Trabalho Final: Análise de Dados de Produção com Python

1. Objetivo

Você vai atuar como se fosse um(a) analista de PCP (Planejamento e Controle da Produção) que recebeu dados brutos de uma linha de produção fictícia e precisa responder, com base em evidência numérica e visual: **onde a fábrica está perdendo eficiência, e o que vale a pena investigar primeiro?**

Este trabalho usa as bibliotecas `pandas` (manipulação de dados) e `matplotlib` (visualização), além de reforçar as seguintes habilidades: especificar um problema antes de codar, usar IA de forma crítica e documentada, e testar o próprio código antes de confiar nele.

---

2. Arquivos fornecidos

Você receberá dois arquivos CSV, referentes a um mês de operação de 4 máquinas (M1 a M4):

`producao_mensal.csv` — uma linha por peça produzida (1.281 registros)

Coluna	Descrição
`id_peca`	identificador da peça
`data`	data de produção (AAAA-MM-DD)
`maquina`	máquina que produziu a peça (M1 a M4)
`operador`	operador responsável
`turno`	Manhã, Tarde ou Noite
`timestamp`	data/hora exata da produção
`tempo_ciclo_seg`	tempo de ciclo da peça, em segundos
`status`	Aprovado, Retrabalho ou Descarte

Coluna	Descrição
`defeito`	tipo de defeito identificado (vazio quando o status é Aprovado)

`paradas_mensal.csv` — uma linha por parada de máquina (44 registros)

Coluna	Descrição
`id_parada`	identificador da parada
`data`	data da parada
`maquina`	máquina que parou
`turno`	turno em que ocorreu a parada
`motivo`	motivo da parada (troca de ferramenta, manutenção corretiva, falta de material etc.)
`duracao_min`	duração da parada, em minutos

Os dois arquivos se relacionam pelas colunas `data` e `maquina`. Parte do trabalho exige que vocês combinem as duas fontes, assim como acontece na prática, onde dados de produção e dados de manutenção raramente vêm de uma tabela só.

Carreguem os arquivos com `pd.read_csv("producao_mensal.csv")` e `pd.read_csv("paradas_mensal.csv")`. Cuidado com o tipo da coluna `data`: por padrão o `pandas` vai carregá-la como texto, convertam para `datetime` antes de fazer qualquer análise temporal.

---

3. O que vocês devem entregar

Etapa 0 — Especificação (entrega separada, antes do código)

Antes de abrir o notebook, escrevam um documento curto (meia página é suficiente) respondendo:

- Quais perguntas específicas sua análise vai responder? (mínimo de 4: a lista de perguntas obrigatórias está na Seção 4)

- Para cada pergunta, que tipo de gráfico ou tabela vai responder melhor?
- Que suposições vocês estão fazendo sobre os dados? (Por exemplo: como vocês vão tratar o campo `defeito` vazio: é dado faltante ou é ausência de defeito?)

Essa etapa é entregue e avaliada **separadamente**, antes do restante do trabalho.

Etapa 1 — Preparação dos dados

Carreguem os dois arquivos, corrijam os tipos de dados necessários (especialmente datas) e tratem valores ausentes de forma justificada — não apliquem `dropna()` sem entender o que está sendo descartado e por quê.

Etapa 2 — Análise

Respondam às perguntas obrigatórias da Seção 4, usando `pandas` (`groupby`, `merge`, agregações etc.).

Etapa 3 — Visualizações

Produzam os gráficos obrigatórios da Seção 5, com `matplotlib`. Todo gráfico precisa de título e rótulos de eixo — um gráfico sem esses elementos não comunica nada para quem não foi quem o fez.

Etapa 4 — Log de uso de IA

Se vocês usaram alguma ferramenta de IA (Claude, ChatGPT, Copilot, etc.) em qualquer parte do código, mantenham um registro separado contendo:

- O prompt usado (pode ser resumido, se for muito longo);
- O que a IA gerou;
- O que vocês precisaram corrigir, ajustar ou rejeitar, e por quê.

Isso não é opcional e não é motivo de penalização, pelo contrário, faz parte da nota. O que é penalizado é usar código gerado por IA sem conseguir explicar o que ele faz, ou sem documentar como foi usado.

Etapa 5 — Testes de sanidade

Incluam no código pelo menos 3 verificações (`assert`) que confirmem que os números batem com uma conferência simples e manual. Por exemplo: a soma das peças por status de uma máquina deve ser igual ao total de peças daquela máquina.

Etapa 6 — Relatório final

Um relatório curto (2 a 4 páginas, ou um notebook comentado que cumpra esse papel) com:

- As respostas às perguntas da Seção 4, acompanhadas dos gráficos correspondentes;
- A interpretação dos resultados: não basta mostrar o gráfico, expliquem **o que ele significa** para alguém que vai tomar uma decisão de produção a partir dele;
- O log de uso de IA (Etapa 4), como seção separada.

---

4. Perguntas obrigatórias que sua análise deve responder

1. Qual máquina tem a maior taxa de retrabalho + descarte? (calculem como proporção, não como contagem bruta: máquinas com mais peças produzidas naturalmente têm mais ocorrências em números absolutos.)
2. O tempo de ciclo varia de forma relevante entre turnos ou entre operadores?
3. Quais motivos de parada consomem mais tempo total ao longo do mês?
4. Existe relação entre dias/turnos com mais paradas e dias/turnos com maior taxa de defeito? (Aqui vocês vão precisar combinar os dois arquivos.)

Opcional, para quem quiser ir além (pontuação extra):
5. Calculem uma aproximação simplificada de OEE (Overall Equipment Effectiveness) por máquina.
6. Algum operador se destaca, positiva ou negativamente, em alguma máquina específica?

O **OEE** (*Overall Equipment Effectiveness*, ou **Eficiência Global do Equipamento**) é um indicador de desempenho industrial usado para medir a eficiência real de máquinas e linhas de produção.

Os Três Pilares do OEE

O cálculo do OEE multiplica três fatores essenciais:

- **Disponibilidade:** mede o tempo em que o equipamento funcionou de fato em relação ao tempo total planejado para a produção, descontando paradas (como quebras ou falta de material).
- **Desempenho (Performance):** compara a velocidade real de produção da máquina com a sua velocidade máxima ideal prevista.
- **Qualidade:** calcula a porcentagem de produtos fabricados que estão sem defeitos em relação ao total de itens produzidos (descontando refugos e retrabalhos)

Como é Calculado

A fórmula básica é:

OEE = Disponibilidade × Desempenho × Qualidade

O resultado é dado em porcentagem. Um OEE de 100% significa que a produção é perfeita: produz apenas peças boas, na velocidade máxima e sem nenhuma parada. Na prática,

valores acima de 85% são considerados de classe mundial para a indústria.

---

5. Gráficos obrigatórios

- **Gráfico de barras**: taxa de retrabalho/descarte por máquina.
- **Boxplot**: distribuição do tempo de ciclo por máquina ou por turno. (também conhecido como diagrama de caixa, é uma representação gráfica da distribuição de um conjunto de dados. Ele é composto por uma "caixa", que vai do primeiro quartil (Q1) ao terceiro quartil (Q3), e uma "mediana", que é representada por uma linha dentro da caixa.)
- **Gráfico de Pareto** (barras + linha acumulada): motivos de parada por tempo total perdido.
- **Série temporal**: produção diária ou taxa de defeito diária ao longo do mês.

---

6. Como vocês serão avaliados

Critério	**Peso**
Especificação prévia (Etapa 0) — clareza das perguntas e das suposições assumidas	15%
Correção técnica do código (`pandas`/`matplotlib`)	25%
Qualidade da interpretação: conectar os números a uma decisão de produção, não só descrever o gráfico	25%
Log de uso de IA: honestidade e criticidade sobre o que foi corrigido	15%
Testes de sanidade (Etapa 5)	10%
Defesa oral	10%

Sobre a defesa oral

Na apresentação (5 a 7 minutos por aluno ou dupla), poderei pedir para vocês explicarem qualquer célula do código, inclusive trechos gerados com apoio de IA. Não é uma pegadinha: é exatamente o tipo de pergunta que qualquer engenheiro vai enfrentar ao revisar um código que não escreveu inteiramente sozinho, seja de um colega, seja de uma IA. Se vocês usaram IA

para gerar uma parte do código, o mínimo esperado é conseguir explicar o que aquela parte faz e por que ela está correta.

---

7. Regras sobre uso de IA

O uso de ferramentas de IA é permitido e até incentivado, mas dentro de duas condições:

1. **Tudo precisa ser documentado** (Etapa 4). Um notebook inteiro gerado por IA e entregue sem esse log será penalizado, independentemente de estar correto.
2. **Vocês precisam ser capazes de explicar qualquer linha do código na defesa oral.** Se a IA gerou algo que vocês não entendem, é sua responsabilidade estudar aquele trecho antes da entrega, não apenas confiar que "funcionou".

Não existe problema algum em pedir ajuda de IA para lembrar a sintaxe de uma função do `pandas` ou do `matplotlib`, isso é uso legítimo e eficiente da ferramenta. O problema é delegar para a IA a parte que é justamente o objetivo do trabalho: decidir **quais perguntas fazer aos dados** e **o que os resultados significam**.

---

8. Formato de entrega

- Notebook (`.ipynb`) ou script (`.py`) com o código completo e comentado;
- Documento de especificação (Etapa 0), em arquivo separado;
- Log de uso de IA (Etapa 4), como seção do relatório ou arquivo separado;
- Relatório final (Etapa 6), em PDF ou Markdown.