

Phase 2 Data Loading & Initial Data Inspection

Load Dataset dengan encoding ISO-8859-1 dan membackup dataset asli untuk presevasi

```
# Load dataset
df = pd.read_csv("../data/raw/data.csv", encoding = 'ISO-8859-1')

# backup data asli
df_original = df.copy()
```

Python

Mengecek informasi dataset

```
# initial information about the dataset

print(f'Number of rows: {df.shape[0]}')
print(f'Number of columns: {df.shape[1]}')
print(f'Name of columns: {df.columns.tolist()}')
print(f'type of data: \n{df.dtypes}')
```

Python

```
Number of rows: 541909
Number of columns: 8
Name of columns: ['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate', 'UnitPrice', 'CustomerID', 'Country']
type of data:
InvoiceNo      object
StockCode      object
Description     object
Quantity       int64
InvoiceDate    object
UnitPrice      float64
CustomerID     float64
Country        object
dtype: object
```

df.shape[0] - Jumlah Baris

- Menampilkan total jumlah record dalam dataset
- Berguna untuk mengetahui ukuran dataset

df.shape[1] - Jumlah Kolom

- Menampilkan total jumlah variabel/fitur dalam dataset
- Membantu memahami dimensi data

df.columns.tolist() - Nama Kolom

- Mengkonversi nama kolom menjadi list Python
- Memudahkan melihat semua variabel yang tersedia

df.dtypes - Tipe Data

- Menampilkan tipe data setiap kolom (int64, float64, object, datetime64, dll)
- Penting untuk:
 - Memahami jenis data yang akan dianalisis
 - Mendeteksi kolom yang perlu konversi tipe data
 - Perencanaan preprocessing data

Output

Number of rows: 541909

Number of columns: 8

Name of columns: ['InvoiceNo', 'StockCode', 'Description', 'Quantity', 'InvoiceDate', 'UnitPrice', 'CustomerID', 'Country']

Type of data:

InvoiceNo int64

StockCode int64

Description object

Quantity int64

InvoiceDate datetime64

UnitPrice float64

CustomerID int64

Country object

dtype: object