

# ***Project Proposal for Data Processing and Analysis Platform***



***Current Date: February 4, 2025***

**Deadline: February 14, 2025**

**Prepared by: Waris Hayat, AI Engineer**

***Company: Artificizen, Lahore Pakistan***

# Contents

<i>Project Proposal for Data Processing and Analysis Platform</i> .....	1
<i>Project: Data Processing and Interaction Platform</i> .....	2
<i>1: Problem Statement</i> .....	2
<i>2: Problem Overview</i> .....	3
<i>3: Proposed Solution</i> .....	3
<i>4: Key Features</i> .....	3
<i>6: Timeline with Milestones (10 Days)</i> .....	5
<i>7: Deliverables</i> .....	5
<i>8: Processing and Interaction Platform Flow</i> .....	6
<i>9: Conclusion</i> .....	6

## *Project: Data Processing and Interaction Platform*

### *1: Problem Statement*

The client requires a platform to handle data processing, analysis, and user interaction efficiently. Given the tight timeline, the focus will be on delivering an MVP with core features that can be expanded later.

## ***2: Problem Overview***

The absence of an integrated platform for data processing and analysis hinders operational efficiency. The client needs a solution that supports secure user login, data summarization, financial calculations, and document management.

## ***3: Proposed Solution***

- The project will deliver an MVP with the following prioritized features:
- User authentication via Gmail and LinkedIn.
- Integration with Google Drive for linking attachments.
- Basic data fetching from APIs.
- Summarization tools for PDFs and WhatsApp chats using OpenAI APIs.
- Financial statement generation (PnL, Sales, Balance Sheets).
- Secure Retrieval-Augmented Generation (RAG) space using OpenAI/Claude.

Advanced features like hybrid search mechanisms, emoji encoding in WhatsApp processing, and dynamic heading modification will be deferred to a future phase.

## ***4: Key Features***

- **User Authentication:** Gmail and LinkedIn login via OAuth 2.0.

- **Google Drive Integration:** Ability to link attachments within workflows.
- **Data Summarization:** Summarize PDFs and WhatsApp text files using OpenAI APIs.
- **Financial Statements:** Pre-calculated Profit & Loss (PnL), Sales, and Balance Sheets with download options in CSV format.
- **Secure RAG Space:** Private access to an OpenAI/Claude-powered summarization tool.
- **Basic Graph Generation:** Simple graphs for data visualization.

## *5: Tech Stack*

- Frontend: React.js (for a responsive user interface).
- Backend: FastAPI (for high-performance REST API development).
- Database: MongoDB (quick setup for unstructured/structured data).
- APIs:
  - OpenAI API (for summarization).
  - Google Drive API (for attachment linking).
- Authentication: OAuth 2.0 for Gmail/LinkedIn login.
- Deployment: AWS services

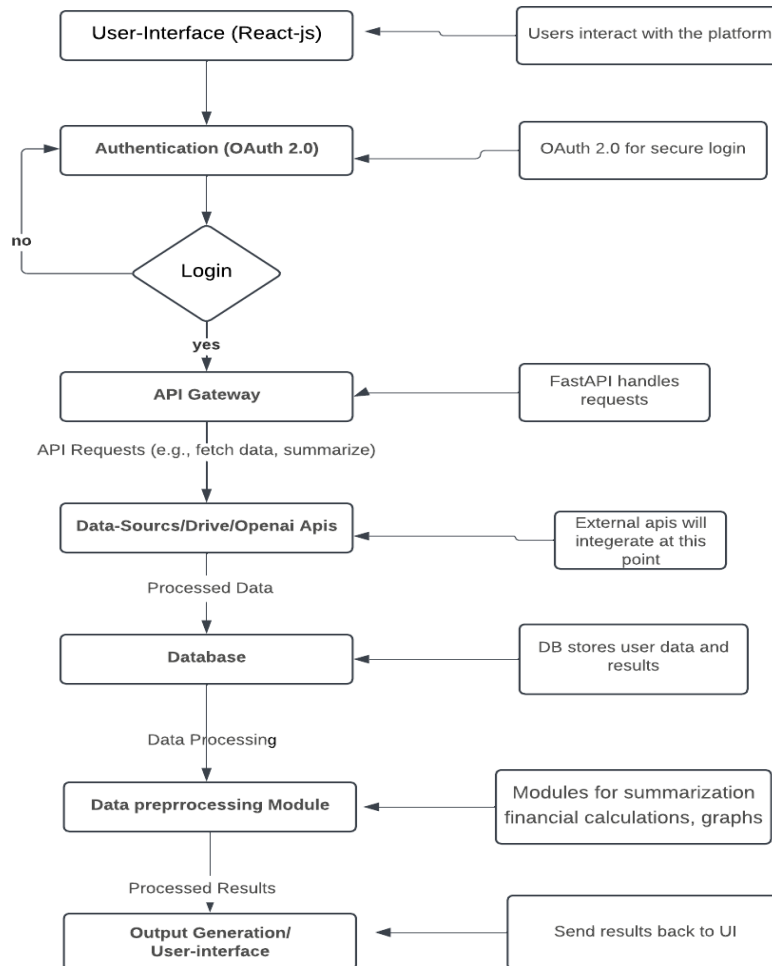
## ***6: Timeline with Milestones (10 Days)***

Day	Task
Day 1	Project kickoff; finalize requirements; set up development environment
Day 2	Implement Gmail/LinkedIn login via OAuth in FastAPI
Day 3	Integrate Google Drive API for attachment linking
Day 4	Develop basic API fetching functionality
Day 5	Build summarization module using OpenAI API
Day 6	Implement financial statement generation module
Day 7	Develop secure RAG space with OpenAI/Claude integration
Day 8	Create simple graph generation feature
Day 9	Frontend integration; finalize UI design for smooth navigation
Day 10	Testing, debugging, and deployment on AWS

## ***7: Deliverables***

- Functional MVP platform with core features implemented.
- Documentation covering:
  - System architecture.
  - API integrations.
  - User guide for core functionalities.
- Deployment of the platform on AWS:

## 8: Processing and Interaction Platform Flow



## 9: Conclusion

This proposal outlines a realistic approach to delivering an MVP within the given timeline of 10 days by leveraging FastAPI's high performance, React.js's speed of development, and AWS's scalability for deployment. The project aims to provide a functional solution that meets the client's immediate needs while allowing for future enhancements.