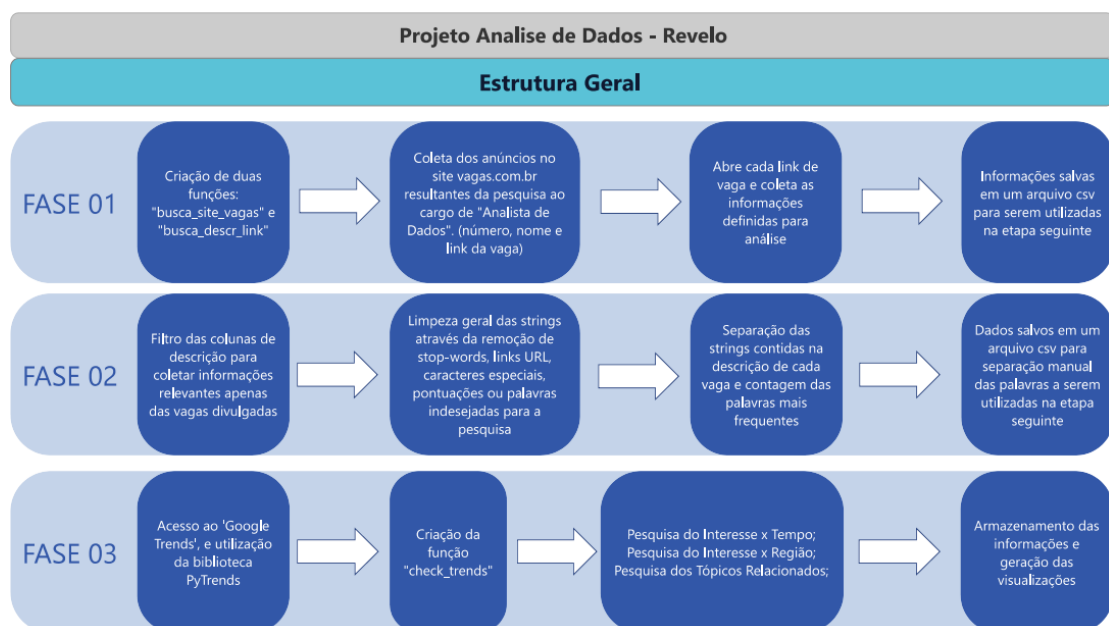




Vamos apresentar a solução desenvolvida por nós três, Raphael Céspedes, Luciana Menges e Wilsner Sakimoto, ao desafio proposto no terceiro projeto da Residência Tech de Análise de Dados da Harve.

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 dividida em três fases.



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.

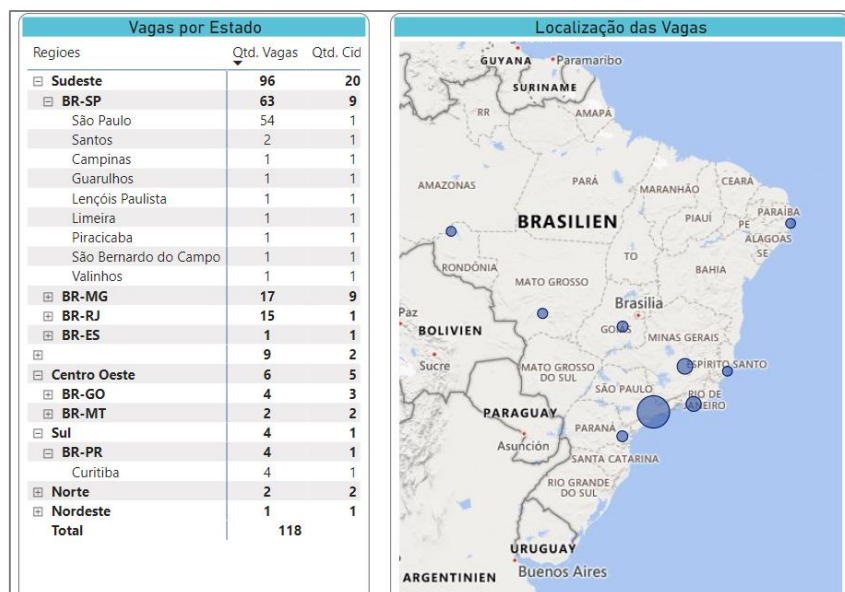
Na **fase 2**, desenvolvida pelo Wilsner, a função trabalhou com a **descrição da vaga**, separadamente das outras, da qual retirou as palavras relevantes e mais frequentes, **fazendo limpeza** nas *strings*, através da remoção de stop-words, links URL, caracteres especiais, pontuações ou palavras indesejadas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

Na última, **fase 3**, o Rafael criou uma função “check_trends”, para **acessar o Google Trends** e pesquisar as palavras-chaves escolhidas e **coletar as informações** relativas à 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 a partir de agora.



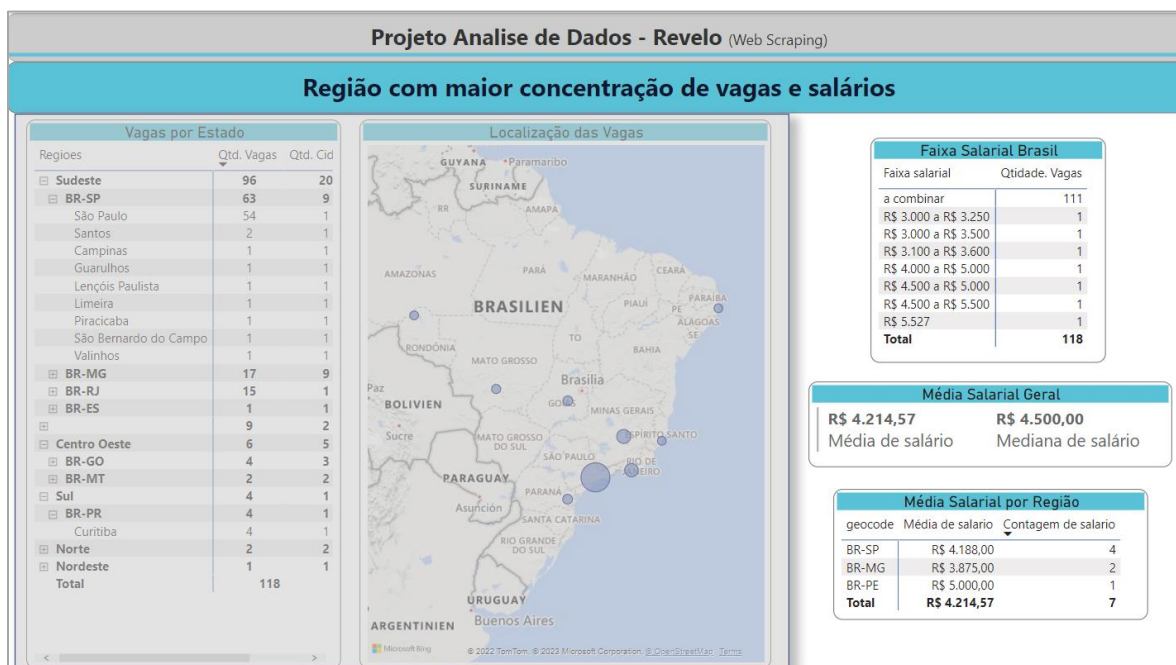
Coletamos na fase 1, ao todo, as informações de 120 vagas de emprego do site “vagas.com.br”, duas delas tivemos que descartar por inconsistências, permanecendo 118, para a nossa análise.

No quadro 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.



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

Vagas por Estado		
Regioes	Qtd. Vagas	Qtd. Cidades
Sudeste	96	20
100% Home Office	9	2
100% Home Office	6	1
Todo o Brasil	3	1
Centro Oeste	6	5
Sul	4	1
Norte	2	2
Nordeste	1	1
Total	118	31



Média Salarial por Região

geocode	Média de salario	Contagem de salario
BR-SP	R\$ 4.188,00	4
BR-MG	R\$ 3.875,00	2
BR-PE	R\$ 5.000,00	1
Total	R\$ 4.214,57	7

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 uma única 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, mesmo sendo uma amostra pequena. Um motivo poderia ser pela falta de mão de obra qualificada, deixando “em aberto” as empresas 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 temos 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.

Projeto Analise de Dados - Revelo (Web Scraping)						
Habilidades & Ferramentas						
Habilidades						
Ind	Word	Frequencia	Retirar	Habilidade	Ferramenta	Tratamento
5	inglês	76		x		
23	negócio	43		x		Inteligência de Negócio
25	indicadores	42		x		indicadores
39	equipe	30		x		
41	rotinas	29		x		rotinas de automação
53	dashboards	26		x		dashboards
65	responsável	25		x		
68	comunicação	24		x		
74	planejamento	23		x		
88	modelagem	20		x		
97	foco	19		x		
105	estatística	18		x		
111	documentação	18		x		
154	inovação	15		x		
164	metodologias	14		x		metodologias ágeis
168	decisões	14		x		
178	etl	13		x		
182	visão	13		x		
193	logística	13		x		
195	learning	12		x		machine learning
204	marketing	12		x		
224	implementação	11		x		
225	ti	11		x		
244	coleta	10		x		Coleta de Dados
245	fluxos	10		x		
249	inteligência	10		x		Inteligência Artificial
268	automatizar	10		x		Rotinas Automação
269	prazos	10		x		
288	extração	9		x		Extração de Dados

Ferramentas						
Ind	Word	Frequencia	Retirar	Habilidade	Ferramenta	Tratamento
14	excel	54			x	excel
19	bi	50			x	power bi
29	banco	39			x	banco de dados
31	office	37			x	pacote office
32	sql	36			x	sql
57	sap	26			x	sap
148	python	15			x	python
185	cloud	13			x	cloud
192	crm	13			x	crm
242	protheus	11			x	protheus
252	analytics	10			x	analytics
271	oracle	10			x	oracle
334	r	8			x	r
335	azure	8			x	azure
351	salesforce	8			x	salesforce
387	linux	7			x	linux

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 2022.

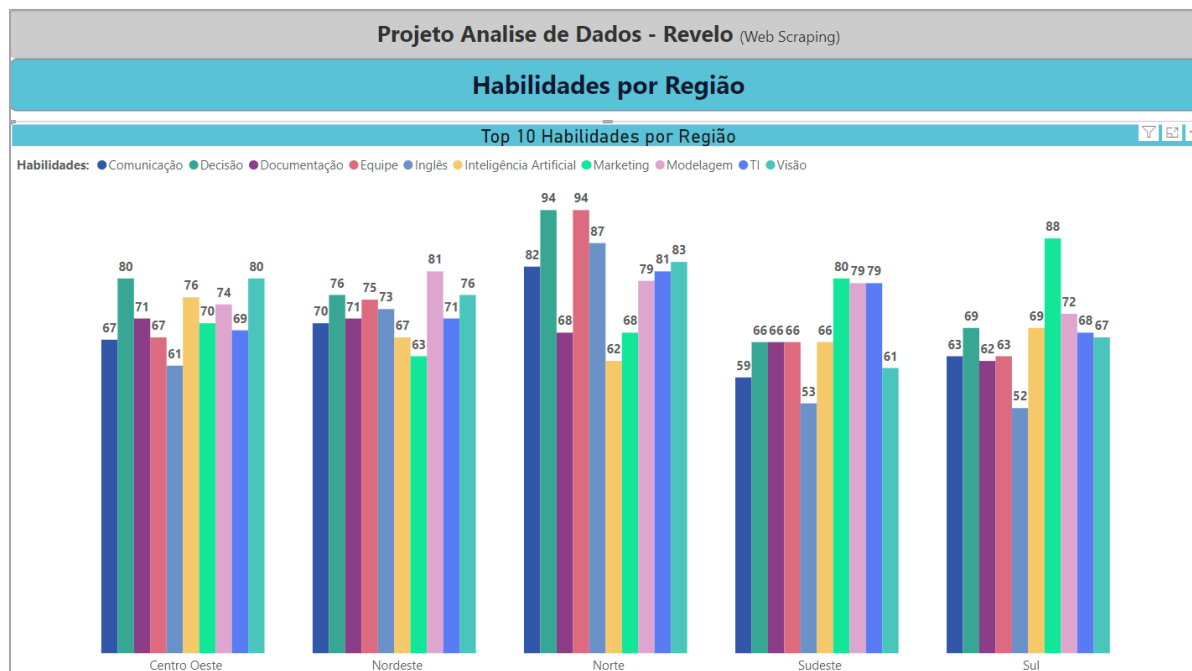
Algumas palavras isoladamente poderiam ter sido utilizadas em vários contextos diferentes, por isso verificamos em qual contexto elas mais apareceram e os definimos manualmente. Por exemplo a palavra “rotinas”, poderia estar relacionada a “rotinas do departamento xy” ou “rotinas de automação”, que também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

A palavra-chave “negócio” não poderia ser utilizada na busca no Google Trends 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”.

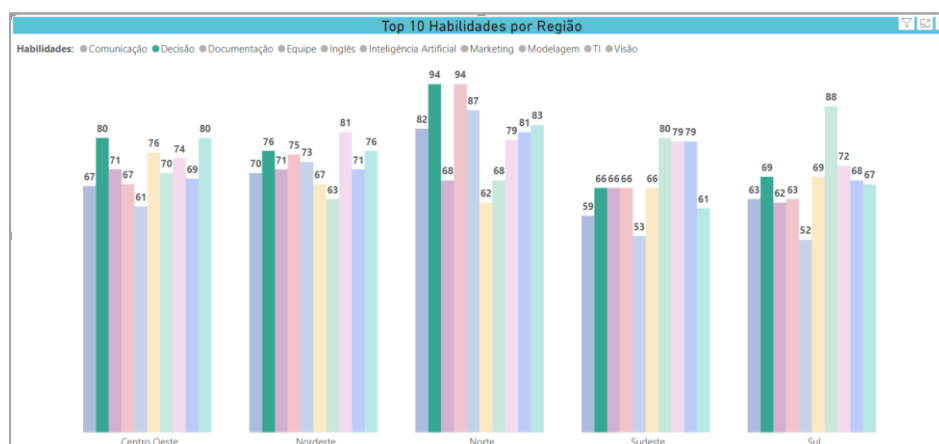
As palavras-chave relativas a “Ferramentas” escolhidas foram: 'Excel', 'Power BI', 'Banco de Dados', 'Pacote Office', 'SQL', 'SAP', 'Python', 'Cloud', 'CRM', 'Protheus', 'Analytics', 'Oracle', 'R', 'Azure', 'Salesforce' e 'Linux'.

As palavras-chave relativas a **“Habilidades”** escolhidas foram: *'Inglês', 'Equipe', 'Responsável', 'Comunicação', 'Planejamento', 'Modelagem', 'Foco', 'Apresentações de Dados', 'Estatística', 'Documentação', 'Inovação', 'Metodologias Ágeis', 'Decisão', 'Prazos', 'Extração de Dados', 'Rotinas', Automação', 'Machine Learning', 'Indicadores', 'Dashboards', 'Logística', 'Marketing', 'Implementação', 'ETL', 'Visão', 'Inteligência Artificial', 'Inteligência de Negócio', 'TI', 'Coleta de Dados' e 'Fluxos'.*

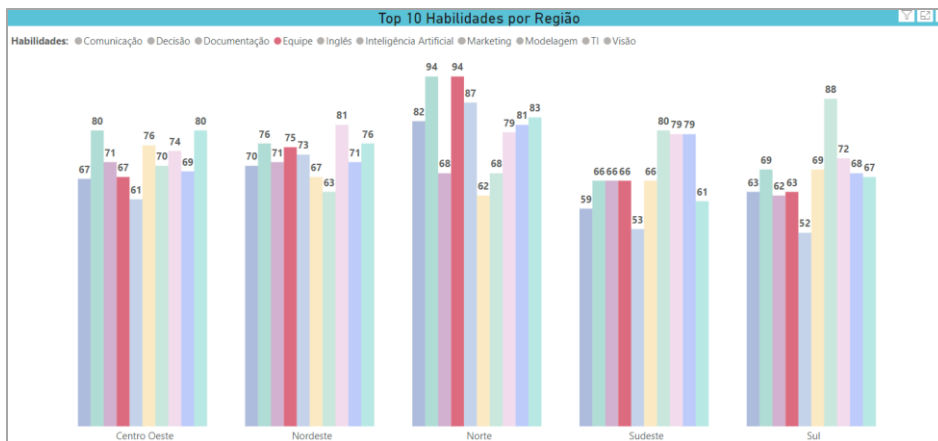
Escolhemos um número maior de palavras, para busca, para não enviesarmos o resultado de alguma forma. E somente para responder cada demanda, selecionamos as dez primeiras.



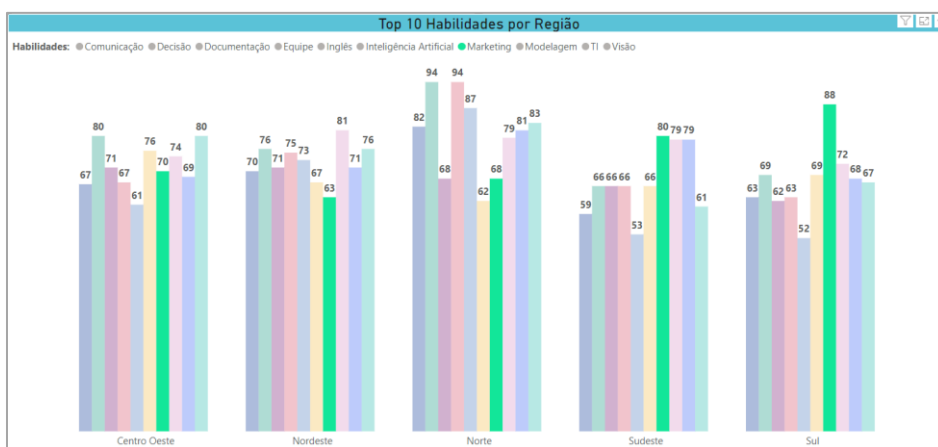
Neste gráfico mostramos as Top dez **“Habilidades”** divididas por região, em ordem alfabética, com os valores medianos para cada um. Não houve grande discrepância, em relação a quais habilidades se destacam entre as top dez.



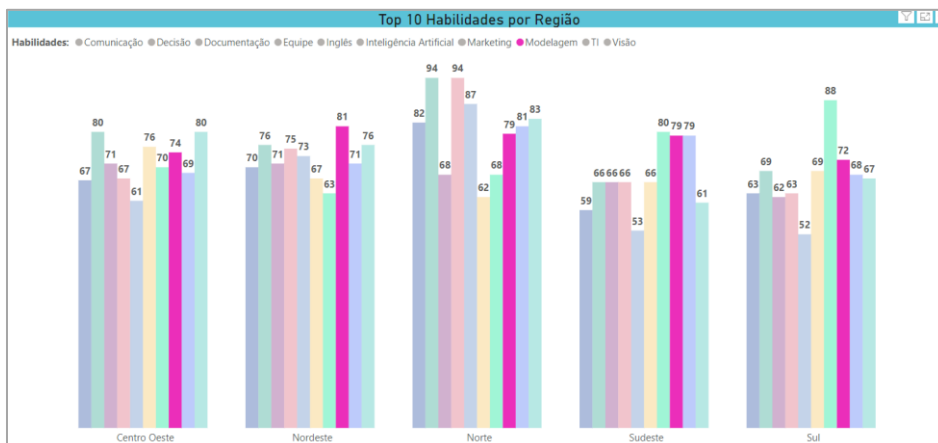
Na região norte houve mais buscas pelas palavras **“Decisão”** e **“Equipe”**. **“Decisão”** ficou também entre as mais procuradas no centro-oeste e nordeste.



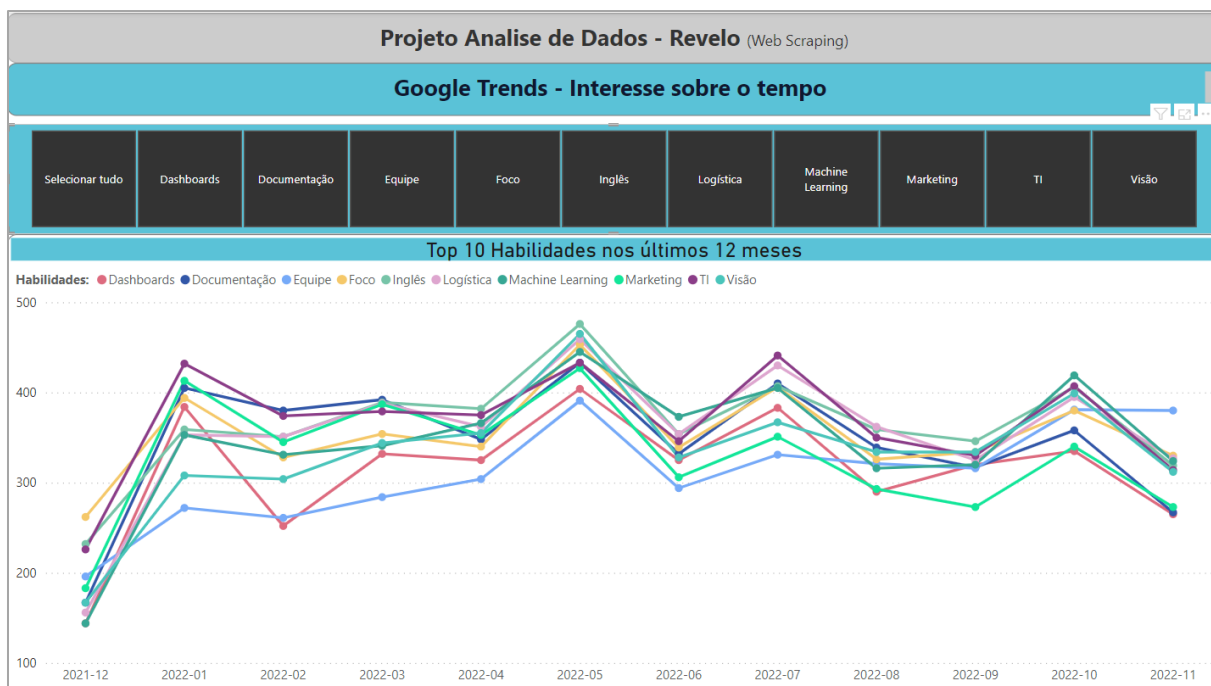
Já a palavra “Equipe” além do Norte se destacou apenas no nordeste, ficando para trás nas buscas nos outros estados.



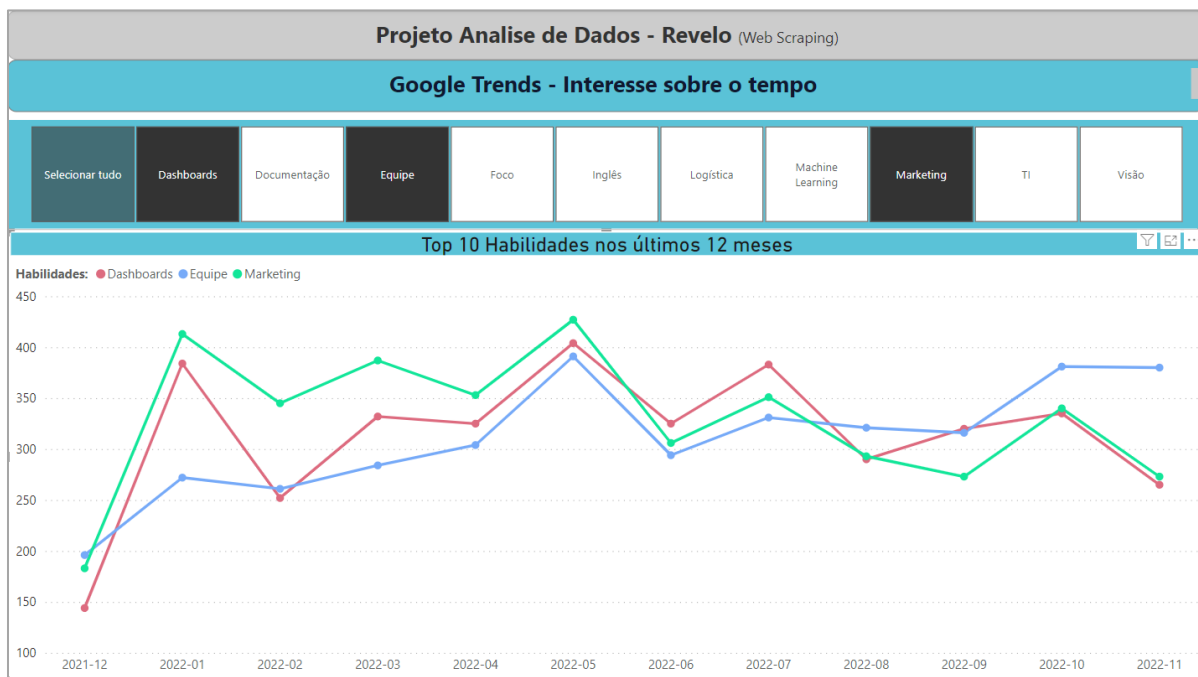
No Sul o destaque foi para a palavra “Marketing”, que teve pontuação similar no Sudeste. E em segundo lugar ficou “Modelagem”



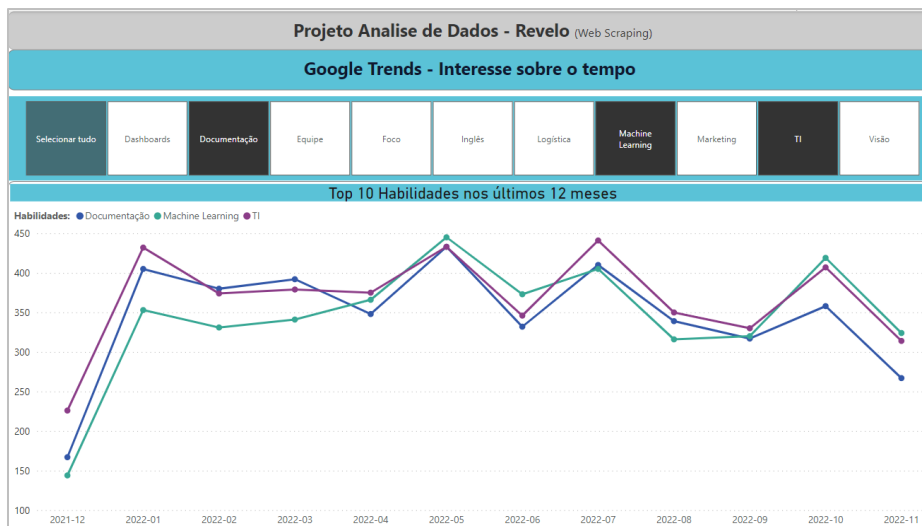
“Modelagem” teve em quase todas as regiões o mesmo nível de busca no período, demonstrando o interesse em todo o país sobre o assunto.



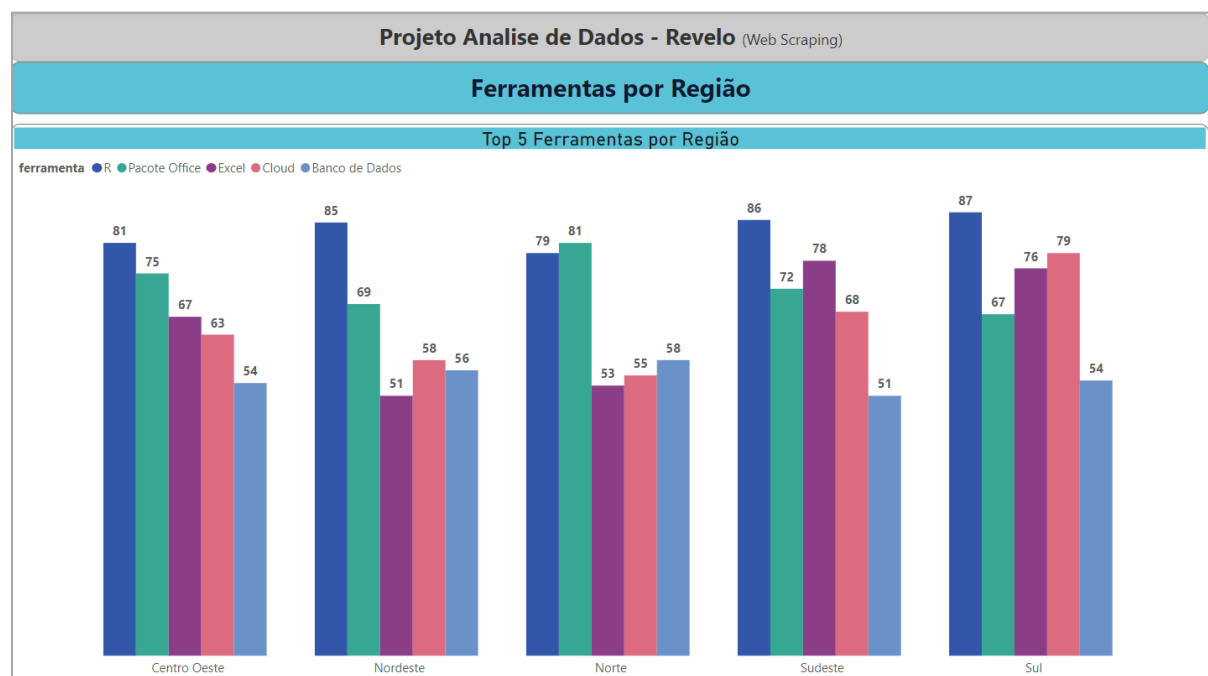
Nesta tela mostramos as dez habilidades mais buscadas ao longo dos últimos doze meses, em valores absolutos mensais, vamos detalhar algumas delas a seguir. Entretanto podemos ver que não há uma grande oscilação entre elas. Nos chama atenção o fato de que todas as dez palavras terminaram o ano de 2021 em baixa, tem o seu pico em maio e depois tem queda em novembro.



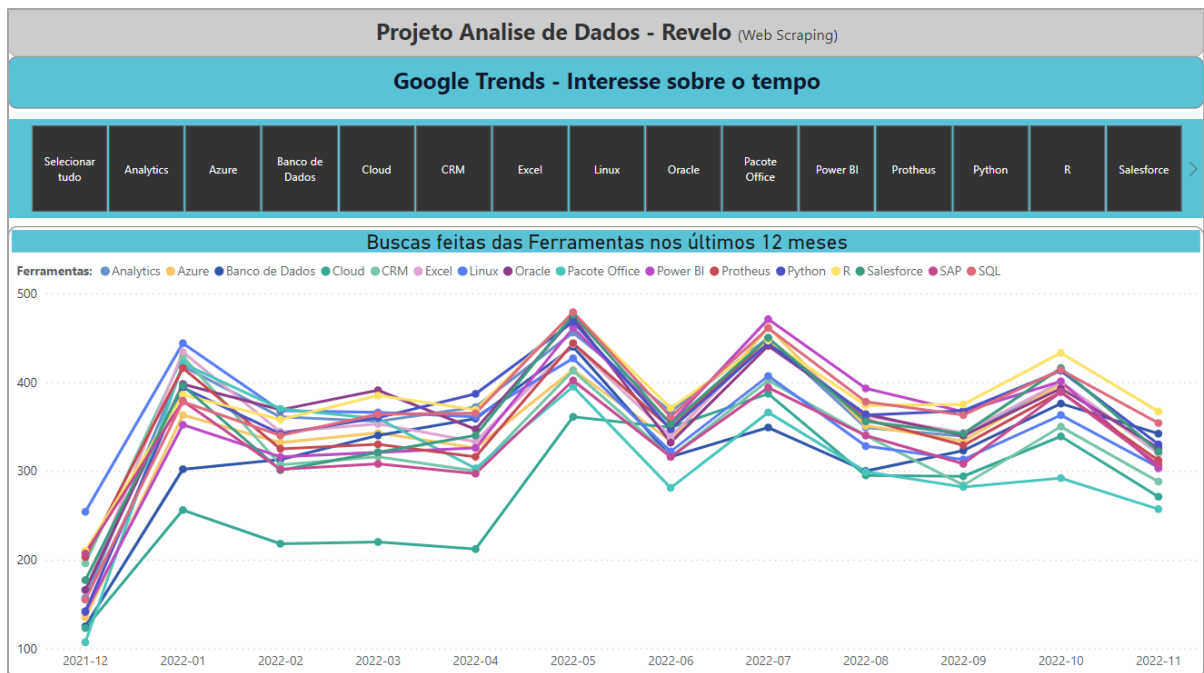
A palavra **“Equipe”** teve uma baixa procura no fim de 2021, com alta sucessiva até maio, depois nova queda, se manteve e finalmente com aumento em outubro, o mesmo de novembro. Esse desempenho poderia ser explicado pela proximidade do final de ano, necessidade de atingimento de metas e consequentemente a melhora do desempenho das equipes. As palavras **“Dashboards”** e **“Marketing”** seguem mais ou menos as mesmas oscilações durante o ano, inclusive acompanham a queda nas buscas em novembro. As pessoas estão mais preocupadas com o desempenho da equipe e não em criar ou desenvolver dashboards, por exemplo.



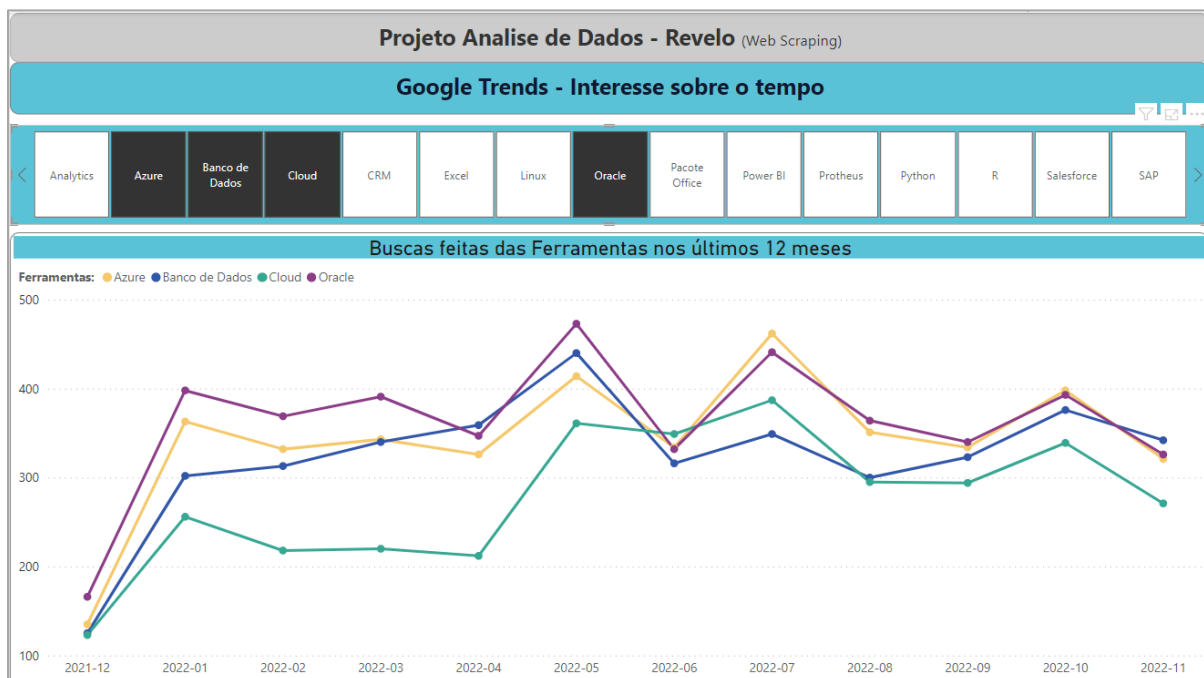
Pegando outras três palavras correlatas vemos o mesmo desempenho ao longo do ano.



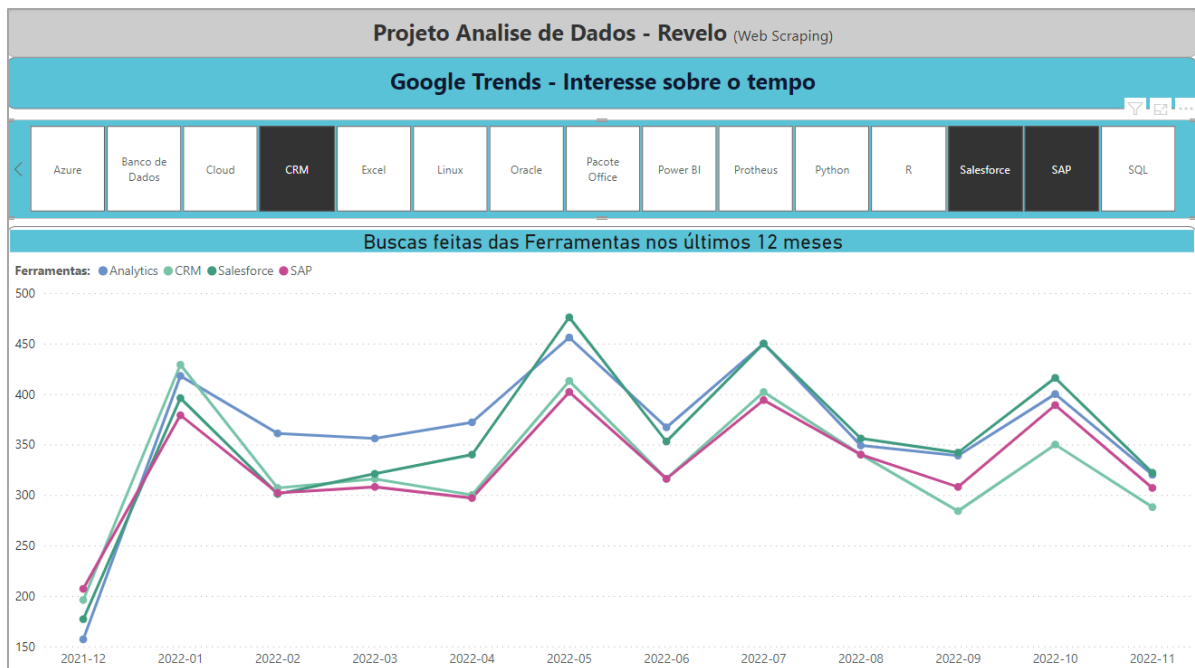
Neste gráfico mostramos as Top dez “Ferramentas” divididas por região, em ordem alfabética, com os valores medianos para cada uma delas. Observamos que a linguagem mais busca é a “R” (colunas em azul escuro), em primeiro lugar em todas as regiões com exceção do Norte onde o “Pacote Office” o ultrapassou. Na região Sul a segunda “Ferramenta” mais procurada foi “Cloud”, seguida por “Excel”. Não mostramos aqui no gráfico, mas a linguagem “Python” ficou entre oitavo e décimo lugar das palavras mais buscadas nestas regiões.



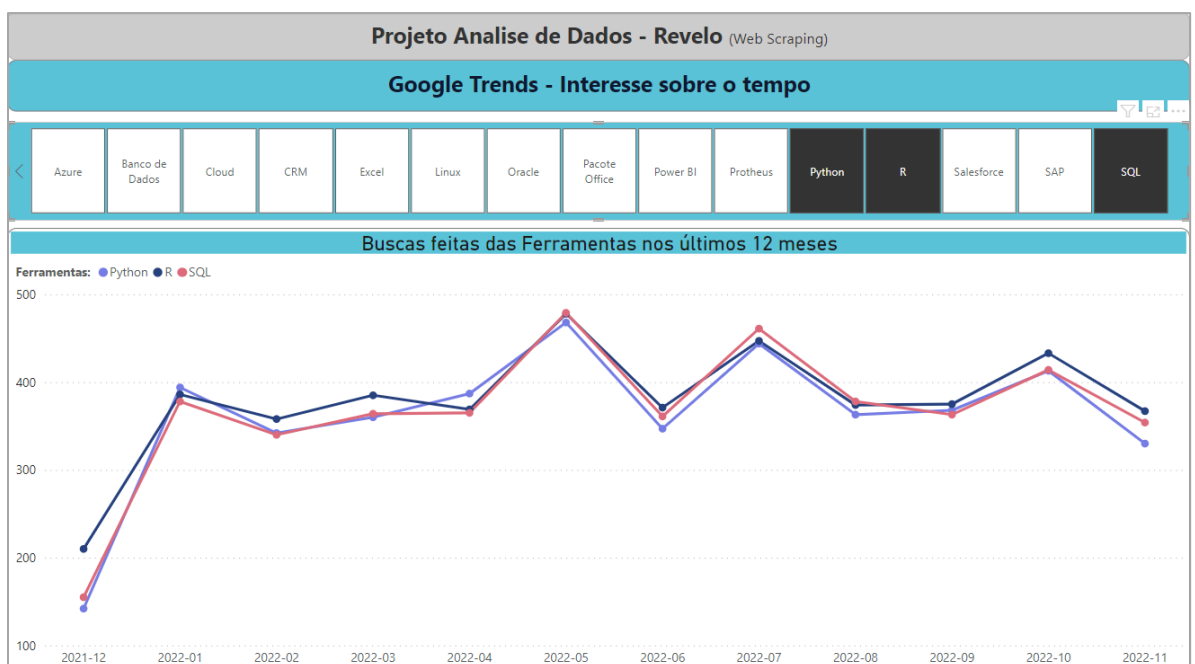
Agora mostramos o comportamento das dezesseis palavras-chave “Ferramentas” ao longo dos últimos doze meses, em valores totais de cada mês. É muito interessante observar como a oscilação ao longo do tempo é bem semelhante entre todas elas, com poucas exceções. Uma delas é a “Cloud”, que podemos ver melhor no slide seguinte.



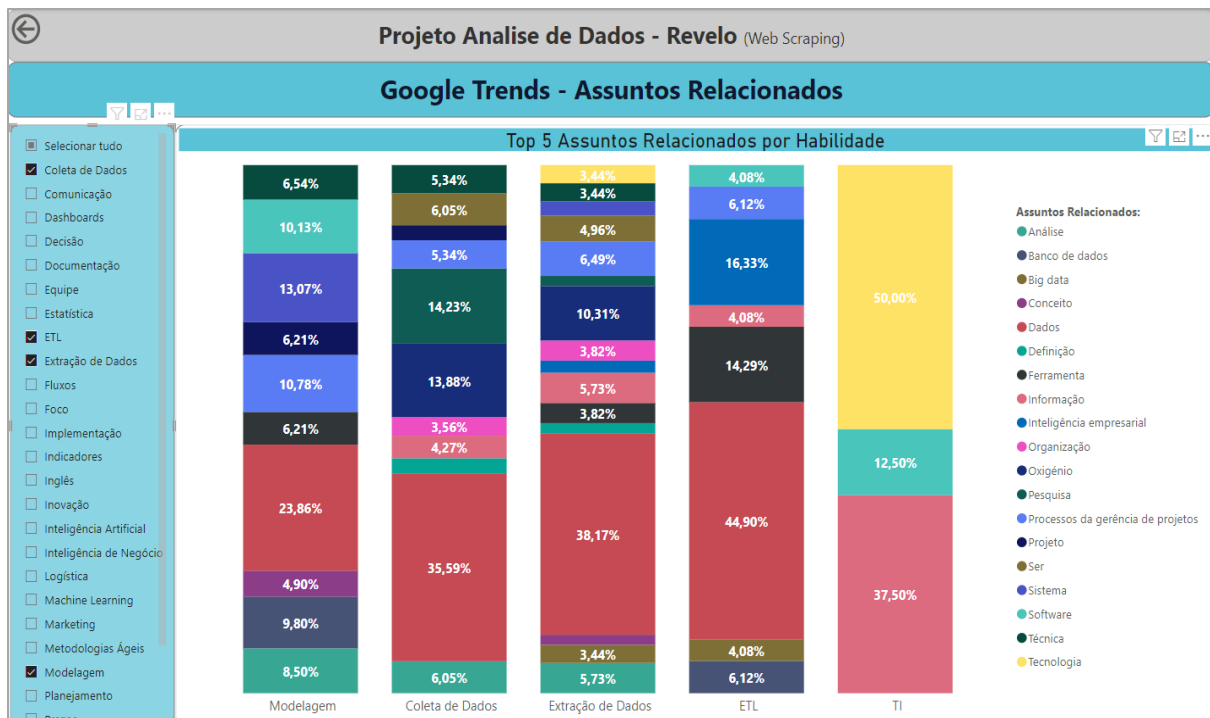
Aqui separamos quatro palavras-chave correlatas com “Cloud”, “Azure”, “Banco de Dados” e “Oracle”. E verificamos como a palavra “Cloud” teve uma procura bem mais baixa do que as outras no início do ano de 2022, mas seguindo a tendência geral no seu maior nível em maio e julho. Concluindo em novembro com queda.



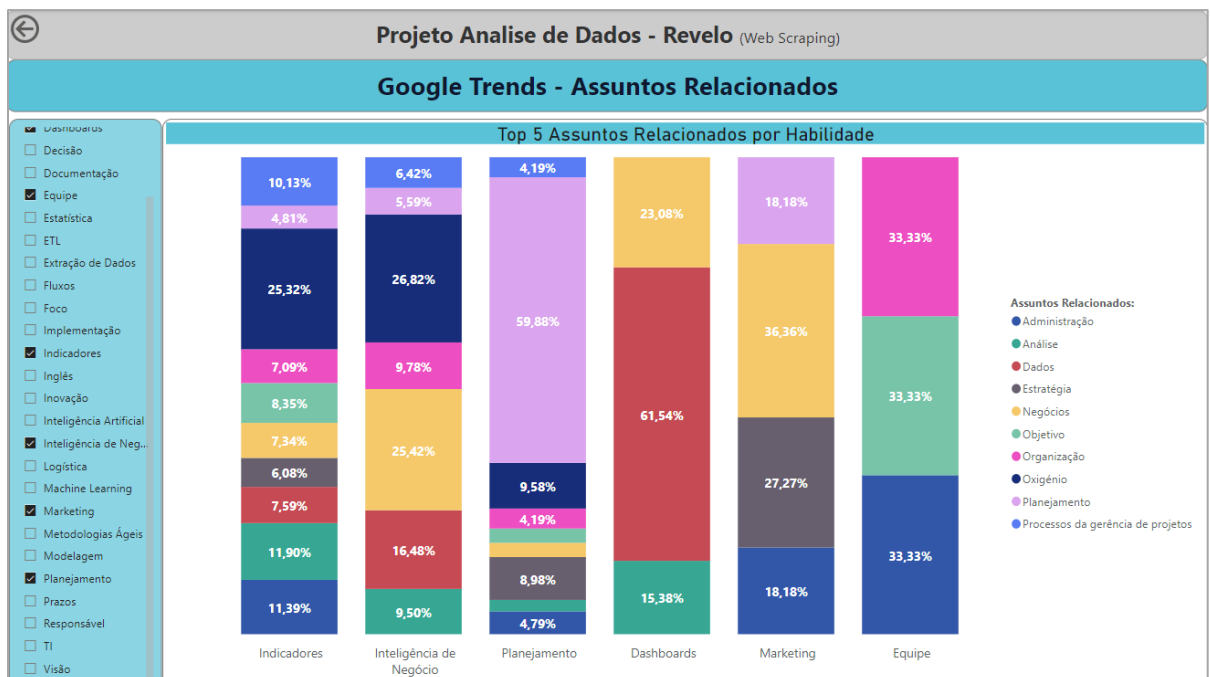
Esse segundo grupo escolhido com palavras-chave correlatas também mostrando a mesma tendência entre “Analytics”, “CRM”, “Salesforce” e “SAP”.



Neste quadro comparamos as linguagens “R”, “Python” e “SQL” e observamos que a quantidade de buscas pelas três são muito próximas. E que “R” é a mais buscada, porém por uma diferença mínima.

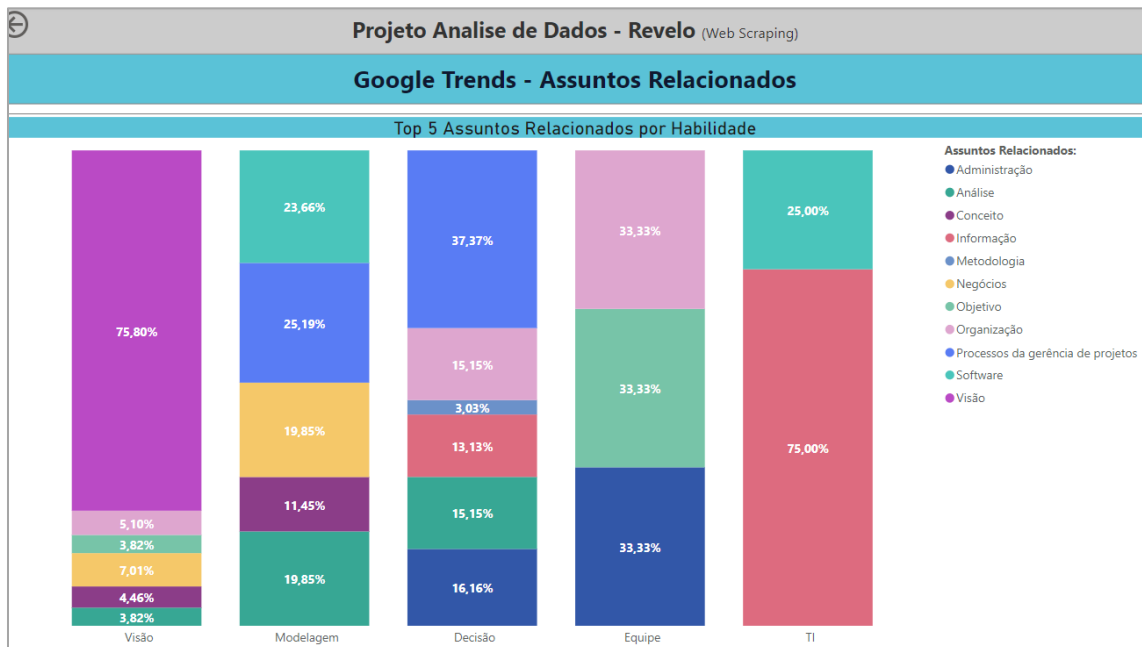


A última demanda que o Product Owner nos solicitou foi para mostrarmos os assuntos relacionados as habilidades buscadas no Google Trends. Nesta tela mostramos cinco habilidades. Por exemplo, com a palavra-chave “Modelagem” em 23,8% das vezes busca-se também “Dados”, que aparece com outras palavras-chave como “Coleta de Dados”, “Extração de Dados” e “ETL”. A palavra-chave “TI” junto com ela busca-se em 50% dos casos “Tecnologia”, em 37,5% Informação e em 12,5% Software.



Aqui temos seis outras habilidades, entre elas “Planejamento”, que está relacionada a ela mesma e outras como “indicadores”, “Inteligência de Negócio” e “Marketing” (Marcada em lilás). Com “Dashboards” busca-se em 61,5% dos casos com “Dados”, em 23% com “Negócios” e em 15,3% com “Análise”. Outro exemplo interessante é a palavra-chave “Equipe” que se apresenta bem dividida,

com 33,33% para cada, entre Administração, Objetivo e Organização. O que acaba reforçando a nossa percepção, comentada anteriormente, de que ela tem uma busca maior no final do ano devido ao fechamento de metas, o que exige mais do desenvolvimento de equipes.



A última demanda que o Product Owner nos solicitou foi para mostrarmos os assuntos relacionados as habilidades buscadas no Google Trends. Nesta tela mostramos cinco habilidades. Como por exemplo, com a palavra-chave "Modelagem" busca-se também Software (23,66%), Processos da gerência de projetos (25,19%), Negócios (19,85%), Conceito (11,45%) e Análise (19,85%). Outro exemplo interessante é a palavra-chave "Equipe" que se apresenta bem dividida, com 33,33% para cada, entre Administração, Objetivo e Organização. O que acaba reforçando a nossa percepção, comentada anteriormente, de que ela tem uma busca maior no final do ano devido ao fechamento de metas, o que exige mais do desenvolvimento de equipes.