# **TUGAS III**

**DATA MINING**



**Disusun Oleh:**

**Giraldo Nainggolan (220441100064)**

**Wisnu Ary Swadana (220441100121)**

**Abib MaulanaAan Nafudi (220441100118)**

**Fairuz Abdullah (220441100070)**

**Dosen Pengampu:**

**Dr.Wahyudi Setiawan, S.Kom, M. Pd.**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**BANGKALAN 2024**

**Studi Kasus: Data Tips Restaurant  
  
Sebuah dataset dari suatu Restaurant memuat  
variabel-variabel berikut:**

* **total\_bill: Total bill (cost of the meal), including  
  tax, in US dollars**
* **tip: Tip (gratuity) in US dollars**
* **sex: Sex of person paying for the meal (0=male,  
  1=female)**
* **smoker: Smoker in party? (0=No, 1=Yes)**
* **day: 3=Thur, 4=Fri, 5=Sat, 6=Sun**
* **time: 0=Day, 1=Night**
* **size: Size of the party**

**Sumber Data:**[**https://www.kaggle.com/ranjeetjain3/seaborn-tips-dataset**](https://www.kaggle.com/ranjeetjain3/seaborn-tips-dataset)

**SOAL:  
1. Adakah tipe variabel yang kurang tepat di  
data tersebut?  
2. Apakah data numeriknya cenderung  
berdistribusi normal?  
3. Apakah ada outlier, noise, missing values,  
dan-atau duplikasi data?  
4. Apakah pelanggan pria dan wanita cenderung  
proporsional (balance)?  
5. Dari data yang ada apakah Pria atau wanita  
ada kecenderungan memberi tips lebih besar?  
6. Dari data yang ada apakah ada kecenderungan  
tips lebih besar di hari-hari tertentu?  
7. Dari data yang ada apakah customer perokok  
cenderung memberi tips lebih besar?  
8. Apakah pola di nomer 5 dan 7 dipengaruhi  
hari?  
9. Pola apalagi yang dapat anda temukan? (misal,  
bisakah anda menyarankan tata letak kursi/meja restaurant dari data ini?)  
10. dari hasil EDA anda saran apa saja yang akan  
anda berikan ke pemilik restaurant?  
11. Skills/kompetensi apa yang terasa sangat  
diperlukan dari latihan ini?**

**Jawaban:**

**1.**

**print("Tipe variabel:")**

**print(df.dtypes)**

**Tipe variabel:**

**total\_bill float64**

**tip float64**

**sex object**

**smoker object**

**day object**

**time object**

**size int64**

**dtype: object**

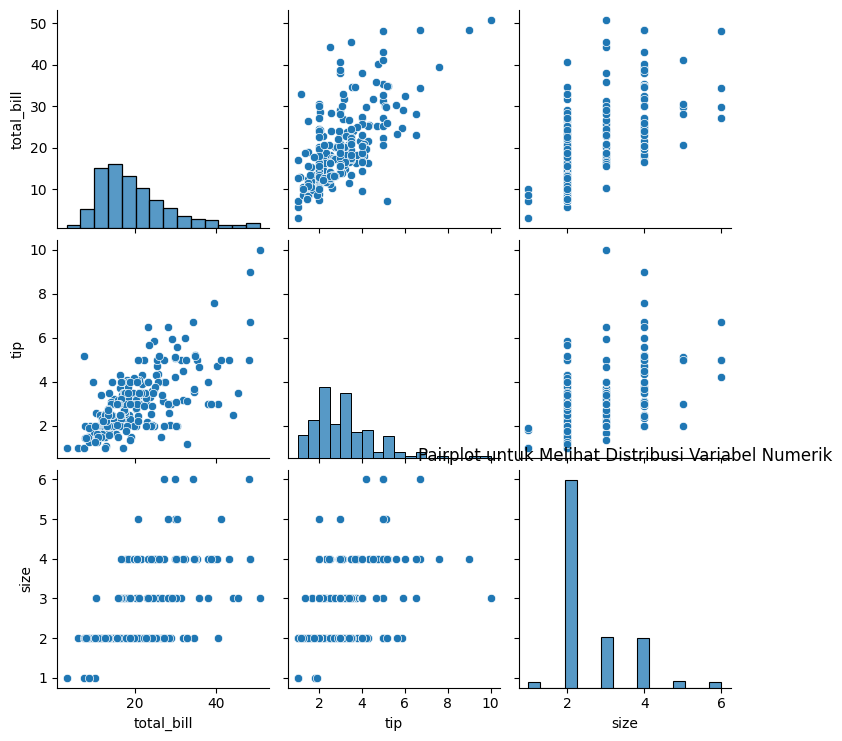
**2.**

**plt.figure(figsize=(10, 8))**

**sns.pairplot(df)**

**plt.title("Pairplot untuk Melihat Distribusi Variabel Numerik")**

**plt.show()**

****

**3.**

**plt.figure(figsize=(10, 8))**

**sns.boxplot(data=df[['total\_bill', 'tip']])**

**plt.title("Box Plot untuk Melihat Outlier")**

**plt.xlabel("Variabel")**

**plt.ylabel("Value")**

**plt.show()**

**# Cek missing values**

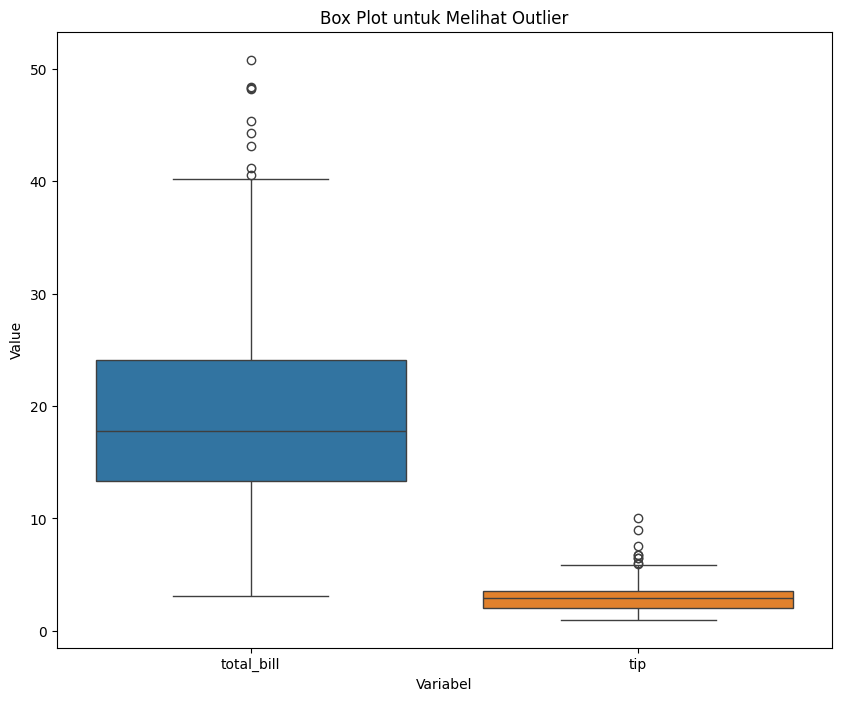
**print("Missing values:")**

**print(df.isnull().sum())**

**# Cek duplikasi data**

**print("Duplikasi data:")**

**print(df.duplicated().sum())**

****

**Missing values:**

**total\_bill 0**

**tip 0**

**sex 0**

**smoker 0**

**day 0**

**time 0**

**size 0**

**dtype: int64**

**Duplikasi Data ; 1**

**4.**

**plt.figure(figsize=(8, 6))**

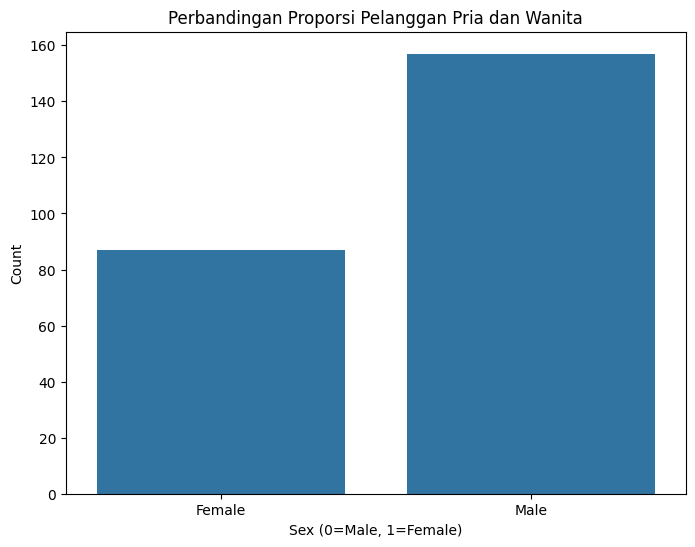
**sns.countplot(data=df, x="sex")**

**plt.title("Perbandingan Proporsi Pelanggan Pria dan Wanita")**

**plt.xlabel("Sex (0=Male, 1=Female)")**

**plt.ylabel("Count")**

**plt.show()**

****

**5.**

**print("Rata-rata tips berdasarkan jenis kelamin:")**

**print(df.groupby("sex")["tip"].mean())**

**Rata-rata tips berdasarkan jenis kelamin:**

**sex**

**Female 2.833448**

**Male 3.089618**

**Na**

**me: tip, dtype: float64**

**6.**

**print("Rata-rata tips berdasarkan hari:")**

**print(df.groupby("day")["tip"].mean())**

**Rata-rata tips berdasarkan hari:**

**day**

**Fri 2.734737**

**Sat 2.993103**

**Sun 3.255132**

**Thur 2.771452**

**Name: tip, dtype: float64**

**7.**

**print("Rata-rata tips berdasarkan status perokok:")**

**print(df.groupby("smoker")["tip"].mean())**

**Rata-rata tips berdasarkan status perokok:**

**smoker**

**No 2.991854**

**Yes 3.008710**

**Name: tip, dtype: float64**

**8.**

**print("Rata-rata tips berdasarkan jenis kelamin dan hari:")**

**print(df.groupby(["sex", "day"])["tip"].mean())**

**Rata-rata tips berdasarkan jenis kelamin dan hari:**

**sex day**

**Female Fri 2.781111**

**Sat 2.801786**

**Sun 3.367222**

**Thur 2.575625**

**Male Fri 2.693000**

**Sat 3.083898**

**Sun 3.220345**

**Thur 2.980333**

**Name: tip, dtype: float64**

**9.   
plt.figure(figsize=(10, 8))**

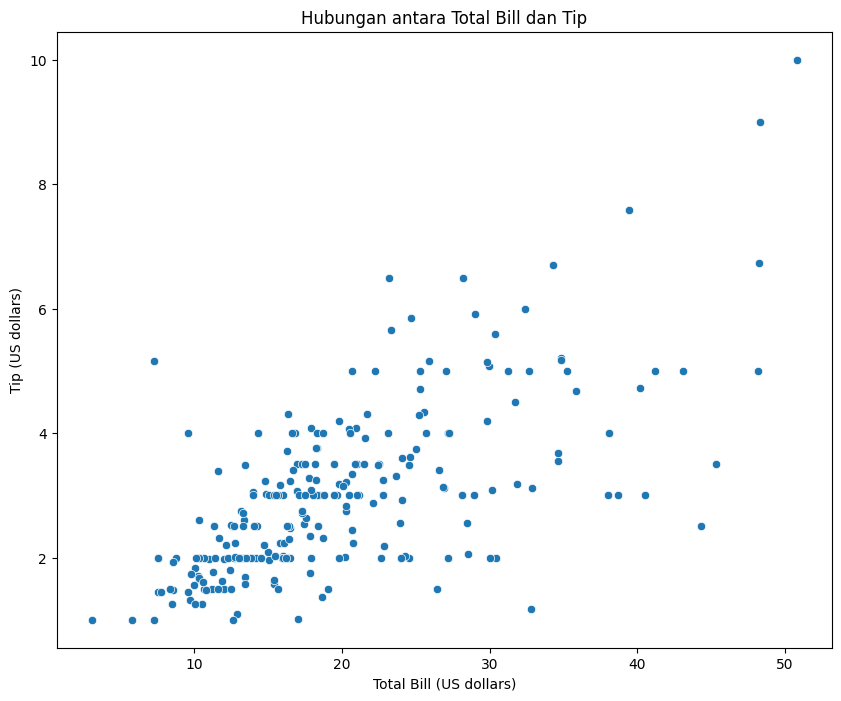
**sns.scatterplot(data=df, x="total\_bill", y="tip")**

**plt.title("Hubungan antara Total Bill dan Tip")**

**plt.xlabel("Total Bill (US dollars)")**

**plt.ylabel("Tip (US dollars)")**

**plt.show()**

****

**10. Menyarankan untuk meningkatkan pelayanan di hari-hari tertentu yang memiliki rata-rata tips lebih rendah.**

**11.  Kemampuan analisis data, pemahaman statistik deskriptif, kemampuan visualisasi data, dan kemampuan komunikasi hasil analisis.**