

# Laporan Praktikum Minggu 2: Data Wrangling & Exploratory Data Analysis (EDA)

Nama: Nurul Hasanah

NIM : 2411070044

URL W&B Project : <https://wandb.ai/nunu05867-stikomelrahma/titanic-eda-2026>

## 1. Pendahuluan

Praktikum ini bertujuan untuk memahami proses Data Wrangling dan Exploratory Data Analysis (EDA) dalam Machine Learning. Data Wrangling adalah proses membersihkan dan menyiapkan data agar dapat digunakan dengan baik. Dataset Titanic digunakan karena mudah dipahami, memiliki data yang tidak lengkap, serta cocok untuk latihan analisis dalam pembelajaran Machine Learning.

## 2. Analisis Data Mentah (Inspeksi)

Berdasarkan hasil inspeksi data menggunakan `df.info()` dan `df.isnull().sum()`, ditemukan beberapa kolom yang memiliki missing values, yaitu:

- Age: 177 data kosong
- Cabin: 687 data kosong
- Embarked: 2 data kosong

## 3. Strategi Pembersihan Data

- Imputasi Age menggunakan median karena data umur tidak merata, sehingga median lebih stabil dibandingkan mean.
- Imputasi Embarked menggunakan modus karena merupakan data kategori, sehingga nilai yang paling sering muncul dianggap paling mewakili.
- Kolom Cabin dihapus karena memiliki terlalu banyak data kosong, sehingga jika diisi dapat menurunkan kualitas data.

## 4. Visualisasi & Temuan Utama

- Survival Rate by Gender: Berdasarkan hasil visualisasi, penumpang perempuan memiliki tingkat keselamatan lebih tinggi dibandingkan penumpang laki-laki. Hal ini sesuai dengan kebijakan evakuasi yang memprioritaskan perempuan dan anak-anak.
- Distribusi Umur: Sebagian besar penumpang berada pada rentang usia dewasa muda hingga dewasa, sedangkan jumlah penumpang anak-anak relatif lebih sedikit.

## 5. Feature Engineering

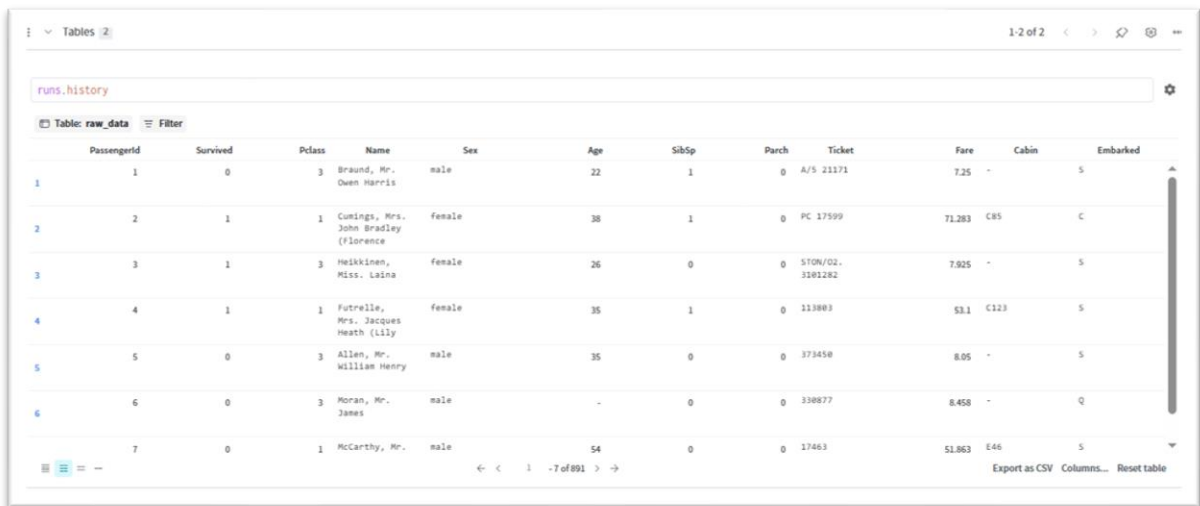
Fitur baru FamilySize dibuat dengan rumus:

$$\text{FamilySize} = \text{SibSp} + \text{Parch} + 1$$

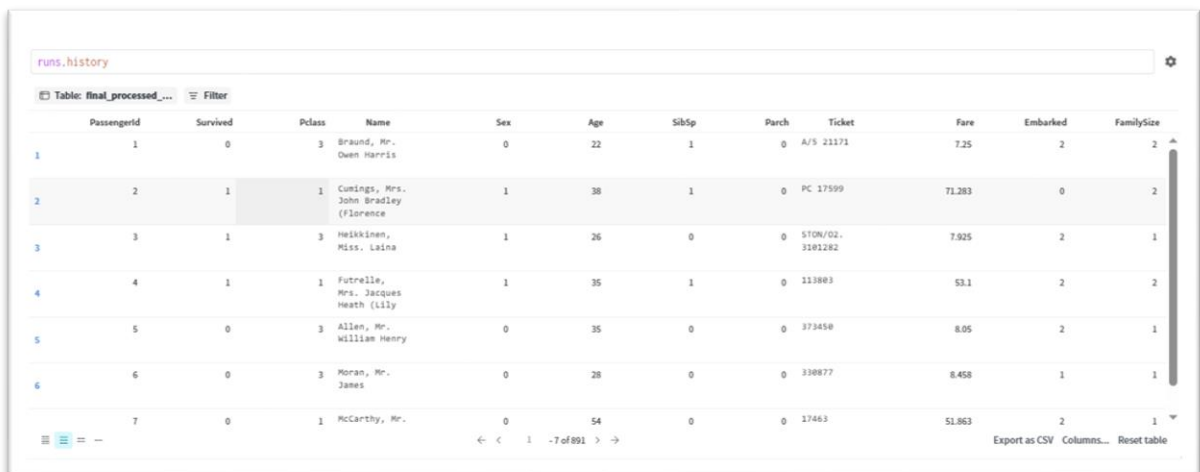
Hasil visualisasi menunjukkan bahwa ukuran keluarga berpengaruh terhadap peluang selamat. Penumpang dengan keluarga kecil memiliki peluang keselamatan yang lebih tinggi dibandingkan penumpang yang bepergian sendiri atau dalam keluarga besar.

Proses encoding dilakukan pada data kategorikal seperti Sex dengan mengubah nilai teks menjadi angka (0 dan 1) agar data dapat diproses oleh algoritma Machine Learning.

## 6. Integrasi Weights & Biases (W&B)



	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25	-	S
2	2	1	1	Cunings, Mrs. John Bradley (Florence)	female	38	1	0	PC 17599	71.283	C85	C
3	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925	-	S
4	4	1	1	Futrelle, Mrs. Jacques Heath (Lily)	female	35	1	0	113803	53.1	C123	S
5	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05	-	S
6	6	0	3	Moran, Mr. James	male	-	0	0	330877	8.458	-	Q
7	7	0	1	McCarthy, Mr.	male	54	0	0	17463	51.863	E46	S



	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	FamilySize
1	1	0	3	Braund, Mr. Owen Harris	0	22	1	0	A/5 21171	7.25	2	2
2	2	1	1	Cunings, Mrs. John Bradley (Florence)	1	38	1	0	PC 17599	71.283	0	2
3	3	1	3	Heikkinen, Miss. Laina	1	26	0	0	STON/O2. 3101282	7.925	2	1
4	4	1	1	Futrelle, Mrs. Jacques Heath (Lily)	1	35	1	0	113803	53.1	2	2
5	5	0	3	Allen, Mr. William Henry	0	35	0	0	373450	8.05	2	1
6	6	0	3	Moran, Mr. James	0	28	0	0	330877	8.458	1	1
7	7	0	1	McCarthy, Mr.	0	54	0	0	17463	51.863	2	1

Pada W&B Tables, ditampilkan data mentah (raw\_data) dan data yang telah diproses (final\_processed\_data). Dari tabel ini, dapat melihat perubahan data setelah dilakukan proses pembersihan, imputasi, dan transformasi fitur.

Fitur filtering dan sorting membantu dalam mencari dan mengelompokkan data dengan lebih cepat dan interaktif, sehingga analisis data menjadi lebih mudah dibandingkan hanya menggunakan tampilan .head() pada notebook.

## **7. Kesimpulan & Refleksi**

Tantangan terbesar dalam praktikum ini adalah memahami proses pembersihan data dan menentukan strategi imputasi yang tepat. Praktikum ini menunjukkan bahwa tahap EDA sangat penting untuk memahami karakteristik data sehingga model yang dibangun nantinya menjadi lebih akurat.