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 precisaria caminhar, em relação ao mercado de vagas de empregos e ferramentas. A Revelo é uma referência em recrutamento de profissionais na área de tecnologia.

A solução proposta pelo grupo foi desenvolvida em linguagem Python, utilizando Spyder como IDE e dividido em três fases.

Diagrama

Descrição gerada automaticamente

Na **fase 1** foram criadas duas funções pela Luciana. 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 desses links, **coleta e salva algumas informações** pré-definidas para a análise posterior.

Foram coletadas nove informações: 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,** desenvolvida pelo Wilsner, a função trabalhou com a **descrição da vaga**, separadamente, da qual retirou 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.

Na última **fase 3,** o Rafaelcriou uma função “check\_trends”, para **acessar o Google Trends** e pesquisar as palavras-chaves escolhidas e **coletar as informações** relativas a tendência de interesse de cada uma delas 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 na fase 1, ao todo, as informações de 120 vagas de emprego do site “vagas.com.br”, duas delas tivemos que descartar por inconsistências, permanecendo 118, para a nossa análise.

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

Mapa

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente

Utilizamos, como base, a cidade indicada em cada vaga, das quais 8% 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 este respeito é de que 94% das vagas constam em salário “a combinar”, no restante aparecem faixas salariais e em outra vaga um valor exato de salário de 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 deixando “em aberto” poderiam alterar a proposta de acordo com o candidato. Ou poderia ser um indicativo de que as empresas não têm um plano de carreira definido previamente e esperam para ver o que os candidatos “querem” e nivelam pelo que acham que deve ser.

Entre as sete vagas, com faixa salarial, fizemos a média, que ficou em R$ 4.215. Dividindo a amostra ao meio em R$ 4.500, ou seja, a mediana. E por fim vemos, que a maior média salarial ficou no estado de Pernambuco, com R$ 5.000.

Interface gráfica do usuário, Tabela

Descrição gerada automaticamente

Na fase 2, como dito anteriormente, foi pego a descrição de cada uma das 118 vagas, feita uma limpeza nos termos e gerado uma lista com as 500 palavras utilizadas com maior frequência. A partir dessa lista selecionamos e classificamos, manualmente as palavras-chaves entre “habilidades” e “ferramentas”, para respondermos os outros backlog´s demandados pelo Product Owner. Ou seja, para a fase 3 precisávamos das dez “habilidades” e “ferramentas” mais solicitadas nas vagas de emprego para verificarmos qual a tendência delas na busca no Google Trends em 2021 e 2022.

Algumas palavras isoladamente poderiam ter sido utilizadas em vários contextos diferentes, por isso verificamos em qual contexto elas mais apareceram e os definimos manualmente. Por exemplo a palavra “rotinas”, poderia estar relacionada a “rotinas do departamento xy” ou “rotinas de automação”, que também está relacionada com “automatizar”. Portanto utilizamos-a na fase seguinte como “rotinas de automação”.

Por esse motivo a palavra-chave “negócio” não poderia ser usada sozinha, porque desta forma traria outro tipo de resultado que não o que nós buscávamos. Então utilizamos a como “Inteligência de negócio”. Outro exemplo, “bi” utilizamos “power bi” ou “banco” como “banco de dados”.