Analisis Data Eksplorasi (Exploratory Data Analysis/EDA)

Analisis Data Eksplorasi (EDA) adalah proses investigasi awal terhadap dataset untuk menemukan pola, anomali, dan karakteristik penting lainnya. Ini adalah langkah penting sebelum melakukan analisis lebih lanjut atau membangun model. EDA membantu Anda memahami data Anda dengan lebih baik, sehingga Anda dapat membuat keputusan yang lebih baik tepat tentang cara membersihkan, mengubah, dan menganalisisnya.

# Mengapa EDA Penting?

EDA membantu Anda:

* **Memahami struktur dan distribusi data:** Anda dapat mengetahui jenis data yang ada, rentang nilainya, dan bagaimana nilai-nilai tersebut didistribusikan.
* **Mengidentifikasi outlier dan anomali:** Nilai ekstrim atau tidak biasa yang dapat mempengaruhi hasil analisis.
* **Menemukan hubungan antar variabel:** Apakah ada korelasi atau ketergantungan antara variabel-variabel yang berbeda?
* **Membuat hipotesis:** EDA dapat memunculkan pertanyaan menarik yang dapat diuji lebih lanjut dengan analisis yang lebih mendalam.

# Alat dan Teknik EDA

EDA dapat dilakukan dengan menggunakan berbagai alat dan teknik, termasuk:

* **Statistika Deskriptif:** Menghitung ukuran pemusatan (mean, median, modus) dan penyebaran (range, varians, deviasi standar) data.
* **Visualisasi Data:** Membuat grafik dan plot seperti histogram, scatter plot, box plot, dan heatmap untuk memvisualisasikan distribusi data dan hubungan antar variabel.
* **Pengujian Hipotesis:** Menguji apakah suatu dugaan tentang data dapat didukung oleh bukti statistik.

# EDA dengan Visual Studio Code dan Pandas

Visual studio Code adalah editor kode yang populer yang juga mendukung *Jupyter Notebook.* Jupyter Notebook adalah lingkungan interaktif yang memungkinkan anda menulis dan menjalankan kode Python, serta melihat output dan visualisasi secara langsung dalam browser. Pandas adalah library Python yang kuat untuk manipulasi dan analisis data.

### Tugas Hari 4:

Lanjutkan dengan dataset Titanic yang telah Anda gunakan pada Hari 3.

#### Tugas:

1. **Statistik Deskriptif:** Gunakan pandas untuk menghitung beberapa statistik deskriptif (mean, median, modus, range, varians, deviasi standar) dari kolom numerik yang relevan (misalnya, Age, Fare). Tampilkan hasilnya.
2. **Visualisasikan Data:** Buat setidaknya 2 jenis plot (misalnya, histogram, box plot, scatter plot) menggunakan Pandas atau library visualisasi lainnya (misalnya, Matplotlib, Seaborn) untuk memvisualisasikan distribusi data dan hubungan antar variabel. Berikan interpretasi singkat tentang plot yang Anda buat.

**Contoh Kode (Pandas):**

import pandas as pd

# Membaca dataset Titanic

titanic\_df = pd.read\_csv("titanic.csv")

# Statistik deskriptif

print(titanic\_df.describe())

# Histogram kolom Age

titanic\_df['Age'].hist()

# Scatter plot antara Age dan Fare

titanic\_df.plot.scatter(x='Age, y=’Fare’)

**Catatan:**

* Pastikan Anda telah menginstall library pandas dan library visualisasi yang di pilih (Matplotlib atau Seaborn) di lingkungan python Anda.

**Selamat mengerjakan tugas!**

### Tugas 4:

1. Statistik Deskriptif:

# Statistika deskriptif

print(titanic\_df['Age'].describe().mean()) # Mean

print(titanic\_df.describe().mean().median()) # Median

titanic\_df['Age'].describe().mean()

print(titanic\_df['Fare'].describe().median) # Median

# range

print(titanic\_df['Age'].rank()) # Apakah ini range? wkwk

print(titanic\_df['Fare'].rank())

# modus

print(titanic\_df['Age'].mod) # Apakah ini modus? I Don't Know!!!

print(titanic\_df['Fare'].mode) # Ouh Shit!!! Module mod sama mode kok sama outputnya?

# varians

print(titanic\_df['Age'].var) # wah-wah ini varians kah? wkwk

print(titanic\_df['Fare'].var)

# deviasi standar

print(titanic\_df['Age'].divide) # Well!!! Kenapa outputnya sama modul var ['age']?

print(titanic\_df['Fare'].std)

1. Visualisasi Data:

# Visualisasi Data

import matplotlib as plt

import seaborn as sns

# scatterplot

titanic\_df.plot.scatter(x='Age', y='Fare')

# histogram

titanic\_df['Age'].hist()

# box plot

titanic\_df.boxplot('Age')

Good Job!

***Belum di update!!!***