

## Panorama Geral da Solução

O plano a seguir foi desenhado especificamente para você, partindo do seu conhecimento atual. Vamos focar em quatro áreas principais:

1. **Maestria em Python para Dados:** Iremos além do básico do Pandas e introduziremos as principais bibliotecas de visualização de dados (Matplotlib e Seaborn), que são essenciais para qualquer analista.
2. **Estatística Aplicada e Storytelling:** Não basta manipular dados; é preciso saber fazer as perguntas certas e comunicar as respostas de forma eficaz. Aqui, vamos focar em aplicar conceitos estatísticos e em criar visualizações que contam uma história.
3. **Projetos Práticos (End-to-End):** A melhor forma de aprender é fazendo. Dedicaremos uma parte significativa do tempo a projetos completos, desde a coleta e limpeza até a análise e apresentação dos resultados.
4. **Introdução ao Machine Learning:** Como um passo natural na evolução de um analista de dados, vamos explorar os fundamentos de Machine Learning para que você possa começar a criar modelos preditivos.

### Suposições:

- Você tem um ambiente Python configurado (por exemplo, com Anaconda, que já inclui Jupyter Notebook, Pandas, etc.).
- Você está familiarizado com o uso de Jupyter Notebooks, que é a ferramenta padrão para análise de dados interativa em Python.

---

## Seu Plano "100 Days of Code" para Análise de Dados

Vamos começar a transformar dados em decisões!

### Fase 1: Fundamentos Sólidos com Pandas e Visualização (Dias 1-30)

Esta fase foi redesenhada para garantir que você domine o Pandas do zero antes de avançar.

- **Dias 1-15: Pandas - Do Básico ao Intermediário**
  - **Dia 1:** Conceitos Fundamentais: O que é uma Series e um DataFrame? Como criar um DataFrame simples a partir de um dicionário Python.

- **Dia 2:** Lendo Dados: A função mais importante! `pd.read_csv()` e `pd.read_excel()`. Explore os parâmetros essenciais como `sep`, `header` e `decimal`.
- **Dia 3:** Primeira Inspeção: Como "sentir" os seus dados? Use `.head()`, `.tail()`, `.info()`, `.shape` e `.describe()` para ter uma visão geral do seu DataFrame.
- **Dia 4:** Selecionando Colunas: Como acessar uma ou mais colunas do seu DataFrame (`df['coluna']`, `df[['coluna1', 'coluna2']]`).
- **Dia 5:** Selecionando Linhas com `.loc[]` e `.iloc[]`: Entenda a diferença crucial entre selecionar por rótulo (`.loc`) e por posição numérica (`.iloc`).
- **Dia 6:** Filtragem com Condições (Boolean Indexing): A técnica mais poderosa para selecionar dados. `df[df['idade'] > 30]`.
- **Dia 7:** Filtragem com Múltiplas Condições: Combine filtros usando `&` (E) e `|` (OU). Não se esqueça dos parênteses! `df[(df['idade'] > 30) & (df['cidade'] == 'São Paulo')]`.
- **Dia 8:** Manipulação Básica: Adicionar uma nova coluna, deletar uma coluna (`.drop()`) e renomear colunas (`.rename()`).
- **Dia 9:** Lidando com Dados Faltantes (NaN): Como encontrar valores nulos com `.isnull().sum()` e as estratégias básicas para lidar com eles: `.dropna()` e `.fillna()`.
- **Dia 10:** Ordenação de Dados: Use `.sort_values()` para classificar seu DataFrame por uma ou mais colunas.
- **Dia 11:** groupby - O Início: Faça sua primeira agrupação. Entenda o que `df.groupby('categoria').mean()` está fazendo por baixo dos panos.
- **Dia 12:** groupby - Agregações Múltiplas: Use `.agg()` para aplicar várias funções de uma vez (ex: média, soma e contagem).
- **Dia 13:** Juntando DataFrames com `pd.merge()`: A versão em Pandas dos JOINS que você conhece do SQL. Pratique `inner`, `left` e `right` joins.
- **Dia 14:** Concatenando DataFrames com `pd.concat()`: Como "empilhar" tabelas que têm a mesma estrutura.
- **Dia 15:** Mini-Projeto de Consolidação: Pegue um dataset simples, leia-o, inspecione, filtre, ordene, faça uma agregação simples e salve o resultado em um novo CSV.

- **Dias 16-25: Visualização de Dados Essencial**

- (Esta seção continua como antes, mas agora você terá uma base de Pandas muito mais forte para criar os gráficos.)
- **Dia 16:** Introdução à Visualização com Matplotlib e Seaborn.
- **Dia 17-20:** Crie gráficos essenciais: linha, barra, histograma, dispersão (scatterplot) e boxplot.
- **Dia 21-23:** Personalize seus gráficos com títulos, rótulos e cores. Aprenda a usar sub-plots para exibir múltiplos gráficos.
- **Dia 24:** Use heatmap para visualizar correlações e pairplot para uma visão geral rápida do dataset.
- **Dia 25:** Crie sua primeira visualização interativa com Plotly.

- **Dias 26-30: Aquisição de Dados Avançada**

- **Dia 26-27:** Conecte Python ao seu banco de dados SQL e execute queries diretamente para o Pandas.
- **Dia 28-30:** Pratique Web Scraping básico com Requests e BeautifulSoup.

---

## **Fase 2: O Kit de Ferramentas do Analista - Estatística e Storytelling (Dias 26-50)**

Agora vamos adicionar a camada de "análise" de fato sobre a manipulação de dados.

- **Dias 26-35: Estatística Descritiva e Inferencial Aplicada**

- **Dia 26:** Medidas de Tendência Central (Média, Mediana, Moda) e Dispersão (Desvio Padrão, Variância, Quartis).
- **Dia 27:** Entendendo distribuições e como visualizá-las com histogramas e curvas de densidade (KDE).
- **Dia 28:** O que é Correlação? Calcule e visualize uma matriz de correlação. Entenda a diferença crucial: **Correlação não implica Causalidade.**
- **Dia 29:** Testes de Hipóteses: O conceito fundamental. O que são hipóteses nula e alternativa? O que é um p-valor?

- **Dia 30:** SciPy: Use a biblioteca scipy.stats para rodar seu primeiro Teste T, comparando a média de dois grupos.
  - **Dias 31-35:** Prática: Pegue um dataset (ex: Titanic do Kaggle) e formule hipóteses. Ex: "A tarifa paga por sobreviventes foi maior que a de não-sobreviventes?". Use os testes estatísticos para respondê-las.
  - **Dias 36-50: Storytelling com Dados e Mini-Projeto**
    - **Dia 36:** Os princípios do Storytelling: Como estruturar sua análise como uma narrativa (contexto, conflito, resolução).
    - **Dia 37:** Escolhendo o gráfico certo para a sua mensagem.
    - **Dia 38:** Limpando suas visualizações: Removendo ruído (linhas de grade, bordas desnecessárias) para focar na informação.
    - **Dia 39:** Usando cores e anotações para guiar a atenção do seu público.
    - **Dia 40:** Montando um relatório em um Jupyter Notebook usando células de Markdown para explicar sua linha de raciocínio.
    - **Dias 41-50: Mini-Projeto de Análise Exploratória (EDA).** Escolha um dataset do seu interesse (ex: dados da COVID-19, dados de aluguel de bicicletas). Faça uma análise exploratória completa, desde a limpeza, passando pela criação de pelo menos 5 visualizações relevantes, e termine com um relatório em Jupyter Notebook resumindo suas principais descobertas.
- 

### Fase 3: Projetos de Análise de Dados do Mundo Real (Dias 51-80)

É hora de sujar as mãos e construir projetos para seu portfólio.

- **Dias 51-65: Projeto 1 - Análise de Vendas de E-commerce**
  - **Objetivo:** Analisar um dataset de vendas para descobrir os produtos mais vendidos, os melhores clientes e os padrões de vendas ao longo do tempo.
  - **Tarefas:**
    - Obtenção e limpeza de dados (tratar valores nulos, formatos de data, etc.).
    - Análise de vendas por produto, categoria e região.

- Análise de séries temporais para identificar sazonalidade.
  - Identificar clientes de maior valor (RFM - Recência, Frequência, Valor Monetário, se possível).
  - **Entregável:** Um Jupyter Notebook detalhado no seu GitHub, apresentando os insights como se fosse para um gerente de vendas.
- **Dias 66-80: Projeto 2 - Análise de Mercado Imobiliário com Web Scraping**
    - **Objetivo:** Analisar o mercado de aluguel ou venda de imóveis em uma cidade de sua escolha.
    - **Tarefas:**
      - Construir um scraper com Requests e BeautifulSoup para coletar dados de um portal imobiliário (preço, área, número de quartos, bairro).
      - Limpar e estruturar os dados coletados.
      - Analisar a distribuição de preços por bairro.
      - Investigar a correlação entre preço e características do imóvel (área, quartos).
      - **Entregável:** Um repositório no GitHub contendo o código do scraper e o notebook com a análise de mercado.
- 

## **Fase 4: Introdução ao Machine Learning e Tópicos Avançados (Dias 81-100)**

Vamos abrir a porta para a análise preditiva.

- **Dias 81-95: Fundamentos de Machine Learning com Scikit-learn**
  - **Dia 81:** O que é Machine Learning? Aprendizagem Supervisionada vs. Não Supervisionada.
  - **Dia 82:** Seu primeiro modelo: Regressão Linear. Prever um valor numérico (ex: prever o preço de um imóvel baseado na sua área).
  - **Dia 83:** Divisão de dados: A importância de separar em conjuntos de treino e teste (train\_test\_split).
  - **Dia 84:** Métricas de avaliação para regressão:  $R^2$ , Erro Quadrático Médio (MSE).

- **Dia 85:** Classificação: Prever uma categoria (ex: prever se um cliente vai cancelar um serviço ou não).
- **Dia 86:** Seu primeiro modelo de classificação: Regressão Logística.
- **Dia 87:** Métricas de avaliação para classificação: Acurácia, Matriz de Confusão, Precisão e Recall.
- **Dia 88:** Pré-processamento e Engenharia de Features: A importância de preparar os dados para os modelos (escalonamento, tratamento de variáveis categóricas).
- **Dia 89-90:** Construa um modelo de classificação do início ao fim com um dataset como o de Churn (cancelamento) de clientes.
- **Dia 91-92:** Aprendizagem Não Supervisionada: O que é clustering? Use K-Means para encontrar segmentos de clientes em seus dados.
- **Dia 93-95:** Revisão e prática: Aplique um modelo simples de regressão ou classificação ao seu projeto anterior.
- **Dias 96-100: Finalização e Próximos Passos**
  - **Dia 96:** Organizando seu Portfólio no GitHub. Como apresentar seus projetos de forma profissional.
  - **Dia 97:** Explorando ferramentas de BI como Power BI ou Tableau. Veja como conectar seu Python a elas.
  - **Dia 98:** Visão geral de Big Data: O que são Spark e Data Lakes?
  - **Dia 99:** Planejando seus próximos passos: Aprofundar em ML, Cloud (AWS, Azure, GCP), Engenharia de Dados?
  - **Dia 100: PARABÉNS!** Você completou o desafio! Revise seus projetos, veja o quanto evoluiu e celebre sua conquista!

### **Dicas para o Sucesso**

1. **Foco na Pergunta:** A ferramenta é o meio, não o fim. Sempre comece com uma pergunta de negócio ou uma curiosidade. "O que eu quero descobrir com estes dados?"
2. **Seja Curioso:** Explore os dados sem um objetivo fixo. A Análise Exploratória de Dados (EDA) é onde as descobertas mais interessantes acontecem.

3. **Kaggle é seu Amigo:** O site Kaggle tem milhares de datasets, notebooks de outros analistas e competições. É uma fonte inesgotável de aprendizado e prática.
4. **Documente Tudo:** Use as células de Markdown do Jupyter Notebook para explicar cada passo da sua análise. Isso ajuda você a raciocinar e facilita a vida de quem vai ler seu trabalho.