

Sia

AI agent for Data Analytics



No coding, No complexity, Simply chat to process your Data

Why we created Sia?



Fragmented data pipelines cost companies significant time & resources.

Before Sia

- ✖ Incoherent Data pipeline
- ✖ Dependency on Manual Expertise
- ✖ Limited Data efficacy & Accessibility
- ✖ Several tools involved & high costs

After Sia

- ✓ Single platform to integrate the data pipeline
- ✓ Automated workflows - Accelerated Analytics
- ✓ Gen AI powered chat to develop data apps
- ✓ Considerable reduction in cost of Analytics

Sia automates Data Analytics and makes it simple and easy to consume.

Introducing Sia



AI Data Scientist



Sia brings a full data team to everyone across the organization, empowering users to build with data, explore trends & co-relations in data, generate dashboards, build ML models, and more,

– ALL THROUGH A SIMPLE CHAT INTERFACE

AI Business Expert

Build a dashboard for me based on the important parameters in the data.

To build a dashboard using the parameters in the dataset, we can create several key visualizations that present important insights derived from the dataset. Here are some components that will provide a comprehensive overview:

Dashboard Components:

1. Count of High Emission Risk by Region:
 - This will show how many vehicles in each region fall into the high emission risk category.
2. Box Plot for CO2 Emissions by Engine Type:
 - Comparing CO2 emissions across vehicle models segmented by their engine types (gasoline, diesel, and hybrid).
3. Scatter Plot of Mileage vs. CO2 Emissions:
 - Explore the relationship between vehicle mileage and CO2 emissions.
4. Distribution of Vehicle Age for High Emission Risk Vehicles:
 - Visualizing the age distribution for vehicles with high emission risk.

Visualization way:

We'll create a dashboard using Matplotlib and Seaborn for visualizations. Here's how to proceed:



AI Data Analyst

Chat With SIA

I want to build a model to predict co2 emissions based on the features available. suggest me how to go about it.

To build a model predicting CO2 emissions based on the provided features, follow this step-by-step process

Step 1 Data Understanding

- Objective Understand what CO2 emissions you want to predict, the features available, and how they relate to CO2 emissions.
- Considerations Identify which features (columns) have a direct impact on CO2 emissions. Key potential predictors include 'Mileage_km', 'Exhaust_Temp_C', 'Fuel_Consumption_L_100km', 'Engine_Type', and others that capture performance or maintenance-related aspects.

Step 2 Data Preparation

- Handling Categorical Variables Features like 'Region', 'Make_Model', and 'Engine_Type' need to be encoded into numerical formats (e.g., one-hot encoding). Consider using the binary columns that are already created for 'Engine_Type' and 'Make_Model' (e.g., 'Engine_Type_Diesel', 'Make_Model_Audi_A4', etc.).
- Feature Selection Identify the most relevant features through exploratory data analysis (EDA) and correlation analysis. Any feature with high multicollinearity or low variability may be removed.
- Scaling and Normalization Some features (e.g., 'Mileage_km', 'Fuel_Consumption_L_100km') might need scaling to ensure they contribute equally to the model. Use standardization or normalization depending on the model choice.

Step 3 Exploratory Data Analysis (EDA)

- Visualize Distributions Plot histograms, scatter plots, or correlation matrices to understand the relationships between features and the target variable 'CO2_Emissions_g_km'.
- Identify Outliers Look for outliers in numerical features, analyze their impact on model performance.

Step 4 Model Selection

- Choose a Prediction Model Depending on your analysis results, you may want to explore different types of regression models (e.g., linear regression, decision tree regression, random forests, or gradient boosting).

SIA Type here...

AI Data Analyst

AL:

What specific characteristics define vehicles classified under high emission risk? Can we identify patterns among them based on model, age, or region?

SIA To identify specific characteristics that define vehicles classified under high emission risk and to explore patterns based on model, age, or region, we can analyze the data visually and statistically. We'll create visualizations to compare high emission risk vehicles across these dimensions and summarize the findings.

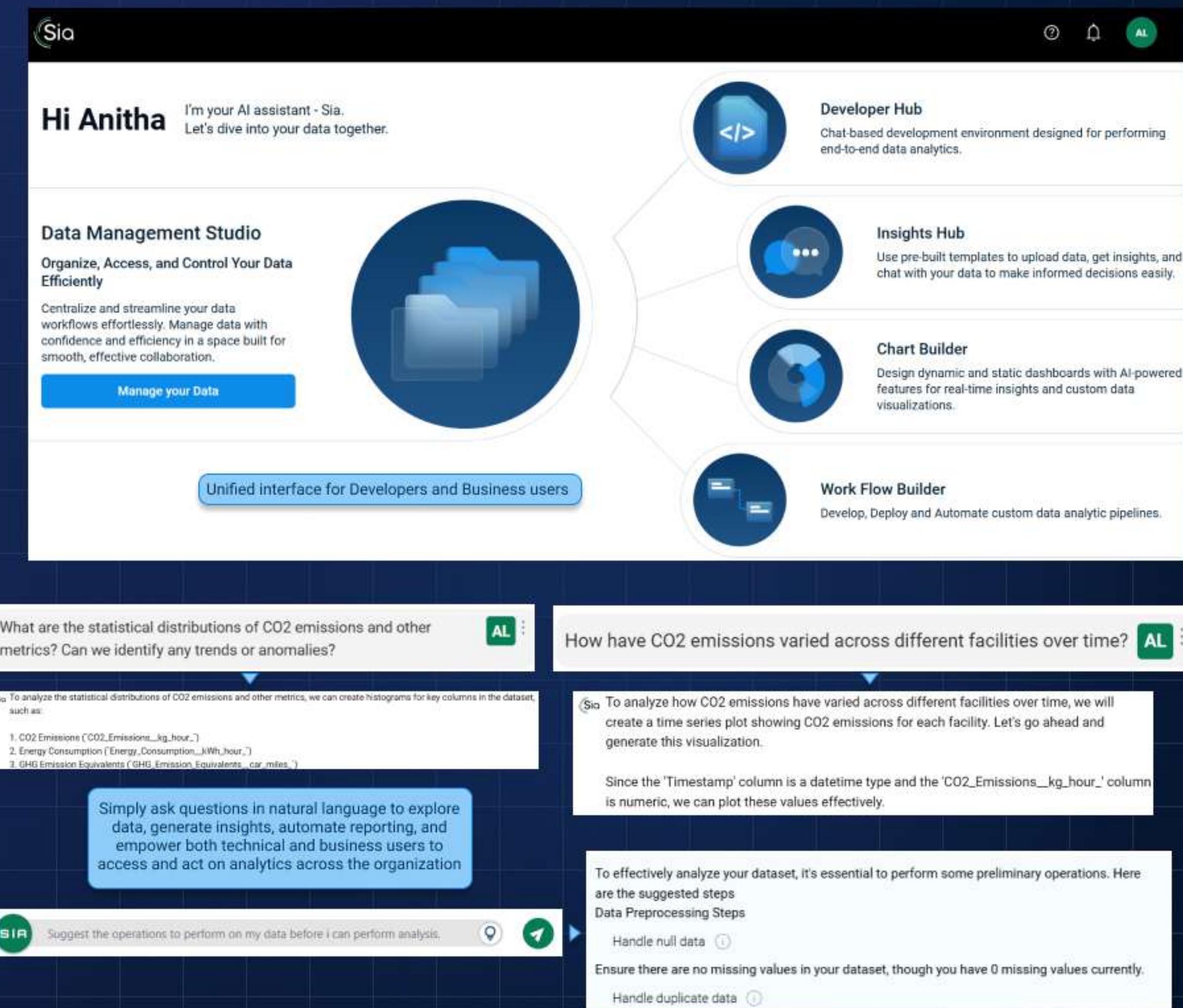
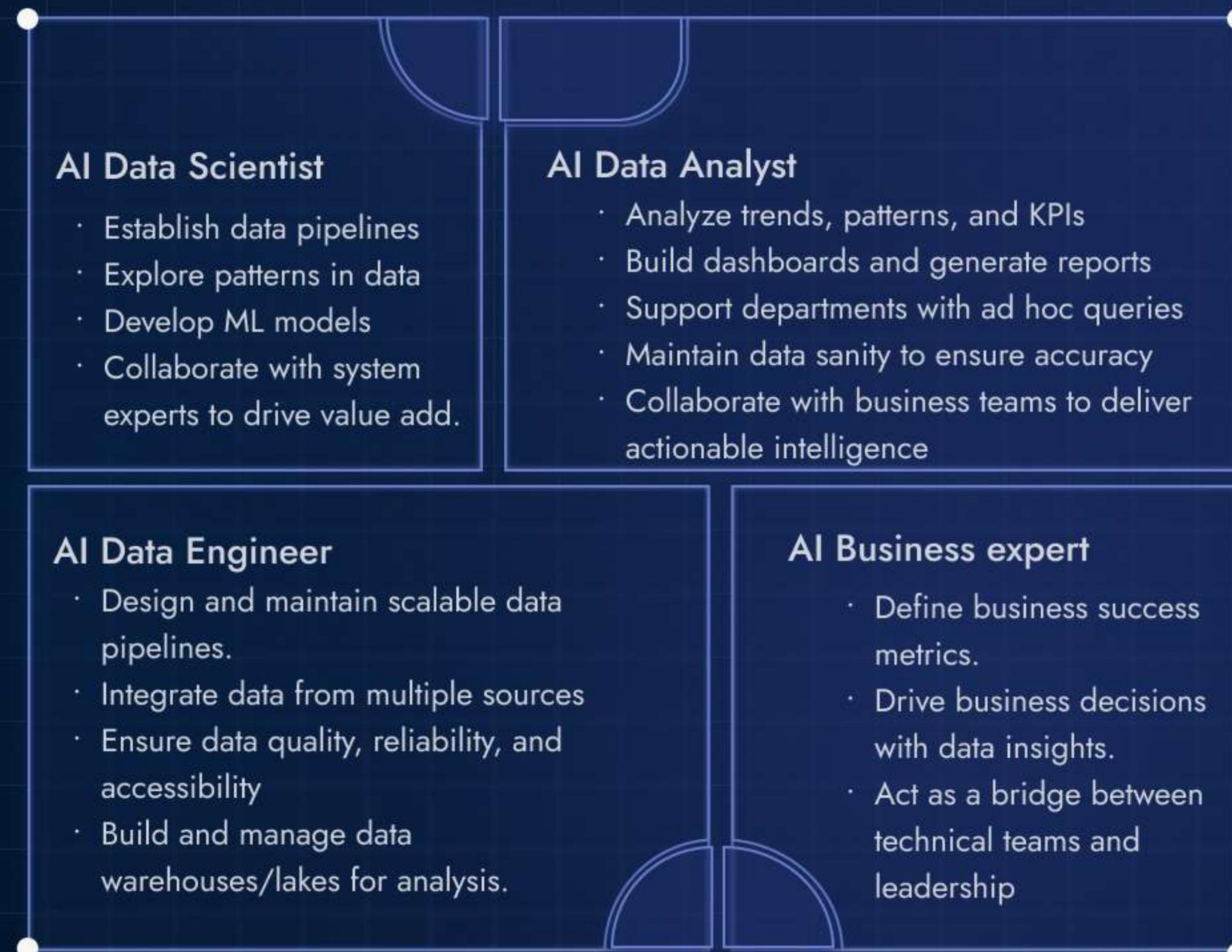
Analysis Plan:

1. Count of High Emission Risk by Region: This will show how many vehicles in each region fall into the high emission risk category.
2. Distribution of Vehicle Age for High Emission Risk Vehicles: We'll visualize the age distribution for vehicles with a high emission risk to see if there is a trend based on age.
3. CO2 Emissions by Make_Model: A box plot will help us identify differences in CO2 emissions between various vehicle models classified under high emission risk.

What does Sia do?



Sia is a AI data team that can ingest, clean, analyze, visualize, predict and prescribe from data with a conversation.



Hi Anitha I'm your AI assistant - Sia. Let's dive into your data together.

Data Management Studio
Organize, Access, and Control Your Data Efficiently
Centralize and streamline your data workflows effortlessly. Manage data with confidence and efficiency in a space built for smooth, effective collaboration.
[Manage your Data](#)

Unified interface for Developers and Business users

Developer Hub
Chat-based development environment designed for performing end-to-end data analytics.

Insights Hub
Use pre-built templates to upload data, get insights, and chat with your data to make informed decisions easily.

Chart Builder
Design dynamic and static dashboards with AI-powered features for real-time insights and custom data visualizations.

Work Flow Builder
Develop, Deploy and Automate custom data analytic pipelines.

What are the statistical distributions of CO2 emissions and other metrics? Can we identify any trends or anomalies? **AL**

To analyze the statistical distributions of CO2 emissions and other metrics, we can create histograms for key columns in the dataset, such as:
1. CO2 Emissions ('CO2_Emissions_kg_hour_')
2. Energy Consumption ('Energy_Consumption_kWh_hour_')
3. GHG Emission Equivalents ('GHG_Emission_Equivalents_car_miles_')

Simply ask questions in natural language to explore data, generate insights, automate reporting, and empower both technical and business users to access and act on analytics across the organization

Suggest the operations to perform on my data before I can perform analysis. **SIA**

How have CO2 emissions varied across different facilities over time? **AL**

To analyze how CO2 emissions have varied across different facilities over time, we will create a time series plot showing CO2 emissions for each facility. Let's go ahead and generate this visualization.

Since the 'Timestamp' column is a datetime type and the 'CO2_Emissions_kg_hour_' column is numeric, we can plot these values effectively.

To effectively analyze your dataset, it's essential to perform some preliminary operations. Here are the suggested steps
Data Preprocessing Steps

- Handle null data
- Handle duplicate data

Ensure there are no missing values in your dataset, though you have 0 missing values currently.

Data Management Studio



Perform all data modifications and prepare the data for Analytics, without coding.

A **System Engineer** will be able to do:



Pull live data seamlessly from any cloud environment for real-time processing.



Validate and pre-process data to ensure it meets analytics-readiness standards.



Generate automated data quality reports highlighting inconsistencies and gaps.



Ensure smooth integration of data pipelines across distributed systems for reliable analytics delivery.

Prepare Files

Concat Dataframes

Merge Dataframes

Pivot Dataframe

Split Dataframe

Groupby / Aggregate

Transpose Dataframe

Unlocking/Flattening multilevel indices

Managing/Converting sparse Dataframe

Setting rows as column names

Setting a column as index

Unpivot Dataframe

Generate Unique IDs/Index no/Row Ids

Transformation template

Reorder/Rearrange Columns

70%

Reduced data engineering/ETL efforts

>90%

Tasks can be done without coding

60-75%

faster data validation

100%



Azure Blob



Google Cloud



FTP



SFTP



HTTP



Databricks



Amazon Red...



Azure Synapse



Google BigQ...



PostgreSQL



MySQL

Developers Suite

Data Development Teams

Conversational AI: Interact with automotive data through natural language, enabling quick operations without writing code or using drag-and-drop tools.

AI-Powered Model Guidance: Get intelligent recommendations to build and deploy predictive models for vehicle performance, maintenance, and customer behavior.

Feature Transparency: Understand the impact and purpose of each feature in your model with clear explanations, aiding trust and refinement.

Accelerated Development: Quickly implement, test, and optimize changes, reducing development cycles and improving time-to-value in automotive applications.

40%

Reduction in development time through automated data processing and model deployment.

20%

Improvement in model performance through guided AI recommendations and explainability features.

The screenshot displays the Sia Developers Suite interface, which is a web-based tool for data analysis and model building. The interface is divided into several sections:

- Top Right:** Sia logo and a "Chat With SIA" section. The "Chat With SIA" section includes a "Hyperparameter" input field, dropdowns for "Problem Type", "Target Column", and "Columns", and buttons for "Reset" and "Submit". It also features a "Mandatory Tuning" toggle switch.
- Left Side:** A sidebar with icons for "Data Processing", "Model Building", "Feature Transparency", and "Development Tools".
- Middle Section:** A "Processing Type" dropdown set to "File", followed by a "Auto Trend Detection Or Removal" section with "Summary", "Plot", and "Comparison" tabs. A tooltip for this section explains that data points are selected at an interval of 10% for better visualization. Below this is a "Sensor Elimination Template" section with "Summary" and "Table" tabs.
- Right Side:** A "Chat With SIA" interface showing a message: "I want to perform feature importance analysis to find top contributing sensors". Below this, a "Selected Option : random forest feature importance" message is shown. A tooltip for the "Sensor Elimination Template" section explains that it uses correlation analysis and feature importance methodologies to identify if a target column can be modeled using remaining columns.

Business Suite



Business Development Teams

Empowers automotive teams to make data-driven decisions independently, without relying on technical experts.

Eliminates the need for complex data engineering or analytics support through an intuitive interface.

Provides easy access to production, performance, and customer data for quick analysis and insights.

Enhances efficiency across design, manufacturing, sales, and service operations through faster decision-making.

70%

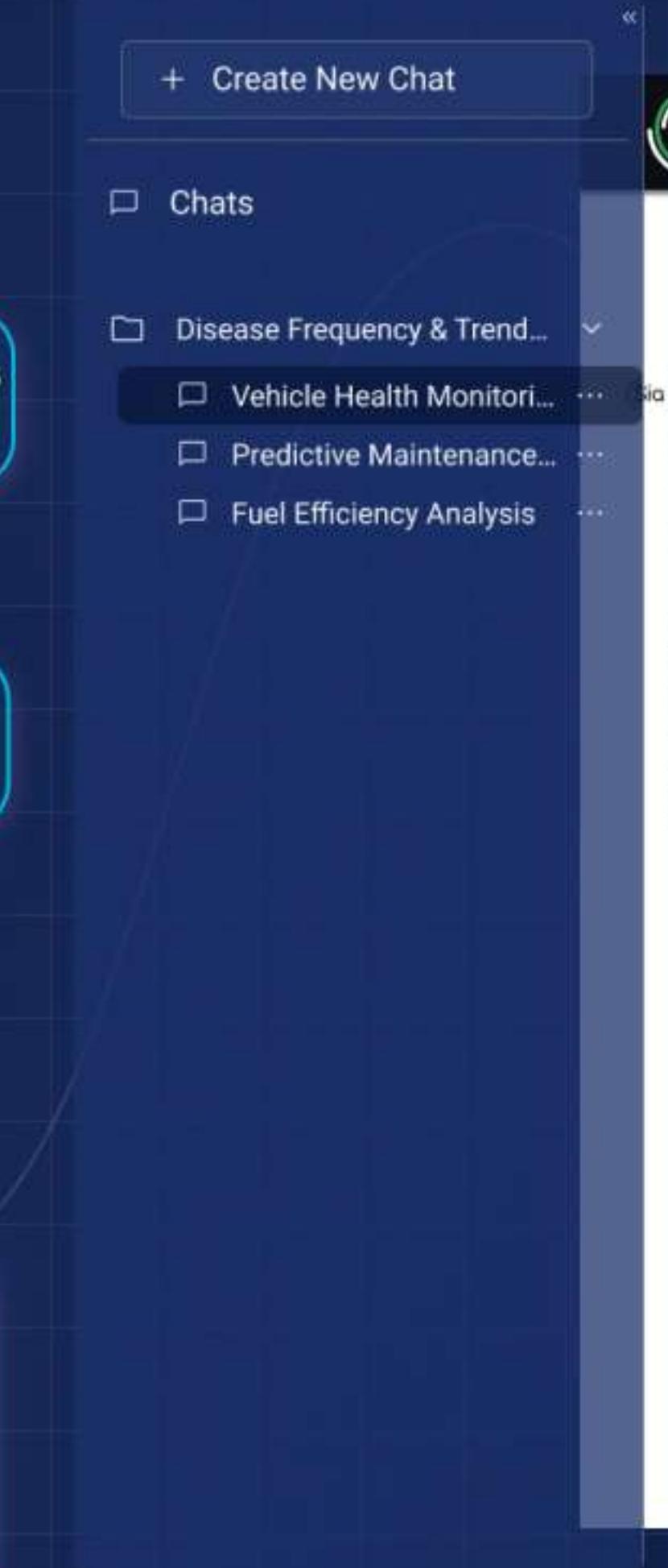
30-40%

20-25%

Faster business decisions through instant visual insights and real-time dashboards.

Increase in operational efficiency by reducing manual reporting and automating data analysis.

Improvement in revenue opportunities by identifying trends and actionable insights faster.



Which vehicle models have the highest risk of emissions non-compliance?

To determine which vehicle models carry the highest risk of emissions non-compliance, we conducted an analysis focusing on the total high emission risks for various vehicle models. The data was organized by the 'Make_Model' and 'High_Emission_Risk' columns, and the results were visualized to highlight models that stand out in terms of their emission risks.

The findings indicate that the top vehicle models with the highest total high emission risks are as follows:

1. Audi A4: 68 high emission risks
2. Ford Fiesta: 67 high emission risks
3. Toyota Prius: 67 high emission risks
4. BMW X5: 62 high emission risks
5. Mercedes C-Class: 51 high emission risks
6. Volkswagen Golf: 46 high emission risks



Sia's Impact



40%

Reduction in data engineering efforts

70%

Automation in data pipeline development and workflows

50%

Reduction in cost of Processing and Handling Data

30%

Increase in organization wide data literacy and productivity.

60%

Improvement in time to insights

Avg Enterprise budget for AI
~\$3 to 5 million pa

Avg Salaries - \$100k to \$150k pa

Sia cost of subscriptions - \$150 to \$350
Minimize resource & Infra costs of data
by at least 30%



Without Sia



With Sia



Company Overview



Our Mission

To simplify development and consumption of Technology, so that people can focus on 'What to Build' instead of 'How to Build'.



Meet the Founders



Karthik, COO



Divya Krishna, CEO

Sumalata, CTO

- Combined experience of over 3 decades in data science
- 10 registered patents, End-to-end project implementation
- Expertise in Cloud, Front-end, and Back-end Technologies

Our Success Stories



- Enhanced innovation and R&D capabilities in a large enterprise, reduced GTM time by over 40%, creating new up-selling opportunities.
- Standardized org-wide data processes and increased efficiency by more than 50% through minimizing manual dependencies and bottlenecks.
- Replaced a legacy analytics tool and reduced costs by more than 70% through auto-assigning of cloud compute resources and AI features.

Our Customers



RICE UNIVERSITY

S-VYASA
Deemed to be University u/s 3 of the UGC Act, 1956
SCHOOL of ADVANCED STUDIES



Backed by Numenica



Contact Us



✉️ info@datastride.ai

📞 +91 87625 25857
+91 84310 42564

🔗 <https://datastride.ai/>
<https://sianalytics.in>

📍 Global Village Tech Park, Sattva Global City,
RR Nagar, Mysore Road, Bengaluru - 59

Thank You