DATA SCIENCE BRAIN
@datasciencebrain

PANDAS CHEAT SHEET for Data Analysis

Save for later reference

01 IMPORTING PANDAS

```
# Import pandas
import pandas as pd
```

02 CREATING DATAFRAMES

```
# From a list
df = pd.DataFrame([['Alice', 25], ['Bob', 30]],
columns=['Name', 'Age'])

# From a dictionary
data = {'Name': ['Alice', 'Bob'], 'Age': [25, 30]}
df = pd.DataFrame(data)

# From a CSV file
df = pd.read_csv('file.csv')
```

03 EXPLORING DATAFRAMES

```
df.head(n=5) # Display first n rows

df.tail(n=5) # Display last n rows

df.describe() # Summary statistics

df.info() # Information about the DataFrame
```

04 INDEXING AND SELECTION

```
df['Name'] # Select a column

df[['Name', 'Age']] # Select multiple columns

df.loc[0] # Select rows by index

df.loc[0, 'Name'] # Select rows and columns by index

df[df['Age'] > 25] # Conditional selection
```

05 DATA CLEANING

df.dropna() # Drop missing values

df.fillna(value) # Fill missing values

df.drop_duplicates() # Drop duplicates

df.rename(columns={'OldName': 'NewName'}, inplace=True) # Rename columns

06 DATA MANIPULATION

df['NewColumn'] = values # Add a new column

df['NewColumn'] = df['OldColumn'].apply(func) # Apply a function to a column

df.groupby('Category').agg({'Column1': 'sum', 'Column2': 'mean'}) # Group by and aggregate

pd.merge(df1, df2, on='key') # Merge DataFrames

07 TIME SERIES

```
# Convert string to datetime
df['Date'] = pd.to_datetime(df['Date'])

# Set datetime as index
df.set_index('Date', inplace=True)

# Resample time series data
df.resample('D').sum()
```

08 PLOTTING

```
# Line plot
df.plot(x='X', y='Y', kind='line')

# Bar plot
df.plot(x='X', y='Y', kind='bar')

# Histogram
df['Column'].plot(kind='hist')
```