
Análise de Dados da Câmara dos Deputados

Objetivo Geral: Este trabalho extensionista consiste em realizar os processos de coleta, preparação e análise de dados sobre os dados abertos da Câmara dos Deputados, que contém dados sobre gastos e atividade parlamentar dos deputados federais.

Quem: Esta atividade deve ser realizada em grupos de 3 ou 4 pessoas. Não serão permitidas entregas individuais.

Quando: Serão realizadas duas apresentações parciais e uma apresentação do trabalho.

- Fase 1: Compreensão do Negócio – Apresentação da questão de pesquisa e etapa de seleção dos dados: [Dia 22/setembro](#)

O que deve ser entregue:

1. Apresentação desenvolvida para essa etapa do trabalho
2. Relatório CRISP-DM referente à etapa de compreensão de negócio
3. Relatório de participação: Documento individual onde cada aluno fará uma reflexão da sua contribuição nesta etapa do trabalho e avaliará o trabalho do grupo como um todo.

- Fase 2: Compreensão dos Dados – Apresentação de uma análise estatística e do planejamento da integração e limpeza dos dados: [Dia 27/outubro](#) (Turma 10) | [Dia 03/novembro](#) (Turma 30)

O que deve ser entregue:

1. Apresentação desenvolvida para essa etapa do trabalho
2. Relatório CRISP-DM referente à etapa de compreensão de negócio e compreensão dos dados
3. Relatório de participação: Documento individual onde cada aluno fará uma reflexão da sua contribuição nesta etapa do trabalho e avaliará o trabalho do grupo como um todo

- Fase 3: Preparação dos Dados – Apresentação dos dashboards: [Dia 01/dezembro](#) (Turma 10) | [Dias 26/novembro e 01/dezembro](#) (Turma 30)

O que deve ser entregue:

1. Link para Github aberto contendo código e notebooks utilizados no trabalho.
2. Versão final do relatório CRISP-DM contendo as três etapas do trabalho
3. Relatório de participação: Documento individual onde cada aluno fará uma reflexão da sua contribuição nesta etapa do trabalho e avaliará o trabalho do grupo como um todo

Onde: Por ser uma atividade extensionista, toda produção do trabalho deve ser disponibilizada de forma *aberta*. Para isto, cada grupo deve ter um repositório no Github contendo seu código e/ou notebooks utilizados para o pré-processamento e análise de dados. Além disso, deve ser enviado o link para o repositório via moodle.

1 Descrição do trabalho

Cada grupo deve escolher um ou mais conjuntos de dados do portal de dados abertos da Câmara dos Deputados, realizar as etapas de pré-processamento, e criar uma visualização para esses dados em formato de dashboard que dê suporte à análise de alguma questão de pesquisa a ser respondida com os dados.

O portal de dados abertos da Câmara dos Deputados pode ser acessado no link a seguir: <https://dadosabertos.camara.leg.br/>. As bases de dados encontram-se em diversos formatos, como API REST, ou arquivos xml, json ou csv, que podem ser baixados diretamente do site. Além disso, há documentação disponível para cada conjunto de dados para que seja possível realizar o entendimento dos dados.

Dentre as principais conjuntos de dados disponíveis, os principais são:

1. Deputados: Dados referentes aos deputados (titulares e suplentes)
2. Dados de Atividade parlamentar: organização política, frentes parlamentares, proposições e votações realizadas
3. Gastos: Informações relativas aos gastos da cota parlamentar de cada deputado.

2 Ferramentas Sugeridas

Sugere-se o uso das ferramentas a seguir:

1. Bibliotecas pandas, scikit-learn e seaborn para o processamento das bases em Python.
2. Tableau ou PowerBI para os dashboards.

3 Sugestões de Trabalhos

Seu grupo é livre para escolher questões de pesquisa e bases de interesse. Entretanto, segue abaixo algumas sugestões de trabalhos.

- Criar métricas de avaliação da atividade parlamentar, considerando os projetos propostos / aprovados, participação nas seções, gasto eficiente de recursos, etc.
- Estudar os fornecedores que prestam serviços para os deputados. Quais as empresas que recebem os maiores valores? As atividades desenvolvidas estão de acordo com a atividade principal da empresa?
- Destrinchar algum tipo de gasto (campo descrição) da cota parlamentar. Será que é possível encontrar outliers dentro desses dados? Quais os deputados que mais realizam esse tipo de gasto?
- Como podemos aumentar os dados de gastos com a cota parlamentar usando as informações das notas fiscais anexadas junto aos dados?