

# Trechos de Código Python Essenciais para Cientistas de Dados: Manipulação e Análise de Dados

Bem-vindo à nossa edição semanal da newsletter de Ciência de Dados da Faculdade de Tecnologia de Votorantim! Nesta edição, exploraremos alguns trechos de código essenciais para a manipulação e análise de dados em Python, utilizando bibliotecas amplamente adotadas como o Pandas e o Scikit-Learn.

Cada trecho de código é acompanhado de uma explicação detalhada e exemplos de como podem ser aplicados na área de Ciência de Dados. Vamos aos códigos...

Nº	Trecho de Código	Explicação	Possíveis Usos em Ciência de Dados
1	<pre>df = pd.read_csv('data.csv')</pre>	Carrega um arquivo CSV em um DataFrame do Pandas.	Importação de dados de diversas fontes para análise e manipulação.
2	<pre>str.upper() str.lower() str.title() str.replace("J", "P")</pre>	Métodos de manipulação de strings para alterar capitalização e substituir substrings.	Limpeza e padronização de dados textuais.
3	<pre>df.dropna(inplace=True), df.fillna(df.mean(), inplace=True) df.drop_duplicates(inplace=True)</pre>	Lida com valores ausentes e remove duplicatas em um DataFrame.	Preparação de dados para análise, garantindo qualidade e consistência.
4	<pre>df.rename(columns={'old_name': 'new_name'}, inplace=True)</pre>	Renomeia colunas de um DataFrame.	Tornar nomes de colunas mais descritivos e amigáveis para análise.
5	<pre>grouped_df = df.groupby('column').agg({'col1': 'sum', 'col2': 'mean'})</pre>	Agrupa dados e aplica funções agregadas.	Análise de dados agregados e geração de estatísticas resumidas.
6	<pre>merged_df = pd.merge(df1, df2, on='common_column', how='inner')</pre>	Mescla dois DataFrames com base em uma coluna comum.	Combinação de diferentes fontes de dados para análise integrada.
7	<pre>pivot = df.pivot_table(values='value', index='index', columns='columns', aggfunc='mean')</pre>	Cria uma tabela dinâmica para resumir dados.	Análise multidimensional e geração de relatórios dinâmicos.
8	<pre>df['new_column'] = df['column'].apply(lambda x: x*2)</pre>	Aplica uma função a uma coluna do DataFrame.	Transformação de dados e criação de novas variáveis derivadas.
9	<pre>mse = mean_squared_error(y_test, predictions) r2 = r2_score(y_test, predictions)</pre>	Calcula métricas de avaliação de modelo, como MSE e R <sup>2</sup> .	Avaliação de performance de modelos de Machine Learning.
10	<pre>import joblib joblib.dump(model, 'model.pkl')</pre>	Salva um modelo treinado em um arquivo.	Persistência de modelos de Machine Learning para uso futuro.
11	<pre>df['date'] = pd.to_datetime(df['date']), df.set_index('date', inplace=True)</pre>	Converte uma coluna para o tipo datetime e define como índice.	Análise de séries temporais e manipulação de dados baseados em tempo.
12	<pre>df['rolling_mean'] = df['column'].rolling(window=12).mean()</pre>	Calcula a média móvel de uma coluna.	Suavização de séries temporais e detecção de tendências.
13	<pre>from sklearn.decomposition import PCA pca = PCA(n_components=2) principal_components = pca.fit_transform(df_scaled)</pre>	Realiza Análise de Componentes Principais (PCA).	Redução de dimensionalidade e visualização de dados.
14	<pre>from sklearn.cluster import KMeans kmeans = KMeans(n_clusters=3) kmeans.fit(df_scaled) df['cluster'] = kmeans.labels_</pre>	Realiza agrupamento K-Means.	Segmentação de dados e descoberta de grupos naturais.
15	<pre>import pandasql as psq query = "SELECT * FROM df WHERE column &gt; value" result = psq.sqldf(query, locals())</pre>	Executa consultas SQL em um DataFrame.	Análise de dados usando sintaxe SQL dentro do ambiente Pandas.

Esperamos que você tenha achado útil esta coleção de trechos de código e suas explicações. A compreensão e aplicação eficaz desses conceitos são fundamentais para a realização de análises de dados robustas e insights valiosos.

Fique atento à próxima edição, onde continuaremos a explorar ferramentas, técnicas e dicas valiosas para aprimorar suas habilidades em Ciência de Dados. Até lá, continue explorando e aprendendo!

Saudações,

Prof. Dr. Dilermando Piva Jr  
Coordenador de Ciência de Dados para Negócios / Fatec Votorantim  
E-mail: f301.cdn@fatec.sp.gov.br