Data Ingestion and Text Summarization API Documentation

Overview

This document outlines the implementation of two key components for our project: a Data Ingestion API and a Text Summarization Tool. The Data Ingestion API allows users to upload CSV files and retrieve basic summary statistics. The Text Summarization Tool condenses news articles into brief summaries using basic NLP techniques.

Assignment 1: Data Ingestion API

Objective

Develop a RESTful API that handles CSV file uploads, stores the data in a database, and provides basic summary statistics.

Endpoints

1. Upload CSV File

- **URL:** `/api/upload/`
- **Method:** `POST`
- **Headers:**
    - `API-KEY: mysecretkey`
- **Body:** `multipart/form-data` with the CSV file
- **Response:**
    - `201 Created` on success
    - `400 Bad Request` if the file is not a valid CSV or is empty

2. Retrieve Summary Statistics

- **URL:** `/api/statistics/`
- **Method:** `GET`
- **Headers:**
    - `API-KEY: mysecretkey`
- **Response:** JSON object with summary statistics for numerical columns.

Implementation Details

1. **Model Definition**
   - `UploadedData` model stores CSV data in a JSON field.
2. **View for Data Upload**
   - Validates file type and content.
   - Converts CSV data to JSON and stores it in the database.
3. **View for Summary Statistics**
   - Retrieves all stored datasets.
   - Calculates mean, median, and mode for numerical columns.
   - Returns results in a structured format.
4. **Decorator for Payload Validation**
   - Ensures that API requests include valid headers and file data.

Setup Instructions

Clone the Repository
bash
Copy code

```bash
git clone https://github.com/yourusername/data-ingestion-api.git
cd data-ingestion-api
```

1.

Create a Virtual Environment (optional)
bash
Copy code

```bash
python -m venv env
source env/bin/activate  # On Windows use `env\Scripts\activate`
```

2.

Install Dependencies
bash
Copy code

```bash
pip install -r requirements.txt
```

3.

Apply Migrations
bash
Copy code

```bash
python manage.py migrate
```

4.

Run the Server
bash
Copy code

```bash
python manage.py runserver
```

5.
6. **Testing Endpoints**

Upload CSV File:
bash
Copy code

```
curl --location 'http://127.0.0.1:8000/api/upload/' \
--header 'API-KEY: mysecretkey' \
--form 'file=@"/path/to/your/file.csv"'
```

○

Retrieve Summary Statistics:
bash
Copy code

```
curl --location 'http://127.0.0.1:8000/api/statistics/' \
--header 'API-KEY: mysecretkey'
```

○

Contact

For further questions or clarifications, please contact Ankit Anand at ankitnnd@outlook.com.