

**Table of Contents**

A screenshot of a computer

Description automatically generated

|  |  |
| --- | --- |
| **Section** | **Title** |
| Introduction | Cittabase Insightify - features |
| 1 | Data Preview and EDA |
| 2 | Model Training and Prediction |
| 3 | AI Questionnaire |

******

***"Cittabase Insightify"*** is a comprehensive platform designed to accelerate the development of prototypes and Proof of Concepts (POCs). If you have data and are looking to conduct swift yet impactful exploratory data analysis (EDA), assess the data’s potential for predictive modelling, or get your queries answered quickly, this tool is your go-to solution. Leveraging cutting-edge technologies and the advanced capabilities of Snowflake Cortex, Insightify empowers users to extract actionable insights with ease.

1. **Data Preview and EDA Interface**: Select the database **INSIGHTIFY\_DB** and Select schema **PUBLIC** under which you can select any table of interest to generate quick visual insights about your data. This intuitive feature enables efficient exploration and understanding of your dataset.
2. **Model Training and Prediction Interface**: Build, train, and run **Classification** and **Regression** models directly on your data. The utility of this is to quickly determine the scope for modelling and the predictive potential of your dataset. *(Ensure that the table is created in the designated database: (***INSIGHTIFY\_DB.PUBLIC***). At this point, the tool supports classification use cases for modelling.*
3. **AI Questionnaire**: For data-driven answers to specific queries, simply upload a YAML file to the staging area (**INSIGHTIFY\_DB.PUBLIC**) and access the ready to use Questions Interface for intelligent responses along with the generated SQL.

*Please note that the Model Builder and the AI Questionnaire options are two independent applications. If the tables are in* ***INSIGHTIFY\_DB. PUBLIC*** *and the corresponding YAML file is uploaded to* ***INSIGHTIFY\_STG****, the AI Questionnaire can be utilized by selecting the relevant YAML file from the dropdown. Therefore, tables can be used solely for AI Questionnaire purposes without the need to build a model, and vice versa. AI Questionnaire leverages Cortex and LLM capabilities.* ***To use AI Questionnaire, ensure that Cortex LLM Functions are enabled for your region.***

If Cortex LLM functions are not enabled, you could consider enabling cross region inference by referring the below link,  
<https://docs.snowflake.com/en/user-guide/snowflake-cortex/cross-region-inference>

This enables inference requests to be processed from a different region where LLMs are available.

**1) Data Preview and EDA Interface:**

This screen provides a preview of the available tables and performs EDA on the selected table.

Follow the instructions given at the beginning to load and give necessary privileges to your table before choosing it from dropdown.

A white rectangular object with black text

Description automatically generated with medium confidence

Use the dropdown menu to select and view the desired table under the dedicated database and schema **(INSIGHTIFY\_DB. PUBLIC)**.

A screenshot of a computer

Description automatically generated

Depending on the type of variables chosen, the tool will automatically plot the respective EDA.

The EDA for the selected table can be displayed under two check boxes - ***Univariate Graph*** and ***Bivariate Graph.***

**Univariate Graph:**

It has a dropdown which consists of the columns from selected table. Upon selecting the column, a visual appears with a plot of the column.

***Few examples of Univariate Chart Representations:***

***Bar Chart***

A graph of blue and white bars

Description automatically generated

***Pie Chart***  
A pie chart with different colors

Description automatically generated

A pie chart appears if the selected feature is categorical, whereas a bar chart appears for a numerical feature.

**Bivariate Graph:**

Unlike the Univariate Graph, it has two dropdowns containing columns from the chosen table. Upon selecting the columns, a visual appears representing the bi-variate graph.

***Few examples of Bi-Variate Chart Representations:***

***Stacked Bar Chart***

A stacked bar chart is plotted if both the selected features are categorical or one of the features is categorical and the other is continuous.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

***Scatter Plot***

A scatter plot is plotted if the selected features are of continuous datatype.

A screenshot of a computer

Description automatically generated

**2) Model Builder**

In this option, the potential or feasibility for modelling your data is explored. Let’s say you have a dataset which you want to initially validate for fitment to model, this option could be leveraged. The model is built, and results are displayed in the UI as well as stored in the respective prediction table. Classification and Regression ML models using **PyCaret Auto ML** python module and leverages Random Forest for building models and predictions effectively.

**Step 1:** To get started, first select the method of ML model you need to use.

*Note: Default Database is* ***INSIGHTIFY\_DB*** *and Schema is* ***PUBLIC.***

A screenshot of a computer

Description automatically generated

**Step 2:** Select the table you want to use as input for building the classification model in this example.

A screenshot of a computer

Description automatically generated

**Step 3:** Choose the dependent variable (target) and independent variables (features) from your selected dataset, then click on the submit button to proceed.

Once you submit the selections, a finalized and processed version of your dataset will be created and made ready for model training purposes.

A screenshot of a computer

Description automatically generated

**Step 4:** Initiate the training process by clicking the 'Train & Evaluate' button, which will allow the model to learn patterns and relationships from the provided dataset. The Training table is created.

The Training process starts and executes. After the training process is completed successfully, the classification model will be trained, and its performance evaluation metrics will be displayed for your review.

A screenshot of a computer

Description automatically generated

A structured template or framework for the prediction dataset will also be automatically created with the name as <SELECTED TABLE\_NAME>\_PREDICTION, ensuring proper formatting for new data inputs.

**Step 5:** Before running predictions, make sure to load the newly created prediction table with prediction-ready data so that accurate results can be generated by the trained model.



A screenshot of a computer

Description automatically generated

**Step 6:** After selecting the prediction table from the drop down above, a data preview will be shown.

**Step 7:** On clicking the “Predict” button, the predicted output for the new data will be displayed.

A white screen with many text

Description automatically generated with medium confidence

The “PREDICTED\_SUSPICIOUS\_FLAG” column indicated the Predicted Dependent Category, and the column “predicted\_score” indicates the probability at which the Dependent Category is predicted.

**3) AI Questionnaire**

This interface enables users to interact by asking questions about their tables and gets answers as well as the generated SQL query. This leverages Snowflake Cortex Analyst API calls integrated with the Semantic Structure of the table information.

*AI Questionnaire leverages Cortex and LLM capabilities.* ***To use AI Questionnaire, ensure that Cortex LLM Functions are enabled for your region.***

If LLM functions are not enabled, you will get the following screen in AI Questionnaire option,

A screenshot of a chat

Description automatically generated

You could consider enabling cross region inference by referring the below link,  
<https://docs.snowflake.com/en/user-guide/snowflake-cortex/cross-region-inference>

This enables inference requests to be processed from a different region where LLMs are available.

As mentioned in the permissions SDK, “**IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE”** is required.

Snowflake Cortex Analyst empowers organizations to simplify and enhance data analysis by leveraging **semantic models**. These models serve as structured representations of data, capturing its meaning and relationships in a way that aligns seamlessly with business concepts and user queries. Within Snowflake Cortex Analyst, semantic models are defined using YAML files, which specify dimensions, metrics, hierarchies, and relationships across datasets. This structured approach enables users to perform natural language querying and analytical tasks effortlessly, with Cortex Analyst interpreting user inputs and mapping them to the underlying data for actionable insights. The semantic model answers the questions asked and gives a SQL query for the user given input.

Cortex Analyst requires a semantic model (YAML) of the data to be uploaded to stage, so it can be leveraged to ask questions. The user can adopt their preferred method of choice for generating the YAML file. (If you would like to know more about the content of the YAML file, please refer to the ***Guidelines for Semantic Model YAML Files.doc***).

Additionally, we have also provided a reference from quick starts to create a Semantic Model generator below.

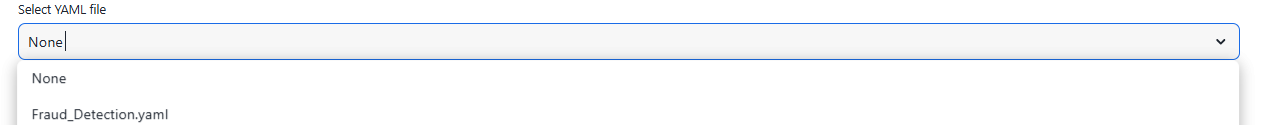
[Refer to create a Yaml file](https://quickstarts.snowflake.com/guide/semantic_file_generation_for_cortex_analyst/index.html?index=..%2F..index#2)

1. Create a YAML file with the required table(s) configuration.
2. Once the Yaml file is created, upload it to the stage **INSIGHTIFY\_STG,** which is created under schema **INSIGHTIFY\_DB. PUBLIC.**

A screenshot of a chat

Description automatically generated

1. Navigate to the AI Questionnaire option and verify that the uploaded YAML file is listed.



1. You can then enter your question in the user input dialogue box, and hit the “Enter” key, the system will generate the desired output along with the corresponding SQL query for better clarity and understanding of the result.

A screenshot of a computer

Description automatically generated