Data Analyst Agent - Setup Instructions

Requirements.txt

```
streamlit==1.28.0
pandas==2.0.3
numpy==1.24.3
matplotlib==3.7.2
seaborn==0.12.2
plotly==5.15.0
scikit-learn==1.3.0
scipy==1.11.1
requests==2.31.0
python-docx==0.8.11
PyPDF2==3.0.1
openpyxl==3.1.2
Pillow==10.0.0
```

Installation & Setup

1. Install Dependencies

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

2. Get Together.ai API Key

- 1. Visit <u>Together.ai</u>
- 2. Create an account
- 3. Navigate to API section
- 4. Generate an API key
- 5. Important: Only use the model (meta-llama/Llama-4-Maverick-17B-128E-Instruct-FP8)

3. Run the Application

Option A: Jupyter Notebook (Core Agent)

```
bash
jupyter notebook data_analyst_agent.ipynb
```

Option B: Streamlit Web App

streamlit run streamlit_app.py

Usage Guide

1. Using Jupyter Notebook

```
python
# Initialize the agent
agent = DataAnalystAgent("your-api-key-here")

# Load a file
result = agent.load_file("your_data.csv")
print(result)

# Get automatic insights
insights = agent.generate_insights()
print(insights)

# Create visualizations
agent.auto_visualize()

# Ask questions
answer = agent.answer_question("What are the main trends in this data?")
print(answer)
```

2. Using Streamlit App

- 1. Enter your Together.ai API key in the sidebar
- 2. Upload your file (CSV, Excel, Text, Word, PDF, or Image)
- 3. Explore different tabs:
 - Overview: Data preview and basic info
 - Analysis: Statistical analysis
 - Visualizations: Auto and custom charts
 - Q&A: Ask questions about your data
 - **Insights**: Al-generated insights

Features

Intelligent Agent Capabilities

Multi-format Support: CSV, Excel, TXT, DOCX, PDF, Images

- Automated Analysis: Statistical summaries, correlation analysis, outlier detection
- Smart Visualizations: Auto-generated charts based on data types
- **Conversational Q&A**: Ask follow-up questions with context retention
- Al-Powered Insights: Uses Llama-4-Maverick model for deep analysis

📊 Data Analysis Features

- Descriptive statistics
- Correlation analysis
- Outlier detection (IQR method)
- Missing value analysis
- Categorical data profiling
- Distribution analysis

Visualization Capabilities

- Automatic chart generation
- Correlation heatmaps
- Distribution plots
- Category bar charts
- Custom scatter plots, histograms, box plots
- Interactive Plotly charts

Q&A System

- Context-aware responses
- Conversation history
- Follow-up question support
- Data-driven answers with evidence

File Type Support

File Type	Analysis Type	Features
CSV/Excel	Structured Data	Full statistical analysis, visualizations, correlations
TXT/DOCX/PDF	Text Analysis	Content analysis, Q&A, text insights
Images	Visual Analysis	Basic image analysis and description

Model Specification

• Model: (meta-llama/Llama-4-Maverick-17B-128E-Instruct-FP8)

Provider: Together.ai

• **Temperature**: 0.3 (for consistent analytical responses)

Max Tokens: 2000 (configurable)

Architecture

Core Components

1. **DataAnalystAgent Class**: Main agent with LLM integration

2. File Processors: Handles multiple file formats

3. Statistical Analyzer: Performs comprehensive data analysis

4. Visualization Engine: Creates charts based on data types

5. **Q&A System**: Context-aware question answering

6. **Insight Generator**: Al-powered data insights

Design Principles

• Intelligence over Simplicity: Built as an agent, not just a chatbot

• Context Retention: Maintains conversation history for follow-ups

Automated Decision Making: Chooses appropriate analysis methods

• Comprehensive Coverage: Handles both structured and unstructured data

• User-Friendly Interface: Clean Streamlit UI with intuitive navigation

Example Use Cases

Business Analytics

- Sales data analysis
- Customer segmentation insights
- Performance trend analysis
- Market research data exploration

Research & Academia

- Research data analysis
- Statistical hypothesis testing
- Survey response analysis
- Literature review (PDF analysis)

Personal Projects

• Personal finance tracking

- Health data analysis
- Social media analytics
- Project documentation analysis

Troubleshooting

Common Issues

- 1. API Key Error: Ensure you're using the correct Together.ai API key
- 2. **Model Access**: Verify you have access to the Llama-4-Maverick model
- 3. File Upload Issues: Check file size limits and format support
- 4. **Memory Issues**: Large files may require chunking for analysis

Performance Tips

- Use CSV format for large structured datasets
- Break down large text files into smaller sections
- Use specific questions for better AI responses
- Regularly clear conversation history for better performance

Contributing

This is a comprehensive data analyst agent designed for the specified requirements. The implementation focuses on:

- Process intelligence (30 marks): Advanced agent architecture with context retention
- Data analysis & visualization (30 marks): Comprehensive statistical analysis and auto-visualizations
- Multi-format Q&A (30 marks): Support for various file types with intelligent responses
- Efficient implementation (10 marks): Clean, modular code with user-friendly interface