MAINCRAFTS TECHNOLOGY
www.maincrafts.com
hr@maincrafts.com

📊 Data Science/Analysis with Python Internship- Task 1

Student Performance dataset

📌 Task Description

Your first step into **Data Science**! In this task, you'll explore the **Student Performance dataset** 📘 and use Python to **clean, analyze, and visualize** it.
This will teach you the **workflow every data scientist follows**: load → clean → analyze → visualize → conclude.

🛠 Responsibilities (What to Build)

1. **Load Dataset**
   - Use pandas to read student-mat.csv dataset (Download here).
2. **Explore & Clean Data**
   - Check missing values (.isnull()).
   - Remove duplicates.
   - Inspect dataset shape & dtypes.
3. **Analysis Questions**
   - Average final grade (G3).
   - How many students scored above 15?
   - Is study time correlated with performance?
   - Which gender performs better on average?
4. **Visualizations**
   - Histogram of grades.
   - Scatterplot: study time vs grades.
   - Bar chart: male vs female average score.
5. **Documentation**
   - Use Markdown in Jupyter Notebook to explain each step.

🎯 Skills Gained

- ✅ Data loading & cleaning (Pandas)
- ✅ Statistical analysis with NumPy
- ✅ Data visualization with Matplotlib/Seaborn
- ✅ Correlation analysis
- ✅ Writing reports with Jupyter Notebook

📂 Sample Starter Code

python

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
df = pd.read_csv("student-mat.csv")

# Analysis
print("Average Final Grade:", df['G3'].mean())

# Visualization
plt.hist(df['G3'], bins=10, color='skyblue', edgecolor='black')
plt.title("Distribution of Final Grades")
plt.xlabel("Grade")
plt.ylabel("Frequency")
plt.show()
```

📌 Resources

📺 Free Video Tutorials

- Pandas Crash Course – Corey Schafer
- Data Science Crash Course – FreeCodeCamp

📖 Reference Docs

- Pandas Official Docs
- Matplotlib Docs
- Seaborn Docs

🛠 Tools

- Google Colab → Free cloud-based Jupyter Notebook
- Kaggle → Datasets + practice environment

✅ Deliverable

A Jupyter Notebook `student_analysis.ipynb` with:

- Clean dataset
- Answers to analysis questions
- Visualizations
- Markdown explanations