

# 📖 আমার ডেটা সায়েন্স ও মেশিন লার্নিং নোটবুক কালেকশন

---

"নিজের শেখার জন্য, নিজের ভাষায়, নিজের মতো করে!"

---

## 👉 সূচনা

এই রিপোর্ট আমার ডেটা সায়েন্স ও মেশিন লার্নিং শেখার জার্নির নোট, কোড, আর উদাহরণ দিয়ে সাজানো। সবকিছু বাংলায় লিখেছি যেন ভবিষ্যতে একবার দেখলেই সব মনে পড়ে যায়।

---

## ১. Pandas ও ডেটা হ্যান্ডলিং (01\_pandas)

### • 01\_list to dataframe.ipynb

- Python list থেকে DataFrame বানানো, ইন্ডেক্স-কলাম কাস্টমাইজ, dict থেকে DataFrame — সব হাতে কলমে।
- 📌 টিপ: `.shape`, `.info()`, `.columns` এগুলো দিয়ে ডেটার গঠন চেক করা যায়।

### • 02\_excel to df.ipynb

- CSV ফাইল থেকে ডেটা আনা, সাবসেটিং, সোর্টিং, বেসিক স্ট্যাটস, করেলেশন, হিটম্যাপ — সব এক জায়গায়।
- 📌 টিপ: `.read_csv()`, `.head()`, `.sort_values()`, `.describe()`, `.corr()` — এগুলো খুব কাজে লাগে।

### • 03\_pandas profiling.ipynb

- ডেটার অটো রিপোর্ট — এক ক্লিকে পুরো ডেটার সারাংশ।
- 📌 টিপ: `pandas_profiling` দিয়ে ডেটার সমস্যা, missing value, outlier সব ধরা যায়।

### • 04\_Importing Google Sheets using Python Pandas.ipynb

- Google Sheets থেকে ডেটা আনার সহজ উপায়, gspread দিয়ে।
- 📌 টিপ: Google API credentials লাগবে, শিট শেয়ার করতে হবে সার্ভিস অ্যাকাউন্টে।

### • Screen Time Data.csv

- মোবাইল স্ক্রিন টাইমের ডেটাসেট, নিজের ডেটা এনালাইসিসের জন্য।
- 

## ২. ভেরিয়েবল ও ডেটার ধরন (02\_types of var)

### • 01\_Types of Variables in Data Science.ipynb

- ভেরিয়েবলের ধরন, মেজারমেন্ট স্কেল, ডেটার role — সব বাংলায়।
  - 📌 টিপ: ডেটার ধরন বুঝে এনালাইসিস করলে ভুল কম হয়।
-

## ৩. লিনিয়ার রিগ্রেশন ও সম্পর্কিত টপিক (03\_linear regression)

- **01\_Regression Analysis.ipynb** — লিনিয়ার রিগ্রেশন, ডেটা ভিজুয়ালাইজেশন, মডেল ফিটিং, প্রেডিকশন।
  - **02\_Loss & Cost Functions in Linear Regression.ipynb** — লস ফাংশন (MSE, MAE), মডেল ইভালুয়েশন।
  - **03\_Coefficient of Determination(R-squared).ipynb** — R-squared মান, মডেলের পারফরম্যান্স।
  - **04\_Assignment 01 - Salary Prediction.ipynb** — বাস্তব ডেটা দিয়ে স্যালারি প্রেডিকশন।
  - **05\_Assignment\_car\_risk\_analysis.ipynb** — গাড়ি চালানোর রিস্ক এনালাইসিস।
  - **06\_Linear Regression with Multiple Variables.ipynb** — মাল্টিপল ভেরিয়েবল নিয়ে রিগ্রেশন।
  - **07\_pearson Correlation coefficient.ipynb** — পিয়ারসন করেলেশন ক্যালকুলেশন।
  - **08\_Polynomial Regression.ipynb** — নন-লিনিয়ার ডেটার জন্য পলিনোমিয়াল রিগ্রেশন।
  - **ডেটাসেট:**
    - `nasdaq100.csv`, `linear_data.csv`, `nonlinear_data.csv`, `car driving risk analysis.csv`, `Salary Data.csv`
    - টিপ: `.fit()`, `.predict()`, `.score()`, `train_test_split` — এগুলো স্কিলার্নের বেসিক ফাংশন।
- 

## ৪. ফিচার ইঞ্জিনিয়ারিং ও ডেটা প্রিপ্রসেসিং (04\_Feature Engineering)

- **01\_Label Encoder in Machine Learning.ipynb** — ক্যাটাগরিক্যাল ডেটাকে লেবেল এনকোডিং।
  - **02\_One-Hot Encoding in Machine Learning.ipynb** — ওয়ান-হট এনকোডিং।
  - **03\_Binary Encoder Explained in Machine Learning.ipynb** — বাইনারি এনকোডিং।
  - **04\_Ordinal Encoder.ipynb** — অর্ডিনাল ডেটার জন্য এনকোডিং।
  - **05\_Min Max Normalization.ipynb** — ডেটা স্কেলিং (০-১)।
  - **06\_Standardization Z-Score.ipynb** — ডেটা স্ট্যান্ডার্ডাইজেশন (মিন=০, SD=১)।
  - টিপ: এনকোডিং আর স্কেলিং ছাড়া মডেলিং ঠিকমতো হয় না!
- 

## ৫. প্রজেক্ট: মাল্টিপল লিনিয়ার রিগ্রেশন (05\_project\_multiple\_linear\_regression)

- **01\_Profit Prediction using Multiple Linear Regression.ipynb**
    - একাধিক ইনপুট (Marketing Spend, Administration, Transport, Area) থেকে Profit প্রেডিকশন।
    - One-hot encoding, train-test split, মডেল ফিটিং, স্কোরিং, R-squared — সব এক জায়গায়।
  - **online.csv** — প্রজেক্টের ডেটাসেট।
- 

## ❖ আমার শেখার টিপস

- **প্র্যাকটিস:** কোড নিজে চালাও, নিজের ডেটা দিয়ে চেষ্টা করো।
  - **ভিজুয়ালাইজেশন:** সবসময় গ্রাফ/চার্টে দেখো, চোখে পড়লে মনে থাকে।
  - **ডকুমেন্টেশন:** pandas, sklearn, matplotlib — অফিসিয়াল ডকুমেন্টেশন পড়ো।
  - **প্রশ্ন:** না বুঝলে গুগল/চ্যাটজিপিটি/স্ট্যাকওভারফ্লোতে প্রশ্ন করো।
  - **নোট:** শেখার সময় ছোট ছোট নোট রাখো, পরে কাজে লাগবে।
-

## ❏ শেষ কথা

এই রিপোর্ট আমার শেখার সঙ্গী। সময় পেলে নতুন টপিক, প্রজেক্ট, উদাহরণ যোগ করবো।

শুভকামনা! ✍