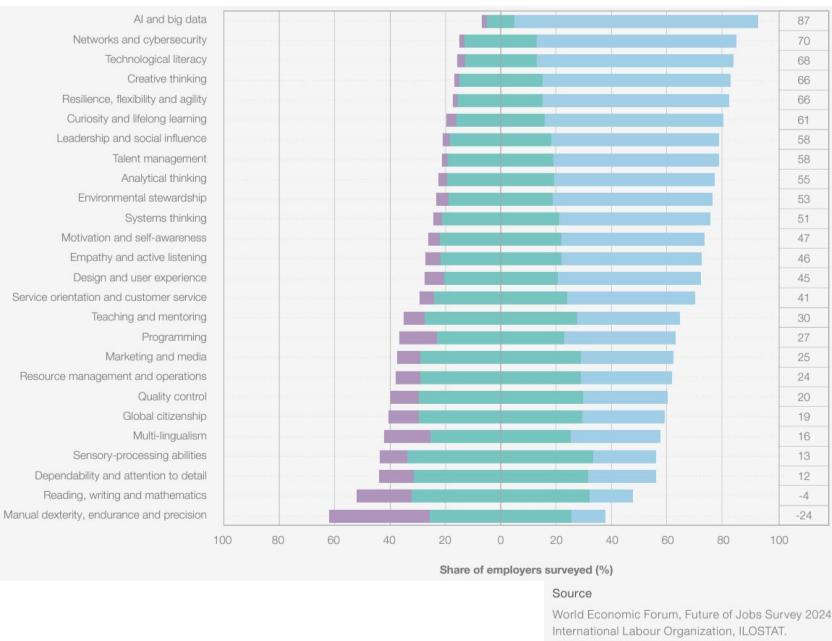

Plano de curso

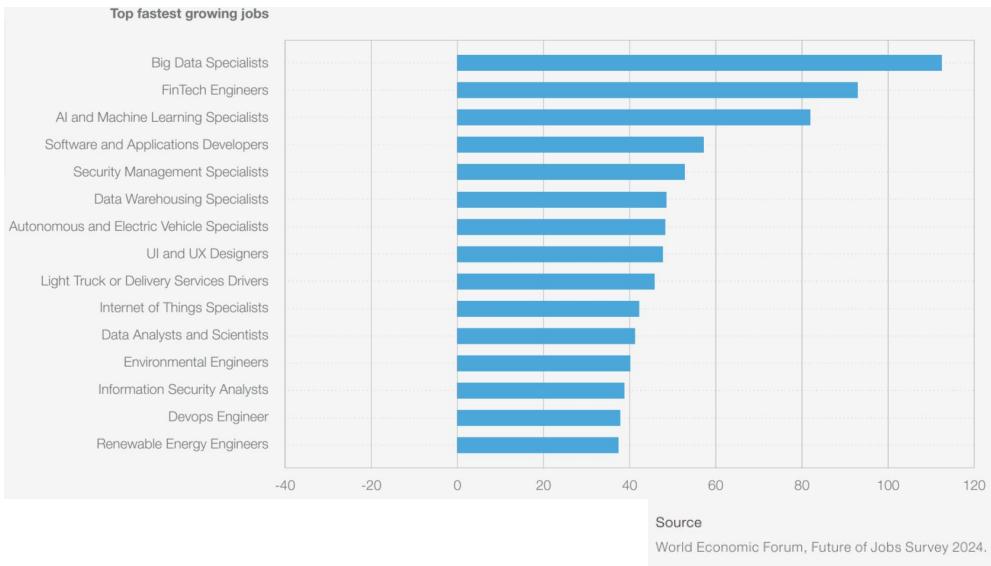
Introdução à Ciência de Dados



Importância de habilidades de 2025 a 2030



Profissões de mais rápido crescimento de 2025 a 2030



Objetivo: Ser capaz de desenvolver e implementar código para caracterizar preliminarmente processos e modelar um banco de dados a partir de dados secundários

Condições:

O estudante terá indicação de **fontes de materiais** para estudo e de arquivos e dados, bem como, acompanhamento e orientação de melhores práticas (ver links no slide seguinte) em sala e oportunidade de usar o laboratório de informática em horários alternativos

Ementa:

1. Captura de dados (de arquivos e api);
2. Tratamento de dados;
3. Visualização de dados;
4. Geração de intervalos de confiança de estatísticas via bootstrapping;
5. Iniciação a banco de dados

Conteúdo

Captura de dados

Manipulação (limpeza) de dados

Redução de séries

Gráficos interativos

Estimativa de intervalos de confiança de estatísticas via Bootstrapping

Fundamentos de Linguagem de consulta estruturada (SQL)

Armazenamento de dados em banco com aplicação de modelagem relacional de objetos

Manipulação e organização de arquivos: binários, csv, zip, xls, json

Uso de dados hierárquicos (abertos): arquivos hdf5, netcdf

Uso de dados abertos: requisição via API

Gráficos estáticos e de séries temporais (data/hora)

Caracterização de disponibilidade de dados (Gráfico de Gantt)

Criação de mapas com espacialização de pontos

Objetivo: Ser capaz de desenvolver e implementar código para caracterizar preliminarmente processos e modelar um banco de dados a partir de dados secundários

Critérios (capacidade, por exemplo, de):

1. Caracterizar a precipitação em uma estação de monitoramento, a partir do hidroweb do SNIRH
(Acesse dados de chuva; Prepare dados em coluna, tendo como índice o par datetime e nível de consistência; Visualize os hietogramas em diferentes discretizações temporais com priorização a dados consistentes; Agregue como hover ao hietograma anual a indicação de número de falhas por ano e a discretização dos dados)
2. Caracterizar a disponibilidade temporal e espacial de dados de uma bacia hidrográfica (Identifique estações de uma região, acesse-os via [API](#) e crie visualização em diagrama de Gantt da abrangência temporal de dados, com indicação de discretização via hover se são dados horários por telemetria ou diários)

3. Comparar os dados obtidos em solo aos obtidos pela estimativa por satélite na missão GPM da Nasa para a mesma região de estudo
4. Apresentar em um mapa a posição das estações do estudo e indique a área (pixel) da informação representada pelo GPM
5. Estimar intervalos de confiança do valor acumulado médio anual de chuva pelas diversas fontes
6. Modelar e armazenar os dados em banco de dados a partir de modelagem relacional de objetos

Projeto: Caracterizar a precipitação em uma estação de monitoramento, a partir do hidroweb do SNIRH

Acessar dados, tratar e os armazenar em dataframe; Gerar visualizações em diferentes discretizações temporais / espaciais; Agregar informações adicionais como hover

Desempenho	Indicadores (Eficiência da descrição do estudo/notebook; Estruturação e eficiência do código (funções/OOP); armazenamento de dados em arquivo; autonomia e independência do(a) estudante; eficiência e diversidade de visualizações de CVDOT do processo; apresentação fundamentada de hipóteses explicativas)
Incipiente	Estudo de difícil leitura e intercalado com código; Código mal estruturado e de difícil entendimento; sem armazenamento de dados em arquivo; desenvolvido a partir da atividade dos colegas; que se limita a descrever poucos aspectos do comportamento do processo; sem apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado
Inferior	Descrição ineficiente intercalada com código desnecessário; Código de difícil entendimento; sem armazenamento de dados em arquivo; desenvolvido com participação intensa dos colegas; que se limita a descrever poucos aspectos do comportamento do processo; sem apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado
Regular	Descrição intercalada com código desnecessário; Código comprehensível; armazena dados em arquivo; desenvolvido a partir de interações com os colegas; que descreve o CVDOT do processo; sem apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado
Razoável	Descrição eficiente intercalada com código desnecessário; Código bem estruturado; armazena dados de forma estruturada com data em arquivo; desenvolvido com mínima participação externa; que descreve o comportamento típico do processo e como este varia ao longo do tempo, em suas diferentes discretizações; apresenta hipóteses explicativas para o comportamento apresentado
Superior	Descrição eficiente que apresenta apenas código que se quer deixar manipulável; código muito bem estruturado e auto-explicativo; armazena dados de forma estruturada com data em arquivo; desenvolvido de forma autônoma e com mínima participação externa; que descreve o comportamento típico do processo e como este varia ao longo do tempo, em suas diferentes discretizações; enriquecido com apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado

Projeto: Caracterizar a precipitação em uma estação de monitoramento, a partir do hidroweb do SNIRH

Acessar dados, tratar e os armazenar em dataframe; Gerar visualizações em diferentes discretizações temporais / espaciais; Agregar informações adicionais como hover

Eficiência da descrição do estudo/notebook:

1. Estudo de difícil leitura e intercalado com código desnecessário
2. Descrição eficiente intercalada com código desnecessário
3. Descrição eficiente que apresenta apenas código que se quer deixar manipulável

Estruturação e eficiência do código (funções/OOP)

1. Código sem funções e de difícil entendimento
2. Código com funções ou com variáveis auto-explicativas e expressões curtas
3. Código com funções, com variáveis auto-explicativas e expressões curtas

Armazenamento de dados em arquivo

1. Não armazena dados em arquivo
2. Armazena dados em arquivo
3. Armazena dados de forma estruturada por data em arquivo

Autonomia e independência do(a) estudante:

1. Desenvolvido com participação intensa dos colegas (inclui professor)
2. Desenvolvido a partir de interações com os colegas
3. Desenvolvido de forma autônoma e com mínima participação externa

Eficiência e diversidade de visualizações de CVDOT do processo:

1. Limita-se a descrever aspectos do comportamento do processo
2. Descreve o CVDOT do processo
3. Descreve o comportamento típico do processo e como este varia ao longo do tempo, em suas diferentes discretizações

Apresentação fundamentada de hipóteses explicativas:

1. Não apresenta hipóteses explicativas para o comportamento apresentado
2. Apresenta hipóteses explicativas para o comportamento apresentado
3. Enriquecido com apresentação de hipóteses explicativas para o comportamento apresentado a partir de outros dados

Desempenho

- Incipiente (1,1,1,1,1): máximo = 3,0
- Inferior (1,1,1,2,2,1): máximo = 5,0
- Regular (2,2,2,2,2,1): máximo = 7,0
- Razoável (2,2,3,2,3,2): máximo = 9,0
- Superior (3,3,3,3,3,3): máximo = 10,0

Projeto: Caracterizar a disponibilidade temporal e espacial de dados de uma bacia hidrográfica

Identifique fontes, acesse-as via API e crie diagrama de Gantt e mapa, com indicação de discretização temporal e percentual de falhas via hover

Desempenho	Indicadores (Eficiência da descrição do estudo/notebook; Estruturação e eficiência do código (funções/OOP); acesso eficiente de dados via API; autonomia e independência do(a) estudante; eficiência e diversidade de visualizações de dados; indica discretização e percentual de falhas via hover)
Incipiente	Estudo de difícil leitura e intercalado com código; Código mal estruturado e de difícil entendimento; não acessa dados via API; desenvolvido a partir da atividade dos colegas; que se limita a descrever disponibilidade temporal ou espacial; não indica discretização, nem percentual de falhas via hover
Inferior	Descrição ineficiente intercalada com código desnecessário; Código de difícil entendimento; não acessa dados via API; desenvolvido com participação intensa dos colegas; que se limita a descrever disponibilidade temporal ou espacial; indica discretização ou percentual de falhas via hover
Regular	Descrição intercalada com código desnecessário; Código comprehensível; acessa dados via API; desenvolvido a partir de interações com os colegas; que descreve disponibilidade temporal e espacial; indica discretização e percentual de falhas via hover
Razoável	Descrição eficiente intercalada com código desnecessário; Código bem estruturado; acessa dados via API; desenvolvido com mínima participação externa; que descreve disponibilidade temporal e espacial; indica discretização e percentual de falhas via hover
Superior	Descrição eficiente que apresenta apenas código que se quer deixar manipulável; código muito bem estruturado e auto-explicativo; acessa dados necessários via API; desenvolvido de forma autônoma e com mínima participação externa; que descreve disponibilidade temporal e espacial; indica discretização e percentual de falhas via hover

Projeto: Caracterizar a disponibilidade temporal e espacial de dados de uma bacia hidrográfica

Identifique fontes, acesse-as via API e crie diagrama de Gantt e mapa, com indicação de discretização temporal e percentual de falhas via hover

Eficiência da descrição do estudo/notebook:

1. Estudo de difícil leitura e intercalado com código desnecessário
2. Descrição eficiente intercalada com código desnecessário
3. Descrição eficiente que apresenta apenas código que se quer deixar manipulável

Estruturação e eficiência do código (funções/OOP)

1. Código sem funções e de difícil entendimento
2. Código com funções ou com variáveis auto-explicativas e expressões curtas
3. Código com funções, com variáveis auto-explicativas e expressões curtas

Acesso eficiente de dados via API

1. Não acessa dados via API
2. Acessa dados via API

Autonomia e independência do(a) estudante:

1. Desenvolvido com participação intensa dos colegas (inclui professor)
2. Desenvolvido a partir de interações com os colegas
3. Desenvolvido de forma autônoma e com mínima participação externa

Eficiência e diversidade de visualizações de dados:

1. Não descreve disponibilidade temporal, nem espacial
2. Limita-se a descrever disponibilidade temporal ou espacial
3. Descreve disponibilidade temporal e espacial

Indica discretização e percentual de falhas via hover:

1. Não indica discretização, nem percentual de falhas via hover
2. Indica discretização ou percentual de falhas via hover
3. Indica discretização e percentual de falhas via hover

Desempenho

- Incipiente (1,1,1,1,1): máximo = 3,0
- Inferior (1,1,1,2,2,1): máximo = 5,0
- Regular (2,2,2,2,2,2): máximo = 7,0
- Razoável (2,2,2,3,3,3): máximo = 9,0
- Superior (3,3,2,3,3,3): máximo = 10,0

Projeto: Comparar os dados obtidos em solo aos obtidos pela estimativa por satélite na missão GPM da Nasa para a mesma região de estudo

Acessar dados, tratar e os armazenar em dataframe; Escolher uma estação de monitoramento em função da sincronia de dados; Gerar visualizações em diferentes discretizações temporais / espaciais

Desempenho	Indicadores (Eficiência da descrição do estudo/notebook; Estruturação e eficiência do código (funções/OOP); Caracterização da precipitação a partir de dados da Nasa quanto ao CVDOT; autonomia e independência do(a) estudante; Eficiência e diversidade de visualizações de CVDOT das diferenças de estimativa; apresentação fundamentada de hipóteses explicativas)
Incipiente	Estudo de difícil leitura e intercalado com código; Código mal estruturado e de difícil entendimento; que se limita a descrever poucos aspectos da precipitação; desenvolvido a partir da atividade dos colegas; que se limita a descrever poucos aspectos de diferenças de estimativa; sem apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado
Inferior	Descrição ineficiente intercalada com código desnecessário; Código de difícil entendimento; que se limita a descrever poucos aspectos do comportamento do processo; desenvolvido com participação intensa dos colegas; que se limita a descrever poucos aspectos das diferenças de estimativa; sem apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado
Regular	Descrição intercalada com código desnecessário; Código comprehensível; que descreve o CVDOT do processo; desenvolvido a partir de interações com os colegas; que descreve o CVDOT das diferenças de estimativa; sem apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado
Razoável	Descrição eficiente intercalada com código desnecessário; Código bem estruturado; que descreve o comportamento típico do processo e como este varia ao longo do tempo, em suas diferentes discretizações; desenvolvido com mínima participação externa; que descreve diferenças típicas de estimativa e como esta varia ao longo do tempo, em suas diferentes discretizações; apresentação hipóteses explicativas para o comportamento apresentado
Superior	Descrição eficiente que apresenta apenas código que se quer deixar manipulável; código muito bem estruturado e auto-explicativo; que descreve o comportamento típico do processo e como este varia ao longo do tempo; desenvolvido de forma autônoma e com mínima participação externa; que descreve diferenças típicas de estimativa e como esta varia ao longo do tempo, em suas diferentes discretizações; enriquecido com apresentação de hipóteses explicativas a partir de outros dados para o comportamento apresentado

Projeto: Comparar os dados obtidos em solo aos obtidos pela estimativa por satélite na missão GPM da Nasa para a mesma região de estudo

Acessar dados, tratar e os armazenar em dataframe; Escolher uma estação de monitoramento em função da sincronia de dados; Gerar visualizações em diferentes discretizações temporais / espaciais

Eficiência da descrição do estudo/notebook:

1. Estudo de difícil leitura e intercalado com código desnecessário
2. Descrição eficiente intercalada com código desnecessário
3. Descrição eficiente que apresenta apenas código que se quer deixar manipulável

Estruturação e eficiência do código (funções/OOP)

1. Código sem funções e de difícil entendimento
2. Código com funções ou com variáveis auto-explicativas e expressões curtas
3. Código com funções, com variáveis auto-explicativas e expressões curtas

Caracterização da precipitação a partir de dados da Nasa quanto ao CVDOT

1. Limita-se a descrever aspectos do comportamento do processo
2. Descreve o CVDOT do processo
3. Descreve o comportamento típico do processo e como este varia ao longo do tempo, em suas diferentes discretizações

Autonomia e independência do(a) estudante:

1. Desenvolvido com participação intensa dos colegas (incluso professor)
2. Desenvolvido a partir de interações com os colegas
3. Desenvolvido de forma autónoma e com mínima participação externa

Eficiência e diversidade de visualizações de CVDOT das diferenças de estimativa:

1. Limita-se a descrever aspectos do comportamento do processo
2. Descreve o CVDOT das diferenças de estimativa
3. Descreve o comportamento típico das diferenças de estimativa e como este varia ao longo do tempo, em suas diferentes discretizações

Apresentação fundamentada de hipóteses explicativas:

1. Não apresenta hipóteses explicativas para o comportamento apresentado
2. Apresenta hipóteses explicativas para o comportamento apresentado
3. Enriquecido com apresentação de hipóteses explicativas para o comportamento apresentado a partir de outros dados

Desempenho

- Incipiente (1,1,1,1,1): máximo = 3,0
- Inferior (1,1,2,1,2,1): máximo = 5,0
- Regular (2,2,2,2,2,1): máximo = 7,0
- Razoável (2,2,3,2,3,2): máximo = 9,0
- Superior (3,3,3,3,3,3): máximo = 10,0

Projeto: Apresentar em um mapa a posição das estações do estudo e indicar a área (pixel) da informação representada pelo GPM

Identifique coordenadas geográficas das estações e dos pixels mais próximo, estime limites de pixel e os apresente em mapa

Desempenho	Indicadores (Eficiência da descrição do estudo/notebook; Estruturação e eficiência do código (funções/OOP); Descrição eficiente de dados em mapa; Autonomia e independência do(a) estudante; Indica abrangência temporal e percentual de falhas via hover)
Incipiente	Estudo de difícil leitura e intercalado com código; Código mal estruturado e de difícil entendimento; não descreve dados em legenda; desenvolvido a partir da atividade dos colegas; que se limita a descrever disponibilidade temporal ou espacial; não indica abrangência temporal, nem percentual de falhas via hover
Inferior	Descrição ineficiente intercalada com código desnecessário; Código de difícil entendimento; mal descreve dados em legenda; desenvolvido com participação intensa dos colegas; indica abrangência temporal ou percentual de falhas via hover
Regular	Descrição intercalada com código desnecessário; Código comprehensível; descreve dados em legenda; desenvolvido a partir de interações com os colegas; indica abrangência temporal e percentual de falhas via hover
Razoável	Descrição eficiente intercalada com código desnecessário; Código bem estruturado; descreve dados em legenda; desenvolvido com mínima participação externa; indica abrangência temporal e percentual de falhas via hover
Superior	Descrição eficiente que apresenta apenas código que se quer deixar manipulável; código muito bem estruturado e auto-explicativo; descreve dados em detalhe na legenda; desenvolvido de forma autônoma e com mínima participação externa; indica abrangência temporal e percentual de falhas via hover

Projeto: Apresentar em um mapa a posição das estações do estudo e indicar a área (pixel) da informação representada pelo GPM

Identifique coordenadas geográficas das estações e dos pixels mais próximo, estime limites de pixel e os apresente em mapa

Eficiência da descrição do estudo/notebook:

1. Estudo de difícil leitura e intercalado com código desnecessário
2. Descrição eficiente intercalada com código desnecessário
3. Descrição eficiente que apresenta apenas código que se quer deixar manipulável

Estruturação e eficiência do código (funções/OOP)

1. Código sem funções e de difícil entendimento
2. Código com funções ou com variáveis auto-explicativas e expressões curtas
3. Código com funções, com variáveis auto-explicativas e expressões curtas

Descrição eficiente de dados em mapa

1. Não descreve dados em legenda
2. Mal descreve dados em legenda
3. Descreve dados em legenda em detalhe

Autonomia e independência do(a) estudante:

1. Desenvolvido com participação intensa dos colegas (inclui professor)
2. Desenvolvido a partir de interações com os colegas
3. Desenvolvido de forma autônoma e com mínima participação externa

Indica abrangência temporal e percentual de falhas via hover:

1. Não indica abrangência temporal, nem percentual de falhas via hover
2. Indica abrangência temporal ou percentual de falhas via hover
3. Indica abrangência temporal e percentual de falhas via hover

Desempenho

- Incipiente (1,1,1,1,1): máximo = 3,0
- Inferior (1,1,2,1,2): máximo = 5,0
- Regular (2,2,2,2,2): máximo = 7,0
- Razoável (2,2,3,3,3): máximo = 9,0
- Superior (3,3,3,3,3): máximo = 10,0

Projeto: Estimar intervalos de confiança do valor acumulado médio anual de chuva pelas diversas fontes

Acessar dados anuais, gerar série de diferenças síncronas, gerar reamostragens, médias e estimar percentis; Gerar visualizações

Desempenho	Indicadores (Eficiência da descrição do estudo/notebook; Estruturação e eficiência do código (funções/OOP); Eficiência de visualizações de intervalos de confiança de diferenças de dados síncronos via bootstrapping; autonomia e independência do(a) estudante)
Incipiente	Estudo de difícil leitura e intercalado com código; Código mal estruturado e de difícil entendimento; não gera visualização de diferenças de dados síncronos; desenvolvido a partir da atividade dos colegas
Inferior	Descrição ineficiente intercalada com código desnecessário; Código de difícil entendimento; não gera visualização de diferenças de dados síncronos; desenvolvido com participação intensa dos colegas
Regular	Descrição intercalada com código desnecessário; Código comprehensível; gera visualização de diferenças de dados síncronos; desenvolvido a partir de interações com os colegas
Razoável	Descrição eficiente intercalada com código desnecessário; Código bem estruturado; gera visualização de diferenças de dados síncronos; desenvolvido com mínima participação externa
Superior	Descrição eficiente que apresenta apenas código que se quer deixar manipulável; código muito bem estruturado e auto-explicativo; gera visualização de diferenças de dados síncronos e descreve como intervalo interquartil se relaciona ao zero; desenvolvido de forma autônoma e com mínima participação externa

Projeto: Estimar intervalos de confiança do valor acumulado médio anual de chuva pelas diversas fontes

Acessar dados anuais, gerar série de diferenças síncronas, gerar reamostragens, médias e estimar percentis; Gerar visualizações

Eficiência da descrição do estudo/notebook:

1. Estudo de difícil leitura e intercalado com código desnecessário
2. Descrição eficiente intercalada com código desnecessário
3. Descrição eficiente que apresenta apenas código que se quer deixar manipulável

Estruturação e eficiência do código (funções/OOP)

1. Código sem funções e de difícil entendimento
2. Código com funções ou com variáveis auto-explicativas e expressões curtas
3. Código com funções, com variáveis auto-explicativas e expressões curtas

Eficiência de visualizações de intervalos de confiança de diferenças de dados síncronos via bootstrapping

1. Não gera visualização de diferenças de dados síncronos
2. Gera visualização de diferenças de dados síncronos
3. Gera visualização de diferenças de dados síncronos e descreve como intervalo interquartil se relaciona ao zero

Autonomia e independência do(a) estudante:

1. Desenvolvido com participação intensa dos colegas (incluso professor)
2. Desenvolvido a partir de interações com os colegas
3. Desenvolvido de forma autônoma e com mínima participação externa

Desempenho

- Incipiente (1,1,1,1): máximo = 3,0
- Inferior (1,1,2,1): máximo = 5,0
- Regular (2,2,2,2): máximo = 7,0
- Razoável (2,2,3,2): máximo = 9,0
- Superior (3,3,3,3): máximo = 10,0

Projeto: Modelar e armazenar os dados em banco de dados a partir de modelagem relacional de objetos

Gerar modelo relacional de dados; Gerar aplicações de armazenar, consultar, atualizar e apagar dados do banco

Desempenho	Indicadores (Eficiência da descrição do estudo/notebook; Estruturação e eficiência do código (funções/OOP); Opera ações no banco; Autonomia e independência do(a) estudante)
Incipiente	Estudo de difícil leitura e intercalado com código; Código mal estruturado e de difícil entendimento; não opera ações no banco ou não é relacional; desenvolvido a partir da atividade dos colegas
Inferior	Descrição ineficiente intercalada com código desnecessário; Código de difícil entendimento; não opera ações no banco ou não é relacional; desenvolvido com participação intensa dos colegas
Regular	Descrição intercalada com código desnecessário; Código comprehensível; opera ações no banco relacional; desenvolvido a partir de interações com os colegas
Razoável	Descrição eficiente intercalada com código desnecessário; Código bem estruturado; opera ações no banco relacional; desenvolvido com mínima participação externa
Superior	Descrição eficiente que apresenta apenas código que se quer deixar manipulável; código muito bem estruturado e auto-explicativo; opera ações no banco relacional; desenvolvido de forma autônoma e com mínima participação externa

Projeto: Modelar e armazenar os dados em banco de dados a partir de modelagem relacional de objetos

Gerar modelo relacional de dados; Gerar aplicações de armazenar, consultar, atualizar e apagar dados do banco

Eficiência da descrição do estudo/notebook:

1. Estudo de difícil leitura e intercalado com código desnecessário
2. Descrição eficiente intercalada com código desnecessário
3. Descrição eficiente que apresenta apenas código que se quer deixar manipulável

Estruturação e eficiência do código (funções/OOP)

1. Código sem funções e de difícil entendimento
2. Código com funções ou com variáveis auto-explicativas e expressões curtas
3. Código com funções, com variáveis auto-explicativas e expressões curtas

Opera ações no banco

1. Não opera ações no banco
2. Não opera ações no banco ou não é relacional
3. Opera ações no banco relacional

Autonomia e independência do(a) estudante:

1. Desenvolvido com participação intensa dos colegas (inclui professor)
2. Desenvolvido a partir de interações com os colegas
3. Desenvolvido de forma autônoma e com mínima participação externa

Desempenho

- Incipiente (1,1,1,1): máximo = 3,0
- Inferior (1,1,2,1): máximo = 5,0
- Regular (2,2,3,1): máximo = 7,0
- Razoável (2,2,3,2): máximo = 9,0
- Superior (3,3,3,3): máximo = 10,0

Gestão de aprendizado e comunicação

Para professor

- Necessidade de permanente atualização
- Necessidade de permanente abertura à interação (com discentes, docentes e monitores e profissionais)
- Necessidade de usar frequentemente a criatividade para reagir na aula às demandas por conhecimento
- Dificuldade para vencer a cultura cômoda de imposição

Para discente

- Necessidade de assumir sua parcela de responsabilidade
- Revisão permanente de atitudes
- Debater e propor soluções com frequência
- Coragem de se submeter a experiências
- Revisar estratégias para aprender sozinho(a)