Tela de celular com foto de mulher

Descrição gerada automaticamente com confiança baixa

Vamos apresentar a solução desenvolvida por nós três, Raphael Céspedes, Luciana Menges e Wilsner Sakimoto, ao desafio proposto no terceiro projeto da Residência Tech de Análise de Dados da Harve.

O Projeto consistia em criar robôs em web scrapping para ler dados e criar associações com dados de APIs e gerar apontamentos para onde a empresa Revelo precisa caminhar, em relação ao mercado de vagas de empregos e ferramentas. A Revelo é uma referência em recrutamento de profissionais na área de tecnologia.

Solução proposta pelo grupo:

Diagrama

Descrição gerada automaticamente

Este projeto foi desenvolvido em linguagem Python, utilizando Spyder como IDE e dividido em três fases.

Na **fase 1** foram criadas duas funções. A primeira, com o objetivo principal, de coletar o link das vagas resultantes da pesquisa ao cargo de “Analista de Dados”, no site “vagas.com.br”. A segunda função, abre cada um destes links, coleta e salva algumas informações pré-definidas para a análise posterior.

As informações coletadas foram nove: título da vaga, número de identificação, nome da empresa contratante, nível exigido, data da publicação, salário, local de atuação, descrição da vaga e da empresa.

Na **fase 2** foi trabalhado separadamente com a descrição da vaga, da qual retiramos as palavras relevantes e mais frequentes, fazendo limpeza nas *strings* através da remoção de stop-words, links URL, caracteres especiais, pontuações ou palavras indesejadas para a pesquisa. Gerando desta forma uma lista de palavras-chaves, que detalharemos mais adiante.

A última fase 3 criamos uma nova função “check\_trends”, para acessar o Google Trends e pesquisar as palavras-chaves escolhidas e coletar as informações relativas a cada uma delas do interesse no tempo, por região e tópicos relacionados. Estas informações foram armazenadas e utilizadas para gerar as visualizações que mostraremos na sequência.

Mapa

Descrição gerada automaticamente

Coletamos ao todo 120 links de vagas de emprego, dois deles tivemos que descartar por inconsistências, permanecendo 118, dos quais mostraremos as conclusões aqui.

No mapa podemos visualizar a distribuição destas vagas pelo país, das quais 81% se concentram no Sudeste, sendo São Paulo com o maior volume, seguida por Minas Gerais, Rio de Janeiro e Espirito Santo. A segunda região com maior número de vagas é o Centro Oeste (Goiás e Mato Grosso) e depois o Sul, apenas com quatro vagas indicadas para Curitiba.

Mapa

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente

Utilizamos a cidade indicada em cada vaga, como base, 8% delas indicavam 100% Home Office, ou seja, trabalho remoto.

Mapa

Descrição gerada automaticamente

Do lado direito da tela temos os indicadores de salários, a segunda demanda indicada pelo Product Owner. O detalhe mais interessante a esse respeito é de que 94% das vagas o salário é “a combinar”, no restante aparecem faixas salariais e em uma vaga um valor exato de salário, no caso R$5.527,00. O que demonstrou para nós como as empresas têm receio em mostrar abertamente o salário que oferecem. Um motivo poderia ser pela falta de mão de obra qualificada, então eles se dispõem em pagar mais do que o candidato deseja, para rete-lo. Ou que as empresas não tem um plano de carreira definido previamente e esperam para ver o que os candidatos “querem”.

Entre as sete vagas com faixa salarial fizemos a média salarial, que ficou em R$ 4215. Dividindo a amostra em R$ 4.500, ou seja, a mediana. E por fim mostramos, que destas a maior média salarial ficou no estado de Pernambuco.

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente