API Ingestion

* It is a process of getting the data from various resources over the Internet.
* API stands for Application Programme Interface.
* It allows systems to communicate, transfer, and process data efficiently over the internet

**Requests module:** requests module in python is a simple and user friendly library for making http requests. It is useful in sending data to the server, retrieving data from the server and to interact with Web API’s.

* Install this library by using command **pip install requests**

The main useful methods in requests module are:

* **POST** : Used to submit new data.
* **GET** : Used to retrieve data.
* **PUT** : Used to update existing data.
* **DELETE** : Used to remove data.

**Key Components:**

* **Request**: The message sent by the client to the server. It contains:
  + **HTTP Method** —GET, POST, PUT, DELETE.
  + **Headers** —for authentication, content type, etc..
  + **Body** —data needed for POST, PUT, PATCH requests.
* **Response**: The message sent by the server to client to confirm that the request was processed.
  + **Status Code** —200 OK, 201 Created, 400 Bad Request….
  + **Response Body** —usually in JSON or XML format.
  + **Error Messages**

**https://api.restful-api.dev/objects?id=3&id=5&id=10**

### 

### **URI Path (Uniform Resource Identifier Path):**

### The URI Path is the part of the URL that specifies the specific resource you want to interact with.

### It comes after the domain name

* In the above url URI path is **/objects**

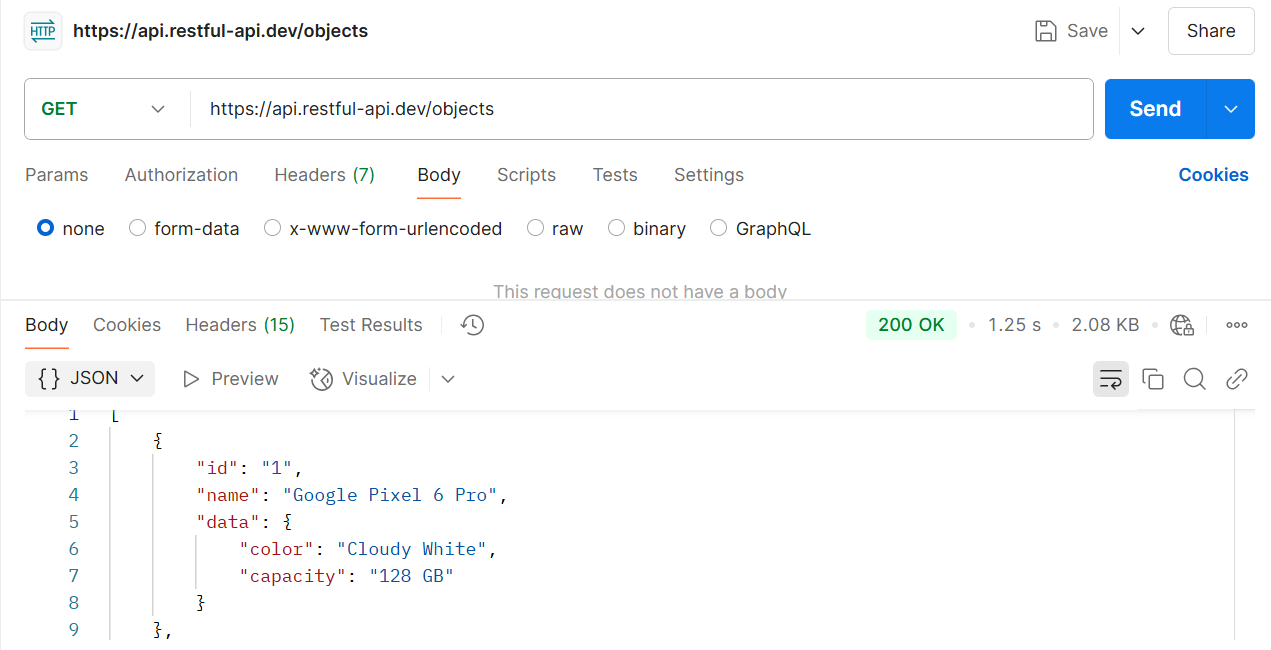
### 

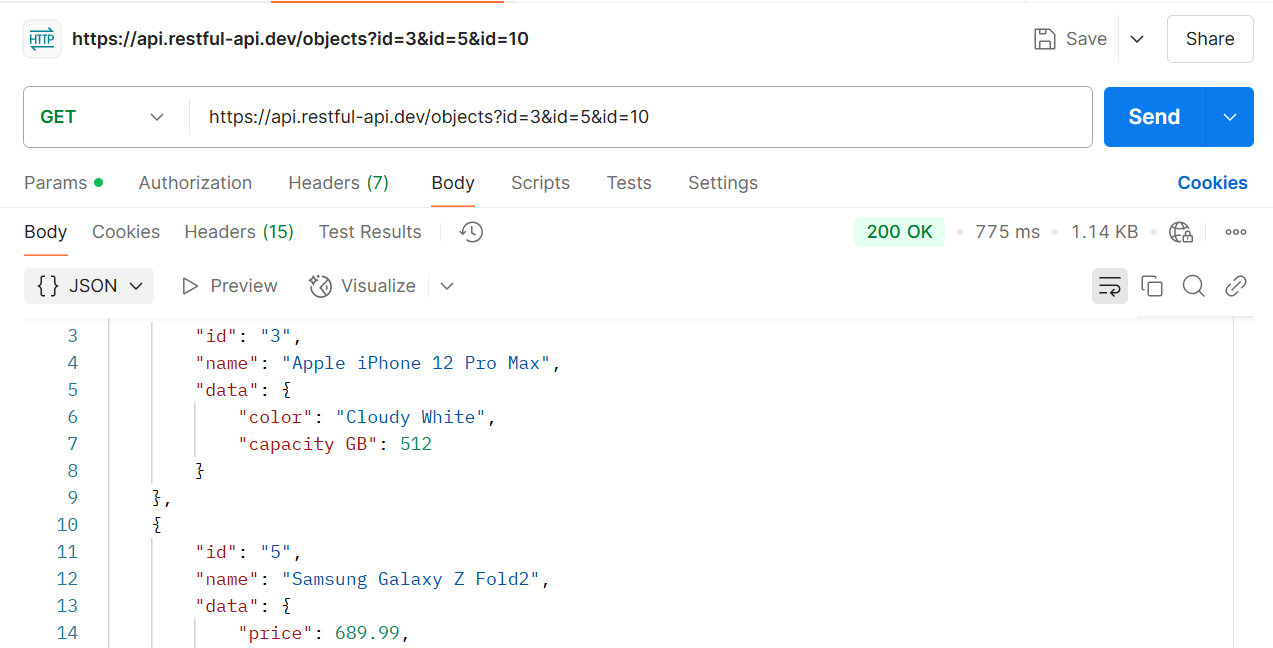
**Query Parameters:**

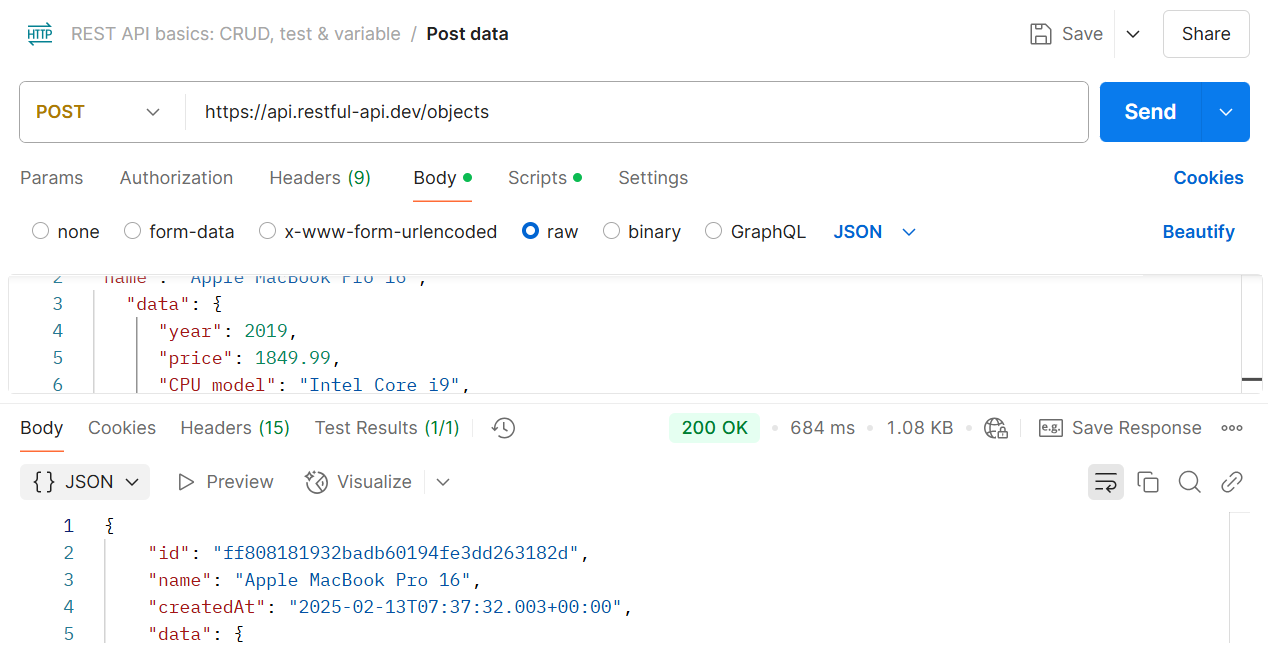
* Query Parameters provide additional information to the server
* They follow the URI path and start with a **?** symbol, with each parameter separated by **&** symbol**.**
* In the above url query parameters are **id=3&id=5&id=10**

**Status codes :** Status codes are part of the response from a server to a client, indicating the result of the request.

* 200 OK: Successful request.
* 201 Created: New resource created.
* 400 Bad Request: Invalid request.
* 401 Unauthorized: Authentication required.
* 403 Forbidden: Access denied.
* 404 Not Found: Resource not found.
* 500 Internal Server Error: Server encountered an error.







**Data Profiling**

It is a process of analyzing the data to understand it’s structure, content, quality, inter relationships between the data by examining it’s characteristics like data types, missing values, inconsistencies with the goal of identifying potential issues before further processing.

* **Techniques:**
  + **Column profiling:** Analyzing individual columns for data quality issues within a table
  + **Cross-column profiling:** Examining relationships and dependencies between different columns
  + **Cross-table profiling:** Comparing data across multiple tables

**When to use data profiling:**

* Before loading data into a data warehouse .
* During data migration projects
* When integrating data from multiple sources

**Why Data profiling:**For improved data quality, informed decision-making, and better data governance within an organization by ensuring your data is accurate and reliable before using it for analysis or reporting.