



Laporan Praktikum Minggu 2: Data Wrangling & EDA

Nama: Siti Yolanda Hareniza

NIM: 2411070032

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

1. Pendahuluan

Jelaskan secara singkat tujuan praktikum minggu ini. Apa itu *Data Wrangling* dan mengapa dataset Titanic digunakan sebagai standar pembelajaran?

Jawaban: Praktikum ini bertujuan memahami proses Data Wrangling dan Exploratory Data Analysis (EDA) sebagai tahap awal sebelum pemodelan Machine Learning. Data Wrangling meliputi pembersihan data, penanganan missing values, dan transformasi fitur agar data siap dianalisis. Dataset Titanic digunakan karena memiliki data numerik dan kategorikal serta mengandung missing values, sehingga sesuai untuk latihan pembersihan dan eksplorasi data.

2. Analisis Data Mentah (Inspeksi)

Berdasarkan hasil ` .info()` dan ` .isnull().sum()`, sebutkan kolom mana saja yang memiliki *missing values* dan berapa jumlahnya?

Jawaban:

Berdasarkan hasil pemeriksaan menggunakan .info() dan .isnull().sum(), ditemukan missing values pada:

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

Kolom lainnya tidak memiliki missing values.

3. Strategi Pembersihan Data

Imputasi Age: Karena data numerik berpotensi memiliki outlier, sehingga median lebih stabil dibanding mean.

Imputasi Embarked: Karena merupakan data kategorikal dan tidak dapat dihitung menggunakan rata-rata.

Penghapusan Cabin: Karena memiliki lebih dari 70% missing values, sehingga kurang representatif jika dipertahankan.

4. Visualisasi & Temuan Utama

Survival Rate by Gender: Tingkat keselamatan penumpang perempuan lebih tinggi dibanding laki-laki berdasarkan pada grafik yang ditampilkan.

Distribusi Umur: Menunjukkan mayoritas penumpang berada pada rentang usia 20–40 tahun, dengan jumlah anak-anak dan lansia lebih sedikit.

5. Feature Engineering

Fitur `FamilySize` : FamilySize=SibSp+Parch+1

Analisis: Hasil analisis menunjukkan bahwa penumpang dengan ukuran keluarga kecil cenderung memiliki peluang selamat lebih tinggi dibanding yang sendirian atau memiliki keluarga sangat besar.

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

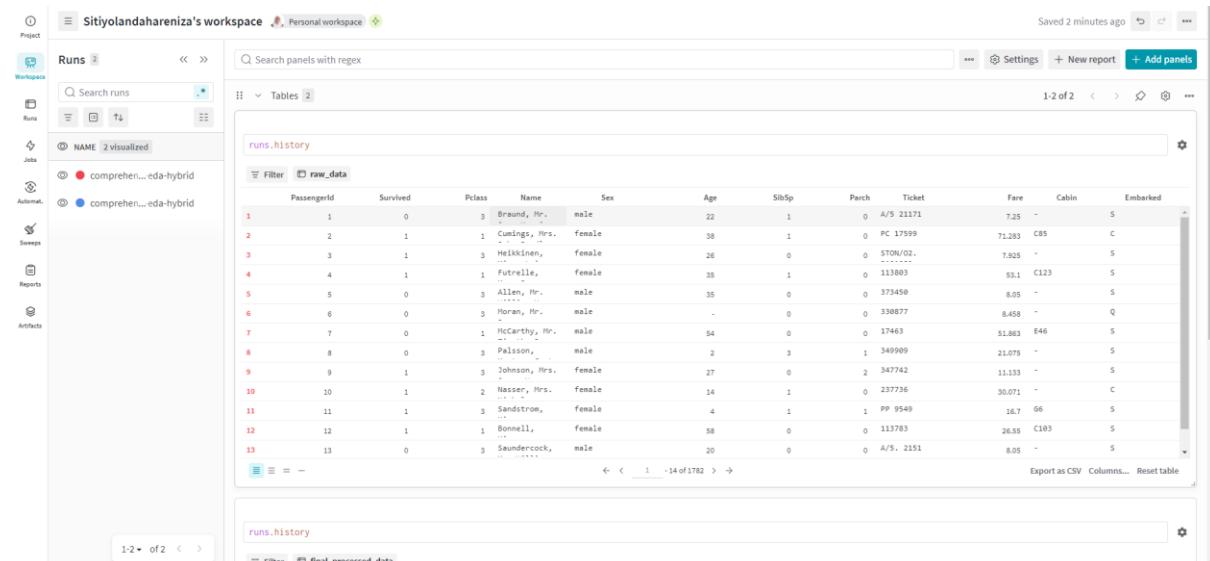
Encoding: Data kategorikal seperti Sex diubah menjadi bentuk numerik (encoding) karena algoritma Machine Learning hanya dapat memproses data numerik.

6. Integrasi Weights & Biases (W&B)

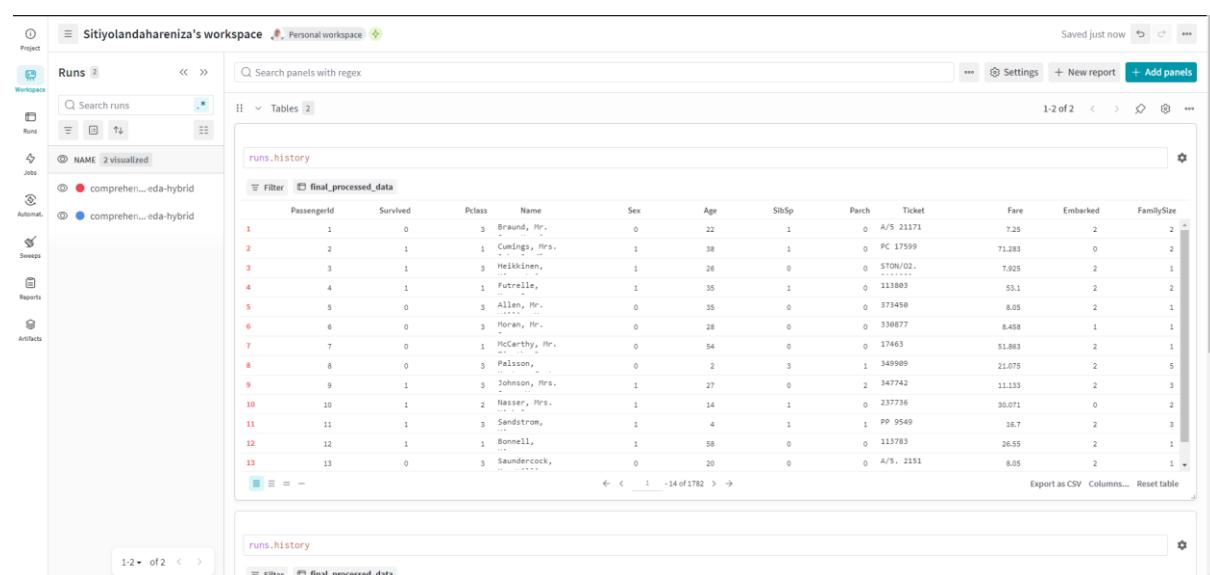
Lampirkan screenshot atau jelaskan apa saja yang Anda lihat di **W&B Tables**.

Bagaimana fitur *filtering* dan *sorting* di W&B membantu Anda memahami data secara interaktif dibandingkan hanya menggunakan ` .head()` di notebook?

Jawaban:



The screenshot shows the W&B Tables interface within a workspace titled "Sitiyolandahareniza's workspace". The left sidebar contains project navigation with sections like Project, Workspace, Runs, Jobs, Automat, Sweeps, Reports, and Artifacts. The main area displays two tables: "raw_data" and "final_processed_data". The "raw_data" table is currently selected, showing 14 rows of data from the Titanic dataset. The columns include PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, and Embarked. The "final_processed_data" table is also visible below it. The interface includes a search bar at the top and various filtering and sorting options for the data.



This screenshot shows the same W&B Tables interface as the previous one, but the "final_processed_data" table is now selected. It displays the same 14 rows of data as the "raw_data" table, but with additional columns: Embarked and FamilySize. The "raw_data" table is still present below it. The interface remains consistent with the first screenshot, featuring a search bar and various data manipulation tools.

The screenshot shows a DataCamp workspace interface. On the left, there's a sidebar with project navigation (Project, Workspace, Runs, Jobs, Automate, Sweeps, Reports, Artifacts) and a search bar. The main area has two tables displayed side-by-side. The top table is titled 'runs.history' and is filtered from 'raw_data' to show rows where 'col14 != "male"'. The bottom table is also titled 'runs.history' and is filtered from 'final_processed_data' to show rows 1 through 25. Both tables have columns for index, ID, Name, Sex, Age, and other travel-related variables.

W&B Tables digunakan untuk membandingkan data mentah dan data yang telah diproses. Fitur filtering dan sorting memungkinkan eksplorasi data secara interaktif, sehingga memudahkan identifikasi pola dan evaluasi hasil pembersihan data dibandingkan hanya menggunakan tampilan .head() di notebook.

7. Kesimpulan & Refleksi

Apa tantangan tersulit yang Anda hadapi dalam praktikum ini? Apa kesimpulan utama Anda mengenai pentingnya tahap EDA sebelum masuk ke tahap pemodelan (**Modeling**)?

Jawaban: Tahap Data Wrangling dan EDA merupakan langkah fundamental dalam proses Machine Learning. Penanganan missing values, transformasi fitur, serta eksplorasi visual membantu memastikan data berada dalam kondisi optimal sebelum tahap pemodelan. Kualitas data yang baik akan meningkatkan keandalan dan performa model yang dibangun.