1. Apresentação
   1. Meu nome é Gabriel e vou apresentar meu projeto final da formação cientista de dados
2. Visão geral
   1. Os objetivos do meu projeto são: . . .
   2. As etapas do projeto foram:
      1. Extração dos dados utilizando web scrapping
      2. Transformação dos dados
      3. Carregamento dos dados em arquivo CSV e RData
      4. Aplicação de algoritmos de cluster e classificação
      5. Desenvolvimento de dashboard
3. Solução técnica
   1. Para o desenvolvimento do projeto foram utilizadas as seguintes ferramentas:
      1. QGIS, PostgreSQL e PostGIS: foram utilizadas para montar a base de dados geográfica, para elaborar mapas
      2. R: foi utilizado para fazer o web scrapping e todo pre-processamento dos dados e também para aplicar algoritmo de cluster
      3. Python: foi utilizado para desenvolver modelos de classificação
      4. Para o desenvolvimento da plataforma web foram utilizadas as tecnologias HTML 5, CSS, JS, JQUERY, D3.JS, HIGCHARTS e BOOTSTRAP
   2. O processo do ETL foi executado da seguinte forma:
      1. EXTRAÇÃO: foi realizado um web scrapping em sites de vagas de emprego e de avaliações de empresas
      2. TRANSFORMAÇÃO: foram aplicadas várias limpezas e transformações nos dados, dentre elas:
         1. Remoção de caracteres especiais;
         2. Novas variáveis para armazenas skills, idioma, graduação referentes a vaga de emprego, estado, cidade das vagas de emprego e das avaliações sobre as empresas
         3. Conversão das datas de postagem das vagas de emprego e revisões
      3. LOAD: após os processos anteriores, estes dados foram armazenados em CSV para ser utilizado nos dashboards e RData um formato mais leve para disponibilizar para a comunidade;
   3. Descrição dos dados
      1. A base de dados de vagas de emprego é formada por 3.260 registros contendo 13 campos
      2. A base de dados de reviews é formada por 109.606 registros contendo 17 campos
      3. Para desenvolver os dashboard os campos utilizados foram:
         1. Para vagas de emprego: . . .
         2. Para reviews: ...
   4. DATA MINING
      1. Nesse projeto foram aplicados duas técnicas de minieração de dados:
         1. CLUSTER: foi utilizado para agrupar vagas de emprego que possuam requisitos semelhantes
         2. CLASSIFICAÇÃO: foi utilizada para classificara partir do comentário, se aquele funcionário/ex-funcionário recomenda ou não a empresa.
   5. Para fazer cluster de skill de vagas de emprego, dentre as 13 variaveis, foram utilizadas as variáveis: state, skills, edu...
      1. O parâmetro K, valor de cluster, foi de 4;
   6. Para desenvolver o modelo de classificação de reviews, dentre 17 variaveis, foi utilizada a variável Title como variável independente e a variável recommend como variável dependente
   7. Foram desenvolvidos diferentes modelos, como, NB, SVC, SVM e DL, o que obteve melhor resultado foi o modelo de ... com ....
   8. Agora irei apresentar o sistema desenvolvido que está disponível no GITHUB
4. Solução PROPOSTA . . .