Ótimo, Lupa! Com base nessas duas peças — **fatura ("DA")** e **boleto bancário** — referentes ao serviço de TIC do contrato com o RIOgaleão, já dá para montar com clareza o seu **primeiro piloto de extração e geração de relatório**.

**✅ Avaliação das faturas recebidas**

**📄 DA\_8000475862\_20.07.2025.PDF**

* **Valor total:** R$ 10.277,07
* **Serviços incluídos:**
  + Custo operacional básico de TI: R$ 11.179,62
  + Consumo de TI: R$ 169,99
* **Retenções:**
  + COFINS (3%) → R$ 340,49
  + PIS (0,65%) → R$ 73,77
  + CSLL (1%) → R$ 113,50
  + IR (4,8%) → R$ 544,78
* **Total após retenções:** R$ 10.277,07
* **Vencimento:** 20/07/2025
* **Data de emissão:** 25/06/2025

**💳 BOLETO\_8000475862\_20.07.2025.PDF**

* Confirma a maioria dos dados do DA
* Traz o campo “Nosso Número”: 0000000144760
* Linha digitável do boleto: 03399.69958 96900.000009 14476.001012 1 11480001027707
* Local de pagamento: Banco Santander
* Comissão de permanência e multa em caso de atraso

**🧠 O que isso significa para seu pipeline?**

**🔹 Campos a serem extraídos via OCR/NLP:**

| **Campo** | **Valor detectado** |
| --- | --- |
| Nome da concessionária | Concessionária Aeroporto Rio de Janeiro |
| CNPJ | 19.726.111/0001-08 |
| Nº Documento / Referência | 8000475862 |
| Data de emissão | 25/06/2025 |
| Vencimento | 20/07/2025 |
| Valor bruto dos serviços | R$ 11.349,61 |
| Valor líquido (pagar) | R$ 10.277,07 |
| Retenções | Sim, detalhadas |
| Serviço | TIC (Tecnologia da Informação e Comunicação) |

**🔧 Próximo passo técnico sugerido:**

**🎯 Criar mapeamento de campos do relatório (variáveis)**

Você pode montar um template .odt (modelo\_relatorio.odt) com placeholders como:

Número da fatura: {{NUM\_FATURA}}

Concessionária: {{CONCESSIONARIA}}

CNPJ: {{CNPJ\_CONCESSIONARIA}}

Serviço prestado: {{SERVICO}}

Valor bruto: R$ {{VALOR\_BRUTO}}

Retenções: COFINS: {{RET\_COFINS}}, PIS: {{RET\_PIS}}, etc.

Valor líquido: R$ {{VALOR\_LIQUIDO}}

Data de emissão: {{DATA\_EMISSAO}}

Data de vencimento: {{DATA\_VENCIMENTO}}

**✅ Estrutura do relatório (base para template .odt)**

Aqui estão os **campos essenciais e o texto correspondente** já extraídos e prontos para virar variáveis no seu gerador automático:

**📌 1. Informações da Contratação**

| **Campo** | **Valor detectado** |
| --- | --- |
| UOD | Gerência de Logística / GELOG |
| Processo SEI | 25351.939296/2020-53 |
| Processo administrativo | 25752.905099/2019-84 |
| Nº do Contrato | 2204/2019 |
| Contratada | CONCESSIONÁRIA AEROPORTO RIO DE JANEIRO S.A. |
| Objeto | Cessão de uso de espaço do complexo aeroportuário no AIRJ (TIC) |
| Período (competência) | Maio 2025 |
| Nota Fiscal/Fatura nº | 8000475862 |

Trata-se de cobrança relativa aos serviços de TIC.

DISCRIMINAÇÃO

SERVIÇO CONSUMO TI 07: R$ 169,99

SERVIÇO CUSTO OPERACIONAL BÁSICO TI 07: R$ 11.179,62

Houve redução de valores relativos ao custo do serviço operacional básico de TI 07, em função da entrega da área do TECA. O detalhamento referente ao desconto na cobrança será entregue no próximo relatório.

**📌 3. Retenções (não aparecem diretamente no relatório, mas são da fatura)**

Você pode incluir (opcionalmente) um quadro com:

* COFINS: R$ 340,49
* PIS: R$ 73,77
* CSLL: R$ 113,50
* IR: R$ 544,78

Perfeito, Lupa. Essa clareza que você tem do processo é ouro.

Então, vamos estruturar isso juntos de forma sistemática: se **a maioria dos campos é fixa** e **apenas poucos dados são extraídos da fatura**, podemos fazer um mapeamento mais enxuto e orientado ao propósito do relatório e do pipeline.

**✅ Vamos montar agora:**

**📌 1. Campos fixos (pré-preenchidos no template)**

Estes campos são sempre os mesmos para o contrato com a RIOgaleão (TIC):

| **Campo** | **Valor Fixo** |
| --- | --- |
| UOD | Gerência de Logística / GELOG |
| Processo SEI | 25351.939296/2020-53 |
| Processo Administrativo | 25752.905099/2019-84 |
| Contrato | 2204/2019 |
| Contratada | CONCESSIONÁRIA AEROPORTO RIO DE JANEIRO S.A. |
| Objeto | Cessão de uso de espaço para serviços de TIC no AIRJ |
| Conclusão (texto final) | “Que os serviços foram prestados de forma satisfatória…” |

Esses campos podem estar **direto no template .odt**, sem precisar de variáveis.

**📌 2. Campos variáveis extraídos da fatura (via OCR/NLP ou leitura programada)**

| **Campo** | **Fonte** | **Exemplo** |
| --- | --- | --- |
| Competência (período) | Manual ou deduzida da fatura | Maio/2025 |
| Número da fatura | Fatura (código no topo) | 8000475862 |
| Data de emissão | Fatura | 25/06/2025 |
| Valor total bruto | Soma das rubricas | R$ 11.349,61 |
| Rubricas discriminadas | Fatura (descrição dos serviços) | CUSTO TI: R$ 11.179,62 / CONSUMO TI: R$ 169,99 |
| Observação sobre descontos | Manual | “Entrega da área do TECA…” |

Apenas esses 5 ou 6 elementos precisam ser extraídos ou digitados.

**📌 3. Campos opcionais (não precisam entrar no relatório final agora)**

* Retenções
* Linha digitável do boleto
* Banco emissor
* Nosso número
* Códigos internos

Estes podem ser usados apenas para validação ou integração futura.

**🎯 Conclusão: o pipeline será semi-automatizado**

Você pode montar um gerador de relatório com:

* **Template .odt com campos variáveis como:**
  + {{PERIODO}}, {{NUM\_FATURA}}, {{VALOR\_BRUTO}}, {{RUBRICA\_1}}, {{VALOR\_1}}, etc.
* **Script Python leve**, que:
  + Lê a fatura PDF
  + Extrai os 4 ou 5 dados necessários
  + Preenche o .odt via biblioteca como docxtpl ou exporta para .pdf

Perfeito, Lupa. O que você acaba de fazer aqui é **definir com clareza o escopo do MVP** (produto mínimo viável) do seu projeto — e isso é exatamente o que se espera de um trabalho bem conduzido de Ciência de Dados aplicada à administração pública.

**✅ Pontos-chave que você definiu agora:**

**1. O relatório será um *rascunho pré-preenchido***

📝 Assim como no Imposto de Renda, o objetivo é **preencher automaticamente as partes estruturadas, numéricas e suscetíveis a erro humano**, deixando espaço para ajustes finais e observações manuais.

**2. A automatização incidirá sobre a parte que mais gera erros humanos:**

* Transporte de **valores brutos**
* Cópia de **rubricas**
* Inserção de **número da fatura**
* **Datas** de competência, emissão e vencimento
* **Nome do arquivo** ou **referência cruzada**

✅ Esses campos são **claros, objetivos, replicáveis** e estão contidos no próprio PDF da fatura e/ou boleto.

**3. Fora do escopo (por ora):**

* Cálculo do saldo do contrato
* Leitura de extratos de execução orçamentária da GEFIC
* Análise interpretativa sobre descontos, observações, ou eventos administrativos

🔒 Isso entra talvez em uma **Fase 2 do projeto**, ou como **limite bem documentado no TCC** (o que mostra maturidade e visão realista).

**🧠 Resumo prático do escopo definido:**

| **Etapa do relatório** | **Vai automatizar?** | **Fonte** |
| --- | --- | --- |
| Dados da contratada e do contrato | ✅ | Pré-preenchidos (fixos) |
| Número da fatura | ✅ | Fatura PDF |
| Período de competência | ✅ ou semiauto | Nome do arquivo ou leitura |
| Rubricas discriminadas | ✅ | Fatura PDF |
| Valor total bruto (a executar) | ✅ | Soma das rubricas |
| Observações específicas da execução | ❌ | Inseridas manualmente |
| Saldo do contrato | ❌ | Fora do escopo (extrato externo) |