

Relatório de Análise de Query - Vagas SP vs. Remoto

Análise de Álgebra Linear com TF-IDF

4 de novembro de 2025

1 Descrição do Dataset

A análise foi realizada utilizando um conjunto de dados público da plataforma Kaggle.

Arquivo:engenheiro_de_dados_6k.csv

Conteúdo:O dataset contém 1.614 vagas para a posição de Engenheiro de Dados (Data Engineer) no Brasil (LinkedIn, 2023). Para esta análise, os 'documentos' são os textos da coluna 'job_description'.

2 Metodologia da Análise

O objetivo deste trabalho é comparar a aderência de dois mercados de trabalho (São Paulo e Remoto) a um perfil de vaga ideal, definido pelo usuário (a 'query').

2.1 Definição da Query e Documentos

A Query (Perfil Ideal):

O usuário forneceu o seguinte texto em inglês descrevendo seu perfil ideal:

data engineer with experience in python and sql

Os Documentos (Os Grupos):

Os documentos da base foram divididos em dois grupos:

- \textbf{Grupo 1 (São Paulo):} 756 documentos.

- \textbf{Grupo 2 (Remotas):} 180 documentos.

2.2 Processamento (Álgebra Linear)

1. **TF-IDF:**A Query e todos os Documentos de um grupo são transformados em vetores numéricos que representam a importância de cada palavra-chave.

2. **Similaridade de Cosseno (Análise de Ângulos):**É calculada a similaridade (cosseno do ângulo) entre o vetor da Query e *cada um* dos vetores de documentos do grupo. Um ângulo de 0° (cosseno = 1) significa que a vaga é idêntica à query.

3 Resultados da Análise Comparativa

A seguir, são apresentados os 3 documentos mais similares (menor ângulo) à Query, separados por cada grupo.

3.1 Top 3 - Vagas em São Paulo

Rank 1 (ID do Documento no CSV: 1216):

\textbf{Similaridade (S):} 0.3576 | \textbf{Ângulo (Graus):} 69.05°

Trecho do Documento:

Na Turing ajudamos desenvolvedores a construírem a carreira de seus sonhos e encontrar as melhores vagas de trabalho online com salário em dólar em empresas de tecnologia dos EUA. Mais de 400 empresas incluindo Johnson & Johnson Dell Disney Volvo etc. já contrataram desenvolvedores com a Turing. A U.S.-based organization that is committed to offering excellent healthcare research and me...

Rank 2 (ID do Documento no CSV: 409):

\textbf{Similaridade (S):} 0.3576 | \textbf{Ângulo (Graus):} 69.05°

Trecho do Documento:

Na Turing ajudamos desenvolvedores a construírem a carreira de seus sonhos e encontrar as melhores vagas de trabalho online com salário em dólar em empresas de tecnologia dos EUA. Mais de 400 empresas incluindo Johnson & Johnson Dell Disney Volvo etc. já contrataram desenvolvedores com a Turing. A U.S.-based organization that is committed to offering excellent healthcare research and me...

Rank 3 (ID do Documento no CSV: 266):

\textbf{Similaridade (S):} 0.2430 | \textbf{Ângulo (Graus):} 75.94°

Trecho do Documento:

Área De Atuação Informática / TI / TecnologiaLocalização: Diferencial São Paulo-SPAtribuições: Certificação Azure Fundamentals 900 Certificação Azure Data EngineerQue Iremos Avaliar Experiência com desenvolvimento de Data Lakes em Cloud; Conhecimento de plataformas Big Data em ambiente de nuvem e banco de dados tradicionais (e.g. Azure Cosmos DB); Experiência Azure Data Lake Storage; Azure Data...

3.2 Top 3 - Vagas Remotas

Rank 1 (ID do Documento no CSV: 311):

\textbf{Similaridade (S):} 0.3032 | \textbf{Ângulo (Graus):} 72.35°

Trecho do Documento:

Na Turing ajudamos desenvolvedores a construírem a carreira de seus sonhos e encontrar as melhores vagas de trabalho online com salário em dólar em empresas de tecnologia dos EUA. Mais de 400 empresas incluindo Johnson & Johnson Dell Disney Volvo etc. já contrataram desenvolvedores com a Turing. A U.S.-based organization that is committed to offering excellent healthcare research and me...

Rank 2 (ID do Documento no CSV: 447):

\textbf{Similaridade (S):} 0.3032 | \textbf{Ângulo (Graus):} 72.35°

Trecho do Documento:

Na Turing ajudamos desenvolvedores a construírem a carreira de seus sonhos e encontrar as melhores vagas de trabalho online com salário em dólar em empresas de tecnologia dos EUA. Mais de 400 empresas incluindo Johnson & Johnson Dell Disney Volvo etc. já contrataram desenvolvedores com a Turing. A U.S.-based organization that is committed to offering excellent healthcare research and me...

Rank 3 (ID do Documento no CSV: 409):

\textbf{Similaridade (S):} 0.3032 | \textbf{Ângulo (Graus):} 72.35°

Trecho do Documento:

Na Turing ajudamos desenvolvedores a construir a carreira de seus sonhos e encontrar as melhores vagas de trabalho online com salário em dólar em empresas de tecnologia dos EUA. Mais de 400 empresas incluindo Johnson & Johnson Dell Disney Volvo etc. já contrataram desenvolvedores com a Turing. A U.S.-based organization that is committed to offering excellent healthcare research and me...

4 Discussão dos Resultados

A Descoberta: A análise da query revelou um 'cluster' (agrupamento) de vagas dominado pela empresa Turing, tanto no mercado de SP quanto no Remoto.

A Causa: Isso é evidenciado pelos scores de similaridade e ângulos idênticos (ex: 0.3032 / 72.35°), provando que múltiplas vagas (Ranks 45, 60, 55) são, na verdade, cópias exatas do mesmo documento de texto.

A Implicação (Ruído nos Dados): Isso também expõe um 'ruído' nos dados: o uso de texto 'boilerplate' (o parágrafo de introdução da Turing) faz com que essas vagas pareçam muito similares entre si e dominem o topo do ranking.

A Prova: O algoritmo, no entanto, funciona como esperado. Outros resultados, como a vaga de 'Azure Data Engineer' (score 0.2430), provam que o sistema está corretamente ranqueando os documentos, mas que as vagas da Turing são matematicamente mais similares à query.