratamento de Valores Nulos

Na terceira etapa, foram aplicadas regras específicas para o tratamento de valores nulos, garantindo coerência sem distorcer as informações.

Na base Order, as colunas extraInfo e scheduledAt permaneceram nulas quando apropriado, já que indicam pedidos sem metadados ou sem agendamento.

Na base Campaign, campos como badge, description, name e type tiveram valores ausentes substituídos por "Nao informado", evitando lacunas em relatórios.

Na base CampaignQueue, os campos message e response também receberam o preenchimento "Nao informado", enquanto sendAt permaneceu nulo para representar mensagens ainda não enviadas.

Por fim, na base Customer, o campo externalCode teve seus nulos substituídos por "Nao informado", enquanto os campos enrichedAt e enrichedBy permaneceram vazios, indicando clientes que ainda não foram enriquecidos.

Essa abordagem priorizou manter o significado dos nulos quando eles representam ausência legítima de informação (por exemplo, mensagens pendentes), ao mesmo tempo em que padronizou os demais valores para análise e visualização.


Figura 10 – Trecho de código mostrando a aplicação de regras de preenchimento e tratamento de valores nulos.

### 03.4. Padronização e Criação de Novos Campos

Após o tratamento de nulos, iniciou-se a etapa de padronização e derivação de novos campos.

Primeiramente, o campo storeId foi padronizado para letras maiúsculas em todas as tabelas que o continham, garantindo consistência com os identificadores dos estabelecimentos.

Em seguida, foram criadas colunas descritivas de status (status_desc) para traduzir códigos numéricos em rótulos textuais.

Na base Customer, os códigos 1 e 2 foram transformados em “Ativo” e “Inativo”; na base Campaign, em “Rascunho”, “Agendada”, “Ativa” e “Concluida”; e na base CampaignQueue, em “Pendente”, “Enviada”, “Entregue”, “Lida”, “Erro” e “Cancelada”.

Além disso, foram criadas colunas de mês (_mes) em todas as bases para permitir a análise temporal dos registros.

Na base Order, o mês foi extraído da coluna createdAt.

Na base Campaign, o campo _mes foi derivado da nova coluna createdAt_filled, que utiliza o valor de updatedAt quando createdAt está ausente.

Na base CampaignQueue, o campo _mes foi calculado a partir das colunas sendAt, scheduledAt e createdAt, priorizando a data mais representativa.

Por fim, na base Customer, foi criado o campo isEnriched_text, que converte valores booleanos (True/False) em texto (“Enriquecido” ou “Nao Enriquecido”), melhorando a legibilidade das informações em painéis analíticos.

Essas padronizações garantiram a coerência entre as tabelas e facilitaram a geração de indicadores consolidados.

```
4) Padronização e criação de novos campos

[11]
# Transforma storeId maiúsculo em todas as bases que possuem essa coluna
for df in (order_cl, campaign_cl, campaign_queue_cl):
    if "storeId" in df.columns:
        df["storeId"] = df["storeId"].astype(str).str.strip().str.upper()

[12]
# Campo textual para enriquecimento de clientes
if "isEnriched" in customer_cl.columns:
    customer_cl["isEnriched_text"] = customer_cl["isEnriched"].map(
        {True: "Enriquecido", False: "Nao Enriquecido"}
    )

[13]
# Status legíveis (sem acentos para evitar problemas de encoding)
map_status_customer = {1: "Ativo", 2: "Inativo"}
map_status_campaign = {1: "Rascunho", 2: "Agendada", 3: "Ativa", 4: "Concluida"}
map_status_queue = {
    1: "Pendente",
    2: "Enviada",
    3: "Entregue",
    4: "Lida",
    5: "Erro",
    6: "Cancelada",
}

if "status" in customer_cl.columns and customer_cl["status"].dtype != "O":
    customer_cl["status_desc"] = customer_cl["status"].map(map_status_customer).fillna("Nao informado")

if "status" in campaign_cl.columns and campaign_cl["status"].dtype != "O":
    campaign_cl["status_desc"] = campaign_cl["status"].map(map_status_campaign).fillna("Nao informado")

if "status" in campaign_queue_cl.columns and campaign_queue_cl["status"].dtype != "O":
    campaign_queue_cl["status_desc"] = campaign_queue_cl["status"].map(map_status_queue).fillna("Nao informado")

[14]
# Criação de campo de mês (_mes) em todas as bases

# Order: mês a partir de createdAt
if "createdAt" in order_cl.columns:
    order_cl["_mes"] = order_cl["createdAt"].dt.to_period("M").astype(str)

# Campaign: preencher createdAt nulo com updatedAt e criar _mes
if "createdAt" in campaign_cl.columns:
    if "updatedAt" in campaign_cl.columns:
        campaign_cl["createdAt_filled"] = campaign_cl["createdAt"].fillna(campaign_cl["updatedAt"])
    else:
        campaign_cl["createdAt_filled"] = campaign_cl["createdAt"]
    campaign_cl["_mes"] = pd.to_datetime(
        campaign_cl["createdAt_filled"], errors="coerce"
    ).dt.to_period("M").astype(str)

# CampaignQueue: mês considerando sendAt / scheduledAt / createdAt
if "sendAt" in campaign_queue_cl.columns:
    base_dt = campaign_queue_cl["sendAt"]
    if "scheduledAt" in campaign_queue_cl.columns:
        base_dt = base_dt.fillna(campaign_queue_cl["scheduledAt"])
    if "createdAt" in campaign_queue_cl.columns:
        base_dt = base_dt.fillna(campaign_queue_cl["createdAt"])
    campaign_queue_cl["_mes"] = pd.to_datetime(
        base_dt, errors="coerce"
    ).dt.to_period("M").astype(str)
```

Figura 11 – Padronização de campos, criação de colunas de status e derivação do campo mensal (_mes).

### 03.5. Inferência de Gênero (Customer)

Uma etapa específica foi dedicada ao enriquecimento da base Customer, com foco na inferência de gênero.

Para complementar registros incompletos e padronizar os valores existentes, foi utilizada a biblioteca gender_guesser, que infere o gênero a partir do primeiro nome.

Primeiro, os nomes foram normalizados e o primeiro nome de cada cliente foi extraído. Em seguida, os valores válidos de gênero foram mantidos, enquanto os registros com campos vazios ou inválidos passaram pelo processo de inferência.

O algoritmo retorna rótulos como “male”, “female”, “mostly_male”, “mostly_female” ou “unknown”, que foram convertidos em “M”, “F” ou “Nao informado”.

Os resultados foram armazenados nas colunas gender_inferred e gender_final, consolidando todos os casos em uma versão padronizada denominada gender_clean.

Esse processo aumentou significativamente a qualidade da base, permitindo segmentações mais precisas e mantendo consistência em todos os registros.

```
[15] ✓ 6s
pip install gender-guesser

Collecting gender-guesser
  Downloading gender_guesser-0.4.0-py2.py3-none-any.whl.metadata (3.0 kB)
Downloading gender_guesser-0.4.0-py2.py3-none-any.whl (379 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 379.3/379.3 kB 18.9 MB/s eta 0:00:00
Installing collected packages: gender-guesser
Successfully installed gender-guesser-0.4.0

[16] ✓ 0s
try:
    import gender_guesser.detector as gender
    det = gender.Detector(case_sensitive=False)
    pode_inferir = True
except Exception:
    print("gender_guesser não disponível. Prosseguindo sem inferência automática.")
    det = None
    pode_inferir = False

def pegar_primeiro_nome(nome):
    if pd.isna(nome):
        return None
    texto = str(nome).strip()
    if not texto:
        return None
    return texto.split()[0]

def traduzir_guess(g):
    if g in ["male", "mostly_male"]:
        return "M"
    if g in ["female", "mostly_female"]:
        return "F"
    return "Nao informado"

if "gender" in customer_cl.columns and "name" in customer_cl.columns:
    customer_cl["gender"] = customer_cl["gender"].str.upper()
    validos = customer_cl["gender"].isin(["M", "F", "O"])
    mascara_invalido = ~validos

    if pode_inferir and mascara_invalido.any():
        primeiros = customer_cl.loc[mascara_invalido, "name"].apply(pegar_primeiro_nome)
        palpites = primeiros.apply(
            lambda n: det.get_gender(n) if isinstance(n, str) else "unknown"
        )
        inferido = palpites.apply(traduzir_guess)
        customer_cl.loc[mascara_invalido, "gender_inferred"] = inferido
        customer_cl.loc[mascara_invalido, "gender_final"] = np.where(
            inferido.isin(["M", "F"]), inferido, "Nao informado"
        )
    else:
        customer_cl.loc[mascara_invalido, "gender_final"] = "Nao informado"

    customer_cl.loc[validos, "gender_final"] = customer_cl.loc[validos, "gender"]
    customer_cl.loc[customer_cl["gender"] == "O", "gender_final"] = "O"
    customer_cl["gender_clean"] = customer_cl["gender_final"]

    print("Distribuição (gender_clean):")
    print(customer_cl["gender_clean"].value_counts(dropna=False))

Distribuição (gender_clean):
gender_clean
F              344
M              309
O              263
Nao informado   84
Name: count, dtype: int64
```

Figura 12 – Trecho de código e saída da inferência de gênero utilizando a biblioteca gender_guesser.

### 03.6. Validações Finais

A etapa final do tratamento de dados teve como principal objetivo confirmar a integridade e a qualidade dos quatro conjuntos de dados tratados: Order, Campaign, CampaignQueue e Customer. Essa validação garantiu que, após todos os processos de limpeza, enriquecimento e padronização, os dados estivessem consistentes, completos e prontos para integração com o banco de dados e utilização em análises e dashboards.

Para essa verificação, foi criada uma função chamada resumo_base(), responsável por apresentar um resumo estatístico de cada tabela. A função exibe o nome da base, o número de linhas e colunas, a contagem de registros duplicados e a quantidade de valores nulos por coluna. Essa análise foi essencial para identificar eventuais falhas e assegurar a completude e coerência das informações tratadas.

Os resultados da execução demonstraram que nenhuma das tabelas possui registros duplicados, evidenciando que o processo de normalização foi bem-sucedido. Na tabela Order, observou-se que todos os campos foram devidamente preenchidos, exceto a coluna scheduledAt, que apresentou 1.729 valores nulos correspondentes a pedidos não agendados — uma condição esperada para esse tipo de registro. Nenhuma outra coluna apresentou valores ausentes, e todas as variáveis numéricas e textuais estavam devidamente formatadas e padronizadas.

Na tabela Campaign, também não foram detectados registros duplicados, e a única coluna com ausência de valores foi createdAt, que apresentou 1.224 registros nulos referentes a campanhas antigas sem data de criação registrada. Esse comportamento é considerado aceitável, uma vez que esses casos foram compensados com a criação da nova coluna createdAt_filled, que preenche automaticamente a data de criação com base no valor de updatedAt. As demais colunas, incluindo storeId, status_desc e _mes, estavam completas e coerentes com o restante da estrutura.

Já na base CampaignQueue, verificou-se que todos os campos principais, como campaignId, customerId, status e createdAt, estavam completos. A única exceção foi a coluna sendAt, que apresentou 1.712 valores nulos. Esses nulos representam mensagens ainda não enviadas, sendo, portanto, um comportamento esperado. O relatório de integridade também confirmou que não há duplicações nem inconsistências em colunas de identificação ou relacionamento.

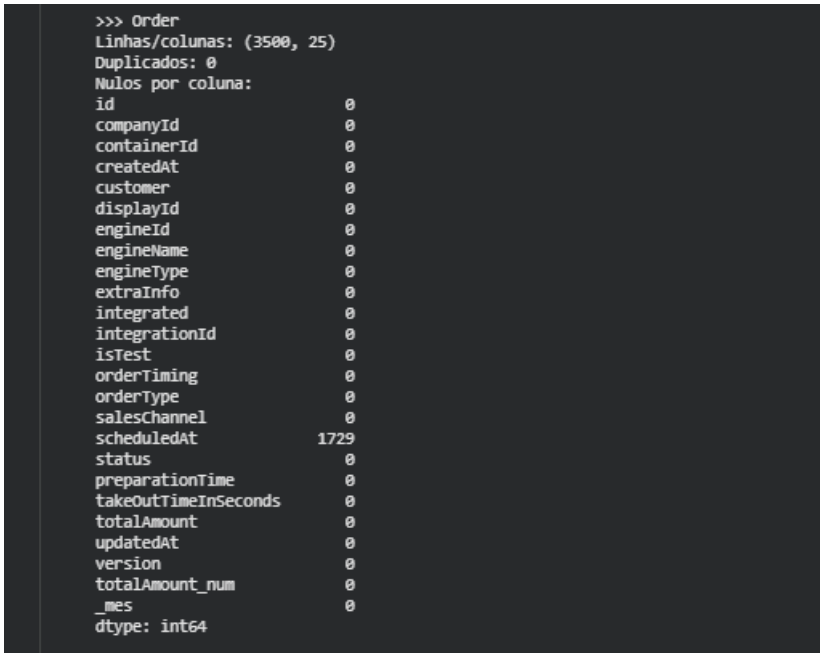
Por fim, na base Customer, observou-se que os campos enrichedAt e enrichedBy apresentaram 496 valores nulos, representando clientes que ainda não haviam passado por processos de enriquecimento de dados. Além disso, o campo gender_inferred apresentou 741 nulos, o que indica casos em que a inferência de gênero, realizada pela biblioteca gender_guesser, não conseguiu determinar o valor com confiança. Apesar dessas ausências, os resultados gerais foram considerados satisfatórios, uma vez que os campos críticos — como id, name, taxId, email, phone e companyId — estavam completos e corretamente padronizados.

Além da análise de nulos e duplicados, foram realizadas múltiplas verificações de integridade referencial entre as bases tratadas, assegurando a coerência dos vínculos entre entidades relacionadas. Primeiramente, verificou-se se todos os campaignId existentes na tabela CampaignQueue estavam devidamente presentes na tabela Campaign, e se todos os customerId da CampaignQueue possuíam correspondência na tabela Customer. Ambas as checagens retornaram zero inconsistências, confirmando que todas as relações entre campanhas, clientes e mensagens foram mantidas corretamente.

Em seguida, a integridade foi validada também entre as demais tabelas: a base Order foi comparada com Customer para garantir que todos os pedidos estejam vinculados a clientes existentes; Order e Customer foram verificados em relação à tabela Estabelecimentos para confirmar que os identificadores de loja (companyId) são válidos; e a tabela Campaign foi checada para assegurar que todos os storeId correspondam a estabelecimentos cadastrados. Em todas essas verificações, o resultado foi igualmente zero registros inválidos, indicando consistência total entre as tabelas.

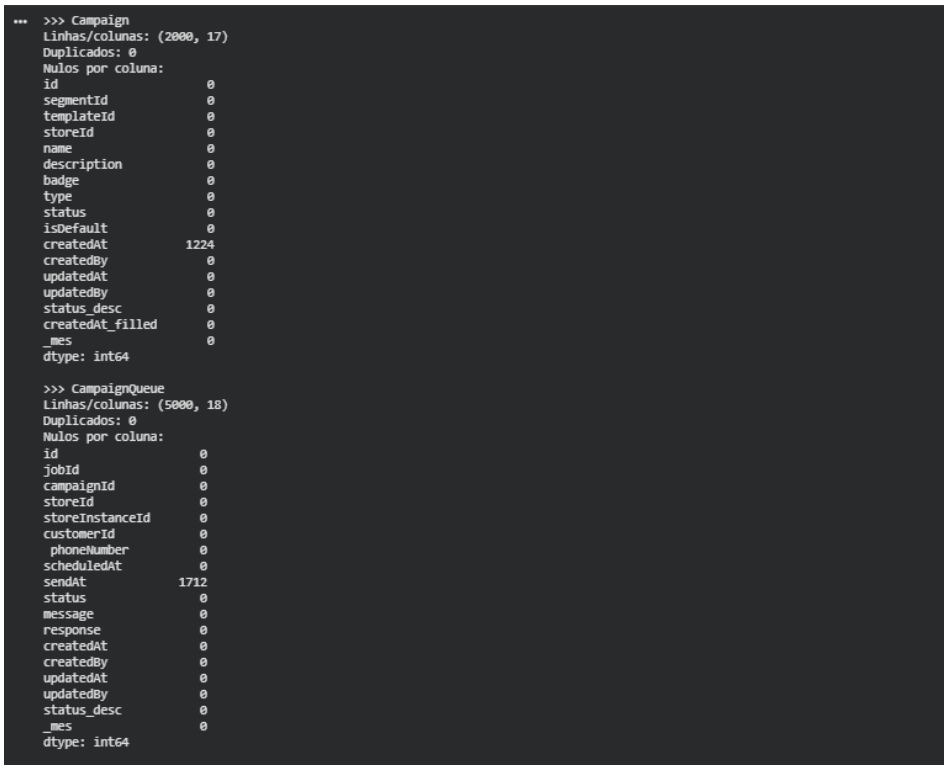
Com esses resultados, conclui-se que todas as cinco bases — Customer, Order, Campaign, CampaignQueue e Estabelecimentos — encontram-se consistentes, sem registros duplicados, com valores nulos apenas em campos esperados e com todos os relacionamentos entre tabelas devidamente íntegros. Esse cenário reforça a confiabilidade e qualidade dos dados tratados, garantindo que o processo de preparação foi eficaz e que as informações estão totalmente aptas para a fase de integração no banco MySQL e posterior análise no dashboard.


Figura 13 – Execução do código de verificação de nulos, duplicados e integridade referencial.



```
>>> Order
Linhas/colunas: (3500, 25)
Duplicados: 0
Nulos por coluna:
id                          0
companyId                   0
containerId                 0
createdAt                   0
customer                    0
displayId                   0
engineId                    0
engineName                  0
engineType                  0
extraInfo                   0
integrated                  0
integrationId               0
isTest                      0
orderTiming                 0
orderType                   0
salesChannel                0
scheduledAt              1729
status                      0
preparationTime             0
takeOutTimeInSeconds        0
totalAmount                 0
updatedAt                   0
version                     0
totalAmount_num             0
_mes                        0
dtype: int64
```

Figura 14 – Análise de nulos e duplicados na base Order.



```
>>> Campaign
Linhas/colunas: (2000, 17)
Duplicados: 0
Nulos por coluna:
id                   0
segmentId            0
templateId           0
storeId              0
name                 0
description          0
badge                0
type                 0
status               0
isDefault            0
createdAt         1224
createdBy            0
updatedAt            0
updatedBy            0
status_desc          0
createdAt_filled     0
_mes                 0
dtype: int64

>>> CampaignQueue
Linhas/colunas: (5000, 18)
Duplicados: 0
Nulos por coluna:
id                   0
jobId                0
campaignId           0
storeId              0
storeInstanceId      0
customerId           0
 phoneNumber         0
scheduledAt          0
sendAt            1712
status               0
message              0
response             0
createdAt            0
createdBy            0
updatedAt            0
updatedBy            0
status_desc          0
_mes                 0
dtype: int64
```

Figura 15 – Análise de nulos e duplicados nas bases Campaign e CampaignQueue.

```
>>> Customer
Linhas/colunas: (1000, 22)
Duplicados: 0
Nulos por coluna:
id                   0
name                 0
taxId                0
gender               0
dateOfBirth          0
status               0
externalCode         0
isEnriched           0
enrichedAt         496
enrichedBy         496
createdAt            0
createdBy            0
updatedAt            0
updatedBy            0
phone                0
email                0
companyId            0
isEnriched_text      0
status_desc          0
gender_inferred    741
gender_final         0
gender_clean         0
dtype: int64

CampaignQueue com campaignId inexistente: 0
CampaignQueue com customerId inexistente: 0

Order com customer inexistente: 0
Order com companyId inexistente em estabelecimentos: 0

Customer com companyId inexistente em estabelecimentos: 0
Campaign com storeId inexistente em estabelecimentos: 0
```

Figura 16 – Análise de nulos e duplicados na base Customer e validação de integridade.

### 03.7. Exportação dos dados tratados

Por fim, as bases tratadas foram exportadas novamente em formato CSV, concluindo o processo de preparação.

Foram gerados quatro novos arquivos: Order_clean.csv, Campaign_clean.csv, CampaignQueue_clean.csv e Customer_clean.csv, todos gravados com o separador “;” e sem índice.

Esses arquivos foram posteriormente utilizados em duas etapas fundamentais do projeto:

(1) importação para o banco de dados MySQL através do Table Data Import Wizard, que criou as tabelas definitivas no schema cannoli.

e (2) uso como fonte de dados para os dashboards interativos desenvolvidos em React com Chart.js, permitindo a visualização de KPIs e métricas de forma dinâmica.

Essa etapa marcou o encerramento do ciclo de tratamento e garantiu que os dados estivessem completamente estruturados e prontos para uso em aplicações analíticas.


Figura 17 – Exportação final dos arquivos tratados e confirmação da geração dos CSVs limpos.

## 04. CARGA DAS BASES TRATADAS NO MYSQL

Após a finalização do processo de tratamento, padronização e enriquecimento, os dados foram preparados para serem integrados em um ambiente de banco de dados relacional, de forma a permitir análises mais rápidas, consultas otimizadas e conexão direta com a aplicação Cannoli Intelligence Dashboard. Para isso, foi utilizada a ferramenta Table Data Import Wizard, disponível no MySQL Workbench, que possibilita a importação direta dos arquivos CSV tratados para o banco de dados, automatizando a criação das tabelas e a inserção dos registros.

Os arquivos Order_clean.csv, Customer_clean.csv, Campaign_clean.csv, CampaignQueue_clean.csv e Estabelecimentos.csv foram carregados para o schema cannoli, resultando na criação de cinco tabelas principais correspondentes às bases tratadas. Cada tabela foi validada para assegurar a correspondência entre os campos originais e os tipos de dados atribuídos no MySQL, como VARCHAR, INT, DECIMAL e DATE. Esse processo garantiu que todos os registros mantivessem integridade e compatibilidade com as etapas posteriores de análise e visualização, além de preservar a estrutura lógica aplicada durante o tratamento dos dados.

Durante a carga das tabelas tratadas, também foram configuradas duas novas estruturas complementares, essenciais para a operação do sistema e para o uso de inteligência artificial. A primeira é a tabela campaign_ai_sugestoes, criada para armazenar os resultados gerados pelo modelo de aprendizado de máquina desenvolvido com o algoritmo RandomForestClassifier. Essa tabela contém os campos campaignId, storeId, status_previsto, confianca, grupo, modelo_versao e gerado_em, permitindo registrar as previsões de desempenho das campanhas. Por meio dela, o sistema pode exibir sugestões automáticas de otimização, como ajustar, pausar ou reforçar campanhas de marketing conforme o comportamento previsto. Essa integração entre dados tratados e resultados preditivos amplia as possibilidades analíticas do projeto, unindo análise descritiva e inteligência artificial em um único ambiente de dados.

A segunda tabela adicional criada foi a usuarios, responsável pelo gerenciamento de acessos ao dashboard e pela autenticação dos diferentes perfis de usuários. Cada registro nessa tabela contém o identificador do usuário, nome, e-mail, senha (armazenada de forma criptografada), perfil de acesso e, quando aplicável, o identificador do estabelecimento vinculado. O campo perfil define o nível de permissão dentro do sistema, diferenciando administradores (ADMIN) — que podem visualizar todas as informações — dos usuários de estabelecimentos (ESTABELECIMENTO), que têm acesso apenas aos dados da própria loja. Essa estrutura garante controle, segurança e privacidade no acesso às informações,

além de permitir que a autenticação seja diretamente integrada à aplicação web desenvolvida em React.

Ao término da importação, o banco de dados cannoli consolidou um total de sete tabelas: order, customer, campaign, campaign_queue, estabelecimentos, campaign_ai_sugestoes e usuarios. Essa estrutura unificada reflete toda a arquitetura lógica do projeto, integrando dados operacionais, preditivos e administrativos em um único ambiente relacional. O processo de importação assegurou integridade referencial entre as tabelas, consistência dos dados e compatibilidade total com o sistema React e os futuros módulos analíticos.

Com isso, o ambiente de dados foi preparado para servir como base do ecossistema Cannoli Intelligence, permitindo o consumo das informações tanto por consultas SQL quanto por aplicações externas. O formato relacional adotado garante escalabilidade e flexibilidade, possibilitando futuras expansões do sistema com novas tabelas, previsões ou integrações com modelos de inteligência artificial. Assim, a carga das bases tratadas no MySQL representa o encerramento da etapa de preparação e a transição para a fase analítica e de visualização dos resultados.

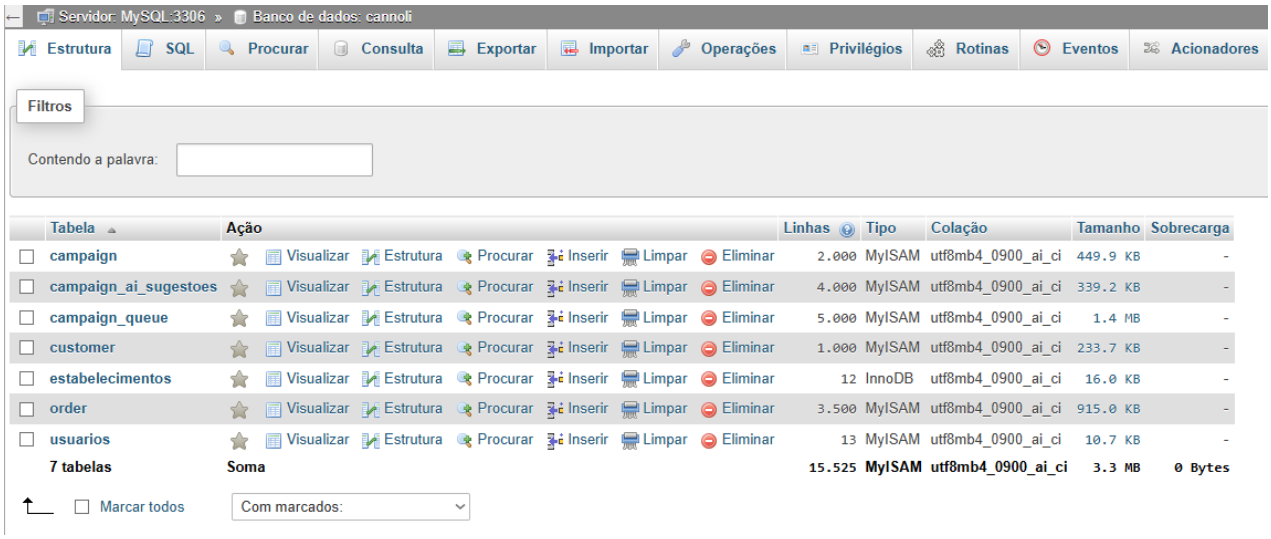



Figura 18 – Estrutura final das tabelas no banco de dados cannoli, mostrando a integração das bases tratadas e das tabelas adicionais de IA e usuários.

## 05. CONCLUSÃO

Portanto, com a finalização desta etapa, o projeto Cannoli Intelligence atingiu um marco essencial na construção de sua base de dados. As informações, antes dispersas e inconsistentes, foram transformadas em um conjunto coeso, limpo e padronizado, pronto para sustentar análises de alto nível. O processo de tratamento e enriquecimento automatizado, aliado à integração das tabelas no ambiente relacional do MySQL, estabeleceu uma estrutura sólida e confiável para o armazenamento e processamento dos dados.

Esse novo modelo de dados permite a geração de indicadores de desempenho (KPIs) e a construção de dashboards interativos, que poderão ser utilizados tanto por administradores quanto por estabelecimentos parceiros para acompanhar métricas em tempo real. Além disso, a padronização e o controle de integridade garantem consistência entre as diferentes fontes, tornando as análises mais precisas e as decisões mais assertivas.

Assim, a conclusão desta fase consolida o ambiente analítico do Cannoli Intelligence, transformando dados brutos em informação estratégica. Essa estrutura de dados, agora robusta e escalável, servirá como base para as próximas etapas do projeto — incluindo o desenvolvimento da camada de visualização, os módulos de inteligência artificial e as funcionalidades de recomendação automatizada — assegurando qualidade, confiabilidade e valor agregado às análises e à tomada de decisão.