Project Report: College Recommendation App with AI

---

1. Objective

The aim of this project is to design and develop an AI-powered College Recommendation System using OpenAI's large language models (LLMs). The system guides users step-by-step in identifying colleges that suit their academic scores, location preferences, and desired branches. It enhances decision-making by providing personalized recommendations through conversational AI and supports both static form-based and dynamic chat-based interfaces.

2. System Design

2.1 Architecture Overview

The system is divided into two major flows:

1. Static Form-Based Recommendation System
    - Users answer predefined questions one by one (e.g., location, branch, 12th marks).
    - OpenAI API processes user inputs using a custom system prompt.
    - Recommendations are returned after moderation and prompt construction.
2. Conversational Chatbot
    - Simulates a human-like chat interface where users can type freely.
    - Chat is moderated and forwarded to the OpenAI model.
    - Maintains session context and builds a dialogue history for continuous conversation.

2.2 Components

- Frontend (HTML + CSS)
    - chat.html: Chat-based UI for real-time interaction.
    - index.html: Step-by-step Q&A format UI.
    - static/style.css: Maintains styling for chat.html.
    - static/style_1.css: Maintains styling for index.html.
- Backend (Flask Python App)
    - routes.py: Manages HTTP routes for /, /chatbot, /chatbot/message, and /restart.
    - openai_utils.py: Encapsulates OpenAI API calls and moderation logic.
    - data_utils.py: Loads and cleans college data from CSV.
- OpenAI Services Used
    - Moderation API: Ensures safe content.
    - ChatCompletion API: Generates personalized recommendations using GPT-4o-mini.
- Data
    - College data in .csv format includes fields like College Name, Type, Location, Rank, Fees, and Packages.

3. Implementation

3.1 Prompt Engineering

System prompts are carefully crafted to guide GPT towards relevant and structured answers. For example:

You are a college recommendation expert. Based on the user's inputs, suggest the top 3 suitable colleges with a summary of their key features.

User inputs are appended as part of the `messages` parameter in the OpenAI API call.

3.2 Moderation Handling

Before sending any input to OpenAI:

- `openai.moderation.create()` is called.
- If flagged, the user receives a warning and is asked to rephrase.
- This ensures ethical and safe interactions.

3.3 Flask Routes and Their Roles

- `/`: Home route initiates the form-based session.
- `/chatbot`: Renders the chat interface.
- `/chatbot/message`: Receives message, moderates it, passes to OpenAI, and returns a JSON response.
- `/restart`: Clears session and redirects to home.

3.4 Data Preprocessing

- Columns like "Highest Package (INR)" and "Fees" are converted into Lakhs.
- Null or empty values are filled with a space for consistency.

4. What Problem Does It Solve?

- Manual College Search is often confusing due to too many options and vague filters.
- This system simplifies the process by:
    - Interactively collecting preferences.
    - Instantly generating insights from AI.
    - Offering personalized results, not just filtered lists.
- It is ideal for high school students and parents during the admission season.

5. Challenges Faced

Challenge	Solution
Designing a good system prompt	Iterated prompts to guide model to structured responses
Managing moderation flags	Implemented pre-checks using OpenAI moderation API
Maintaining chat context	Stored user sessions and history using Flask session
Data inconsistencies in CSV	Applied preprocessing and normalization
Handling multi-turn conversations	Used message history as context for GPT

6. Lessons Learned

- Prompt design plays a vital role in controlling the output from GPT.
- OpenAI moderation is essential for user safety.
- Handling user state in Flask is critical for conversational memory.
- UI/UX is equally important as backend logic in AI tools.
- Chat-based interfaces improve user engagement compared to static forms.

7. Future Enhancements

- Add a feedback mechanism to improve model performance based on real user ratings.
- Connect with real-time college data APIs for updated information.
- Add multilingual support for broader reach.

8. Tools and Technologies

Tool	Purpose
OpenAI API (GPT-4o-mini)	Conversational and recommendation logic
Flask	Web framework for routing and server
HTML/CSS	UI design
Pandas	Data loading and transformation
GitHub Desktop	Version control and collaboration

9. Folder Structure

```
├── app/
│   ├── static/
│   │   └── style.css
│   ├── templates/
│   │   └── index.html
│   │   └── chatbot.html
│   ├── __init__.py
│   ├── chatbot.py
│   └── routes.py
├── data/
│   └── college_data.csv
├── utils/
│   ├── common_utils.py
│   ├── data_utils.py
│   └── openai_utils.py
├── venv/
├── .env
├── main.py
└── requirements.txt
```

└─── README.md

10. Setup Instructions

10.1. Clone the Repository

git clone https://github.com/dev-android1/CollegeAssistant
cd CollegeAssistant

10.2. Create a Virtual Environment

python3 -m venv venv
source venv/bin/activate

10.3. Install Requirements

pip install -r requirements.txt

10.4. Add OpenAI API Key

Create a `.env` file in the root directory and add your OpenAI key:
OPENAI_API_KEY1=your_openai_api_key

10.5. Run the Application

python main.py

We have 2 version of app

a. open[http://localhost:5000/] - This will ask predefined questions. The user can skip any question. Atleast one input is required. Finally, we will show them matched college list

b. open [http://localhost:5000/chatbot] - This will show a chat window where user can enter question in natural language and api will process it and will display data.

11. System Prompt

You are a helpful college recommendation assistant.
Your task is to analyze the student's preferences and return the best-matching colleges in a structured JSON format.

Instructions:
- You are provided with a list of colleges and their attributes in JSON.
*- Recommend colleges that **match the student's city or town exactly**.*
*- If none are found, look for colleges in **nearby cities within the same state**.*
*- If still none are found, return top colleges in the **state or region** as a fallback.*
*- Be flexible with **misspellings, vague, or colloquial location names**.*
*- Use **relevance to branch, location proximity, tuition cost, and satisfaction**—not just rank.*
*- Return the output **strictly in a list of JSON objects**, no extra explanation or intro text.*

Expected Output Format:
```json
[
{{
"College": "...",

```
    "Type": "...",
    "Location": "...",
    "Rank": ...,
    "Branches": "...",
    "Avg Package (Lakh)": ...,
    "High Package (Lakh)": ...,
    "Notable Alumni": "...",
    "Student Satisfaction (/10)": ...,
    "Hostel": "...",
    "Facilities": "...",
    "Placements": "...",
    "Tuition (Lakh)": ...,
    "Hostel (Lakh)": ...,
    "Scholarships": "...",
    "Exams": "...",
    "Cutoff Marks": ...,
    "12th % Needed": ...
  }},
  ...
]
```

Example 1:
Student Query: "I want colleges in Pune for computer science."

Expected Output:

```
[
  {{
    "College": "MIT Pune",
    "Type": "Private",
    "Location": "Pune",
    "Rank": 34,
    "Branches": "Computer Science, IT",
    "Avg Package (Lakh)": 6.2,
    "High Package (Lakh)": 18.0,
    "Notable Alumni": "John Doe",
    "Student Satisfaction (/10)": 8.2,
    "Hostel": "Available",
    "Facilities": "Library, Labs, Wi-Fi",
    "Placements": "Strong in CS",
    "Tuition (Lakh)": 6.5,
    "Hostel (Lakh)": 1.2,
    "Scholarships": "Merit-based and Need-based",
    "Exams": "JEE, MHT-CET",
    "Cutoff Marks": 88,
    "12th % Needed": 75
  }}
]
```

Example 2:
Student Query: "Looking for CS colleges near Indore."

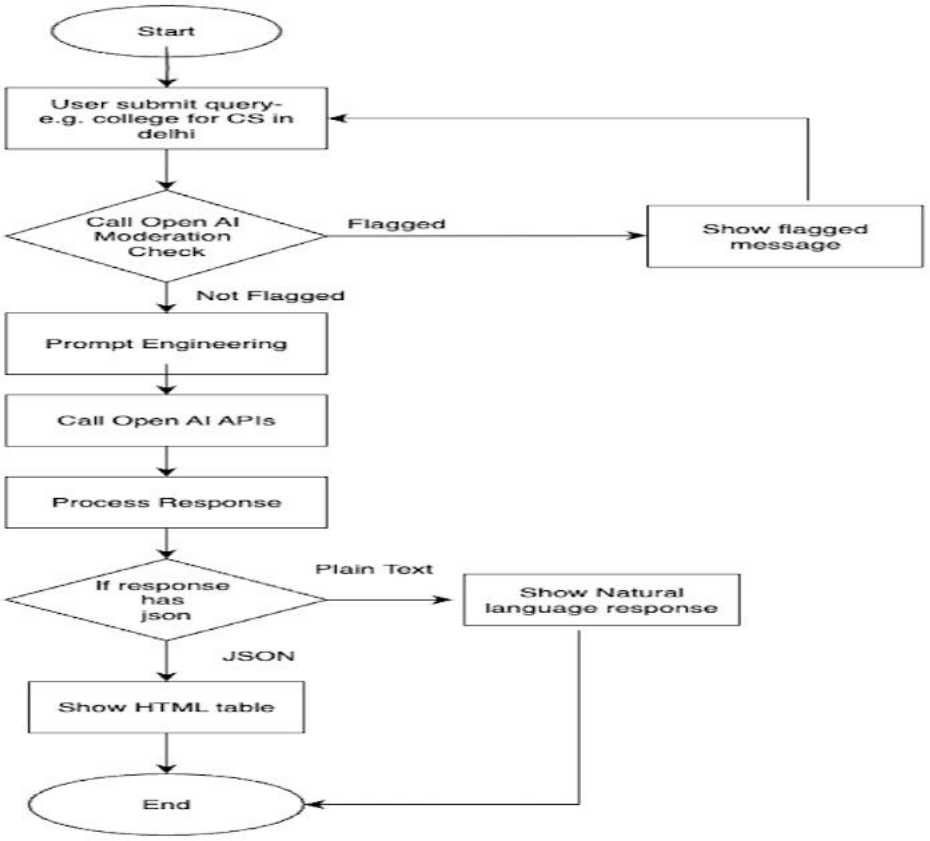
Expected Output:

```
[
  {{
    "College": "IIIT Bhopal",
    "Type": "Government",
    "Location": "Bhopal",
    "Rank": 42,
    "Branches": "Computer Science",
    "Avg Package (Lakh)": 8.1,
    "High Package (Lakh)": 22.0,
    "Notable Alumni": "Jane Smith",
    "Student Satisfaction (/10)": 9.0,
    "Hostel": "Yes",
    "Facilities": "Gym, Labs, Cafeteria",
    "Placements": "Excellent",
```

```
    "Tuition (Lakh)": 5.4,
    "Hostel (Lakh)": 1.0,
    "Scholarships": "SC/ST and Minority",
    "Exams": "JEE",
    "Cutoff Marks": 85,
    "12th % Needed": 70
  }}
]
```

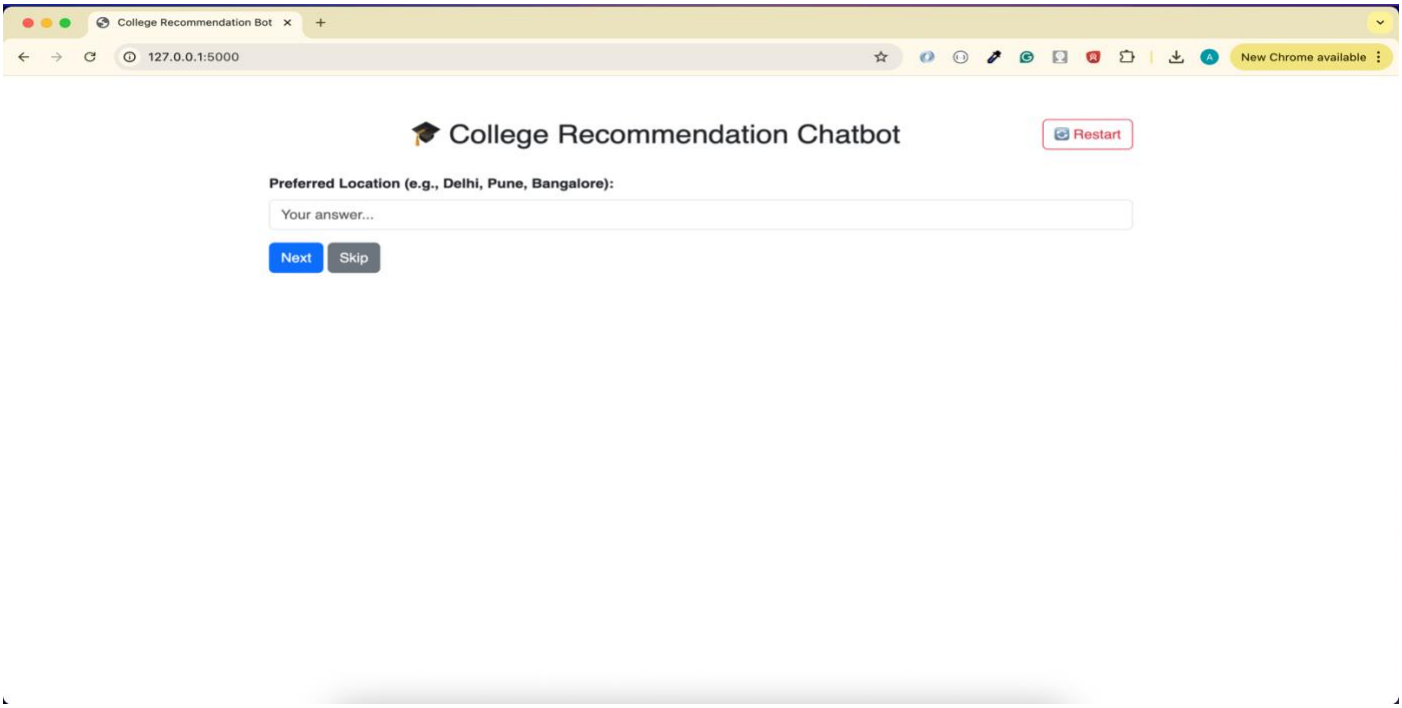
Now analyze the following college dataset and respond in the above format:
{college_data_json}

12. Flow Chart

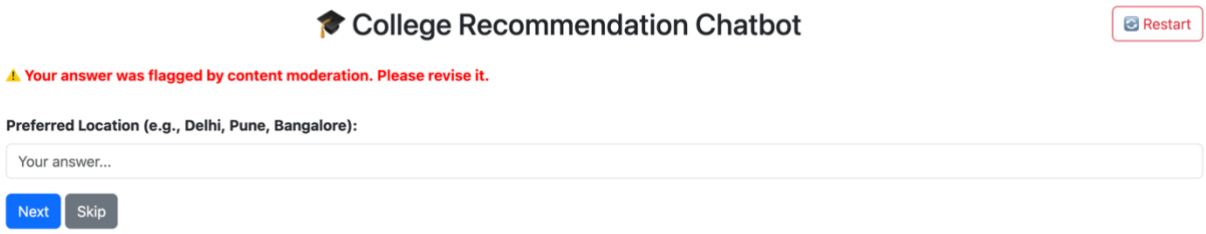


13. Screenshots

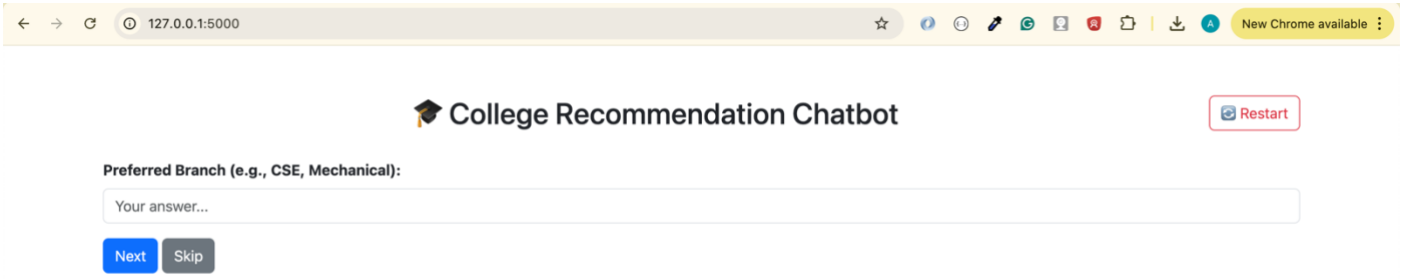
Step -1 (http://localhost:5000/)



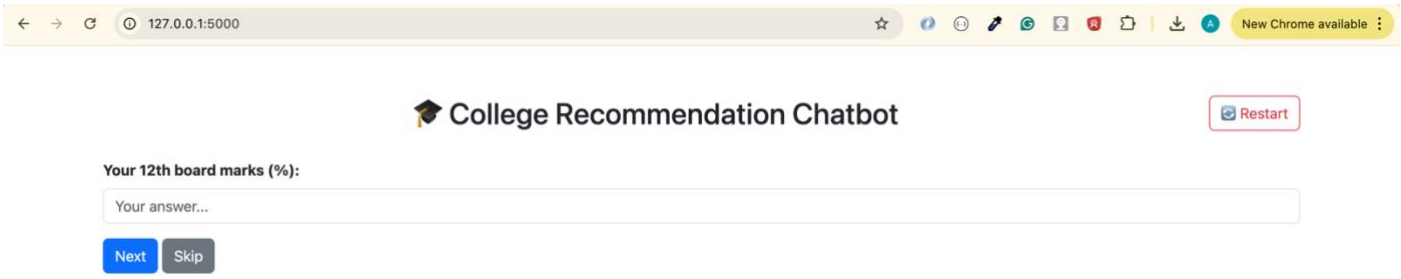
For Flagged Input – By Moderation API



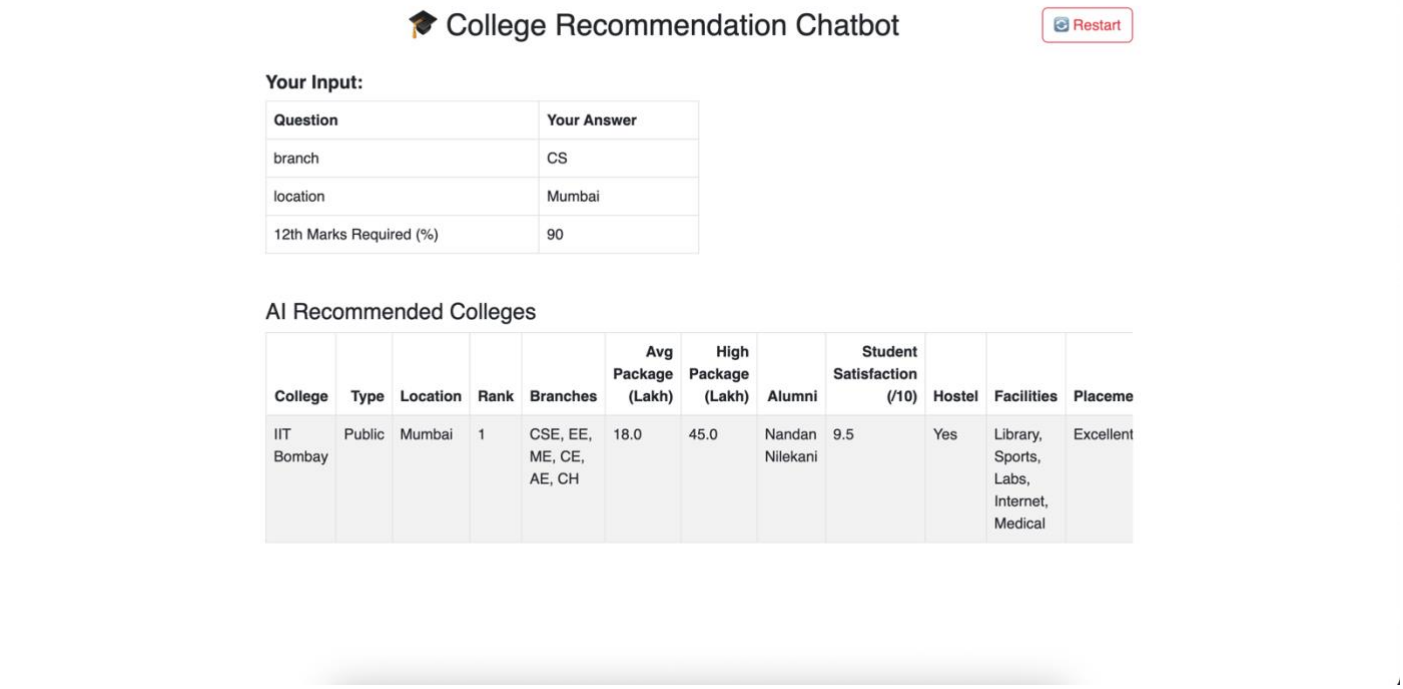
Step -2



Step -3

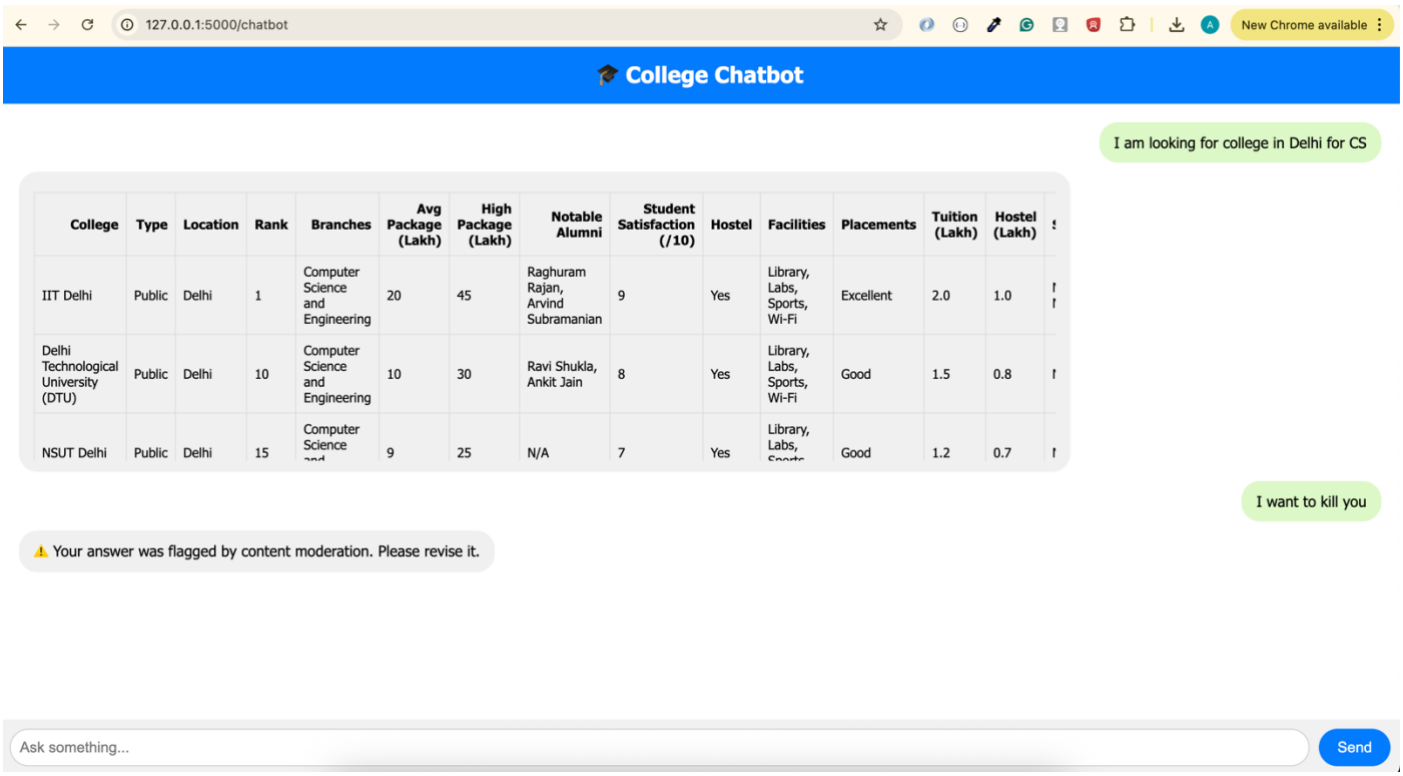


Output-



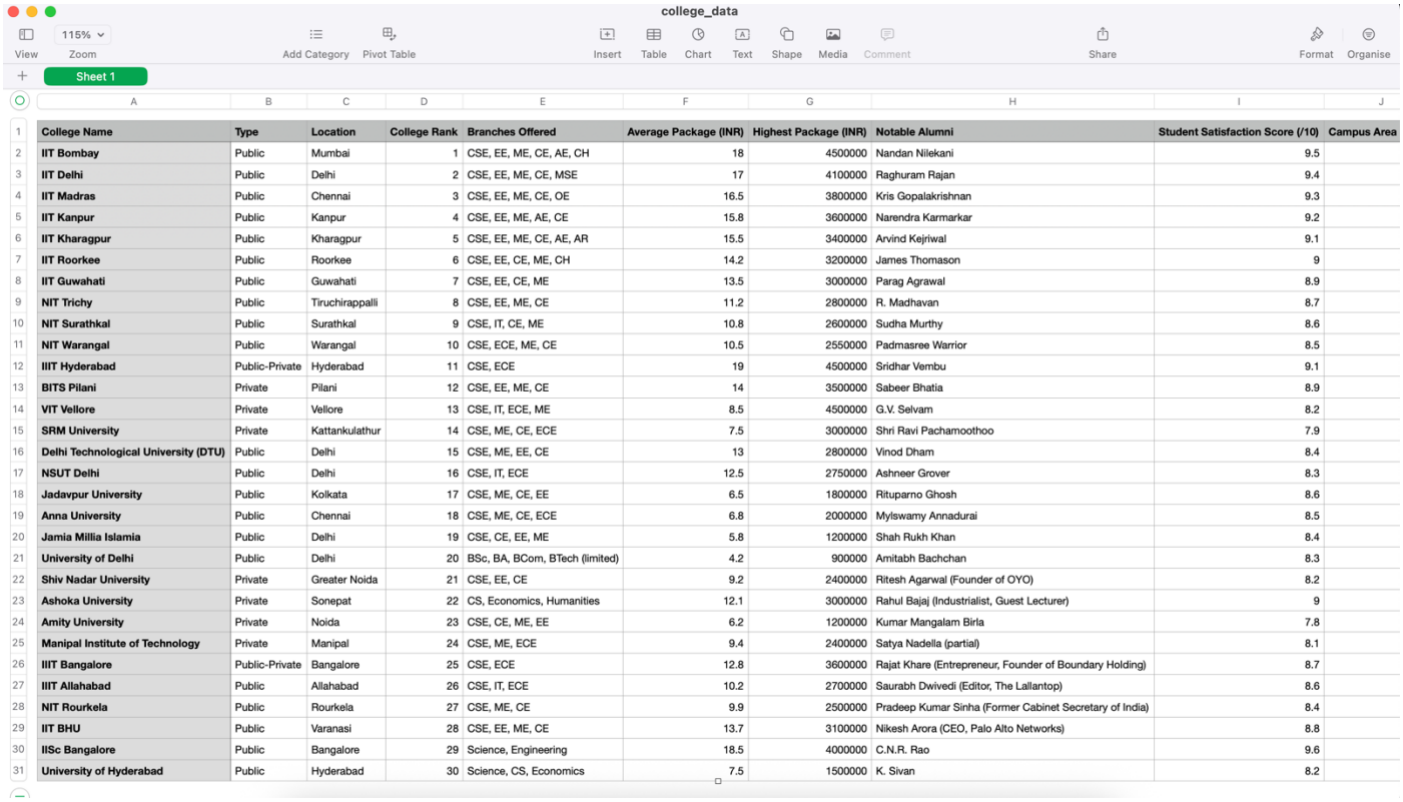
Version 2 Message Conversational view (http://localhost:5000/chatbot)

Include real + flagged input



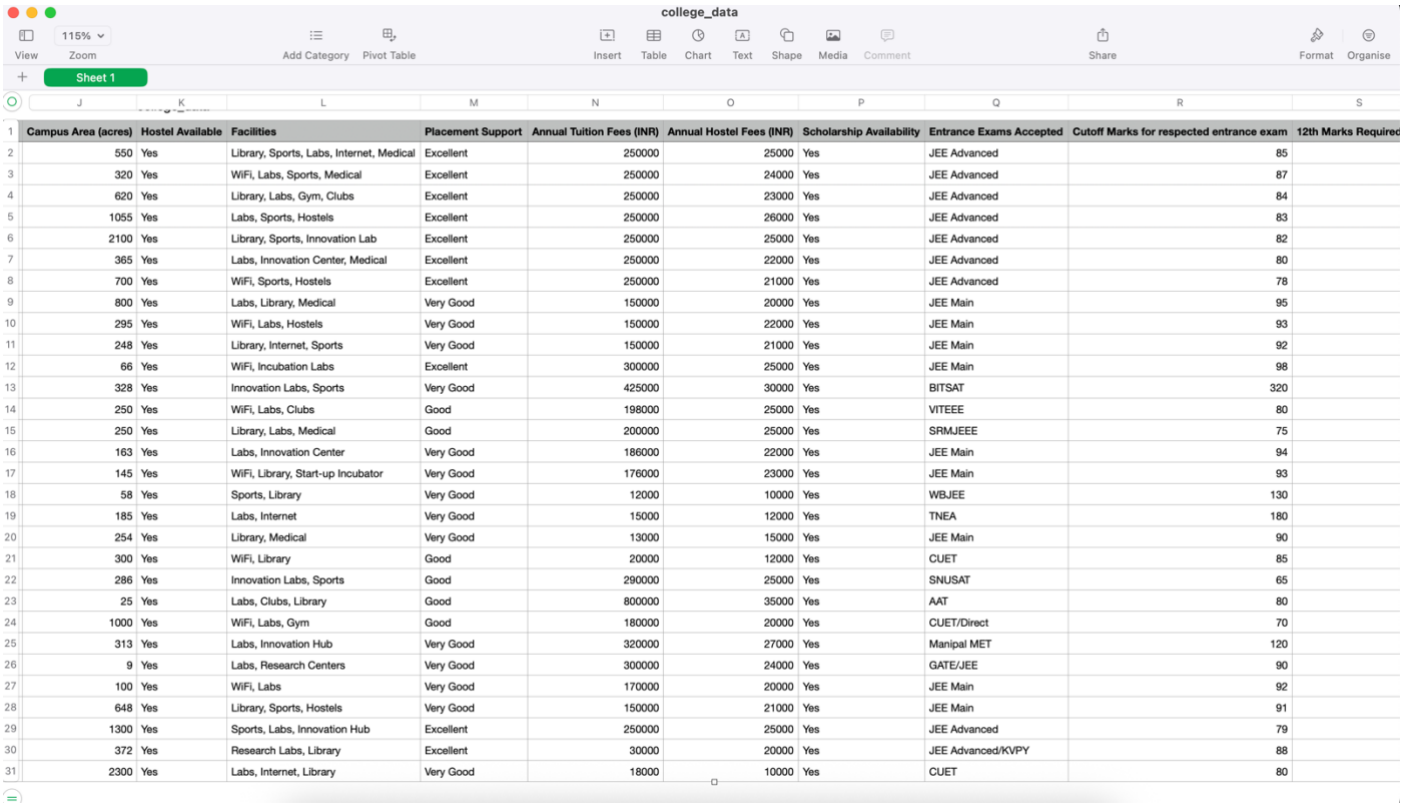
College	Type	Location	Rank	Branches	Avg Package (Lakh)	High Package (Lakh)	Notable Alumni	Student Satisfaction (/10)	Hostel	Facilities	Placements	Tuition (Lakh)	Hostel (Lakh)
IIT Delhi	Public	Delhi	1	Computer Science and Engineering	20	45	Raghuram Rajan, Arvind Subramanian	9	Yes	Library, Labs, Sports, Wi-Fi	Excellent	2.0	1.0
Delhi Technological University (DTU)	Public	Delhi	10	Computer Science and Engineering	10	30	Ravi Shukla, Ankit Jain	8	Yes	Library, Labs, Sports, Wi-Fi	Good	1.5	0.8
NSUT Delhi	Public	Delhi	15	Computer Science	9	25	N/A	7	Yes	Library, Labs,	Good	1.2	0.7

Data Sample (first few columns)



College Name	Type	Location	College Rank	Branches Offered	Average Package (INR)	Highest Package (INR)	Notable Alumni	Student Satisfaction Score (/10)	Campus Area
IIT Bombay	Public	Mumbai	1	CSE, EE, ME, CE, AE, CH	18	4500000	Nandan Nilekani	9.5	
IIT Delhi	Public	Delhi	2	CSE, EE, ME, CE, MSE	17	4100000	Raghuram Rajan	9.4	
IIT Madras	Public	Chennai	3	CSE, EE, ME, CE, OE	16.5	3800000	Kris Gopalakrishnan	9.3	
IIT Kanpur	Public	Kanpur	4	CSE, EE, ME, AE, CE	15.8	3600000	Narendra Karmarkar	9.2	
IIT Kharagpur	Public	Kharagpur	5	CSE, EE, ME, CE, AE, AR	15.5	3400000	Arvind Kejriwal	9.1	
IIT Roorkee	Public	Roorkee	6	CSE, EE, CE, ME, CH	14.2	3200000	James Thomason	9	
IIT Guwahati	Public	Guwahati	7	CSE, EE, CE, ME	13.5	3000000	Parag Agrawal	8.9	
NIT Trichy	Public	Tiruchirappalli	8	CSE, EE, ME, CE	11.2	2800000	R. Madhavan	8.7	
NIT Surathkal	Public	Surathkal	9	CSE, IT, CE, ME	10.8	2600000	Sudha Murthy	8.6	
NIT Warangal	Public	Warangal	10	CSE, ECE, ME, CE	10.5	2550000	Padmasree Warrior	8.5	
IIIT Hyderabad	Public-Private	Hyderabad	11	CSE, ECE	19	4500000	Sridhar Vembu	9.1	
BITS Pilani	Private	Pilani	12	CSE, EE, ME, CE	14	3500000	Sabeer Bhatia	8.9	
VIT Vellore	Private	Vellore	13	CSE, IT, ECE, ME	8.5	4500000	G.V. Selvam	8.2	
SRM University	Private	Kattankulathur	14	CSE, ME, CE, ECE	7.5	3000000	Shri Ravi Pachamoothoo	7.9	
Delhi Technological University (DTU)	Public	Delhi	15	CSE, ME, EE, CE	13	2800000	Vinod Dham	8.4	
NSUT Delhi	Public	Delhi	16	CSE, IT, ECE	12.5	2750000	Ashneer Grover	8.3	
Jadavpur University	Public	Kolkata	17	CSE, ME, CE, EE	6.5	1800000	Rituparno Ghosh	8.6	
Anna University	Public	Chennai	18	CSE, ME, CE, ECE	6.8	2000000	Mylswamy Annadurai	8.5	
Jamia Millia Islamia	Public	Delhi	19	CSE, CE, EE, ME	5.8	1200000	Shah Rukh Khan	8.4	
University of Delhi	Public	Delhi	20	BSc, BA, BCom, BTech (limited)	4.2	900000	Amitabh Bachchan	8.3	
Shiv Nadar University	Private	Greater Noida	21	CSE, EE, CE	9.2	2400000	Ritesh Agarwal (Founder of OYO)	8.2	
Ashoka University	Private	Sonepat	22	CS, Economics, Humanities	12.1	3000000	Rahul Bajaj (Industrialist, Guest Lecturer)	9	
Amity University	Private	Noida	23	CSE, CE, ME, EE	6.2	1200000	Kumar Mangalam Birla	7.8	
Manipal Institute of Technology	Private	Manipal	24	CSE, ME, ECE	9.4	2400000	Satya Nadella (partial)	8.1	
IIIT Bangalore	Public-Private	Bangalore	25	CSE, ECE	12.8	3600000	Rajat Khare (Entrepreneur, Founder of Boundary Holding)	8.7	
IIIT Allahabad	Public	Allahabad	26	CSE, IT, ECE	10.2	2700000	Saurabh Dwivedi (Editor, The Lallantop)	8.6	
NIT Rourkela	Public	Rourkela	27	CSE, ME, CE	9.9	2500000	Pradeep Kumar Sinha (Former Cabinet Secretary of India)	8.4	
IIT BHU	Public	Varanasi	28	CSE, EE, ME, CE	13.7	3100000	Nikesh Arora (CEO, Palo Alto Networks)	8.8	
IISc Bangalore	Public	Bangalore	29	Science, Engineering	18.5	4000000	C.N.R. Rao	9.6	
University of Hyderabad	Public	Hyderabad	30	Science, CS, Economics	7.5	1500000	K. Sivan	8.2	

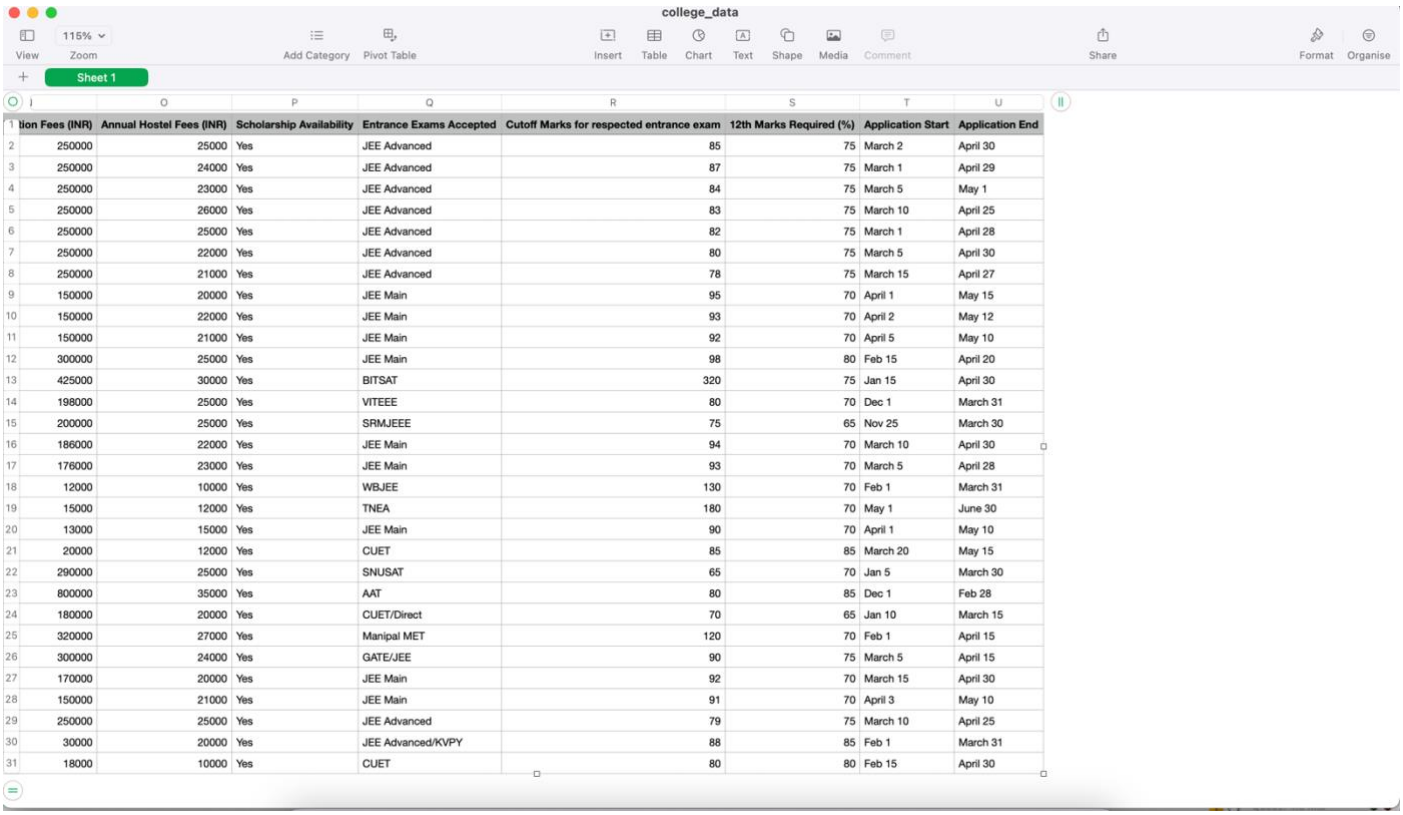
Data Sample (next few columns)



college_data

Campus Area (acres)	Hostel Available	Facilities	Placement Support	Annual Tuition Fees (INR)	Annual Hostel Fees (INR)	Scholarship Availability	Entrance Exams Accepted	Cutoff Marks for respected entrance exam	12th Marks Required
550	Yes	Library, Sports, Labs, Internet, Medical	Excellent	250000	25000	Yes	JEE Advanced	85	
320	Yes	WiFi, Labs, Sports, Medical	Excellent	250000	24000	Yes	JEE Advanced	87	
620	Yes	Library, Labs, Gym, Clubs	Excellent	250000	23000	Yes	JEE Advanced	84	
1055	Yes	Labs, Sports, Hostels	Excellent	250000	26000	Yes	JEE Advanced	83	
2100	Yes	Library, Sports, Innovation Lab	Excellent	250000	25000	Yes	JEE Advanced	82	
365	Yes	Labs, Innovation Center, Medical	Excellent	250000	22000	Yes	JEE Advanced	80	
700	Yes	WiFi, Sports, Hostels	Excellent	250000	21000	Yes	JEE Advanced	78	
800	Yes	Labs, Library, Medical	Very Good	150000	20000	Yes	JEE Main	95	
295	Yes	WiFi, Labs, Hostels	Very Good	150000	22000	Yes	JEE Main	93	
248	Yes	Library, Internet, Sports	Very Good	150000	21000	Yes	JEE Main	92	
66	Yes	WiFi, Incubation Labs	Excellent	300000	25000	Yes	JEE Main	98	
328	Yes	Innovation Labs, Sports	Very Good	425000	30000	Yes	BITSAT	320	
250	Yes	WiFi, Labs, Clubs	Good	198000	25000	Yes	VITEEE	80	
250	Yes	Library, Labs, Medical	Good	200000	25000	Yes	SRMJEEE	75	
163	Yes	Labs, Innovation Center	Very Good	186000	22000	Yes	JEE Main	94	
145	Yes	WiFi, Library, Start-up Incubator	Very Good	176000	23000	Yes	JEE Main	93	
58	Yes	Sports, Library	Very Good	12000	10000	Yes	WBJEE	130	
185	Yes	Labs, Internet	Very Good	15000	12000	Yes	TNEA	180	
254	Yes	Library, Medical	Very Good	13000	15000	Yes	JEE Main	90	
300	Yes	WiFi, Library	Good	20000	12000	Yes	CUET	85	
286	Yes	Innovation Labs, Sports	Good	290000	25000	Yes	SNUSAT	65	
25	Yes	Labs, Clubs, Library	Good	800000	35000	Yes	AAT	80	
1000	Yes	WiFi, Labs, Gym	Good	180000	20000	Yes	CUET/Direct	70	
313	Yes	Labs, Innovation Hub	Very Good	320000	27000	Yes	Manipal MET	120	
9	Yes	Labs, Research Centers	Very Good	300000	24000	Yes	GATE/JEE	90	
100	Yes	WiFi, Labs	Very Good	170000	20000	Yes	JEE Main	92	
648	Yes	Library, Sports, Hostels	Very Good	150000	21000	Yes	JEE Main	91	
1300	Yes	Sports, Labs, Innovation Hub	Excellent	250000	25000	Yes	JEE Advanced	79	
372	Yes	Research Labs, Library	Excellent	30000	20000	Yes	JEE Advanced/KVPY	88	
2300	Yes	Labs, Internet, Library	Very Good	18000	10000	Yes	CUET	80	

Data Sample (remaining columns)



college_data

ion Fees (INR)	Annual Hostel Fees (INR)	Scholarship Availability	Entrance Exams Accepted	Cutoff Marks for respected entrance exam	12th Marks Required (%)	Application Start	Application End
250000	25000	Yes	JEE Advanced	85	75	March 2	April 30
250000	24000	Yes	JEE Advanced	87	75	March 1	April 29
250000	23000	Yes	JEE Advanced	84	75	March 5	May 1
250000	26000	Yes	JEE Advanced	83	75	March 10	April 25
250000	25000	Yes	JEE Advanced	82	75	March 1	April 28
250000	22000	Yes	JEE Advanced	80	75	March 5	April 30
250000	21000	Yes	JEE Advanced	78	75	March 15	April 27
150000	20000	Yes	JEE Main	95	70	April 1	May 15
150000	22000	Yes	JEE Main	93	70	April 2	May 12
150000	21000	Yes	JEE Main	92	70	April 5	May 10
300000	25000	Yes	JEE Main	98	80	Feb 15	April 20
425000	30000	Yes	BITSAT	320	75	Jan 15	April 30
198000	25000	Yes	VITEEE	80	70	Dec 1	March 31
200000	25000	Yes	SRMJEEE	75	65	Nov 25	March 30
186000	22000	Yes	JEE Main	94	70	March 10	April 30
176000	23000	Yes	JEE Main	93	70	March 5	April 28
12000	10000	Yes	WBJEE	130	70	Feb 1	March 31
15000	12000	Yes	TNEA	180	70	May 1	June 30
13000	15000	Yes	JEE Main	90	70	April 1	May 10
20000	12000	Yes	CUET	85	85	March 20	May 15
290000	25000	Yes	SNUSAT	65	70	Jan 5	March 30
800000	35000	Yes	AAT	80	85	Dec 1	Feb 28
180000	20000	Yes	CUET/Direct	70	65	Jan 10	March 15
320000	27000	Yes	Manipal MET	120	70	Feb 1	April 15
300000	24000	Yes	GATE/JEE	90	75	March 5	April 15
170000	20000	Yes	JEE Main	92	70	March 15	April 30
150000	21000	Yes	JEE Main	91	70	April 3	May 10
250000	25000	Yes	JEE Advanced	79	75	March 10	April 25
30000	20000	Yes	JEE Advanced/KVPY	88	85	Feb 1	March 31
18000	10000	Yes	CUET	80	80	Feb 15	April 30

REF –

GIT REPO URL – https://github.com/androidev1/CollegeAssistant