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! 🚀

Prompt:

i am using sqllite onlie how to fetch the files from github of other repository to run the files

Response:

Great! You're using [SQLite Online](https://sqliteonline.com/). Here's how you can fetch SQL files from a GitHub repo and run them in SQLite Online:

✅ Step-by-Step Guide to Run SQL from GitHub in SQLite Online

🔹 Step 1: Get the Raw URL of the SQL file on GitHub

1. Navigate to the .sql file in the GitHub repository.
2. Click the “Raw” button on the top-right of the file.
3. Copy the raw file URL.

Example:

https://raw.githubusercontent.com/username/repo-name/branch-name/path/to/file.sql

🔹 Step 2: Open the Raw URL in a browser

Open the copied raw GitHub link in your browser. You’ll see plain SQL code.

🔹 Step 3: Copy the SQL code

Select all (Ctrl + A) → Copy (Ctrl + C) the SQL code.

🔹 Step 4: Paste it into [SQLite Online](https://sqliteonline.com/)

1. Go to <https://sqliteonline.com/>.
2. In the left-hand SQL editor pane, paste the SQL code you copied.
3. Click the Run (▶️) button on top.

✅ Bonus: If the file contains CREATE TABLE, INSERT, SELECT etc., everything should execute directly.

⚠️ Note:

* SQLite might not support all SQL syntax used in MySQL/PostgreSQL, so make sure the SQL is SQLite-compatible.
* If you’re using FOREIGN KEY constraints or specific data types, double-check that SQLite supports them.

If you give me the GitHub link, I can even check the file and help you adapt it for SQLite if needed!

**Refined prompt:**

Dataset: Chinook Database (Music Store) You can run it on SQLite Online or import into any SQL IDE. Tasks: write 2 tasks queries. Submit a .sql file with queries and output screenshots. this is my question but when i opened the link for the dataset there was no dataset in it but github with sql scripts explain the question and how i have to do it

**Response from chat gpt:**

Got it! You're asked to work with the Chinook Database (a popular sample database for SQL practice, similar to a music store) — and your task is to:

1. Write 2 SQL queries (tasks) based on the database.
2. Submit a .sql file containing the queries.
3. Take output screenshots of your queries being run.
4. Submit both the file and screenshots.

But you mentioned that when you opened the dataset link, there was no dataset, only GitHub with SQL scripts.

✅ Let's break this down:

🔸 What is the Chinook Database?

The Chinook Database is a sample database with tables like:

* Album
* Artist
* Customer
* Employee
* Invoice
* InvoiceLine
* Track
* Genre
* MediaType
* Playlist

It simulates a digital music store: albums, tracks, customers, etc.

🔸 Why was there no dataset, only scripts?

That’s actually correct — you’re expected to use the SQL script to create the database.

On GitHub, you'll typically find:

* A .sql file (e.g., Chinook\_Sqlite.sql)
* This script contains all the CREATE TABLE and INSERT INTO commands
* You run this script to create and populate the Chinook database.

✅ Here's exactly how to proceed using [SQLite Online](https://sqliteonline.com/):

✳️ Step 1: Download the Chinook SQLite SQL script

* GitHub link (example):  
  <https://github.com/lerocha/chinook-database>
* Navigate to:  
  ChinookDatabase/DataSources/Chinook\_Sqlite.sql
* Click Raw → Copy the entire content

✳️ Step 2: Paste into [SQLite Online](https://sqliteonline.com/)

1. Go to <https://sqliteonline.com/>
2. Choose "SQLite" as the language from dropdown (top left)
3. Paste the entire SQL script from GitHub into the editor
4. Click Run – this will create and populate the database

✳️ Step 3: Write your 5 SQL task queries

Here are 5 example queries you can use or modify:

-- 1. List all customers from the USA

SELECT \* FROM Customer WHERE Country = 'USA';

-- 2. Find the total number of tracks for each genre

SELECT g.Name AS Genre, COUNT(t.TrackId) AS TotalTracks

FROM Genre g

JOIN Track t ON g.GenreId = t.GenreId

GROUP BY g.Name;

✳️ Step 4: Take screenshots of the query results

* After running each query, take a screenshot of the output (use Snipping Tool or screenshot shortcut).
* Save each as Query1.png, Query2.png, etc.

✳️ Step 5: Save your queries in a .sql file

Use any text editor (like Notepad or VSCode):

1. Paste all your 5 queries in one file
2. Save it as chinook\_tasks.sql

✅ Final Submission:

You should now submit:

1. chinook\_tasks.sql — the file with your 5 SQL queries
2. Query1.png to Query5.png — screenshots of each query result

Want help writing or checking your SQL file before you submit? Just paste it here and I’ll review it!

**MY Feedback**: The text was neatly generated and not what I wanted, 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:**

The response generated was not aligned with what I was expecting. It included unnecessary elements like hashtags and emojis, which made the explanation look informal and cluttered. The text was not clearly arranged and added extra content that I did not request. I asked ChatGPT to revise the response to better match my preferences by making it visually and textually more appealing, while also removing sentences I found unnecessary.