🎓 Portfolio Building 2 Guideline: Data Science Portfolio — End-to-End ETL with Public API

This project allows students to showcase their ability to build a data pipeline from extraction to insight using modern tools. It emphasizes real-world workflow, good coding practices, and storytelling through dashboards.

---

1. 🎯 Project Overview

Objective:
Build a portfolio project that demonstrates the end-to-end process of collecting data from a public API, processing and storing it in a cloud data warehouse, and visualizing the insights through an interactive dashboard.

---

2. 🔁 Project Workflow

2a. Extract Data from a Public API

- Research and select a public API.
    - Public API
    - Get more dataset
- Access and retrieve data using Python.
- Handle authentication (if required).
- Document API structure, parameters, and rate limits.

2b. Clean, Transform, and Analyze the Data

- Clean and prepare the data.

- Normalize, parse, categorize, and add derived fields.
- Conduct exploratory and descriptive analysis.

2c. Store Cleaned Data in Google BigQuery

- Create a new Google Cloud project and BigQuery dataset.
- Upload the cleaned dataset into BigQuery using Python.
- Ensure the data is well-structured and queryable.

2d. Build a Dashboard with Looker Studio / Tableau

- Connect Looker Studio to BigQuery.
- Create meaningful, interactive visualizations.
- Include filters, slicers, or date selectors.

---

3. 📦 Final Deliverables

3a. Presentation Slides

- Summary of project steps and findings.
- Include objectives, workflow, key insights, and dashboard visuals.

3b. Interactive Dashboard

- Hosted on Looker Studio.
- At least 3–5 visualizations and user interactivity.

---

4. ⚙️ Optional Enhancements

4a. Modular Python Scripts

- Organize code into reusable functions and files.

4b. Automation with Prefect or Equivalent

- Automate ETL pipeline using a workflow orchestration tool.

4c. Integrate Google Sheets

- Use Google Sheets for lightweight transformation or validation.

---

5. 🔍 Types of Analysis (Choose 1 or More)

5a. Descriptive Analysis

- Patterns, distributions, and summary statistics.

5b. LRFM Analysis

- Analyze Length, Recency, Frequency, Monetary (if applicable).

5c. Behavioral Analysis

- Segment and understand user or customer behavior.

5d. Text Analysis

- Apply basic NLP techniques for insight from text fields.

🧱 Suggested Project Structure

```
/etl_portfolio/
├── README.md                # Project overview and setup instructions
├── data_extraction.py       # Script to fetch data from API
├── data_cleaning.py         # Script for cleaning and transforming data
├── load_to_bigquery.py      # Script to load data into Google BigQuery
├── dashboard_link.txt       # URL to the published Looker Studio dashboard
├── presentation_slides.pdf  # Final presentation summarizing the project
└── requirements.txt         # List of Python packages used
```

✅ Evaluation Criteria

Category	Description
Technical Implementation	API integration, data cleaning, transformation, and loading
Analytical Rigor	Depth and clarity of insights
Code Quality	Clean, modular, and well-documented code
Dashboard Design	Insightful, interactive, and user-friendly visualizations
Communication	Clear and professional final presentation
Bonus Features	Automation, modularization, and integration of external tools