Prompt Engineering

Prompt:

I don’t know how to create a unique and simple readme for students to understand create for one based on the information I have Provided:

Info: **Section 1: Python + Machine Learning (1.5 hours)**

**Dataset:** [Student Performance Dataset (Math & Portuguese)](https://www.kaggle.com/datasets/spscientist/students-performance-in-exams)

**Tasks:**

1. Load the dataset and perform **data cleaning**.
2. Perform **EDA** using visualizations (use seaborn, matplotlib).
3. Predict whether a student will **pass or fail** based on scores and other features.
4. Use **Logistic Regression** or **Random Forest** for classification.
5. Evaluate the model using **accuracy, confusion matrix, F1-score**.
6. Add comments/markdown to explain the steps.

**✅ Section 2: SQL (30–45 mins)**

**Dataset:** [Chinook Database (Music Store)](https://github.com/lerocha/chinook-database)

You can run it on [SQLite Online](https://sqliteonline.com/) or import into any SQL IDE.

**Tasks:**

1. List the top 5 customers by total purchase amount.
2. Find the most popular genre in terms of total tracks sold.
3. Retrieve all employees who are managers along with their subordinates.
4. For each artist, find their most sold album.
5. Write a query to get monthly sales trends in the year 2013.

Submit a .sql file with queries and output screenshots.

**✅ Section 3: Tableau (30 mins)**

**Dataset:** [Airbnb Listings in NYC](https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data)

**Tasks:**

1. Create a dashboard showing:
   * Listings count by neighborhood
   * Price distribution per room type
   * Availability trends
2. Include filters for room type and neighborhood.
3. Publish your dashboard to [Tableau Public](https://public.tableau.com/) and share the link.

**✅ Section 4: Excel (30 mins)**

**Dataset:** [Online Retail Dataset (UCI)](https://archive.ics.uci.edu/ml/datasets/online+retail)

**Tasks:**

1. Clean the data (remove nulls, handle duplicates).
2. Create a pivot table to show **total sales by country and month**.
3. Add formulas to calculate:
   * Average order value
   * % contribution of each country to total sales
4. Highlight top 5 countries by revenue using conditional formatting.
5. Create a chart to visualize monthly revenue trend.

**Response:**

# 🎓 Student Project: Data Analytics & Machine Learning

Welcome to your end-to-end data project! This guide will walk you through tasks in Python, SQL, Tableau, and Excel. Follow each section carefully and complete the steps. Don't forget to comment your code and submit all required files.

---

## ✅ Section 1: Python + Machine Learning (⏱ 1.5 hours)

\*\*Dataset:\*\* Student Performance (Math & Portuguese)

### What You’ll Do:

1. \*\*Load and Clean Data\*\*

- Combine the Math and Portuguese datasets.

- Remove duplicates and handle missing values.

- Convert categorical values to numeric if needed.

2. \*\*Explore the Data (EDA)\*\*

- Use \*\*matplotlib\*\* and \*\*seaborn\*\* for visualizations.

- Show distributions of scores, pass/fail rates, and correlations.

3. \*\*Predict Student Outcome\*\*

- Create a new column: `Pass/Fail` based on final grade (e.g., G3 ≥ 10 = Pass).

- Use \*\*Logistic Regression\*\* or \*\*Random Forest\*\* to predict pass/fail.

- Split the data into training and test sets.

4. \*\*Evaluate the Model\*\*

- Show accuracy, confusion matrix, and F1-score.

✅ \*\*Remember\*\*: Add comments and markdown to explain each step.

---

## ✅ Section 2: SQL (⏱ 30–45 mins)

\*\*Dataset:\*\* Chinook Music Store Database

Run queries using SQLite Online or any SQL tool.

### Tasks:

- Find top 5 customers by total purchases.

- Find the most popular music genre by total tracks sold.

- Show managers and their subordinates.

- Find each artist’s top-selling album.

- Show monthly sales trend for 2013.

✅ \*\*Submit\*\*: `.sql` file with all your queries + screenshots of results.

---

## ✅ Section 3: Tableau (⏱ 30 mins)

\*\*Dataset:\*\* Airbnb Listings in NYC

### Dashboard Requirements:

- Total listings by neighborhood

- Price distribution by room type

- Availability trends

- Add filters for room type and neighborhood

✅ \*\*Publish\*\* your dashboard to \*\*Tableau Public\*\* and share the link.

---

## ✅ Section 4: Excel (⏱ 30 mins)

\*\*Dataset:\*\* Online Retail Dataset (UCI)

### Tasks:

- Clean the data: remove nulls and duplicates.

- Create a pivot table showing total sales by country and by month.

✅ Save your final workbook with pivot tables included.

---

## 🔚 Submission Checklist:

- [ ] Python code with comments and markdown

- [ ] SQL file + screenshots

- [ ] Tableau dashboard link

- [ ] Cleaned Excel file with pivot table

Good luck and have fun exploring! 🚀

MY Feedback: The text was neatly generated but with unnecessary hashtags and emojis with it is not acceptable with the given sentence so I wanted to change it. The text was not cleanly arranged it has added its own version. So asked ChatGPT to make the text and its visual more appealing according to my preferences but removing some sentences which I thought was unnecessary.

This another prompt: To rewrite the above paragraph to have a polished version.

**Prompt:** Rewrite The above feedback.

ChatGPT Response:

Sure! Here's a clearer and more polished version of your sentence:

**"The text was neatly generated, but it included unnecessary hashtags and emojis, which were not suitable for the given context. It also wasn’t cleanly arranged and seemed to follow its own version. So, I asked ChatGPT to revise the text and visuals to better match my preferences, while removing some sentences I found unnecessary."**

Let me know if you want it more formal, casual, or in a different tone!