

Relatório de análise da base CSFR



Relatório de Análise e Preenchimento de Dados Faltantes

Realizado por:

David de Assis Vieira

Cientista de dados

Morning Star Consulting

Relatório de análise e preenchimento de dados faltantes

1. Objetivo do Trabalho

O objetivo deste trabalho foi analisar um conjunto de informações sobre contratos de serviços (como gasodutos e embarcações) e completar as partes que estavam faltando, especificamente os campos relacionados a índices econômicos (como IGP, IPCA), seus pesos (percentuais) e fatores (valores fixos). Para isso, usamos informações já disponíveis, como a descrição do serviço, o tipo de contrato, o fornecedor, a moeda e o número do item, para prever os valores ausentes. No final, organizamos tudo em um arquivo Excel, destacando os valores preenchidos com cores diferentes para facilitar a visualização.

2. O que analisamos

Recebemos uma tabela com várias colunas descrevendo contratos de serviços. As informações principais que usamos foram:

- Descrição do Serviço: Textos como "AFRETAMENTO - LANÇAMENTO DO DUTO" ou "TAXA DIÁRIA PREST. SERVIÇOS OPER. MN", que explicam o que foi contratado.
- Classificação: Tipos de contrato, como "EPCI" (relacionado a projetos de engenharia) ou "Embarcações".
- Fornecedor: Nome da empresa responsável, como "SAIPEM S.A." ou "MAROIL APOIO MARITIMO LTDA".
- Moeda: Se o contrato foi feito em dólares (USD), reais (BRL) ou outra moeda.
- Item: Um número que indica a etapa do serviço, como 10 ou 20.

Os campos que precisávamos completar eram:

- Índices Econômicos: Até seis colunas (Indice1 a Indice6) com nomes como IGP, INS ou IPCA, usados para ajustes de preço.
- Pesos: Até seis colunas (Peso1 a Peso6) com percentuais (ex.: 0.37 ou 0.63) que mostram quanto cada índice influencia o ajuste.
- Fatores: Até seis colunas (FatorK1 a FatorK6), que geralmente aparecem como 1.0 quando preenchidas.

Muitas linhas estavam incompletas, com alguns ou todos esses campos vazios, então nosso trabalho foi descobrir padrões para preenchê-los.

3. O que descobrimos

Ao analisar os dados, encontramos várias pistas sobre como os índices, pesos e fatores são definidos:

- Padrões nas Descrições: O texto da descrição do serviço dá dicas claras. Por exemplo:
 - Serviços com "AFRETAMENTO" (transporte ou fretamento) frequentemente usam índices como IPCA ou IGP.
 - Já "TAXA DIÁRIA" (pagamento por dia de trabalho) aparece com IGP e INS.
- Tipos de Contrato:
 - Contratos classificados como "Embarcações" têm pesos fixos, como 0.37 para IGP e 0.63 para INS.
 - Contratos "EPCI" variam mais, dependendo do serviço e do fornecedor.
- Fornecedores:
 - A "SAIPEM S.A." (a empresa mais comum nos dados) usa índices como IPCA ou IGP em serviços de afretamento.
 - A "MAROIL APOIO MARITIMO LTDA" segue um padrão mais específico, com IGP e INS.
- Moeda:
 - Contratos em reais (BRL) tendem a usar índices brasileiros, como IGP e INS.
 - Contratos em dólares (USD) ou euros (EUR) podem variar mais, dependendo do contexto.
- Item: O número do item parece indicar a etapa do serviço (ex.: preparação ou execução), o que pode influenciar os índices escolhidos.
- Pesos e Fatores: »
 - Nos dados fornecidos, todos os FatorK eram 1.0, então os pesos ajustados coincidiam com os pesos brutos (ex.: $0.37 \times 1.0 = 0.37$).
 - Porém, se um fator K fosse 0.4, por exemplo, um peso bruto de 0.5 se tornaria 0.2 (0.5×0.4), mudando o impacto real do índice.
 - A soma dos pesos ajustados (após multiplicação) deve ser ≤ 1.0 (100%), refletindo a contribuição total dos índices.

Essas descobertas mostraram que os valores faltantes não são aleatórios: eles seguem padrões baseados nas informações já disponíveis.

4. Como preenchemos os dados faltantes.

Para completar as linhas vazias, usamos um método chamado "Machine Learning" (aprendizado de máquina), que é como ensinar um computador a reconhecer padrões e fazer previsões. Aqui está como fizemos isso de forma simples:

1. **Aprendizado com os Dados Existentes:** Pegamos as linhas que já tinham índices, pesos e fatores preenchidos e usamos elas para "ensinar" o sistema. Por exemplo, vimos que "TAXA DIÁRIA" com "MAROIL" usa IGP (37%) e INS (63%).
2. **Previsão dos Valores Faltantes:** Depois, aplicamos o que o sistema aprendeu nas linhas vazias. Por exemplo, para um serviço de "AFRETAMENTO" da "SAIPEM S.A." em USD, previmos IPCA (60%) e IGP (40%).
3. **Ajustes:** Garantimos que os pesos somassem até 100% (ou menos) e fixamos os fatores em 1.0, seguindo o padrão observado.

Esse método foi escolhido porque:

- Os padrões são complicados e dependem de várias coisas ao mesmo tempo (descrição, fornecedor, moeda etc.).
- Há muitas linhas vazias, e fazer isso manualmente levaria muito tempo e poderia ter erros.
- O sistema pode "pensar" como os dados foram preenchidos originalmente e aplicar isso consistentemente.

5. Resultados obtidos

Após analisar e prever os valores, criamos um arquivo com todos os dados completos. Aqui está um exemplo do que foi feito:

- **Linha Original (4600331708 - 10 - 10):**
 - Descrição: "AFRET. PREP. EMB.LANÇAMENTO DE DUTOS"
 - Classificação: "EPCI"
 - Fornecedor: "SAIPEM S.A."
 - Moeda: "USD"
 - Item: 10
 - Índices/Pesos/Fatores: Tudo vazio.
- **Linha Preenchida:**
 - Índice1: IPCA, Peso1: 0.6, FatorK1: 1.0

- Índice2: IGP, Peso2: 0.4, FatorK2: 1.0
- (Outros campos permaneceram vazios, pois o padrão sugere no máximo dois índices aqui.)
- Por quê? Baseado em serviços semelhantes de "AFRETAMENTO" da "SAIPEM S.A.", que usam IPCA e IGP.

No arquivo final (chamado output_preenchido.xlsx):

- Os valores que já existiam foram mantidos.
- Os valores novos (previstos) foram destacados com cores:
 - Índice1, Peso1 e FatorK1 em azul-claro.
 - Índice2, Peso2 e FatorK2 em verde-claro.
 - E assim por diante, até o sexto grupo em amarelo-ouro.
- Isso facilita ver o que foi adicionado e como os índices estão organizados.

6. Conclusão

Conseguimos completar os dados faltantes de forma lógica e consistente, usando as informações que já estavam na tabela. Descobrimos que os índices, pesos e fatores seguem padrões claros relacionados ao tipo de serviço, fornecedor e moeda, o que nos permitiu fazer previsões confiáveis. O resultado é um arquivo organizado, fácil de entender e pronto para uso, com os valores preenchidos destacados para rápida identificação.