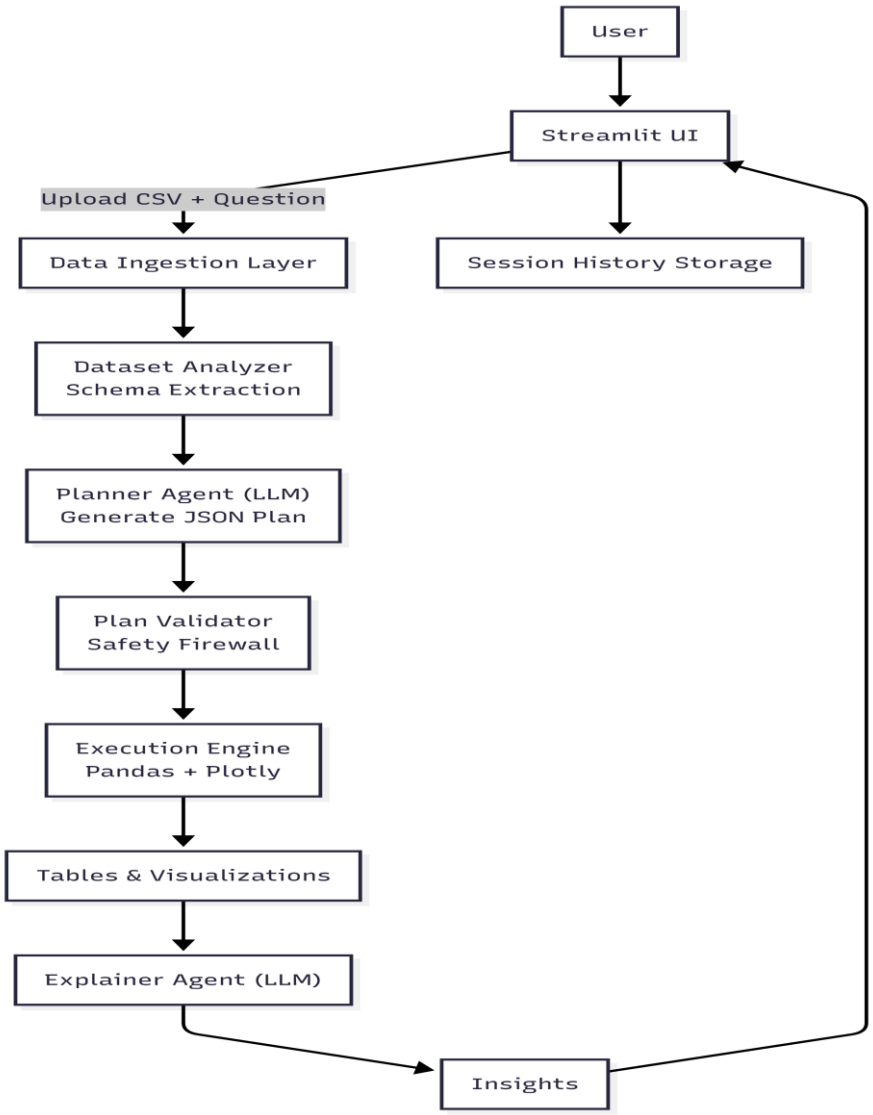
# Technical Report

## AI Data Analyst Agent for CSV-Based Analysis

## 1. System Architecture

The AI Data Analyst Agent is an end-to-end, modular system designed to perform **structured, safe, and reproducible analysis** on user-uploaded CSV datasets. The system follows a **Planner–Validator–Executor–Explainer** architecture and is implemented as an interactive **Streamlit web application**.

The core design principle is a strict separation between **LLM-based reasoning** and **deterministic data execution**, ensuring transparency, traceability, and reproducibility.



# 1.1 High-Level Architecture Overview

**AI Data Analyst Agent**

The AI Data Analyst Agent is an end-to-end analytics system that transforms natural language questions into safe, structured, and reproducible insights from user-uploaded CSV datasets.

The architecture follows a layered pipeline that separates:

- user interaction
- AI reasoning
- safety validation
- deterministic execution
- explanation

This separation ensures transparency, auditability, and reliable analytics.

# 1.2 System Layers

## 1.2.1 User Interaction Layer — Streamlit UI

The Streamlit interface is the entry point of the system.

**It allows users to:**

- upload CSV datasets
- preview data
- ask questions in natural language
- view results and insights
- track session history

This layer handles presentation only — no analytics logic is executed here.

## 1.2.2 Data Ingestion Layer

The ingestion layer converts uploaded CSV files into structured Pandas DataFrames.

**Responsibilities:**

- CSV parsing
- encoding handling
- dataset validation
- in-memory dataset creation

**Output:**

`Validated Pandas DataFrame`
This becomes the source dataset for the analytics pipeline.

### 1.2.3 Dataset Understanding Layer

The Dataset Analyzer extracts schema-level metadata:

- column names
- data types
- dataset structure

Only schema metadata is shared with the AI planner.

Raw dataset values are never exposed to the planner.

This prevents hallucinated analytics.

### 1.2.4 Intelligence Layer — Planner Agent (LLM)

The Planner Agent converts natural language questions into a strict JSON analysis plan.

**It determines:**

- analysis type
- metrics
- grouping
- filters
- ranking rules
- visualization strategy

The planner behaves like a compiler:

```
English → Analytics Instructions
```

It never executes data operations.

### 1.2.5 Safety Layer — Plan Validator

The validator acts as a firewall between AI output and execution.

**It ensures:**

- valid columns
- allowed operations
- schema compliance
- safe visualization rules

Invalid plans are rejected before execution.

### 1.2.6 Execution Layer — Deterministic Engine

The execution engine performs analytics using Pandas and Plotly.

**It handles:**

- filtering
- aggregation
- ranking
- correlation
- distribution analysis
- visualization generation

All operations are deterministic and reproducible.

No AI decisions occur here.

### 1.2.7 Explanation Layer — Explainer Agent (LLM)

The Explainer Agent interprets results and generates human-friendly insights.

**It:**

- summarizes findings
- explains patterns
- avoids technical jargon
- communicates business meaning

It does not modify results — only explains them.

### 1.2.8 Presentation Layer

Final outputs include:

- result tables
- charts
- textual insights
- session history

These are displayed back to the user in Streamlit.

# 1.3 End-to-End Flow

1. User uploads CSV and asks a question
2. Dataset is ingested into Pandas
3. Schema metadata is extracted
4. Planner generates JSON analysis plan
5. Validator checks plan safety
6. Executor runs deterministic analysis
7. Explainer generates insights
8. Results are displayed to the user

## 1.4 Architectural Benefits

This design provides:

- safe AI-driven analytics
- reproducible execution
- zero hallucinated computation
- explainable results
- modular extensibility
- enterprise-ready architecture

# 2. Agent Responsibilities and Interactions

The system follows a multi-agent design where each agent has a single, clearly scoped responsibility. This separation improves modularity, safety, and maintainability.

### 2.1 Planner Agent

**Purpose**

The Planner Agent is responsible for reasoning and decision-making. It translates natural language questions into structured, machine-readable JSON analysis plans without performing any computation.

**Inputs:**

- User's natural language question
- Dataset schema metadata
- Predefined allowed operations and constraints

**Responsibilities:**

- Interpret analytical intent (aggregation, comparison, trends, correlation)
- Identify relevant dataset columns
- Define filters, group-by fields, metrics, sorting, and visualization configuration
- Attach user intent metadata (highest, lowest, both)
- Sanitize and normalize plans to remove unsafe or invalid instructions

**Key Design Characteristics:**

- Uses an LLM only for planning and reasoning
- Outputs JSON only, never executable code
- Does not access raw dataset values

**Interaction:**

- Sends validated JSON plans to the Validation layer

## 2.2 Executor Agent

**Purpose:**

The Executor Agent performs deterministic and reproducible execution of validated analysis plans.

**Inputs:**

- Validated JSON plan
- Original dataset (Pandas DataFrame)

**Responsibilities:**

- Apply filters safely with numeric coercion
- Perform group-by and aggregation operations
- Apply sorting and Top-N logic deterministically
- Generate structured result tables
- Generate Plotly visualizations

**Key Design Characteristics:**

- No LLM usage
- No dynamic code execution
- Only predefined Pandas and Plotly operations are allowed

**Interaction:**

- Outputs results to the Explainer Agent

## 2.3 Explainer Agent

**Purpose:**

The Explainer Agent communicates analytical results in a clear, concise, and business-friendly manner.

**Inputs:**

- User's original question
- Execution results and rankings
- Contextual metadata from the plan

**Responsibilities:**

- Interpret numerical outputs
- Identify meaningful patterns and extremes
- Generate concise insights grounded strictly in execution results

**Key Design Characteristics:**

- Uses an LLM only for explanation
- Prevents hallucination by restricting inputs

### 2.4 Dataset Analyzer (Supporting Component)

**Purpose:**
Provides schema-level understanding of uploaded datasets.

**Responsibilities:**

- Extract column names and data types
- Generate compact schema summaries
- Assist planning and validation stages

## 3. Planning Schema and Execution Safeguards

### 3.1 JSON Planning Schema

Each plan explicitly defines:

- Analysis type
- Filters
- Group-by columns
- Metrics
- Sorting rules
- Visualization configuration
- User intent metadata

### 3.2 Validation Rules

- Allowed operations only
- Valid dataset columns
- Safe aggregation functions
- Visualization constraints

Invalid plans are rejected prior to execution.

### 3.3 Execution Safeguards

- No LLM-generated code is executed
- Only Pandas and Plotly operations are permitted
- Safe numeric coercion is enforced
- Deterministic execution is guaranteed

# 4. End-to-End Application

The Streamlit application provides:

- CSV upload and validation
- Natural language query interface
- Display of dataset preview
- Display of JSON analysis plans
- Tables and visualizations
- Final natural-language insights
- Session-level query history

# 5. Evaluation Protocol

The system was evaluated using a combination of **automated** and **human** evaluation:

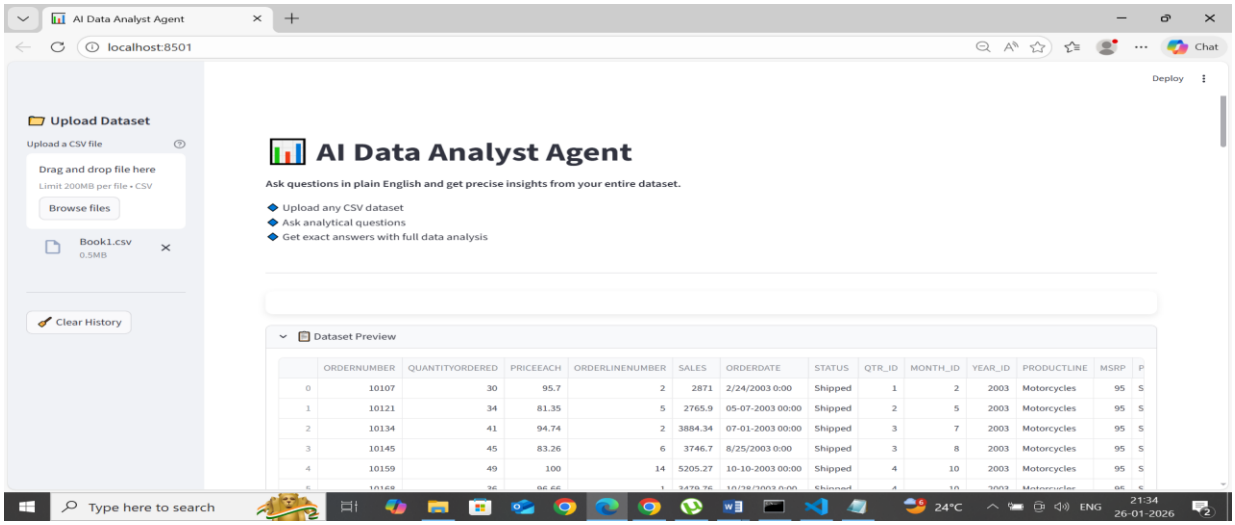
- **Automated evaluation:** verifies schema compliance, execution correctness, and deterministic behavior
- **Human evaluation:** assesses clarity, usefulness, and faithfulness of generated insights

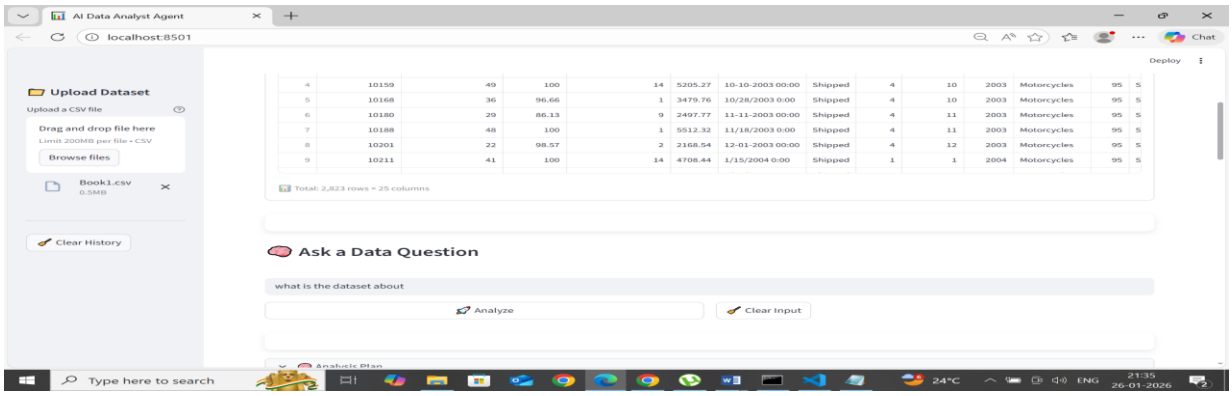
Evaluation configurations are defined declaratively in `experiments/*.yaml`.

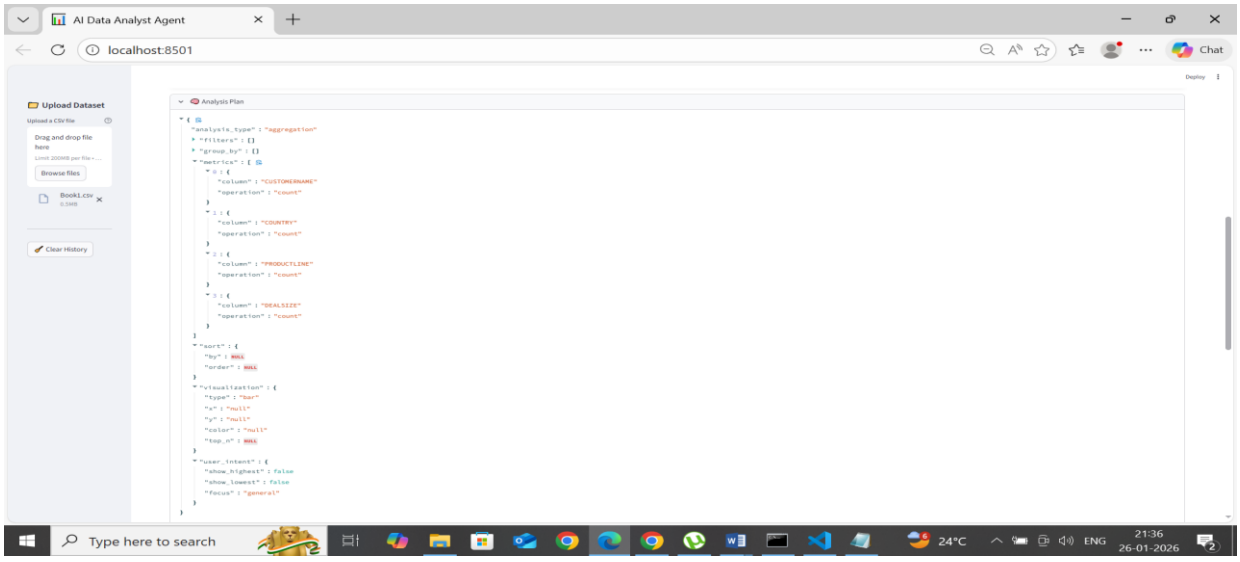
# 6. Appendix: Modular Agent Design – Pseudocode

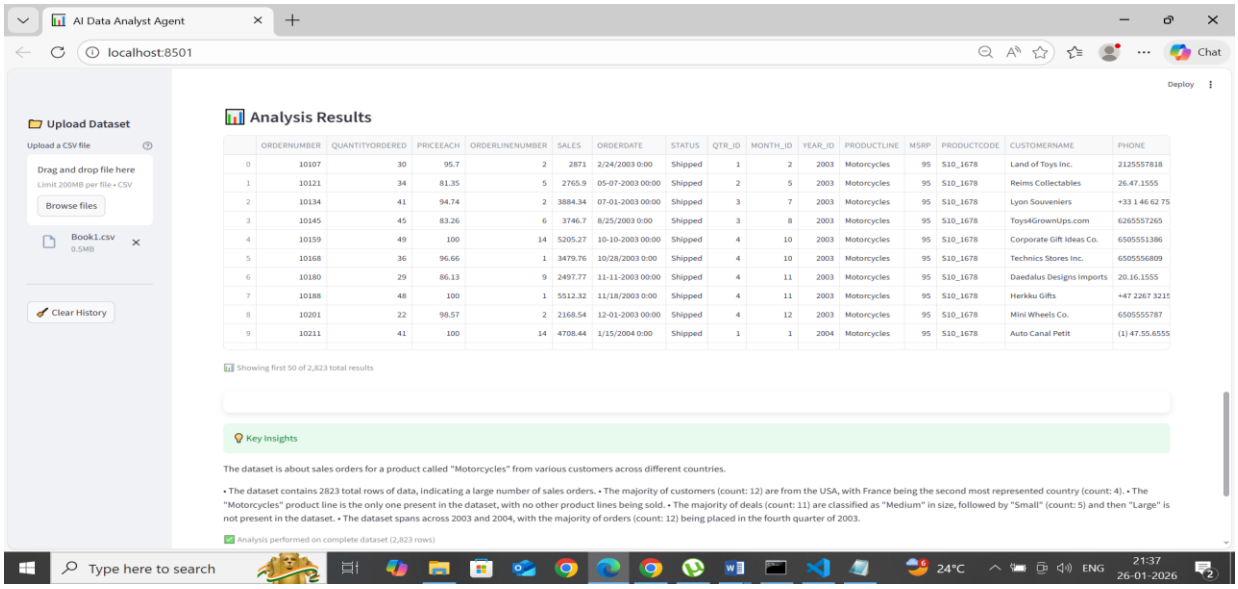
```
START APPLICATION
LOAD Streamlit User Interface
WAIT for user to upload CSV dataset
WAIT for user to enter a natural language question

IF dataset and question are provided:
   SCHEMA ← analyze_dataset(dataset)
   PLAN ← PlannerAgent.generate_plan(SCHEMA, question)
   VALIDATE PLAN against schema and dataset columns

   IF plan is valid:
      RESULTS, CHARTS ← Executor.execute_plan(dataset, PLAN)
      INSIGHTS ← ExplainerAgent.generate_insights(question, RESULTS)
      DISPLAY results, charts, and insights
   ELSE:
      DISPLAY validation error
END
```

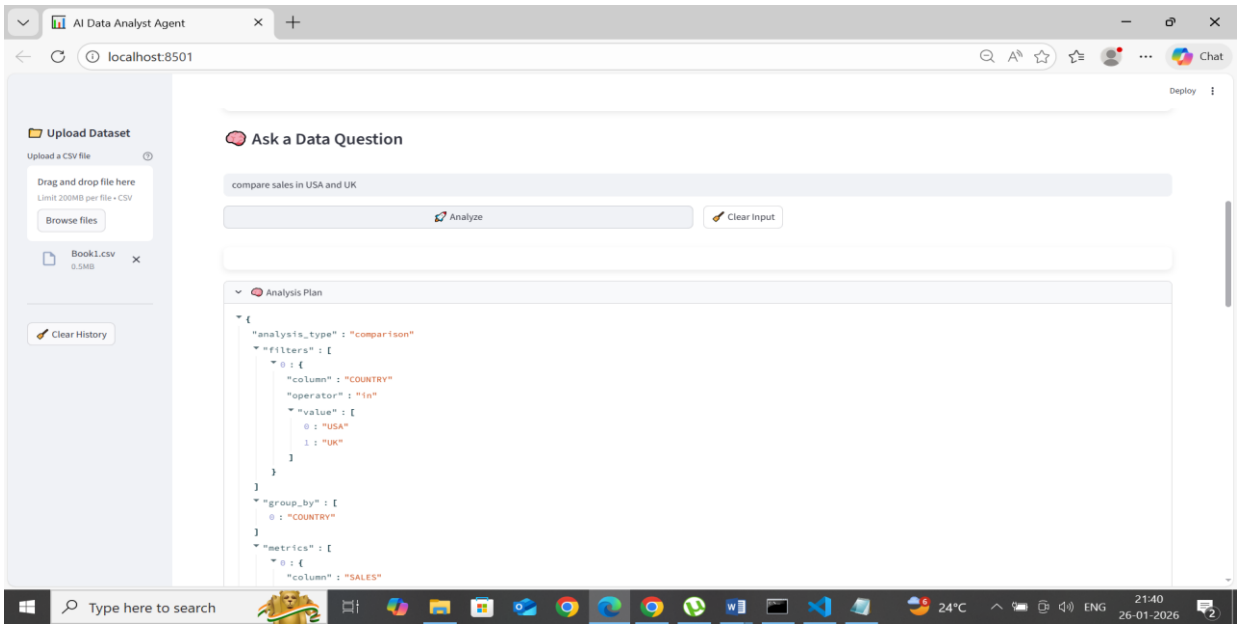
# EXAMPLE USE CASES:

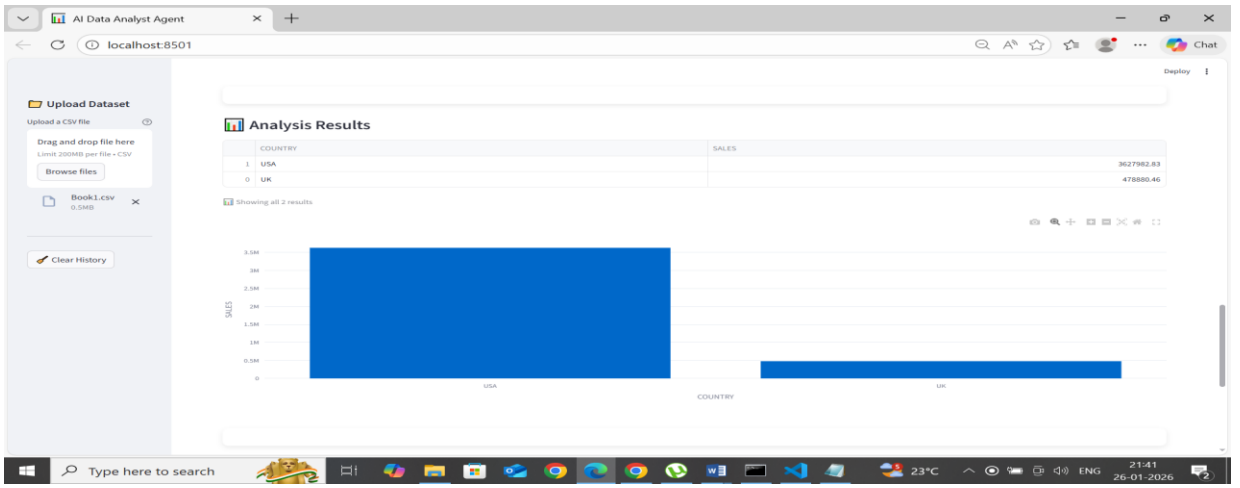


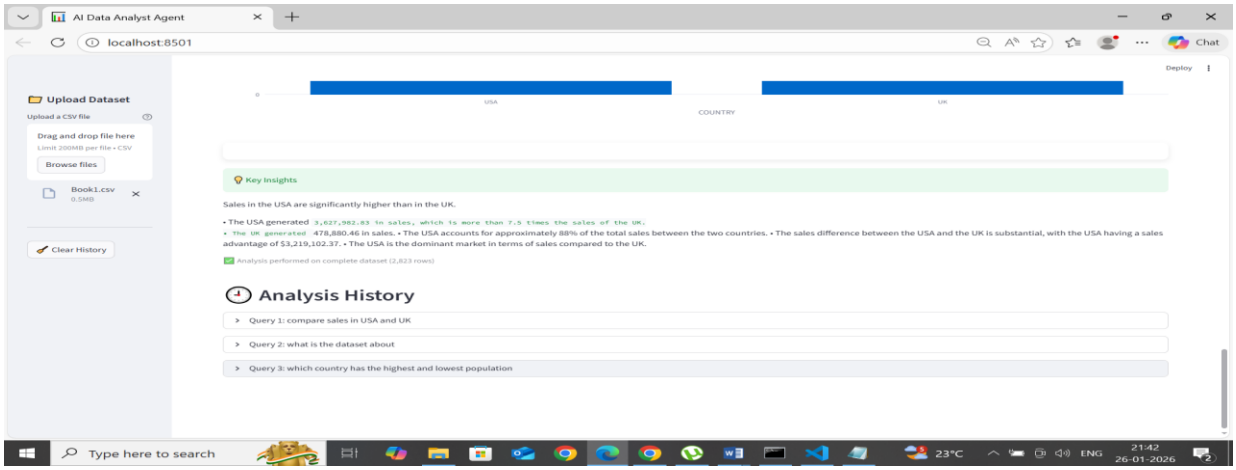



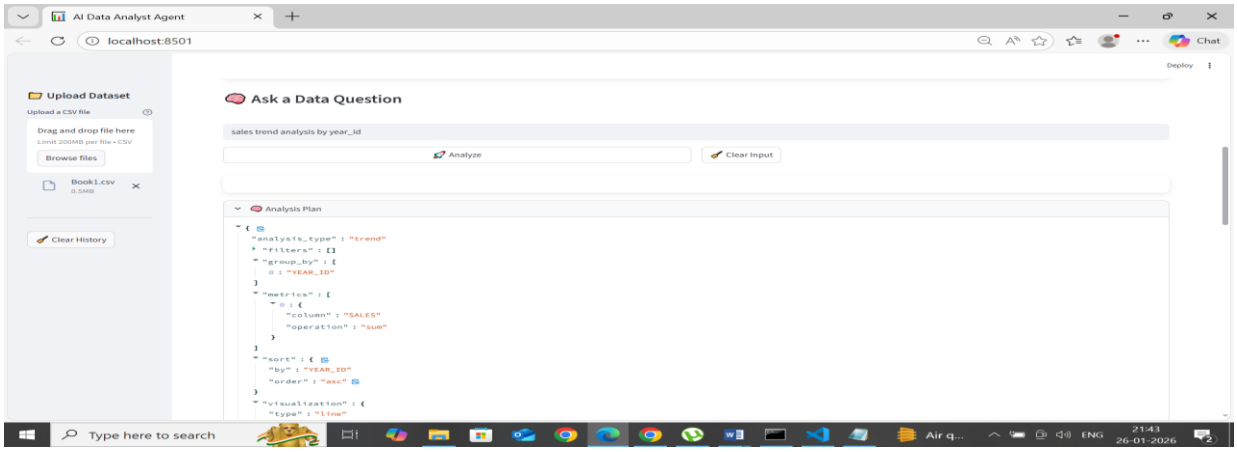



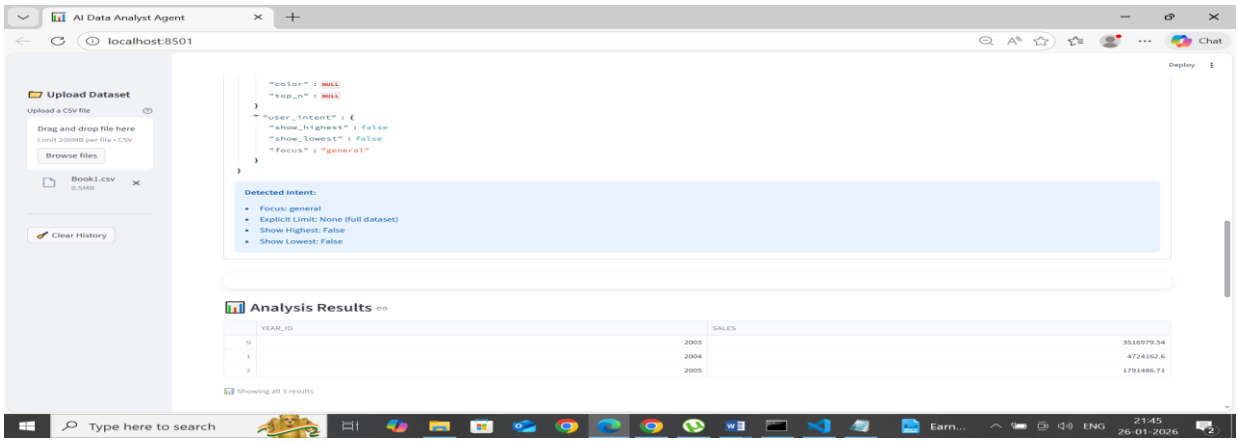

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
Analysis Results
ORDERNUMBER QUANTITYORDERED PRICEEACH ORDERLINENUMBER SALES ORDERDATE STATUS QTR_ID MONTH_ID YEAR_ID PRODUCTLINE MSRP PRODUCTCODE CUSTOMERNAME PHONE
0 10107 30 95.7 2 2871 2/24/2003 0:00 Shipped 1 2 2003 Motorcycles 95 S10_1678 Land of Toys Inc. 2125557818
1 10121 34 81.35 5 2765.9 05-07-2003 00:00 Shipped 2 5 2003 Motorcycles 95 S10_1678 Reims Collectables 26.47.1555
2 10134 41 94.74 2 3884.34 07-01-2003 00:00 Shipped 3 7 2003 Motorcycles 95 S10_1678 Lyon Souveniers +33 1 46 62 75
3 10145 45 83.26 6 3746.7 8/25/2003 0:00 Shipped 3 8 2003 Motorcycles 95 S10_1678 Toys4GrownUps.com 6265557265
4 10159 49 100 14 5205.27 10-10-2003 00:00 Shipped 4 10 2003 Motorcycles 95 S10_1678 Corporate Gift Ideas Co. 6505551386
5 10168 36 96.66 1 3479.76 10/28/2003 0:00 Shipped 4 10 2003 Motorcycles 95 S10_1678 Technics Stores Inc. 6505556809
6 10180 29 86.13 9 2497.77 11-11-2003 00:00 Shipped 4 11 2003 Motorcycles 95 S10_1678 Daedalus Designs Imports 20.16.1555
7 10188 48 100 1 5512.32 11/18/2003 0:00 Shipped 4 11 2003 Motorcycles 95 S10_1678 Herkku Gifts +47 2267 3215
8 10201 22 98.57 2 2168.54 12-01-2003 00:00 Shipped 4 12 2003 Motorcycles 95 S10_1678 Mini Wheels Co. 6505555787
9 10211 41 100 14 4708.44 1/15/2004 0:00 Shipped 1 1 2004 Motorcycles 95 S10_1678 Auto Canal Petit (1) 47.55.6555
Showing first 50 of 2,823 total results
Key Insights
The dataset is about sales orders for a product called "Motorcycles" from various customers across different countries.
• The dataset contains 2823 total rows of data, indicating a large number of sales orders. • The majority of customers (count: 12) are from the USA, with France being the second most represented country (count: 4). • The "Motorcycles" product line is the only one present in the dataset, with no other product lines being sold. • The majority of deals (count: 11) are classified as "Medium" in size, followed by "Small" (count: 5) and then "Large" is not present in the dataset. • The dataset spans across 2003 and 2004, with the majority of orders (count: 12) being placed in the fourth quarter of 2003.
Analysis performed on complete dataset (2,823 rows)
Type here to search
24°C
ENG
21:37
26-01-2026

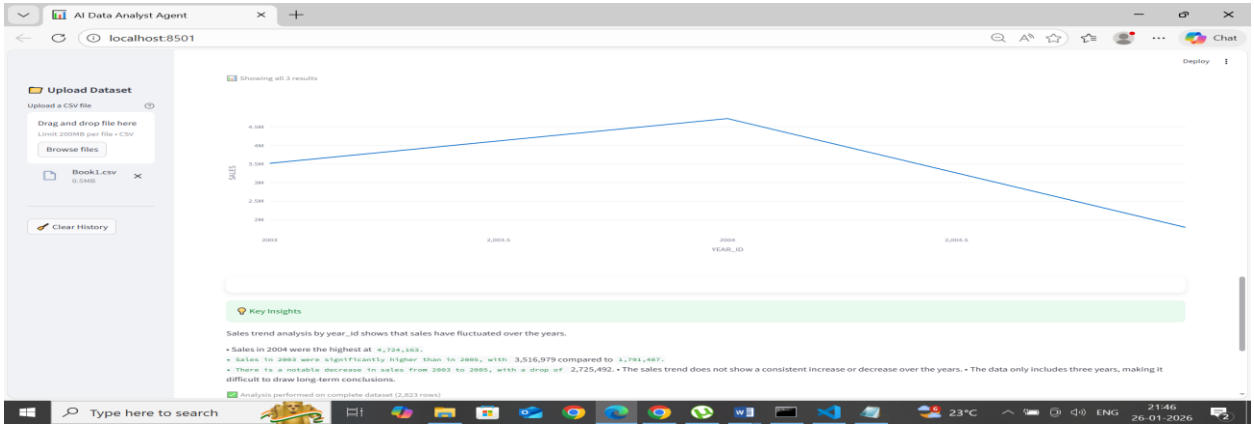

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
Ask a Data Question
compare sales in USA and UK
Analyze
Clear Input
Analysis Plan
{
"analysis_type" : "comparison"
"filters" : [
0 : {
"column" : "COUNTRY"
"operator" : "in"
"value" : [
0 : "USA"
1 : "UK"
]
}
]
"group_by" : [
0 : "COUNTRY"
]
"metrics" : [
0 : {
"column" : "SALES"
Type here to search
24°C
ENG
21:40
26-01-2026

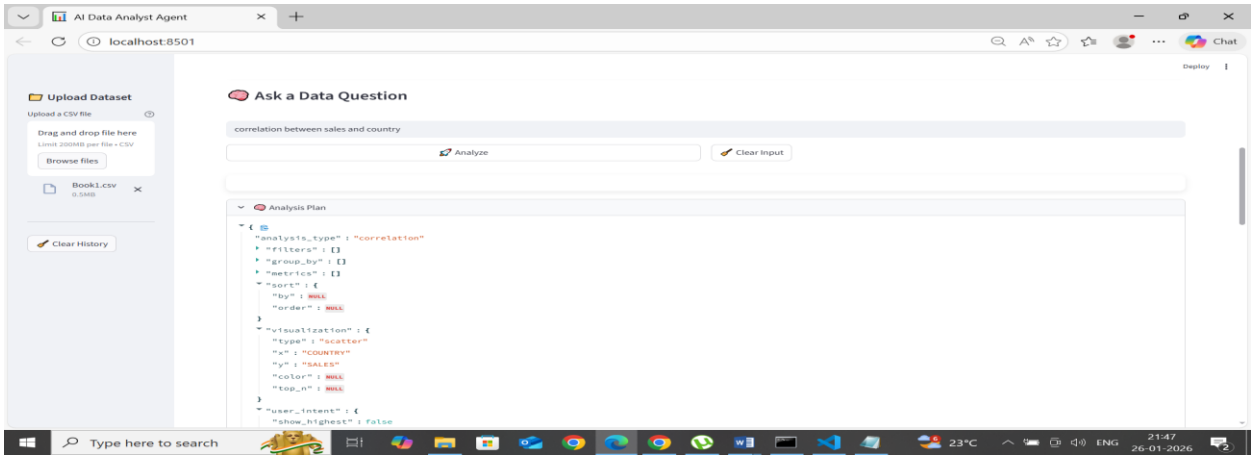

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
Analysis Results
COUNTRY SALES
1 USA 3627982.83
0 UK 478880.46
Showing all 2 results
SALES
COUNTRY
USA
UK
Type here to search
23°C
ENG
21:41
26-01-2026

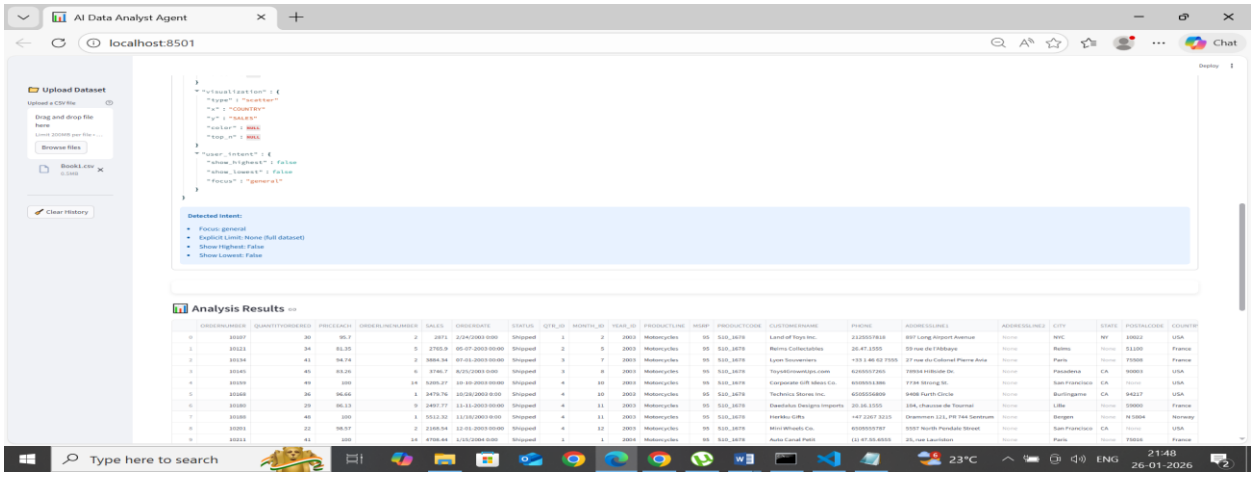

AI Data Analyst Agent
localhost:8501
Upload Dataset
Drag and drop file here
Browse files
Book1.csv
Clear History
USA
UK
COUNTRY
Key Insights
Sales in the USA are significantly higher than in the UK.
Analysis History
Query 1: compare sales in USA and UK
Query 2: what is the dataset about
Query 3: which country has the highest and lowest population


AI Data Analyst Agent
localhost:8501
Upload Dataset
Ask a Data Question
sales trend analysis by year_id
Analyze
Clear Input
Analysis Plan
"analysis_type" : "trend"
"filters" : []
"group_by" : [
"YEAR_ID"
"metrics" : [
"column" : "SALES"
"operation" : "sum"
"sort" : {
"by" : "YEAR_ID"
"order" : "asc"
"visualization" : {
"type" : "line"

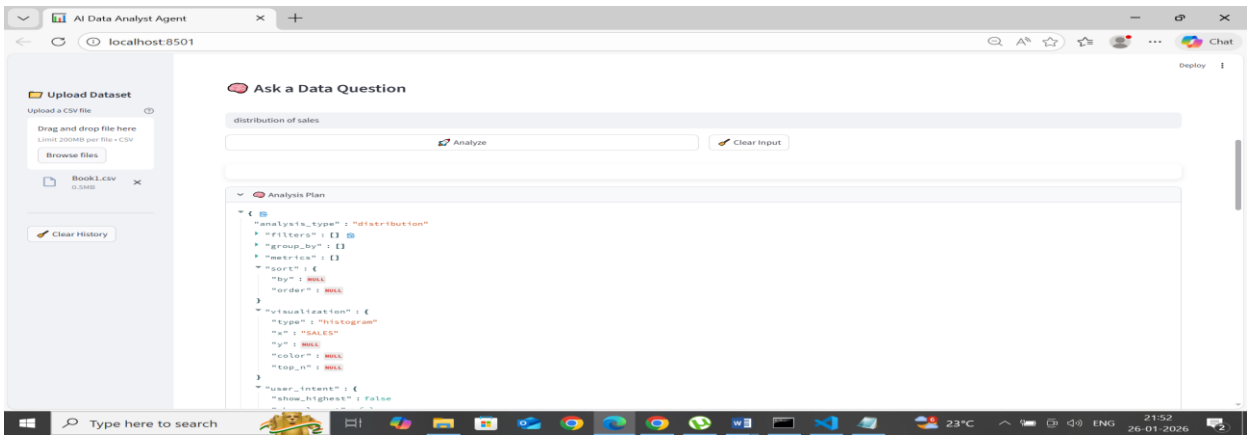

AI Data Analyst Agent
localhost:8501
Upload Dataset
"color" : NULL
"top_n" : NULL
"user_intent" : {
"show_highest" : false
"show_lowest" : false
"focus" : "general"
Detected Intent:
Focus: general
Explicit Limit: None (full dataset)
Show Highest: False
Show Lowest: False
Analysis Results
YEAR_ID
SALES
2003
3516979.54
2004
4724162.6
2005
1791486.71
Showing all 3 results

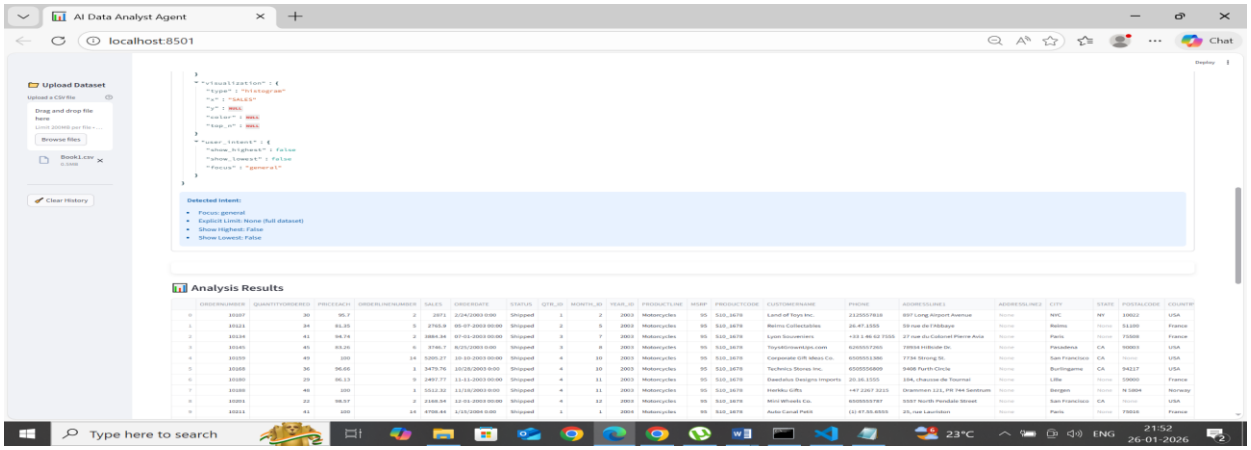

AI Data Analyst Agent
localhost:8501
Upload Dataset
Showing all 3 results
SALES
YEAR_ID
Key Insights
Sales trend analysis by year_id shows that sales have fluctuated over the years.
• Sales in 2004 were the highest at 4,724,162.
• Sales in 2003 were significantly higher than in 2005, with 3,516,979 compared to 1,791,487.
• There is a notable decrease in sales from 2003 to 2005, with a drop of 2,725,492. • The sales trend does not show a consistent increase or decrease over the years. • The data only includes three years, making it difficult to draw long-term conclusions.
Analysis performed on complete dataset (2,823 rows)

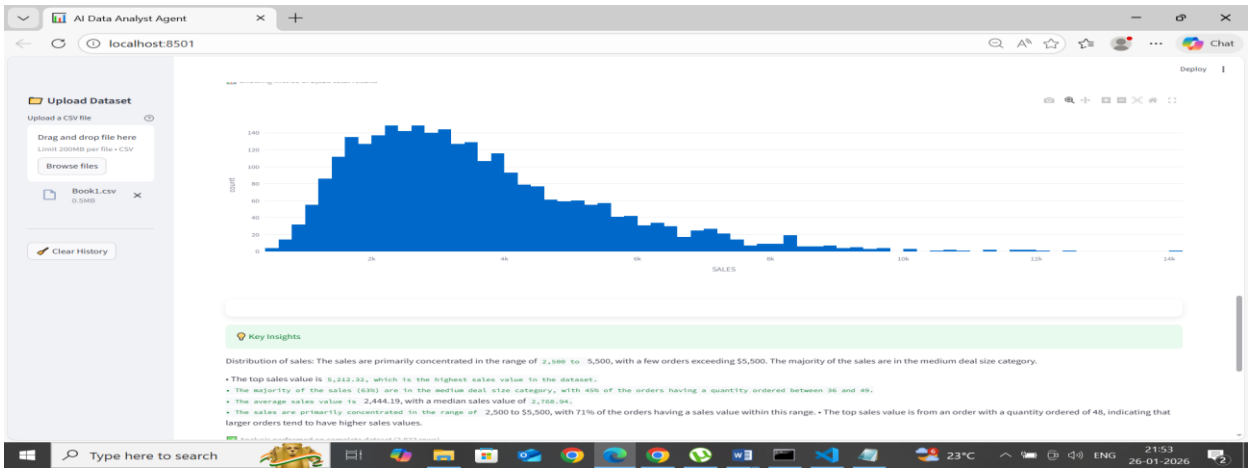

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
Ask a Data Question
correlation between sales and country
Analyze
Clear Input
Analysis Plan
{
"analysis_type" : "correlation"
"filters" : []
"group_by" : []
"metrics" : []
"sort" : {
"by" : NULL
"order" : NULL
}
"visualization" : {
"type" : "scatter"
"x" : "COUNTRY"
"y" : "SALES"
"color" : NULL
"top_n" : NULL
}
"user_intent" : {
"show_highest" : false
Type here to search
23°C
ENG
21:47
26-01-2026

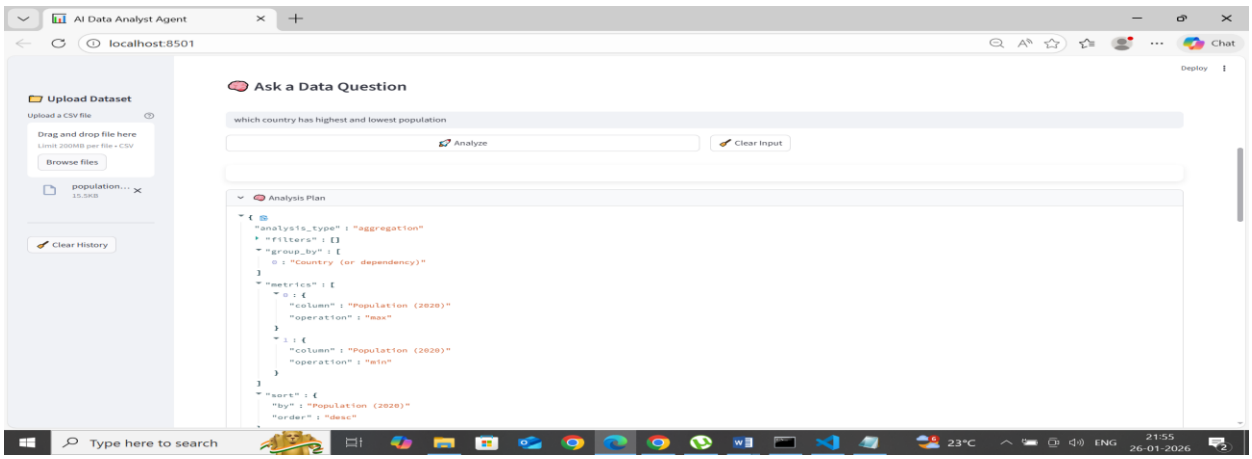

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
"visualization" : {
"type" : "scatter"
"x" : "COUNTRY"
"y" : "SALES"
"color" : NULL
"top_n" : NULL
}
"user_intent" : {
"show_highest" : false
"show_lowest" : false
"focus" : "general"
}
Detected intent:
Focus: general
Explicit Limit: None (full dataset)
Show Highest: False
Show Lowest: False
Analysis Results
Type here to search
23°C
ENG
21:48
26-01-2026

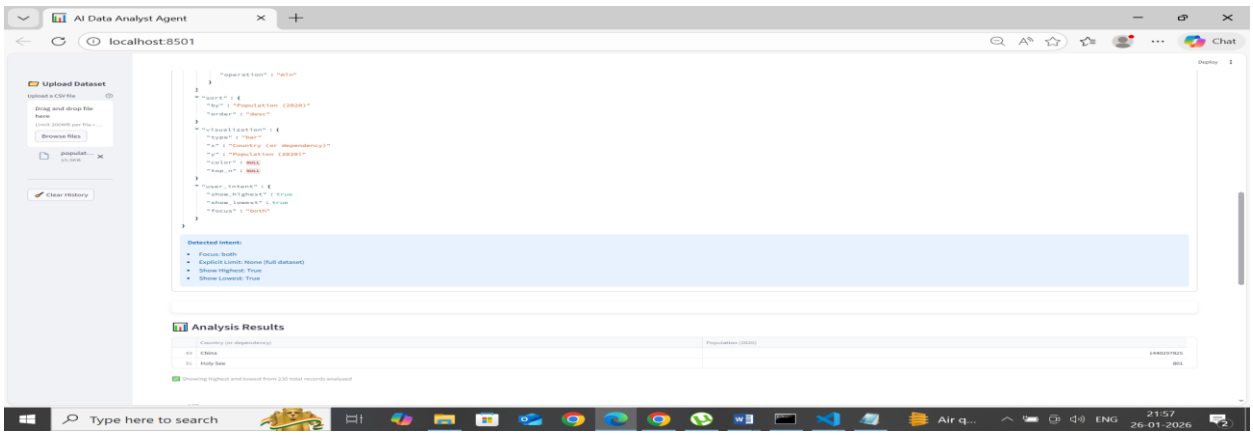

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
Showing first 50 of 2,823 total results
SALES
USA
France
Norway
Australia
Finland
Austria
UK
Spain
Sweden
Singapore
Canada
Japan
Italy
Denmark
Belgium
Philippines
Germany
Switzerland
Ireland
COUNTRY
Key Insights
Direct Answer and key insights: There is a correlation between sales and country, with the USA having the highest average sales. The top countries by average sales are USA, France, and Norway.
• The USA has the highest average sales at 3,434.45, with a total of 7 orders.
• France has an average sales of 3,144.11, with a total of 3 orders. • Norway has an average sales of $5,512.32, with a total of 1 order. • The country with the lowest average sales is not specified in the data, but it is clear that there is a significant variation in sales across different countries. • The correlation between sales and country suggests that there may be regional differences in customer behavior or market conditions that are driving sales.
Type here to search
23°C
ENG
21:49
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
Ask a Data Question
distribution of sales
Analyze
Clear Input
Analysis Plan
{
"analysis_type" : "distribution"
"filters" : []
"group_by" : []
"metrics" : []
"sort" : {
"by" : NULL
"order" : NULL
}
"visualization" : {
"type" : "histogram"
"x" : "SALES"
"y" : NULL
"color" : NULL
"top_n" : NULL
}
"user_intent" : {
"show_highest" : false
Type here to search
23°C
ENG
21:52
26-01-2026

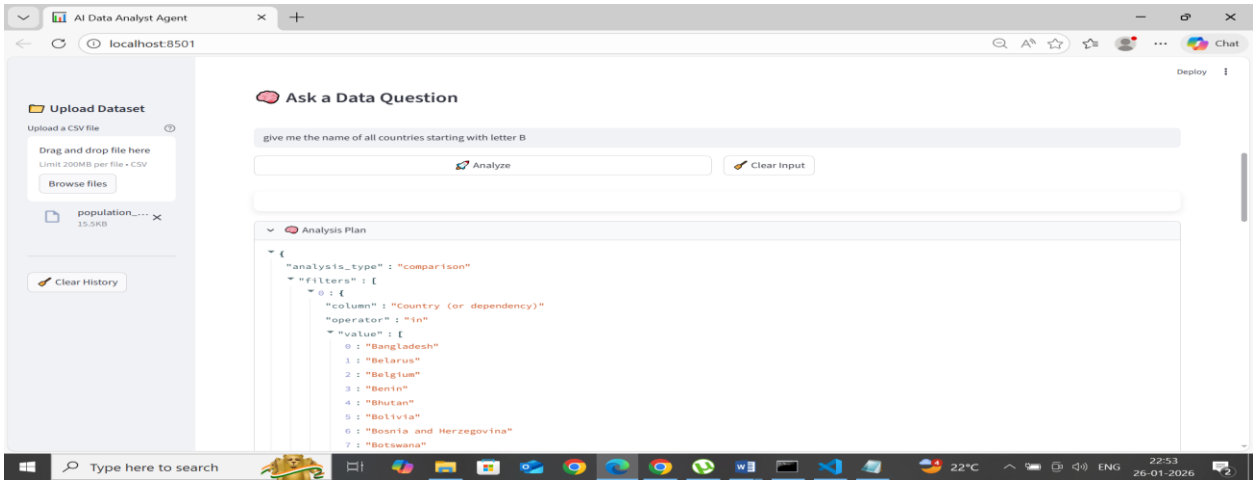

AI Data Analyst Agent
localhost:8501
Chat
Upload Dataset
Clear History
Detected Intent:
Focus: general
Explicit Limit: None (full dataset)
Show Highest: False
Show Lowest: False
Analysis Results
Type here to search
23°C
ENG
21:52
26-01-2026

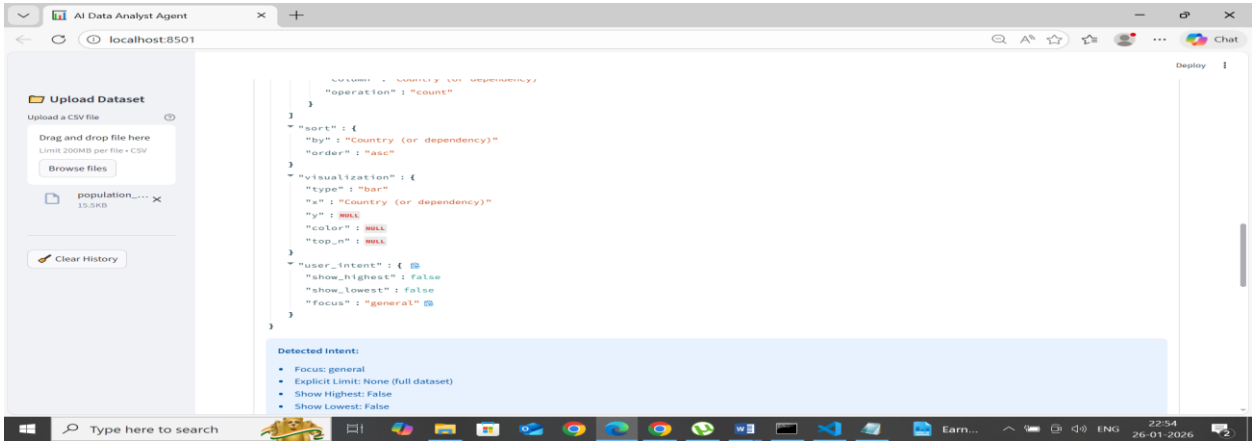

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Book1.csv
0.5MB
Clear History
count
SALES
Key Insights
Distribution of sales: The sales are primarily concentrated in the range of 2,500 to 5,500, with a few orders exceeding $5,500. The majority of the sales are in the medium deal size category.
• The top sales value is 1,212.32, which is the highest sales value in the dataset.
• The majority of the sales (63%) are in the medium deal size category, with 45% of the orders having a quantity ordered between 36 and 49.
• The average sales value is 2,444.19, with a median sales value of 2,768.94.
• The sales are primarily concentrated in the range of 2,500 to $5,500, with 71% of the orders having a sales value within this range. • The top sales value is from an order with a quantity ordered of 48, indicating that larger orders tend to have higher sales values.
Type here to search
23°C
ENG
21:53
26-01-2026

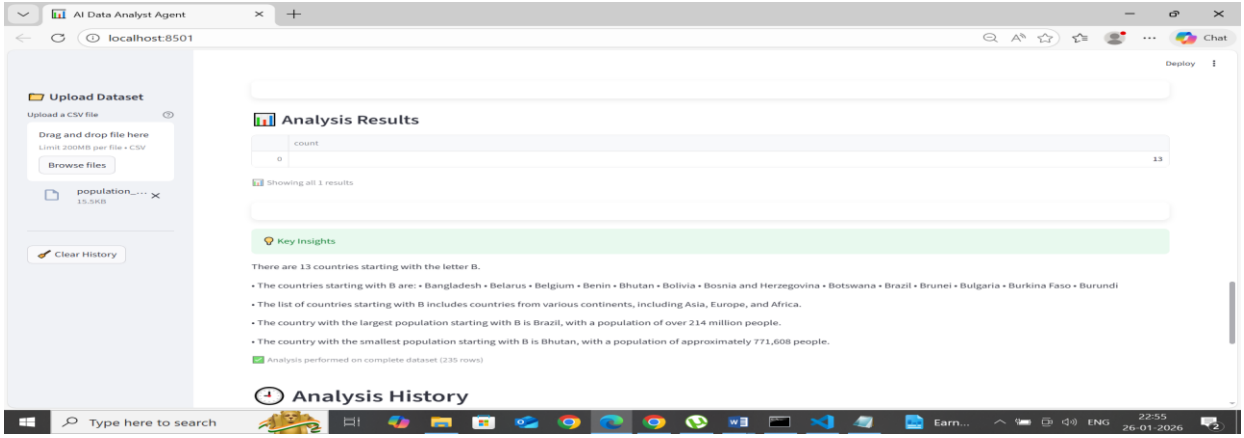

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population...
15.5KB
Clear History
Ask a Data Question
which country has highest and lowest population
Analyze
Clear Input
Analysis Plan
{
"analysis_type" : "aggregation"
"filters" : []
"group_by" : [
0 : "Country (or dependency)"
]
"metrics" : [
0 : {
"column" : "Population (2020)"
"operation" : "max"
}
1 : {
"column" : "Population (2020)"
"operation" : "min"
}
]
"sort" : {
"by" : "Population (2020)"
"order" : "desc"
Type here to search
23°C
ENG
21:55
26-01-2026

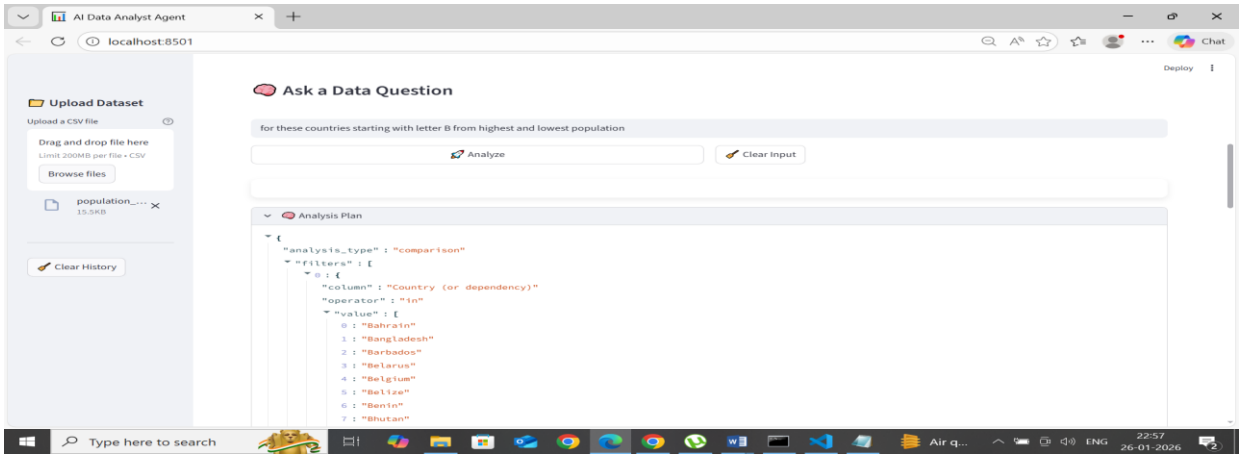

AI Data Analyst Agent
localhost:8501
Chat
Upload Dataset
Clear History
"operation" : "min"
"sort" : {
"by" : "Population (2020)"
"order" : "desc"
"visualization" : {
"type" : "Bar"
"x" : "Country (or dependency)"
"y" : "Population (2020)"
"color" : NULL
"top_n" : NULL
"user_intent" : {
"show_highest" : true
"show_lowest" : true
"focus" : "both"
Detected Intent:
Focus: both
Explicit Limit: None (full dataset)
Show Highest: True
Show Lowest: True
Analysis Results
Type here to search
Air q...
ENG
21:57
26-01-2026

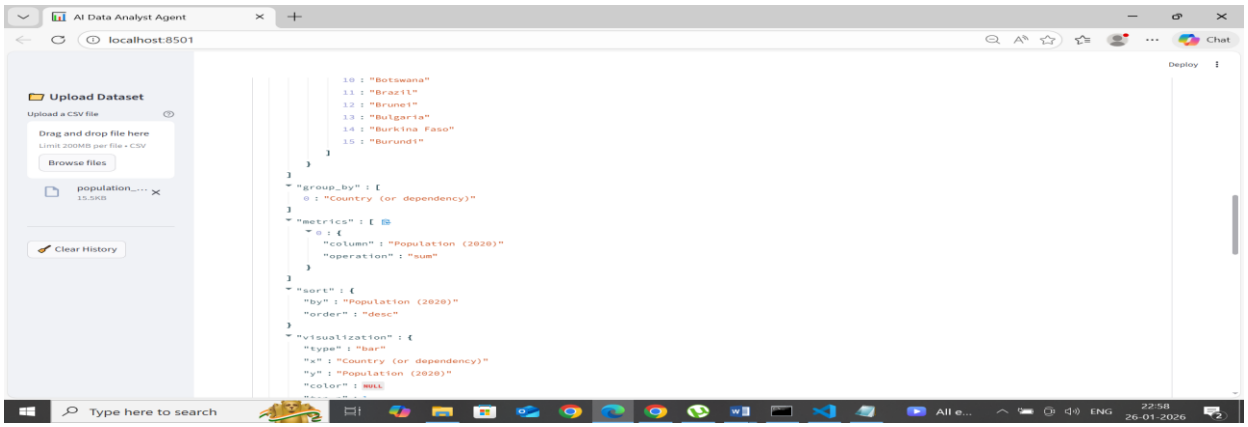

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
Population (2020)
Country (or dependency)
Key Insights
The country with the highest population is China with 1.44 billion people, and the country with the lowest population is the Holy See with 801 people.
• China has a population more than 1.7 million times larger than the Holy See. • The Holy See is a sovereign city-state located within Rome, Italy. • China's population accounts for approximately 18% of the world's total population. • The population difference between China and the Holy See is roughly 1.44 billion people.
Analysis performed on complete dataset (235 rows)
Type here to search
ENG
21:58
26-01-2026

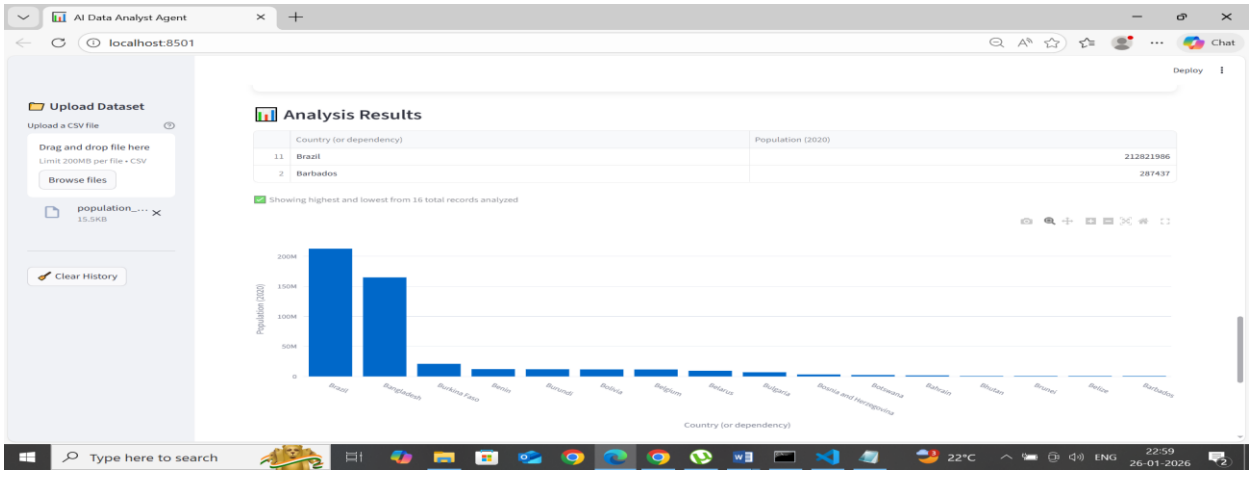

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
Ask a Data Question
give me the name of all countries starting with letter B
Analyze
Clear Input
Analysis Plan
{
"analysis_type" : "comparison"
"filters" : [
0 : {
"column" : "Country (or dependency)"
"operator" : "in"
"value" : [
0 : "Bangladesh"
1 : "Belarus"
2 : "Belgium"
3 : "Benin"
4 : "Bhutan"
5 : "Bolivia"
6 : "Bosnia and Herzegovina"
7 : "Botswana"
Type here to search
22°C
ENG
22:53
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
"operation" : "count"
}
]
"sort" : {
"by" : "Country (or dependency)"
"order" : "asc"
}
"visualization" : {
"type" : "bar"
"x" : "Country (or dependency)"
"y" : NULL
"color" : NULL
"top_n" : NULL
}
"user_intent" : {
"show_highest" : false
"show_lowest" : false
"focus" : "general"
}
}
Detected Intent:
Focus: general
Explicit Limit: None (full dataset)
Show Highest: False
Show Lowest: False
Type here to search
Earn...
ENG
22:54
26-01-2026

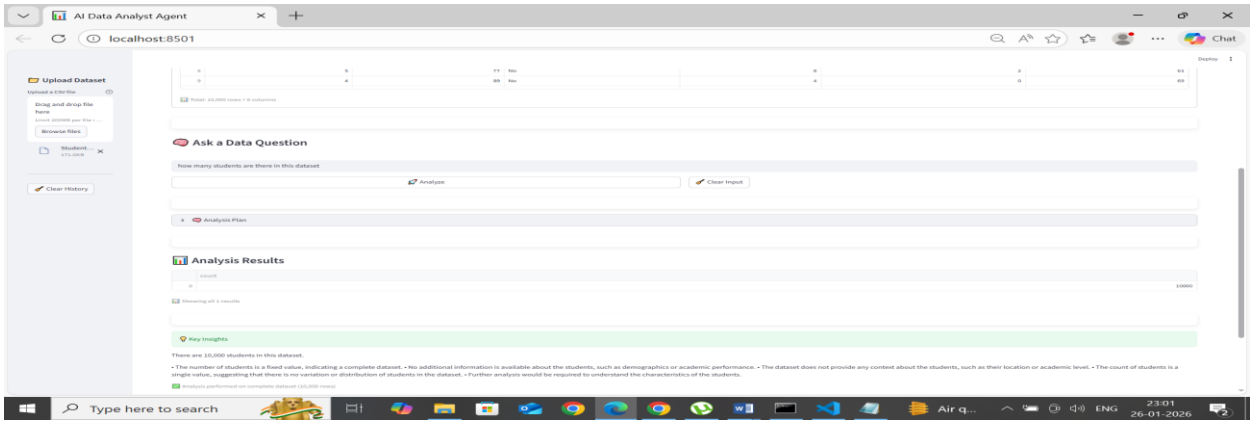

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
Analysis Results
count
0
13
Showing all 1 results
Key Insights
There are 13 countries starting with the letter B.
• The countries starting with B are: • Bangladesh • Belarus • Belgium • Benin • Bhutan • Bolivia • Bosnia and Herzegovina • Botswana • Brazil • Brunei • Bulgaria • Burkina Faso • Burundi
• The list of countries starting with B includes countries from various continents, including Asia, Europe, and Africa.
• The country with the largest population starting with B is Brazil, with a population of over 214 million people.
• The country with the smallest population starting with B is Bhutan, with a population of approximately 771,608 people.
Analysis performed on complete dataset (235 rows)
Analysis History
Type here to search
Earn...
ENG
22:55
26-01-2026

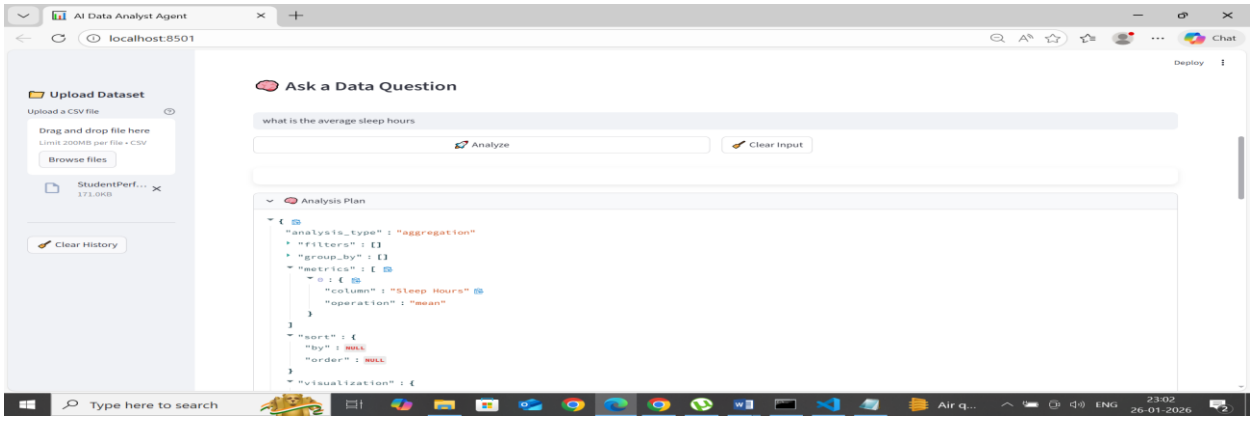

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
Ask a Data Question
for these countries starting with letter B from highest and lowest population
Analyze
Clear Input
Analysis Plan
{
"analysis_type" : "comparison"
"filters" : [
0 : {
"column" : "Country (or dependency)"
"operator" : "in"
"value" : [
0 : "Bahrain"
1 : "Bangladesh"
2 : "Barbados"
3 : "Belarus"
4 : "Belgium"
5 : "Belize"
6 : "Benin"
7 : "Bhutan"
Type here to search
ENG
22:57
26-01-2026

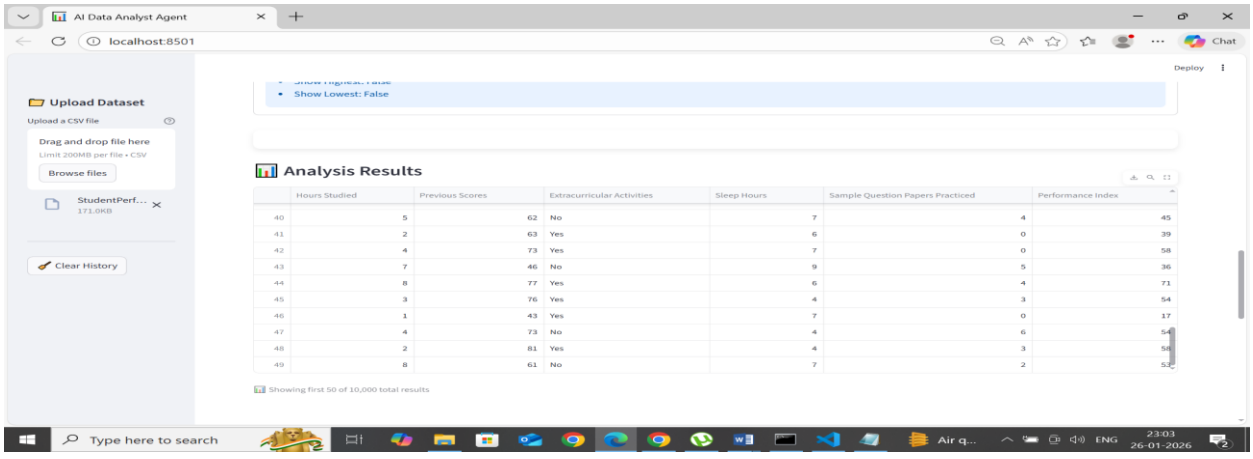

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
10 : "Botswana"
11 : "Brazil"
12 : "Brunei"
13 : "Bulgaria"
14 : "Burkina Faso"
15 : "Burundi"
"group_by" : [
0 : "Country (or dependency)"
"metrics" : [
"column" : "Population (2020)"
"operation" : "sum"
"sort" : {
"by" : "Population (2020)"
"order" : "desc"
"visualization" : {
"type" : "bar"
"x" : "Country (or dependency)"
"y" : "Population (2020)"
"color" : null
Type here to search
ENG
22:58
26-01-2026

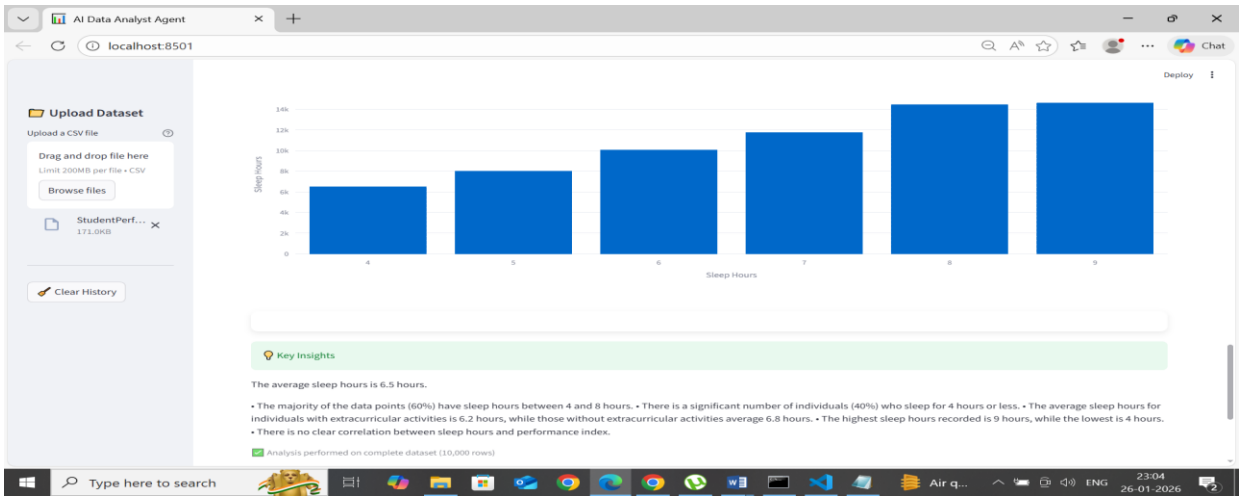

AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
Analysis Results
Country (or dependency)
Population (2020)
11 Brazil 212821986
2 Barbados 287437
Showing highest and lowest from 16 total records analyzed
200M
150M
100M
50M
0
Population (2020)
Brazil
Bangladesh
Burkina Faso
Benin
Burundi
Bolivia
Belgium
Belarus
Bulgaria
Bosnia and Herzegovina
Botswana
Bahrain
Bhutan
Brunei
Belize
Barbados
Country (or dependency)
Type here to search
22°C
ENG
22:59
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
population_....
15.5KB
Clear History
200M
150M
100M
50M
0
Population (2020)
Brazil
Bangladesh
Burkina Faso
Benin
Burundi
Bolivia
Belgium
Belarus
Bulgaria
Bosnia and Herzegovina
Botswana
Bahrain
Bhutan
Brunei
Belize
Barbados
Country (or dependency)
Key Insights
Brazil has the highest population among the countries starting with the letter B, with a population of 212,821,986 in 2020. Barbados has the lowest population, with 287,437 people in 2020.
- The top 5 countries by population starting with the letter B are Brazil, Bangladesh, Belgium, Bolivia, and Belarus. - Brazil's population is more than 730 times larger than Barbados'. - The combined population of the 16 countries starting with the letter B is approximately 4.3 billion people. - The average population of these 16 countries is around 268 million people. - The population of Bangladesh is more than 740 times larger than that of Barbados.
Analysis performed on complete dataset (235 rows)
Type here to search
22°C
ENG
23:00
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Clear History
Ask a Data Question
how many students are there in this dataset
Analyze
Clear Input
Analysis Plan
Analysis Results
Key Insights
There are 10,000 students in this dataset.
Type here to search
ENG
23:01
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
StudentPerf...
171.0KB
Clear History
Ask a Data Question
what is the average sleep hours
Analyze
Clear Input
Analysis Plan
"analysis_type" : "aggregation"
"filters" : []
"group_by" : []
"metrics" : [
"column" : "Sleep Hours"
"operation" : "mean"
"sort" : {
"by" : NULL
"order" : NULL
"visualization" : {
Type here to search
ENG
23:02
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
StudentPerf...
171.0KB
Clear History
Show Lowest: False
Analysis Results
Hours Studied
Previous Scores
Extracurricular Activities
Sleep Hours
Sample Question Papers Practiced
Performance Index
40 5 62 No 7 4 45
41 2 63 Yes 6 0 39
42 4 73 Yes 7 0 58
43 7 46 No 9 5 36
44 8 77 Yes 6 4 71
45 3 76 Yes 4 3 54
46 1 43 Yes 7 0 17
47 4 73 No 4 6 54
48 2 81 Yes 4 3 58
49 8 61 No 7 2 53
Showing first 50 of 10,000 total results
Type here to search
ENG
23:03
26-01-2026


AI Data Analyst Agent
localhost:8501
Chat
Deploy
Upload Dataset
Upload a CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
StudentPerf...
171.0KB
Clear History
Sleep Hours
Key Insights
The average sleep hours is 6.5 hours.
• The majority of the data points (60%) have sleep hours between 4 and 8 hours. • There is a significant number of individuals (40%) who sleep for 4 hours or less. • The average sleep hours for individuals with extracurricular activities is 6.2 hours, while those without extracurricular activities average 6.8 hours. • The highest sleep hours recorded is 9 hours, while the lowest is 4 hours. • There is no clear correlation between sleep hours and performance index.
Analysis performed on complete dataset (10,000 rows)
Type here to search
ENG
23:04
26-01-2026

# 7. Conclusion

This project demonstrates a production-style AI analytics system by combining:

- LLM-based reasoning for planning and explanation
- Deterministic Pandas execution for reliability
- Modular agent design for maintainability

By enforcing deterministic execution and declarative experiment definitions, the system provides strong reproducibility guarantees and aligns with modern AI-powered business intelligence practices.