# Judul

Isi Materi …

## Pertanyaan

1. **Judul**: Integrasi Data dari Beberapa Sumber  
   **Deskripsi**: Buat kode untuk menggabungkan dataset SKD dan Kesehatan dari SPMB Politeknik Statistika STIS menggunakan operasi merge, kemudian analisis hasil integrasi.  
   **Contoh Kode**:

* import pandas as pd  
  import numpy as np  
    
  # Membaca dataset pertama: SKD  
  skd\_url = 'https://media.githubusercontent.com/media/evanightly/codeasy/refs/heads/main/fastapi/datasource/Data%20Hasil%20SPMB%20Politeknik%20Statistika%20STIS%202024/spmb2024\_1\_skd\_polstatstis.csv'  
  df\_skd = pd.read\_csv(skd\_url)  
    
  # Membaca dataset kedua: Kesehatan  
  kesehatan\_url = 'https://media.githubusercontent.com/media/evanightly/codeasy/refs/heads/main/fastapi/datasource/Data%20Hasil%20SPMB%20Politeknik%20Statistika%20STIS%202024/spmb2024\_3\_kesehatan\_polstatstis.csv'  
  df\_kesehatan = pd.read\_csv(kesehatan\_url)  
    
  # Melihat informasi kedua dataset  
  print("Dataset SKD:")  
  print(f"- Jumlah baris: {df\_skd.shape[0]}")  
  print(f"- Jumlah kolom: {df\_skd.shape[1]}")  
  print("- Kolom:", df\_skd.columns.tolist())  
  print("\nDataset Kesehatan:")  
  print(f"- Jumlah baris: {df\_kesehatan.shape[0]}")  
  print(f"- Jumlah kolom: {df\_kesehatan.shape[1]}")  
  print("- Kolom:", df\_kesehatan.columns.tolist())  
    
  # Persiapan untuk merge  
  # Konversi tipe data kolom kunci jika diperlukan  
  df\_skd['no.ujian'] = df\_skd['no.peserta'].astype(str).str[-5:].astype(int)  
    
  # Menggabungkan dataset dengan inner join pada kolom 'no.ujian'  
  df\_merged = pd.merge(  
   df\_skd[['no.ujian', 'jenis.kelamin', 'skd.total', 'skd.hasil', 'nama.prodi', 'lokasi.formasi']],  
   df\_kesehatan[['no.ujian', 'mtk.nilai', 'psi.nilai', 'kes.nilai']],  
   on='no.ujian',  
   how='inner'  
  )  
    
  # Analisis hasil penggabungan  
  print("\nHasil Penggabungan:")  
  print(f"- Jumlah baris: {df\_merged.shape[0]}")  
  print(f"- Jumlah kolom: {df\_merged.shape[1]}")  
  print("- 5 baris pertama:")  
  print(df\_merged.head())  
    
  # Analisis statistik dari data gabungan  
  print("\nStatistik Nilai SKD vs Nilai Psikotes dan Kesehatan:")  
  correlation = df\_merged[['skd.total', 'mtk.nilai', 'psi.nilai', 'kes.nilai']].corr()  
  print(correlation)  
    
  # Menghitung persentase kelulusan berdasarkan lokasi formasi  
  print("\nPersentase Kelulusan per Lokasi Formasi:")  
  location\_stats = df\_merged.groupby('lokasi.formasi').agg({  
   'skd.hasil': lambda x: (x == 'P').mean() \* 100  
  }).rename(columns={'skd.hasil': 'Persentase Lulus (%)'})  
  print(location\_stats.sort\_values(by='Persentase Lulus (%)', ascending=False))
* **Test Case**:
  + self.assertIn("merge", student\_code)
  + self.assertIn("how='inner'", student\_code)
* **Pre-code**:
* import pandas as pd  
  import numpy as np  
    
  # Membaca dataset pertama: SKD  
  skd\_url = 'https://media.githubusercontent.com/media/evanightly/codeasy/refs/heads/main/fastapi/datasource/Data%20Hasil%20SPMB%20Politeknik%20Statistika%20STIS%202024/spmb2024\_1\_skd\_polstatstis.csv'  
    
  # Membaca dataset kedua: Kesehatan  
  kesehatan\_url = 'https://media.githubusercontent.com/media/evanightly/codeasy/refs/heads/main/fastapi/datasource/Data%20Hasil%20SPMB%20Politeknik%20Statistika%20STIS%202024/spmb2024\_3\_kesehatan\_polstatstis.csv'  
    
  # Tulis kode integrasi data Anda di sini