⊘ Checklist Semanal – Python + Machine Learning (6 meses)

℃ Mês 1 – Fundamentos de Python

	J	
Semana 1		
•	☐ Instalar Python + Jupyter Notebook/VS Code ☐ Variáveis, operadores, tipos de dados ☐ Estruturas de controle (if, for, while)	
Semana 2		
•	Funções em Python Estruturas de dados (listas, dicionários, tuplas, sets) Exercícios de lógica	
Semana 3		
•	Introdução à POO (classes, objetos, herança) Introdução ao NumPy (arrays, operações matemáticas)	
Semana 4		
•	 □ Pandas: criação e manipulação de DataFrames □ Operações básicas de limpeza de dados ★ Projeto: análise de dataset simples (ex.: vendas ou notas de alunos). 	

Mês 2 – Fundamentos de Dados

Semana 5

• SQL básico (SELECT, WHERE, ORDER BY)

Semana 6		
 Pandas avançado (merge, groupby, filtros complexos) Exercícios com datasets reais (Kaggle) 		
 Visualização com Matplotlib (gráficos básicos) Visualização com Seaborn (heatmaps, scatterplots) 		
Semana 8		
 Storytelling com dados (tabelas + gráficos) Projeto: dashboard exploratório (ex.: COVID, esportes, mercado financeiro). 		
Mês 3 – Estatística e Probabilidade Semana 9		
 Média, mediana, moda, variância, desvio padrão Distribuições (normal, binomial) 		
_		
• Distribuições (normal, binomial)		
 Distribuições (normal, binomial) Semana 10 Correlação e regressão linear 		
 Distribuições (normal, binomial) Semana 10 Correlação e regressão linear Visualização de correlação em heatmaps 		
 Distribuições (normal, binomial) Semana 10 Correlação e regressão linear Visualização de correlação em heatmaps Semana 11 Probabilidade básica (independência, eventos condicionais) 		

☐ Mês 4 – Machine Learning Básico		
Semana 13		
 Conceitos de ML: treino, teste, overfitting Regressão Linear no scikit-learn 		
Semana 14		
 Regressão Logística (classificação binária) Métricas de avaliação (accuracy, precision, recall, F1) 		
Semana 15		
 Árvores de decisão e Random Forest Visualização de árvores de decisão 		
Semana 16		
• KNN e K-Means (supervisionado e não supervisionado) ★ Projeto: modelo preditivo (ex.: prever preços de casas ou churn de clientes).		
✓ Mês 5 – ML Avançado		
Semana 17		
 Introdução a Redes Neurais (conceitos básicos) Criar primeira rede com TensorFlow ou PyTorch 		
Semana 18		
 Feature Engineering (seleção/criação de variáveis) Normalização e padronização de dados 		
Semana 19		
 Aprendizado não supervisionado avançado (PCA) Clustering aplicado a datasets reais 		
Semana 20		

•	☐ Processamento de linguagem natural (NLP) ☐ Bag of Words + TF-IDF ★ Projeto: classificação de sentimentos em reviews de filmes.
# N	1ês 6 – Consolidação e Portfólio
Seman	na 21
•	Revisão dos conceitos de Python e ML Praticar desafios em Kaggle
Seman	na 22
•	 □ Escolher dataset para projeto final (churn, recomendação, previsão de demanda etc.) □ Planejar pipeline: coleta → análise → modelo → resultados
Seman	na 23
•	☐ Treinar e otimizar modelos ☐ Comparar métricas de desempenho
Seman	na 24
•	Finalizar projeto, publicar no GitHub Criar README com explicação clara

* Resultado

- Portfólio no GitHub com 5-6 projetos de dados/ML.
- Base sólida em Python, estatística, SQL, machine learning.
- Pronto para aplicar a vagas de Analista de Dados Jr / Cientista de Dados Jr / ML Engineer Jr.

Perfeito, Alexandre então vou montar uma lista de **datasets reais** (e gratuitos) para você usar em cada etapa do plano de 6 meses.

☆ Projeto Final: **pipeline completo de ML** resolvendo problema real.

Assim, além do checklist, você terá **materiais prontos para praticar** e construir seu portfólio no GitHub.

Datasets sugeridos por etapa

7 Mês 1 − Fundamentos de Python

Projeto: análise simples de dados

- Dataset Notas de Alunos (Kaggle)
 - ★ Bom para: médias, desvios, gráficos simples (desempenho por gênero, horas de estudo, etc.).

Mês 2 – Fundamentos de Dados

Projeto: dashboard exploratório

- COVID-19 (Our World in Data)
- NBA Players Stats (Kaggle)
 - ★ Bom para: SQL, Pandas, joins, gráficos evolutivos e comparativos.

Mês 3 – Estatística

Projeto: análise estatística aplicada

- Preços de Imóveis (Kaggle)
 - ★ Bom para: correlação, regressão linear, teste de hipóteses (ex.: área do imóvel afeta preço?).

☐ Mês 4 – Machine Learning Básico

Projeto: modelo preditivo

- <u>Customer Churn Dataset (Kaggle)</u>
 - ★ Bom para: classificação (quem vai cancelar serviço).
- Titanic Dataset (Kaggle)
 - ★ Clássico para treinar modelos de classificação.

6 Mês 5 − Machine Learning Avançado

Projeto: NLP + clustering

- IMDB Reviews (Kaggle)
 - ★ Bom para: análise de sentimentos em texto (positivo/negativo).
- Wholesale Customers Dataset (UCI)
 - ★ Bom para: clustering de clientes.

₩ Mês 6 – Projeto Final

Projeto: pipeline completo de ML (escolha um)

- Retail Sales Forecasting (Kaggle)
 - ★ Previsão de vendas.
- Netflix Shows (Kaggle)
 - * Recomendação de filmes.
- Brazilian E-Commerce Dataset (Olist Kaggle)
 - ★ Análise e previsão de comportamento de clientes (churn, entregas, satisfação).

* Como usar

- Use **Pandas** para carregar os CSVs.
- Crie notebooks organizados → cada projeto em uma pasta do GitHub.
- No README, explique o problema, os passos, os gráficos e os resultados (isso valoriza seu portfólio).