

Projeto Aplicado I

Professor Fabio Silva Lopes





Sumário

Diretrizes do Projeto	3
1) Boas-vindas	3
2) Objetivo do projeto	3
3) Etapas do projeto	5
4) Atividades da Etapa 1	6
5) Considerações finais	7

Diretrizes do Projeto 1) Boas-vindas

Seja bem-vindo ao componente curricular "Projeto Aplicado I".

O curso de Ciência de Dados foi concebido considerando uma abordagem prática, muito importante para o desenvolvimento dos skills necessários à formação do cientista de dados. A teoria, quando colocada em prática, estimula nossas habilidades cognitivas e conseguimos mais efetividade no aprendizado.

Por este motivo, teremos quatro componentes de projeto aplicado ao longo do curso. Cada um deles seguirá cobrando a prática de conceitos, técnicas, metodologias aprendidas nas etapas anteriores ou correntes de sua formação.

A ideia é trabalhar com aspectos práticos e reais do cotidiano de um cientista de dados. Enfrentaremos, aqui, questões que exigirão habilidades técnicas, as quais serão comumente demandadas aos profissionais de CD.

Assim, é fundamental seu envolvimento e sua participação nos projetos que proporemos ao longo do curso. Isso complementará sua formação técnica e suas habilidades conhecidas como soft skills, hoje muito apreciadas pelas empresas que poderão te contratar em um futuro próximo.

2) Objetivo do projeto

Neste primeiro trabalho, dividiremos a turma em grupos de quatro alunos para desenvolver um estudo prático sobre os dados de uma empresa, de escolha do grupo de trabalho.

A proposta é escolher uma empresa ou temática de interesse e verificar a disponibilidade de dados que podem ser utilizados no estudo. Podem ser dados públicos ou privados, desde que haja permissão de uso para fins educacionais.

Encorajamos o uso de dados cuja fonte possa ser da empresa que você trabalha, ou de uma empresa pública, a qual disponibiliza dados abertos para pesquisa. Outra possibilidade é explorar sites que disponibilizam dados abertos. Listaremos algumas opções de fontes de dados:

Kaggle: https://www.kaggle.com/>.

· Disponibiliza mais de 50 mil datasets públicos, além de scripts.

World Bank Open Data: https://data.worldbank.org/>.

· Acesso livre para dados sobre desenvolvimento global.

Google Public Data Explorer: https://www.google.com/publicdata/directory.

· Disponibiliza datasets de interesse público.

WHO (World Health Organization) – Open Data Repository: https://www.who.int/data/gho/>.

 Repositório aberto de datasets sobre saúde dos 194 membros da Organização Mundial da Saúde.

Registry of Open Data on AWS – RODA: https://registry.opendata.aws/>.

· Contém datasets públicos, disponibilizados com recursos da AWS.

UCI - Machine Learning Repository: https://archive.ics.uci.edu/ml/index.php.

 Mantido pela Universidade de Irvine, contém mais de 600 datasets para a comunidade de Machine Learning.

IEEE DataPort: https://ieee-dataport.org/>.

· Armazena mais de 2TB de datasets de pesquisa em regime de open access.

Linked Open Data Cloud: https://lod-cloud.net/>.

· Contém mais de 1.200 datasets com links entre eles.

Portal Brasileiro de Dados Abertos: https://dados.gov.br/dataset.

Dados abertos do Governo Federal Brasileiro.

DATA.RIO: https://www.data.rio/>.

Informações sobre a cidade do Rio de Janeiro.

Veja mais opções no link: <https://www.freecodecamp.org/news/https-medium-

-freecodecamp-org-best-free-open-data-sources-anyone-can-use-a65b514b-0f2d/>.

3) Etapas do projeto

Conforme apresentamos na videoaula, nosso projeto terá quatro etapas. As Etapas 1 e 2 serão finalizadas até o início das Provas Intermediárias 1. As Etapas 3 e 4 serão finalizadas no início das Provas Intermediárias 2. São elas:

1- Apresentação do projeto, objetivos, metas e milestones:

· Kick-off do projeto (detalhes na próxima seção).

2- Definição do produto analítico:

- · Elaboração da proposta analítica.
- · Análise exploratória de dados.

3- Apresentação de produtos:

- Como apresentar resultados analíticos.
- Data Storytelling.

4- Encerramento do projeto:

- · Ajustes finais.
- Apresentações.

No início das Etapas 2, 3 e 4, teremos os feedbacks do professor, além de tempo para eventuais ajustes no projeto.

A Etapa 2 consiste em analisar os dados do dataset por meio das técnicas de análise exploratória de dados, executada por meio de scripts em Python.

A Etapa 3 tratará da estruturação dos documentos que serão apresentados no final do projeto, bem como aspectos de narratologia para elaborar a apresentação.

A Etapa 4 deve gerar a apresentação final que consiste em um relatório do projeto, o Github com os materiais utilizados e um vídeo de apresentação gravado

pelo grupo.

O detalhamento de cada etapa será descrito no material de apoio das Aulas 2, 3 e 4, respectivamente.

4) Atividades da Etapa 1

4.1) Montagem dos grupos:

O primeiro passo é organizar um grupo de estudo. Temos um recurso no Moodle para criar grupos. Serão grupos de três ou quatro participantes. Não teremos projetos individuais porque pretendemos trabalhar as habilidades como liderança, empatia e trabalho em equipe. Nas empresas, os profissionais de CD trabalham em squads (equipes). Temos que nos habituar com esta forma de trabalho, na qual compartilhamos conhecimentos e habilidades entre os membros do grupo.

4.2) Escolha da temática:

Na sequência, o grupo deve decidir sobre a temática com a qual trabalhará e o dataset que terá à disposição para o estudo. Aí é necessário um trabalho de garimpo nos links indicados e uma discussão em grupo para definir os rumos do estudo.

4.3) Orientações dos grupos:

Os dois encontros síncronos com o professor poderão contribuir para o alinhamento do projeto e eventuais dúvidas sobre o escopo. Havendo dúvidas sobre a escolha, traga para o encontro as propostas e o professor os ajudará a decidir. Observe que o professor tem um papel importante como orientador do projeto. Por este motivo, teremos mais encontros síncronos ao longo do semestre.

4.4) Comunicação:

Os outros canais de comunicação também devem ser utilizados.

- · Fórum geral.
- · Mensagem para o professor.
- · E-mail para o professor.

Utilizem as plataformas digitais (Teams, Zoom ou Meet) para agendar uma reunião com o grupo. WhatsApp e Discord podem ser bons recursos de comunicação e compartilhamento de dados.

4.5) Organização do material:

Criem um Github para organizar o projeto. Os documentos e resultados poderão ser armazenados e compartilhados lá.

4.6) Cronograma do projeto:

Definam um cronograma mais detalhado, considerando as etapas que foram divulgadas na Videoaula 1, definindo tarefas e responsabilidades para o trabalho não ficar pesado para ninguém. Lembrem-se de que os demais componentes do curso continuam no paralelo e a atenção e dedicação devem ser dosadas para atender a todas as demandas do semestre letivo.

4.7) Entregáveis da etapa:

No segundo encontro síncrono, fechem todos os pontos da etapa com o professor. A entrega da etapa é uma formalização dos itens da lista que segue. O documento deve apresentar:

- Título do trabalho.
- · Membros do grupo.
- Contexto do estudo.
- Referências de aquisição do dataset (origem dos dados, limitações de uso e período da coleta).
- Descrição da origem (informações sobre a organização que gerou os dados e o contexto em que os dados foram gerados).
- Descrição do dataset (o que ele contém, qual é a proposta, quais problemas ou fenômenos foram registrados).

5) Considerações finais

Os casos omissos neste material serão endereçados ao professor, o qual conduzirá os devidos encaminhamentos.

Considerando que o componente é projetual, cada etapa é finalizada com uma entrega, a qual vale nota e somará na média final na seguinte ponderação:

Etapa 1 (1,5) + Etapa 2 (1,5) + Etapa 3 (2,0) + Etapa 4 (5,0)

No caso de o grupo não alcançar a média 6,0, este deverá reapresentar a Etapa 4 com ajustes, a título de substitutiva, atendendo aos prazos de substitutiva do ca-

lendário da Universidade.

A gestão do grupo é responsabilidade do próprio grupo. Casos de desistência ou desentendimento entre os membros deve ser comunicado imediatamente ao professor. Eventuais problemas comunicados somente no final do projeto não serão aceitos como justificativa de não cumprimento das tarefas. É importante lembrar que estamos avaliando as soft skills do grupo como parte do projeto.

Para cada entrega, estejam atentos às rubricas de avaliação disponíveis para a etapa. Elas ajudam a organizar a entrega, antecipando como os itens serão avaliados.

Por fim, esperamos que este projeto ajude em sua formação em Ciência de Dados e traga uma perspectiva de engajamento e motivação para o desenvolvimento das atividades que farão parte de sua vida profissional.

Bons estudos!