

# 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

AI Business Expert

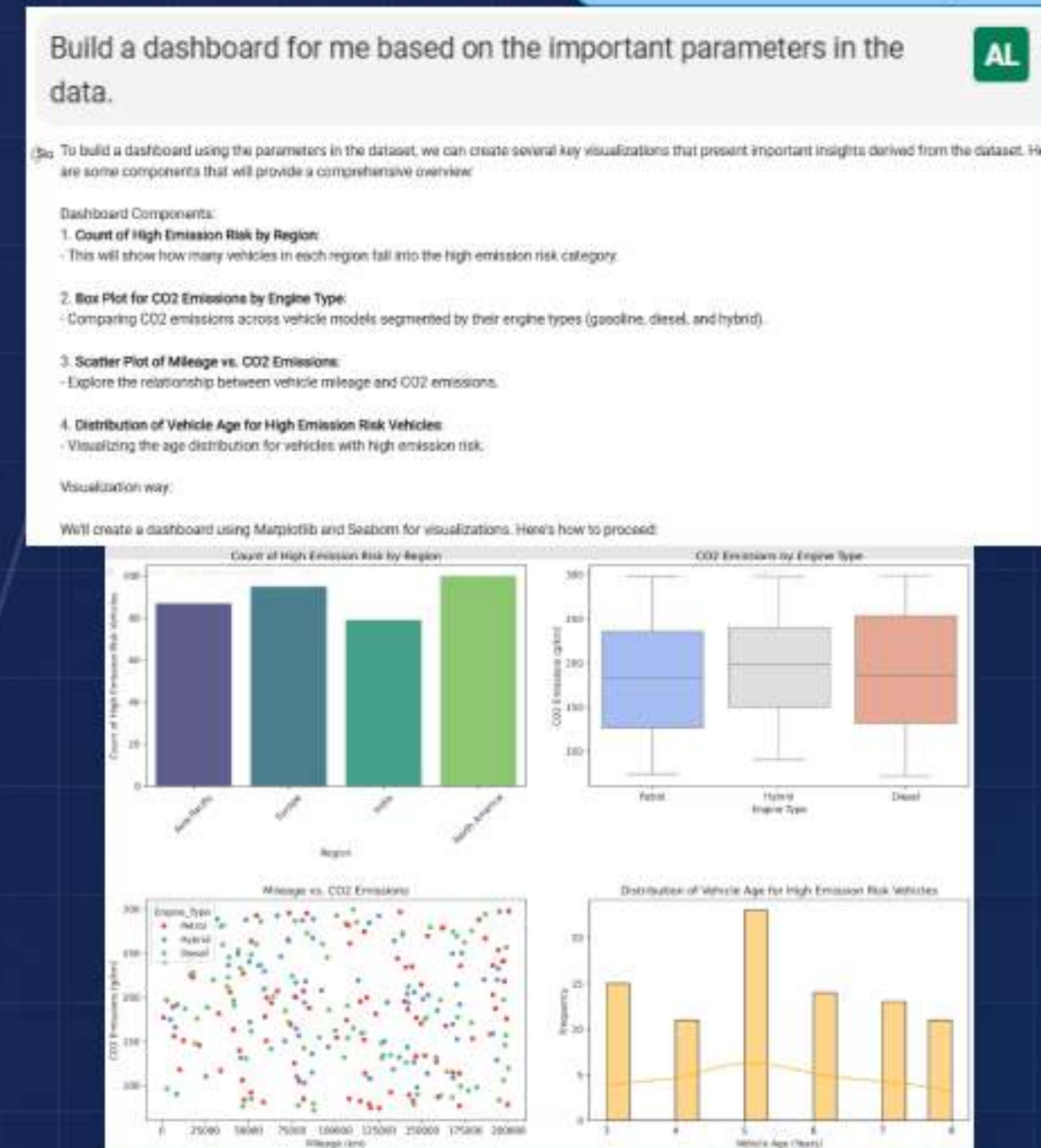


Product Dev teams  
System teams  
R&D teams  
C Suite officials  
Data teams



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



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).



Type here...



AI Data Analyst

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

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.

## AI Data Scientist

- Establish data pipelines
- Explore patterns in data
- Develop ML models
- Collaborate with system experts to drive value add.

## AI Data Analyst

- Analyze trends, patterns, and KPIs
- Build dashboards and generate reports
- Support departments with ad hoc queries
- Maintain data sanity to ensure accuracy
- Collaborate with business teams to deliver actionable intelligence

## AI Data Engineer

- Design and maintain scalable data pipelines.
- Integrate data from multiple sources
- Ensure data quality, reliability, and accessibility
- Build and manage data warehouses/lakes for analysis.

## AI Business expert

- Define business success metrics.
- Drive business decisions with data insights.
- Act as a bridge between technical teams and leadership



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

SIA

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

AL

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

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

Handle duplicate data



# 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. On the left, a sidebar contains icons for data processing, model guidance, feature transparency, and accelerated development. The main area is divided into two sections. The top section, titled 'Auto Trend Detection Or Removal', includes a 'Summary' tab and a 'Plot' tab. The 'Summary' tab shows a list of steps: 1. The function parameters without user inputs are assigned their default values. Refer the documentation to know the default values. 2. For better visualization, data points are selected at an interval of 100 by reducing the number of data points. Only every 100th data point trends more apparent. 3. This function detects the underlying long-term movement or direction. It captures the overall increasing or decreasing pattern in the data, which is useful for various purposes, such as identifying patterns, forecasting future values, and understanding the underlying data. 4. New Columns added to the existing data. New column names will have the target column name suffixed with Trend & Extended Column names remain the same without any change. Below this, there is a 'Sensor Elimination Template' section with a 'Summary' tab and a 'Table' tab. The 'Summary' tab shows a list of steps: 1. The function parameters without user inputs are assigned their default values. Kindly refer the documentation to know the default values. 2. Dataframe is clean. 3. Correlation analysis and feature importance methodologies are used to identify if the given target column can be modelled using remaining columns. On the right, a 'Chat With SIA' window is open. It has a title bar with 'Sia' and a close button. The chat area contains a message: 'Please fill the hyperparameter details to keep analyzing'. Below this, there is a 'Hyperparameter' section with three dropdown menus: 'Problem Type' (None), 'Target Column' (None), and 'Columns' (None). There are 'Reset' and 'Submit' buttons. At the bottom, there is a text input field with the placeholder 'Type here...' and a 'Send' button. The chat area also displays a response from SIA: 'I want to perform feature importance analysis to find top contributing sensors'. Below this, there is a message: 'To perform feature importance analysis and identify the top contributing sensors in your dataset, here are the recommended next steps: Feature Importance Analysis. Auto-run base classification models. This function runs multiple classification models on your data using the "machine\_status" column as the target, providing insights on feature importance.' Below this, there is a message: 'Selected Option : random forest feature importance'. Below this, there is a message: 'The Random Forest Feature Importance function assesses the contribution of each feature in predicting the target variable by analyzing the average decrease in impurity when splitting. This function helps you identify key predictors, guiding feature selection and enhancing model interpretability. Do not forget to check the output window for a thorough summary of the feature importance results derived from your data. Please ensure the hyperparameter details are filled in to keep analyzing.'



# 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%

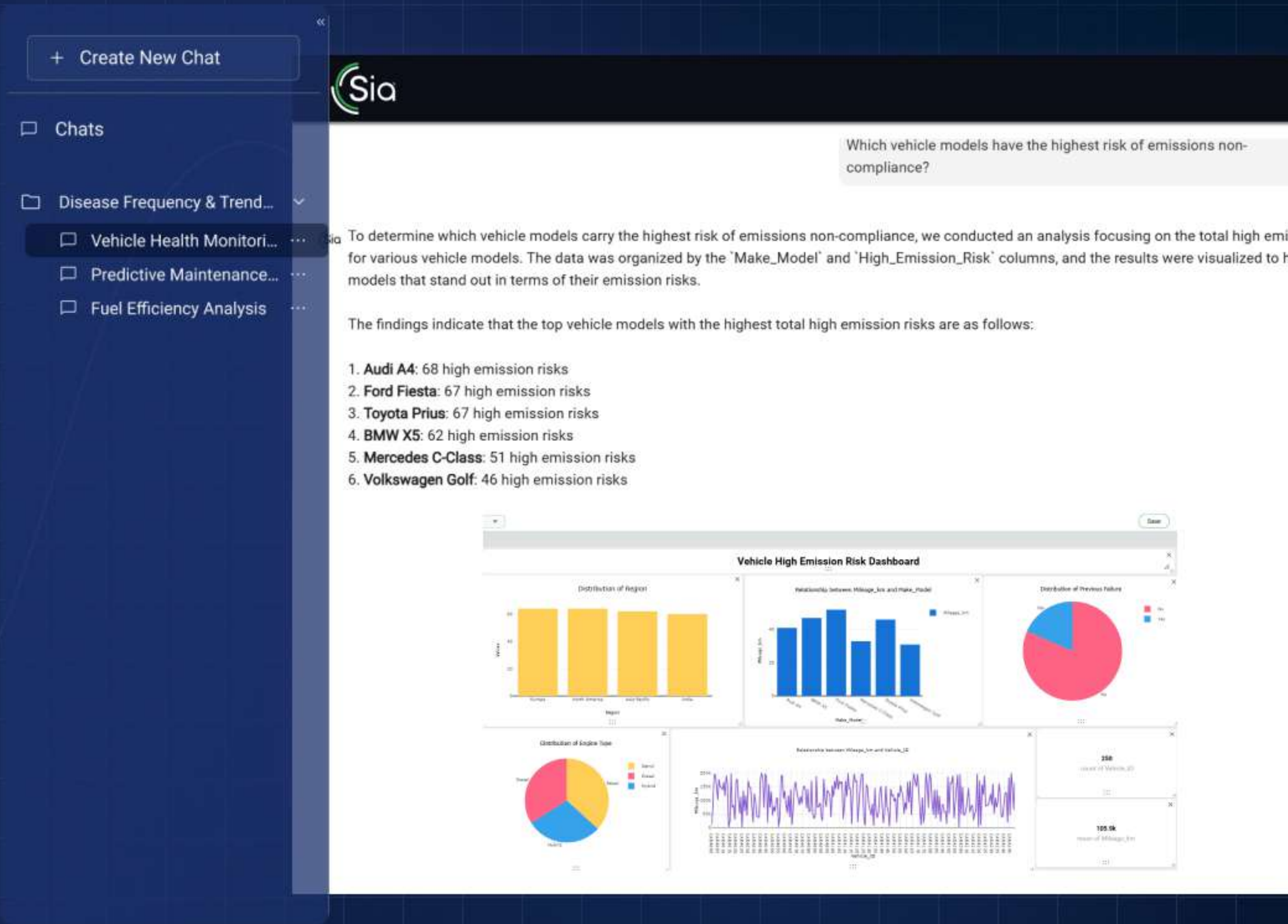
Faster business decisions through instant visual insights and real-time dashboards.

30-40%

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

20-25%

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





# 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



Backed by Numentica





# 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