

98H07-04 Visualização de Dados

Professora Isabel Harb Manssour

Especificação do Projeto Final da Disciplina

Descrição Geral do Projeto

Inicialmente, a turma deverá se reunir em trios (como tem 21 alunos matriculados, teremos exatamente sete trios). Um recurso de escolha de grupo foi disponibilizado no Moodle para facilitar esta tarefa. Depois, cada grupo deverá trabalhar na solução do problema proposto.

O objetivo principal deste projeto é proporcionar aos alunos a oportunidade de projetar e implementar uma solução para um problema proposto por profissionais da Empresa Randon: Apresentar visualizações interativas de Divisões de Remessa, Consumo de Materiais e um retrato do Estoque. O problema será detalhadamente descrito pelos proponentes, que apresentarão diferentes desafios a serem tratados pelos grupos, mas os entregáveis de cada parte do projeto são apresentados a seguir.

Parte I

Descrição

A primeira parte do projeto, simplificada, consiste: na análise e preparação dos dados; na definição das perguntas de negócio; na definição do ambiente de desenvolvimento; e no projeto da solução. Cada grupo irá trabalhar com um objetivo diferente, mas o conjunto de dados será o mesmo.

Considerando o processo CRISP-DM, o *pipeline* de visualização de dados e o processo de *visual analytics*, nesta primeira parte do desenvolvimento do projeto devem ser seguidas as seguintes etapas:

1. Entendimento do negócio e do desafio proposto;
2. Entendimento dos dados através de uma análise exploratória visual;
3. Preparação e pré-processamento de dados, que pode também incluir a coleta de outros dados se necessário;
4. Projeto das representações visuais que serão implementadas.

Espera-se que sejam utilizadas as ferramentas apresentadas nas aulas para auxiliar a analisar as características e a qualidade dos dados, e a necessidade de fazer algum pré-processamento ou enriquecimento dos dados.

A partir do problema proposto e da análise dos dados, devem ser definidas as perguntas de negócio, ou seja, o que se quer descobrir a partir das visualizações propostas. Não há um número mínimo e máximo de perguntas, cada grupo deve elaborar a quantidade de perguntas que achar mais adequada. O grupo também deve analisar as ferramentas e bibliotecas disponíveis, para escolher o ambiente de desenvolvimento do trabalho. Alguns exemplos de sistemas, bibliotecas ou pacotes são: Plotly (<https://plotly.com/python/>), D3 (<https://d3js.org/>), Crossfilter (<http://crossfilter.github.io/crossfilter/>), Altair (<https://altair-viz.github.io/>), C3 (<http://c3js.org/>), NVD3 (<http://nvd3.org/>), DC (<https://dc-js.github.io/dc.js/>), Seaborn (<https://seaborn.pydata.org/>), Bokeh (<https://docs.bokeh.org/en/latest/>), Shiny (<https://shiny.posit.co/>), Streamlit (<https://streamlit.io/>), Tableau (<https://www.tableau.com/>) ou PowerBI (<https://powerbi.microsoft.com/>).

A última etapa da parte I do projeto consiste na definição, justificativa e descrição das visualizações interativas que serão implementadas na parte II do projeto. Cada grupo deve explorar diferentes formas de visualização (*heatmaps*, espiral, gráficos de linha, *small multiples*, etc.), **justificar** as escolhas e descrever como os dados serão utilizados/mapeados, identificando qual pergunta de negócio responde e porque a técnica de visualização escolhida é adequada para analisar os dados.

Os entregáveis da parte I do projeto são:

- **Relatório com as seguintes informações:** (1) Nomes dos componentes do grupo; (2) descrição do conjunto de dados, dos *insights* obtidos com a análise exploratória visual (incluindo gráficos) e do enriquecimento dos dados (se realizado); (3) apresentação das perguntas de negócio; (4) descrição das visualizações e interações que serão implementadas com uma **justificativa** para escolha de cada uma e os dados que serão utilizados; (5) lista das ferramentas, sistemas, linguagens ou bibliotecas que serão utilizadas. Não há um formato pré-definido para o relatório, mas todas as informações especificadas devem ser apresentadas de uma forma objetiva. Cada item deve ser descrito acima deve ser colocado como um título. Recomenda-se o uso de folha tamanho A4 com orientação retrato, uma coluna, fonte tamanho 12 e espaçamento simples de parágrafo. Títulos devem estar com uma fonte maior e em negrito.
- **Pitch:** o grupo deve preparar e apresentar um *pitch* de 8 a 10 minutos. O objetivo é mostrar a análise dos dados, apresentar as perguntas de negócio e as vantagens e a inovação da solução proposta para o problema apresentado, justificando as escolhas feitas.

Data

Apresentação do *Pitch* e entrega do relatório no Moodle: 07/05/2024 (durante o horário de aula).

Critérios de avaliação

Serão considerados para avaliação da parte I os seguintes critérios:

- Relatório: inclusão de todo conteúdo solicitado; relevância e profundidade da análise dos dados; relevância, originalidade, eficácia e **justificativa** das visualizações interativas propostas; capacidade de síntese e escrita.
- *Pitch*: qualidade da apresentação (clara e fluente); respeito ao tempo estabelecido; viabilidade da solução proposta; justificativa apresentada.

Parte II

Descrição

A segunda parte do projeto consiste na implementação, teste e validação da solução proposta na parte I.

Considerando o processo CRISP-DM, o *pipeline* de visualização de dados e o processo de *visual analytics*, inicialmente deve ser feito o mapeamento dos dados para as representações visuais especificadas na parte I. A implementação das visualizações e interações projetadas deve ser feita usando as ferramentas, linguagens ou bibliotecas também definidas na parte I. Estas visualizações interativas devem ser organizadas, obrigatoriamente, em um (ou mais) *dashboard* ou página e permitir a exploração dos dados de maneira eficiente e intuitiva, possibilitando responder as perguntas de negócio especificadas. É importante

explorar diferentes formas de visualização, como *heatmaps*, espiral, gráficos de linha e *small multiples*, além de outros gráficos que possam apresentar informações complementares, tais como informações estatísticas. Também devem ser consideradas as boas práticas que serão vistas em aula para o desenvolvimento das visualizações e do(s) *dashboard(s)*, incluindo os *design patterns* para *dashboards* (<https://dashboarddesignpatterns.github.io/>).

Depois, a solução implementada deve ser testada e avaliada para verificar se responde a todas as perguntas de negócio definidas na parte I do projeto. Nesta etapa pode ser necessário fazer alguns ajustes nas visualizações interativas propostas, pois é importante que elas possibilitem extrair as informações necessárias, permitindo solucionar os desafios propostos.

Os entregáveis da parte II do projeto são:

- **Vídeo com uma demonstração da solução proposta:** deve ser preparado um vídeo de no **máximo 2 minutos** para demonstrar as funcionalidades da solução proposta e como ela permite responder às perguntas de negócio e obter *insights*. O vídeo deve ser colocado em algum drive compartilhado ou no YouTube e o *link* para ele deve ser colocado no relatório.
- **Relatório com as seguintes informações:** (1) Nomes dos componentes do grupo; (2) lista de todas as ferramentas, linguagens ou bibliotecas que foram utilizadas; (3) descrição das visualizações e interações implementadas, demonstrando a facilidade de uso das mesmas e os padrões utilizados para o desenvolvimento do(s) *dashboard(s)*; (4) análise dos dados utilizando as visualizações implementadas, incluindo as respostas para as perguntas de negócio e os *insights* obtidos com o uso da solução implementada; (5) discussão dos benefícios e limitações da solução proposta e apresentação de possíveis melhorias como trabalho futuro; (6) *link* para o vídeo com a demonstração da solução proposta. Não há um formato pré-definido para o relatório, mas as todas as informações especificadas devem ser apresentadas de uma forma objetiva. Recomenda-se o uso de folha tamanho A4 com orientação retrato, uma coluna, fonte tamanho 12 e espaçamento simples de parágrafo. Títulos devem estar com uma fonte maior e em negrito.
- **Pitch:** o grupo deve preparar e apresentar um *pitch* de 8 a 10 minutos. O objetivo é mostrar o funcionamento, as vantagens e a inovação da solução implementada, bem como os resultados obtidos (respostas das perguntas de negócio e *insights* identificados). As limitações da solução também devem ser abordadas.

Data

Apresentação do *Pitch* e entrega do relatório no Moodle: 25/06/2024 (durante o horário de aula).

CrITÉRIOS de avaliação

Serão considerados para avaliação da parte II os seguintes critérios:

- **Vídeo:** demonstração das funcionalidades desenvolvidas e dos *insights* que foram obtidos com a solução proposta; respeito ao tempo estabelecido.
- **Relatório:** inclusão de todo conteúdo solicitado; relevância, originalidade e eficácia das visualizações interativas propostas; relevância e profundidade da análise dos resultados obtidos; qualidade do projeto e consistência com a implementação realizada; capacidade de síntese e escrita.
- **Pitch:** qualidade da apresentação (clara e fluente); respeito ao tempo estabelecido; criatividade e qualidade da solução proposta; demonstração das vantagens da solução proposta através dos *insights* obtidos.