Project Report

Project Title: Exploring open source BI tools (Superset, Metabase)

Project ID: CO28

Project Area: Data Visualization

Name: Prasiddha Bhat

TABLE OF CONTENTS:

1. Problem statement

This is an exploratory project that aims to analyze a crop production dataset from the Department of Agriculture & Farmers Welfare (DA&FW), to provide understandable insights to those in the agricultural field. By exploring open source BI tools like Apache Superset, and Metabase, the project will create intuitive dashboards that visualize and illustrate the key performance indicators of crop production, identify key metrics and trends, and help us make informed decisions.

2. Dataset Overview

2.1 Description of the Crop Production dataset

The crop production dataset provides detailed final estimates of crop production in India over different agricultural seasons from 2013-14 to 2023-24. The dataset specifically focuses on production of many different crops over three seasons: Kharif, Rabi and Summer. It includes details regarding foodgrains, oilseeds and other commercial crops totalling to 34 different crops. Most of the production entries are in lakh tonnes, allowing for the analysis of trends and statistics over time.

2.2 Source of the dataset

The data is sourced from Department of Agriculture & Farmers Welfare (DA&FW). It is a valuable resource for the agricultural folk looking to understand performance, improve agricultural practices and make informed decisions.

2.3 Number of columns and their data types

- Thirteen columns- Crop, Season, 11 years ranging 2013-14 to 2023-24.
- Crops: food grains, oilseeds and other commercial crops
- Seasons- Kharif, Rabi and Summer.
- Years- numeric datatype, representing production in Lakh Tonnes, with some entries marked with ‘@’, ‘$’, or ‘#’ indicating inclusions.
- Crop and Season: categorical

2.4 Data preprocessing

Data preprocessing is used to convert raw data into a clear format as clean data is easier to visualize. This is done by filling missing values, removing noisy data, standardizing or categorizing values and so on.

2.4.1 Replace special characters

It is given that '@' is included in Rabi and '$' is included in Kharif. So, we can replace all '@'s and '$'s by 0. It is also stated as follows:

- #: Cotton Production in Lakh Bales, 1 Bale=170 Kg
- ##: Jute & Mesta Production in Lakh Bales, 1 Bale=180 Kg

Since all other production values are in Lakh tonnes, it is preferable to standardize these values.

1. For cotton production, Lakh Tonne= Lakh Bales*0.17
2. For Jute & Mesta Production, Lakh Tonne= Lakh Bales*0.18

2.4.2 Reshape data

We notice that the years are spread across eleven columns and if we were to make time based analyses, it would be advisable to transform these columns to a longer format under a single combined column, 'Year'.

Use Get & Transform Data and Unpivot 'Year' Columns to transform the dataset into four columns: 'Crop', 'Season' and 'Year' are of categorical values (string datatype), and 'Production' is of numeric values (float datatype).

2.4.3 Standardizing

Convert 'Year' to a numerical category and change it to the standard yyyy-mm-dd date format.

Since the 'total' entries under Season are redundant, we can filter and delete these entries.

Ministry of Agriculture & Farmers Welfare
Department of Agriculture & Farmers Welfare (DA&FW)
Final Estimate of Production of Food Grains

Dated : 25.09.2024
Source: DA&FW
Production in Lakh Tonnes

Crop	Season	Production									
		2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23	2023-24
Rice	Kharif	914.97	913.92	914.13	963.03	971.35	1020.40	1022.77	1052.08	1110.01	1105.12	1132.59
	Rabi	151.49	140.91	129.95	133.96	156.22	144.38	165.94	191.60	184.71	150.04	146.01
	Summer	@	@	@	@	@	@	@	@	@	102.40	99.65
	Total	1066.46	1054.82	1044.08	1096.98	1127.58	1164.78	1188.70	1243.68	1294.71	1357.55	1378.25
Wheat	Rabi	958.50	865.27	922.88	985.10	998.70	1035.96	1078.61	1095.86	1077.42	1105.54	1132.92
Maize	Kharif	171.45	170.14	160.53	189.19	201.18	194.14	194.29	215.55	226.81	236.74	222.45
	Rabi	71.14	71.59	65.14	69.81	86.34	83.02	93.37	100.92	110.49	116.90	120.28
	Summer	@	@	@	@	@	@	@	@	@	27.21	33.92
	Total	242.60	241.73	225.67	259.00	287.53	277.15	287.66	316.47	337.30	380.85	376.65
Barley	Rabi	18.31	16.13	14.38	17.47	17.81	16.33	17.22	16.56	13.71	19.13	16.99
Jowar	Kharif	23.93	23.00	18.16	19.64	22.74	17.35	16.97	19.86	15.98	14.80	15.09
	Rabi	31.49	31.45	24.22	26.04	25.30	17.40	30.75	28.26	25.52	23.22	31.89
	Summer	@	@	@	@	@	@	@	@	@	0.13	0.40
	Total	55.42	54.45	42.38	45.68	48.03	34.75	47.72	48.12	41.51	38.14	47.37
Bajra	Kharif	77.09	91.84	80.67	97.30	92.09	86.64	103.63	108.63	97.81	103.49	96.63
	Summer	$	$	$	$	$	$	$	$	$	10.82	10.53
	Total	77.09	91.84	80.67	97.30	92.09	86.64	103.63	108.63	97.81	114.31	107.16
Ragi	Kharif	19.83	20.61	18.22	13.85	19.85	12.39	17.55	19.98	17.01	16.91	16.70
Small Millets	Kharif	4.30	3.86	3.91	4.42	4.39	3.33	3.71	3.47	3.67	3.84	4.49

Fig. 1: Dataset image

	A	B	C	D	E
1	Crop	Season	Year	Production	
2	Rice	Kharif	2013-01-01	914.97	
3	Rice	Kharif	2014-01-01	913.92	
4	Rice	Kharif	2015-01-01	914.13	
5	Rice	Kharif	2016-01-01	963.03	
6	Rice	Kharif	2017-01-01	971.35	
7	Rice	Kharif	2018-01-01	1020.4	
8	Rice	Kharif	2019-01-01	1022.77	
9	Rice	Kharif	2020-01-01	1052.08	
10	Rice	Kharif	2021-01-01	1110.01	
11	Rice	Kharif	2022-01-01	1105.12	
12	Rice	Kharif	2023-01-01	1132.59	
13	Rice	Rabi	2013-01-01	151.49	
14	Rice	Rabi	2014-01-01	140.91	
15	Rice	Rabi	2015-01-01	129.95	
16	Rice	Rabi	2016-01-01	133.96	
17	Rice	Rabi	2017-01-01	156.22	
18	Rice	Rabi	2018-01-01	144.38	
19	Rice	Rabi	2019-01-01	165.94	
20	Rice	Rabi	2020-01-01	191.6	
21	Rice	Rabi	2021-01-01	184.71	

CROP_dataset

Fig. 2: Dataset after preprocessing

3. Installation Steps

3.1 Prerequisites needed

The prerequisites required are Docker Desktop, MySQL Workbench, MySQL Server, MySQL Shell. Using Docker, we can install Superset and Metabase. Using MySQL, we can connect an sql database to these visualization tools.

3.2 Installation using Docker

3.2.1 Superset Installation using Docker

- In docker, search the Superset image,pull the image and then click on ‘view on hub’. Copy the latest tag id:b8f7669-dev.
- We can now find the Superset container under containers.
- Open cmd.
- docker run -d -p 8080:8088 -e "SUPERSET_SECRET_KEY=mysuperset" --name superset apache/superset:b8f7669-dev
- docker exec -it superset superset fab create-admin --username admin --firstname Superset --lastname Admin --email admin@superset.com --password admin
- docker exec -it superset superset db upgrade
- docker exec -it superset superset load_examples
- docker exec -it superset superset init
- Then go to the link in docker container- Username: admin, Password: admin

3.2.2 Metabase Installation using Docker

- After installing and running docker, copy the pull command into the command prompt, from the open source quick start guide documentation of metabase, to pull the latest image from docker hub.
- We can now start our metabase container with the run command.
- We can now find the Metabase image and container on docker desktop.
- Then go to the link in the docker container. This will take us to the localhost.
- Then go through the wizard to set up metabase.

4. Dashboard Creation and design

When creating user-friendly dashboards, it's important to prioritize a seamless data exploration experience. This includes developing comprehensive, semantic models that ensure consistency throughout the data set while minimizing redundancy. By organizing the data in a meaningful and relatable manner, users can easily navigate and understand the information being displayed. Two fully functional dashboards were created using two powerful BI tools- Apache Superset and Metabase. The dashboards showcase a range of visualizations, filters, and interactions.

Superset is fast, interactive, and provides a multitude of options for users of all skill sets to easily explore and visualize their data, and these options range from a mere line chart to highly sophisticated geospatial charts. Metabase is a fast analytics tool that provides a user friendly interface to let a company explore their data on their own. Metabase connects to over 20 data sources and offers flexibility from free to enterprise-level access. We will also assess the key differences, advanced capabilities, flexibility in creating dashboards.

4.1 Dashboard Creation and design in Superset

Explore data and find insights from interactive dashboards. Superset includes 40+ pre-installed visualizations, drag-and-drop, faster data loading and SQL query support. Users can make intuitive dashboards and conduct advanced data analysis through cross-filters.

4.1.1 Data import and setup

We notice that the ‘Year’ column is not in a Date or temporal format type, we should thus transform it to a standard Date format using MySQL workbench. This way, we can make use of Superset’s multitude of time series charts and make analysis of the trends and statistics of crop production over the years.

```sql
-- Create the schema
CREATE SCHEMA cropsdatabase;

-- Use the newly created schema
USE cropsdatabase;

-- Create the table
CREATE TABLE crops (
    Crop VARCHAR(40),
    Season VARCHAR(20),
    Years VARCHAR(20),
    Production DECIMAL(10, 2)
);
```

Fig. 3: cropsdatabase schema and crops table creation

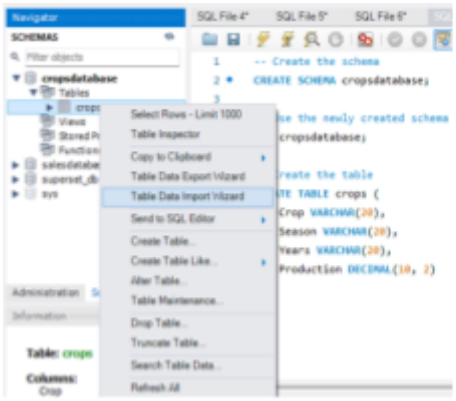

Fig. 4: Browse and import preprocessed dataset

```sql
ALTER TABLE crops
MODIFY COLUMN  Years DATE;
```

Information

Table: crops

Columns:

Crop	varchar(20)
Season	varchar(20)
Years	date
Production	decimal(10,2)

Fig. 5: Modify Years column to DATE type

We now have the ‘Years’ column in the standard date format.

4.1.2 Using MySQL as the Database engine

Go to the Datasets section and click on the Add dataset option.

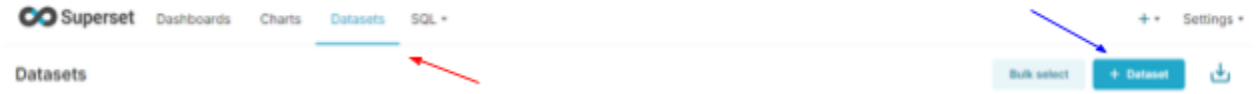

Fig. 6: Import Dataset to Superset

The database selected is MySQL, schema is cropsdatabase and table is crops. It is verified that the 'Years' column is now a DATE datatype. Click on 'Create dataset and create chart' option. This successfully imports our clean data into Superset.

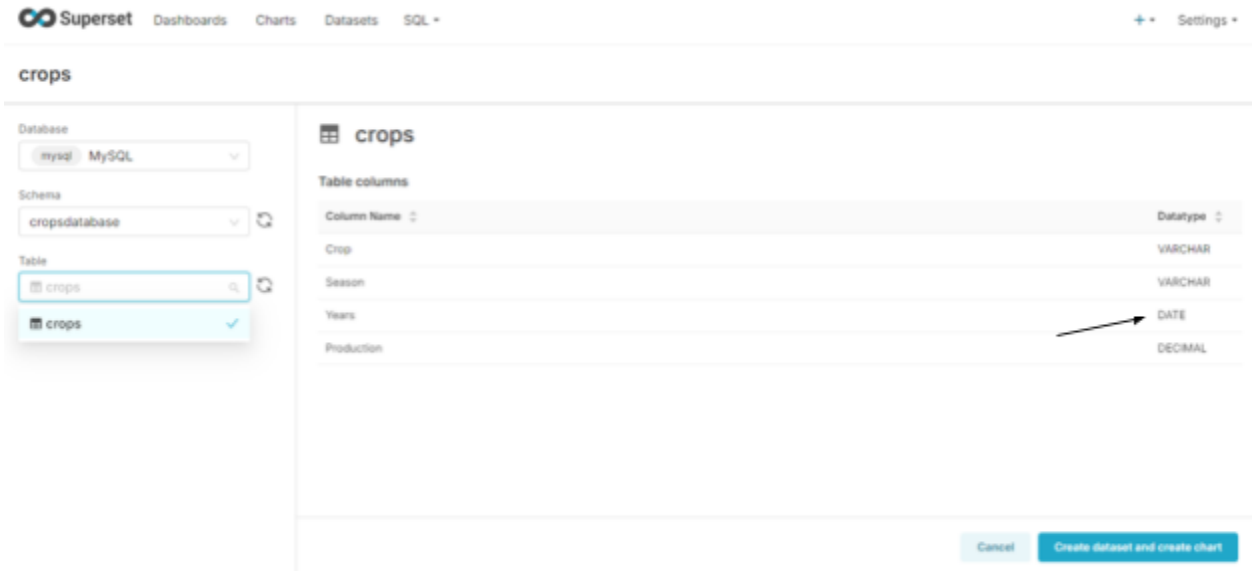

Fig. 7: Superset Dataset Import setup

4.1.3 Dashboard Design

Superset is popular for its beautiful charts and interactive dashboards that can be created quickly. It can be used to make intuitive dashboards for data analysis and exploration workflow and provides drag-and-drop options, simple charts, sophisticated charts, faster data loading, and cross-filters, helping users gain targeted insights.

4.1.3.1 List key metrics and Chart Types used in dashboard

- Table1: Crop and production table, Table 2: Top 10 crops.

The ‘Crop and production table’ summarizes the total production of each crop, and the ‘Top 10 crops’ table highlights the best 10 crops in terms of highest total production. This could help us in identifying the crops that help maximize yield.

- Production changes in crops-Line chart

The ‘Production changes in crops’ line chart effectively highlights variations and growth patterns, making it easier to identify variation in production across different crops.

- Pie chart1: Sum of Production by Crop, Pie chart2: Seasonal contribution to production

The ‘Sum of Production by Crop’ chart represents the proportion of total agricultural production contributed by each crop. The ‘Seasonal contribution to production’ chart shows the distribution of crop production across the three seasons.

- Production change over the years based on season- Area chart

The ‘Production change over the years based on season’ area chart represents the variations in crop production over time, segmented by season. This allows for easy comparison of seasonal trends and highlighting shifts in agricultural output across different years.

- Production according to Seasons per year- Pivot table.

This pivot table summarizes crop production data by organizing it into seasons for each year, helping users to analyze trends, gain seasonal insights, and identify patterns in crop production across different years.

- Two big number charts: Total crops, and total production in lakh tonnes.

These charts highlight the total number of crops, and total production generated over the years studied in the data respectively.

- Best seasonal performance- bar chart

This is a stacked bar chart that illustrates the production of crops in each season over the years, giving insights of the best season of production and its performance over the years.

- Best production over the years- Time-series line chart

This particular chart illustrates production over the years and highlights periods of significant growth or decline.

- Crop production Overview: Bullet chart

Bullet charts showcase the progress of a single metric against a given target. The higher the fill, the closer the metric is to the target. Here, it shows the progress of crop production, allowing for a quick visual assessment of performance.

- Nightingale Rose chart for production over the years

A nightingale rose chart is a polar coordinate chart where the circle is broken into wedges of equal angle, and the value represented by any wedge is illustrated by its area, rather than its radius or sweep angle. This chart illustrates crop production over the years using wedges, where the area of each wedge represents the production volume for that year. The chart thus effectively highlights fluctuations in crop production.

- Production trends of crops across years and seasons- sunburst chart

A sunburst chart uses circles to visualize the flow of data through different stages of a system. This sunburst chart visually represents the hierarchy of crop production, organized by year, season, and crop type. The size of each segment corresponds to the total production for that category, allowing users to easily

navigate through the data and identify patterns in production across different years and seasons, as well as the contribution of individual crops to overall output.

- Gauge chart for seasonal crop production

A gauge chart uses a gauge to showcase progress of a metric towards a target. This gauge chart visually represents the progress of seasonal crop production toward a defined target

- Crops word cloud

A word cloud visualizes the words in a column- Crops- that appear the most often. Bigger font corresponds to higher total production.

- Tree map of crop production

A tree map shows hierarchical relationships of data, with the value represented by area, showing proportion and contribution to the whole. This tree map visualizes the hierarchical relationships of different crops within the overall crop production dataset. Each rectangle represents a crop, with the area of the rectangle corresponding to its production volume, allowing for easy comparison of contributions to total production.

- Best Crop Choice for Each Season: Heatmap

A heatmap visualizes a related metric across pairs of groups. This heatmap visualizes the optimal crop choices for different seasons. It allows for quick identification of the best-performing crops for each season, helping farmers and decision-makers easily discern which crops are most suitable to plant based on seasonal conditions and expected outcomes.

4.1.3.2 Interactive features used

The dashboard on superset has many interactive features like filters, layout changes, customizations, Annotations and so on.

- There are three interactive native filters on this dashboard- Season, Crop and Years, to get targeted insights of the data by selecting desired crop or season or year. This way, users can make informed decisions regarding specific agricultural contexts.

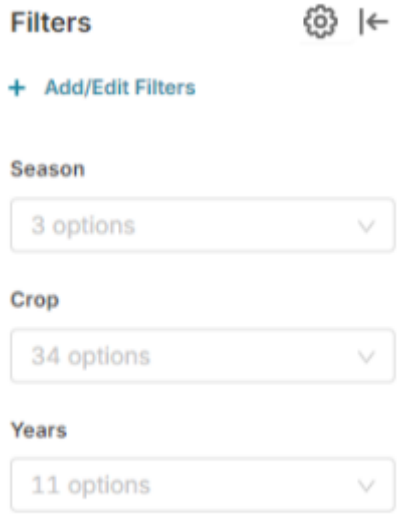

Fig.8: Filters

- There are many layout elements that help organize the dashboard. There are two main tabs: Overview and Exploratory. This structure helps users navigate the dashboard efficiently, accessing both overall performance insights and in-depth data exploration seamlessly. The charts are placed in rows and columns. There is a text/markdown element that contains a definition and the link to the original dataset.
- Superset also provides a way to make customizations to have the dashboard look more appealing. The live CSS editor was used in setting specific font families and sizes for headings, altering background colors for various components, and applying unique styles to big number charts, including gradient backgrounds and images.
- Alongside descriptive titles, annotations are efficient in helping viewers interpret a chart. Annotating makes the chart more informative. Some charts were annotated to highlight periods of production growth and upper bounds.

4.1.3.3 Role based Access used: Gamma role

The Gamma role in Superset is designed for users who primarily consume content. These users have limited access and can view data sources they have explicit permissions for, but cannot modify or add data sources.

To provide access, grant permissions for

- the crops dataset
- the MySQL database
- the cropsdatabase schema
- the filters (Crop, Season, Years).

To edit these permissions, navigate to Security > List Roles and then edit the Gamma role permissions for the same.

The Gamma role is suitable for users who need to view dashboards and charts without the ability to alter or add data sources.

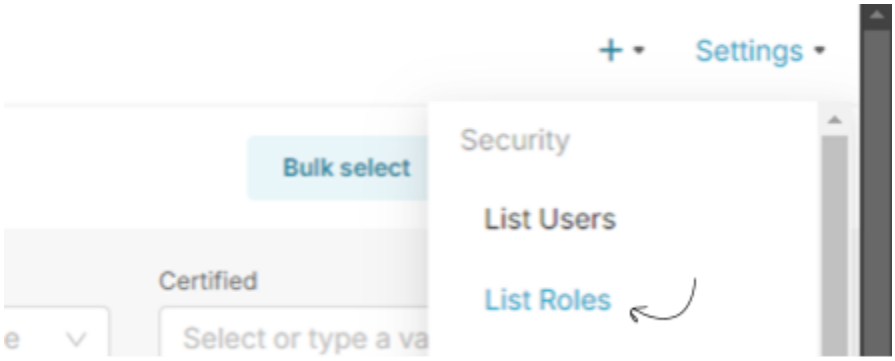

Fig. 9: List Roles

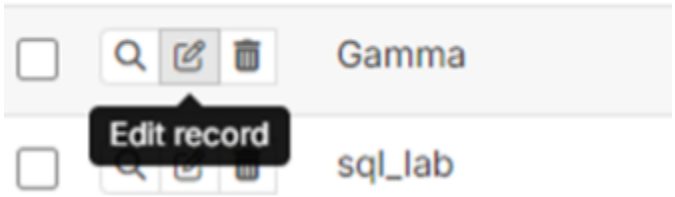

Fig. 10: Gamma role

Fig. 11: Superset Dashboard

Fig. 12: Superset Dashboard continued (1)

Fig. 13: Superset Dashboard continued (2)

Fig. 14: Superset Dashboard continued (3)

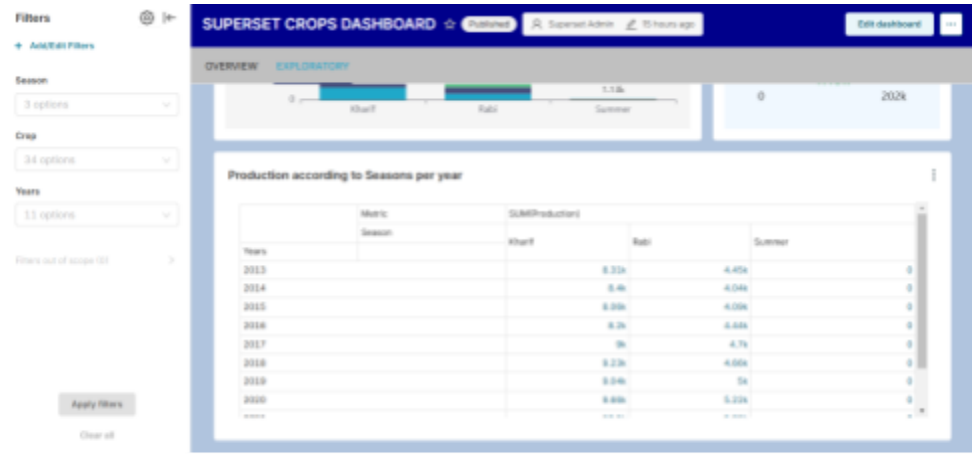

Fig. 15: Superset Dashboard continued (4)

4.2 Dashboard Creation and design in Metabase

Design user-friendly dashboards that enable seamless data exploration, with rich, semantic models for consistency and reduced redundancy. Users can create and share engaging visualizations effortlessly, that cater to both large-scale analytics and routine tasks, all within an intuitive interface.

4.2.1 Data import and setup

As mentioned earlier, since the ‘Year’ column is not in a Date or temporal format type, it is transformed to a standard Date format using MySQL workbench.

```sql
-- Create the schema
CREATE SCHEMA metabase;

-- Use the newly created schema
USE metabase;

-- Create the table
CREATE TABLE crops (
    Crop VARCHAR(40),
    Season VARCHAR(20),
    Years VARCHAR(20),
    Production DECIMAL(10, 2)
);
```

Fig. 16: metabase schema and crops table creation

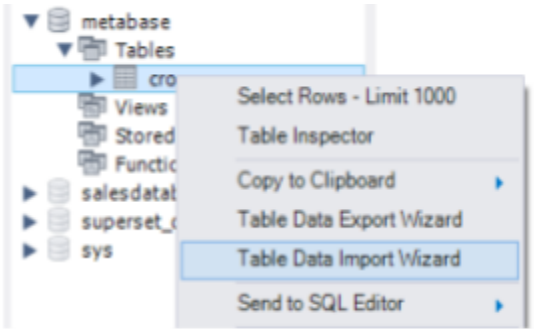

Fig. 17: Browse and import preprocessed dataset

```sql
ALTER TABLE crops
MODIFY COLUMN  Years DATE;
```

Information

Table: crops

Columns:

Crop	varchar(20)
Season	varchar(20)
Years	date
Production	decimal(10,2)

Fig. 18: Modify Years column to DATE type

The ‘Years’ column is now in the standard date format.

4.2.2 Using MySQL as the Database engine

To connect Metabase to our data source, we need to add a database connection.

To add a database connection, click on the gear icon in the top right, and navigate to Admin settings > Databases > Add a database.

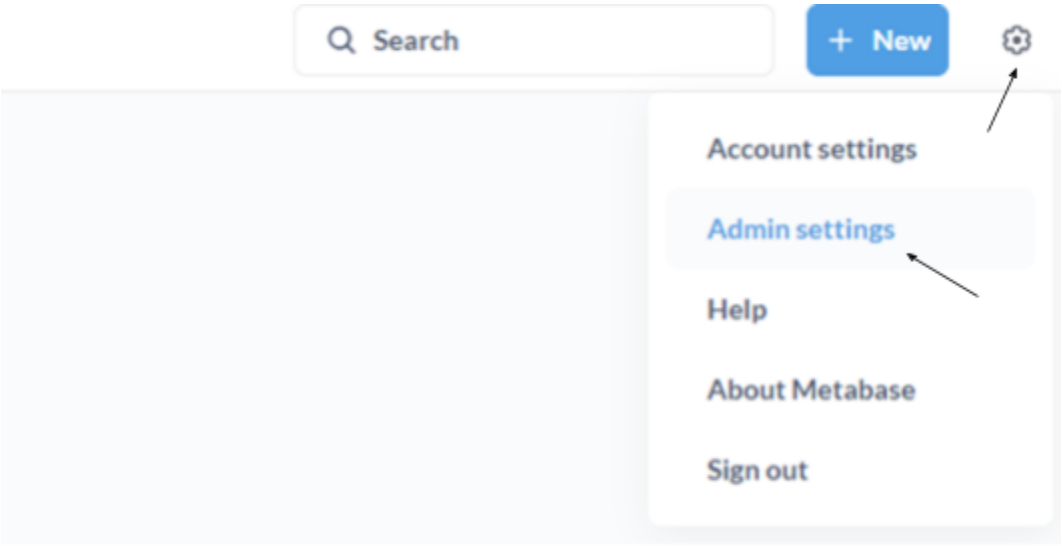

Fig. 19: Go to admin settings on metabase

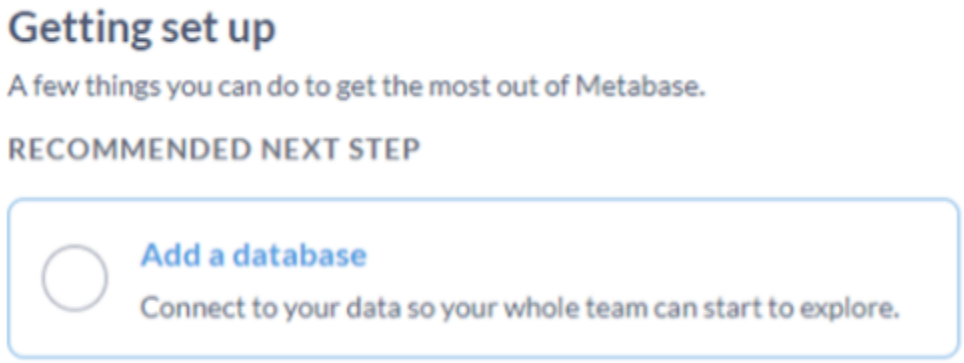

Fig. 20: Add a database

Fill out the fields for the MySQL database, and click Save changes at the bottom.

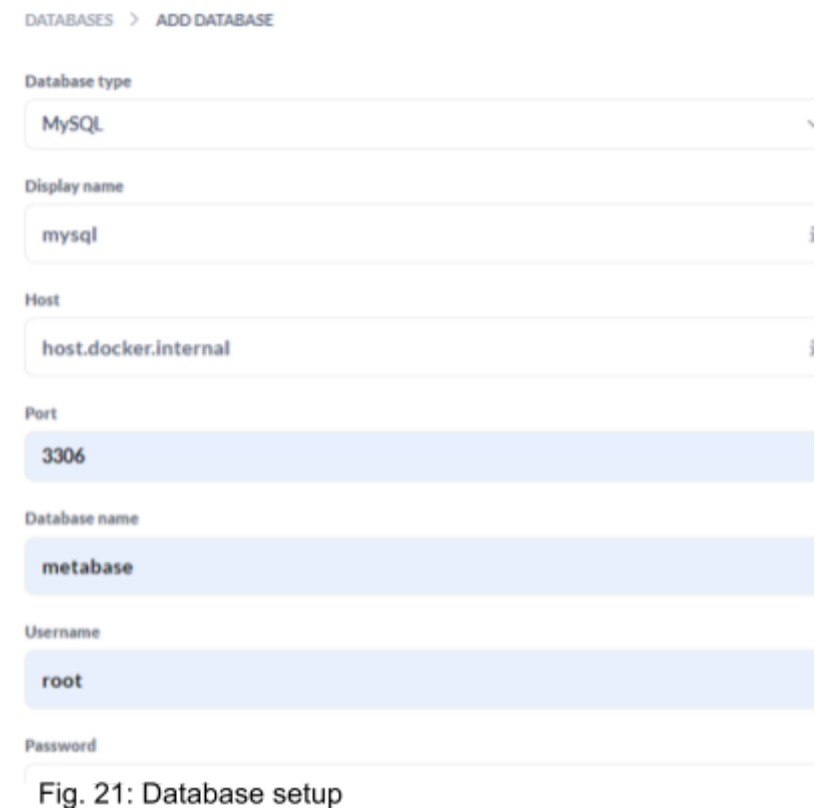

Fig. 21: Database setup

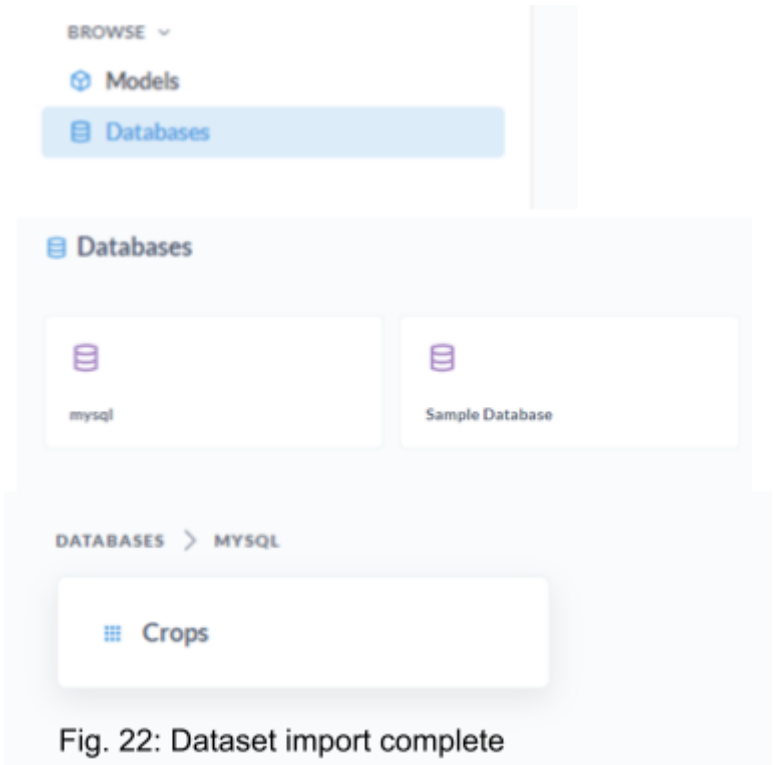

Fig. 22: Dataset import complete

This way, we can easily use MySQL as the database engine and import our data.

4.2.3 Dashboard Design

A smooth data exploration experience including rich, semantic models that ensure consistency and reduce redundancy is prioritized.

The dashboard should enable effortless creation and sharing of engaging visualizations, allowing users to tailor their views to their specific needs. This interface should remain intuitive and accessible, enabling users of all technical backgrounds to gain insights without needing complex coding or SQL queries.

4.2.3.1 List key metrics and Chart Types used in dashboard

- Two big number charts: Total crops, and total production in lakh tonnes.

These charts highlight the total number of crops, and total production generated over the years studied in the data respectively.

- Table1: Crop and production table, Table 2: Top 10 Crops.

The ‘Crop and production table’ summarizes the total production of each crop, and the ‘Top 10 crops’ table highlights the best 10 crops in terms of highest total production. This could help us in identifying the crops that help maximize yield.

- Pie chart1: Sum of Production by Crop, Pie chart2: Seasonal contribution to production

The ‘Sum of Production by Crop’ chart represents the proportion of total agricultural production contributed by each crop. The ‘Seasonal contribution to production’ chart shows the distribution of crop production across the three seasons.

- Production according to Seasons per year- Pivot table.

This pivot table summarizes crop production data by organizing it into seasons for each year, helping users to analyze trends, gain seasonal insights, and identify patterns in crop production across different years.

- Max of Production by Season- Row chart

This row chart shows the highest production values for each season using horizontal bars, making comparison across seasons easier.

- Best seasonal performance- bar chart

This is a stacked bar chart that illustrates the production of crops in each season over the years, giving insights of the best season of production and its performance over the years.

- Average of production by season- waterfall chart

A waterfall chart is a data visualization technique that helps in understanding the cumulative effect of sequentially introduced positive or negative values. This waterfall chart illustrates how each season's changes in average production contribute to the overall total, highlighting seasonal trends.

- Line charts-
    - Production changes in crops: This line chart effectively highlights variations and growth patterns, making it easier to identify variation in production across different crops.
    - Production growth in the last five years: This line chart illustrates production over the last five years highlighting the periods of significant growth or decline.
    - Seasonal performance over the years: This line chart illustrates how production varies across the three seasons over the years. Each line represents one of the three seasons, showing trends and fluctuations in performance. This ensures easy comparison of seasonal performance and highlights seasonal output variations over time.
    - Average Production over the years: This line chart displays the average production levels for each year, highlighting trends and changes in performance over time.
    - Best and Worst Productions each Year: Depicts the maximum and minimum production each year on a line chart, depicting the changes over the years.

- Sugarcane production over the years: This line chart illustrates the performance of sugarcane, which is recognized as the best performing crop. By tracking its production levels over the years, the chart provides valuable insights into trends, growth patterns, and variations, helping to grasp the crop's overall performance and its impact on agricultural yield.

4.2.3.2 Interactive features used

Metabase dashboards offer many interactive features like filters, layout changes, customizations, Annotations and so on. These features together enhance the usability of the dashboards, making data analysis more user-friendly and insightful.

- There are three interactive native filters on this dashboard- Season, Crop and Years, to get targeted insights of the data by selecting desired crop or season or year. This way, users can make informed decisions regarding specific agricultural contexts.

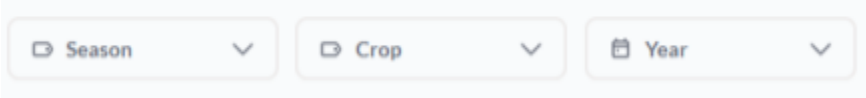

Fig.23: Filters

- There are many layout elements that help organize the dashboard and provide a personalized view. There are two main tabs: Overview and Exploratory. This structure helps users navigate the dashboard efficiently, accessing both overall performance insights and in-depth data exploration seamlessly. The text/markdown element contains a definition and the link to the original dataset.
- There is an option to add a link card which can be used to provide direct access to a specific dataset enhancing accessibility and facilitating deeper data exploration.
- The “Hide this card if there are no results” option allows users to automatically conceal a card when it doesn’t display any data. This feature makes the dashboard look clean and appealing, enhancing the usability by eliminating empty or irrelevant charts from cluttering the display so that users only see meaningful visualizations and metrics.

- The dashboard can be customized with different visualization types, colors, and styles to tailor to user needs and enhance readability.
- Alongside descriptive titles, annotations are efficient in helping viewers interpret a chart. Users can add descriptions or comments directly on the charts to provide context, highlight key insights, or explain data trends. Some charts were annotated to highlight upper bounds of crop production.
- Drill-Down features: Users can click on data points to gain access to more detailed information, allowing deeper analysis of specific parameters.

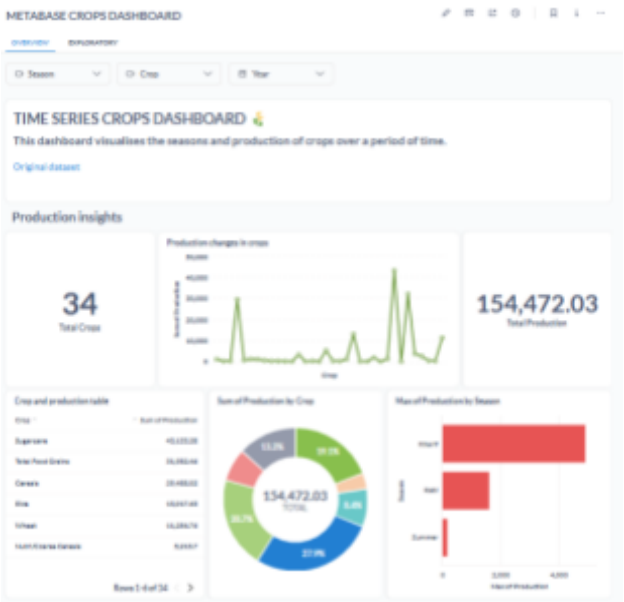

Fig. 24: Metabase Dashboard

Fig. 25: Metabase Dashboard continued(1)

Fig. 26: Metabase Dashboard continued(2)

Fig. 27: Metabase Dashboard continued(3)

Fig. 28: Metabase Dashboard continued(4)

5. Dashboard Embedding using iframe

5.1 Embedding Superset dashboard into HTML

1. Create superset_config.py in Desktop
2. Mount in Docker using command:

- *docker cp superset_config.py superset:/app/pythonpath/superset_config.py*

3. Verify that it has been copied:

- *docker exec -it <container_name_or_id> /bin/bash*

4. Go to the desired path inside docker: cd /app/pythonpath/

- *ls*
- exit (from shell)

5. Restart superset on docker

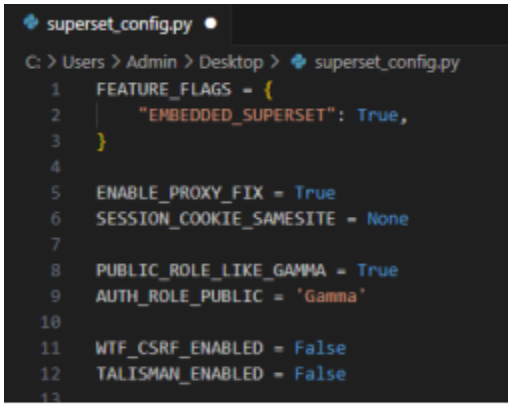

Fig. 29: superset_config.py

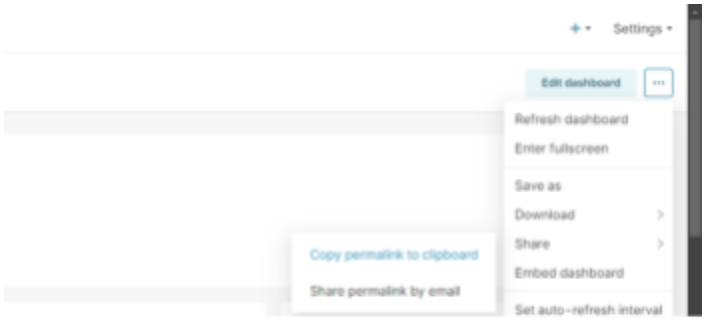

Fig. 30: Permalink to superset dashboard

Use this link inside the iframe in the html file. Once the superset_config file is mounted on Docker, you can run the HTML file to access the embedded dashboard.

5.2 Embedding Metabase dashboard

- Go to Settings > Admin settings > Embedding.
- Toggle the Enable embedding option.
- Go to the dashboard , Click on the sharing icon.Select Embed.
- Copy the Public embed link and add to your HTML file.

5.3 HTML Web page

The web page is designed to display both the embedded dashboards, allowing users to switch between them easily using buttons.

- Gradient Background- transition from light blue to light steel blue.
- Centered title "Embedded Dashboards" with a hover effect changing color to navy.
- Two **buttons** to switch between the two dashboards with a hover effect (changing color).
- Dashboards are embedded within **iframes**

5.4 Running the HTML file locally

Opening the HTML file directly might lead to issues related to Cross-Origin Resource Sharing (CORS) and other browser security restrictions. To avoid these issues, we can use Python's built-in HTTP server to easily host the HTML file locally. This way, both the Metabase dashboard and the HTML file are considered to be from the same host and hence the resources don't get blocked.

- Open cmd prompt, navigate to directory with embedded html file.
- Write: python -m http.server 8000
- Go to localhost:8000
- Open the embedded dashboards html file

Directory listing for /

- EMBED.html

Fig. 31: Serving the HTML File with Python HTTP Server

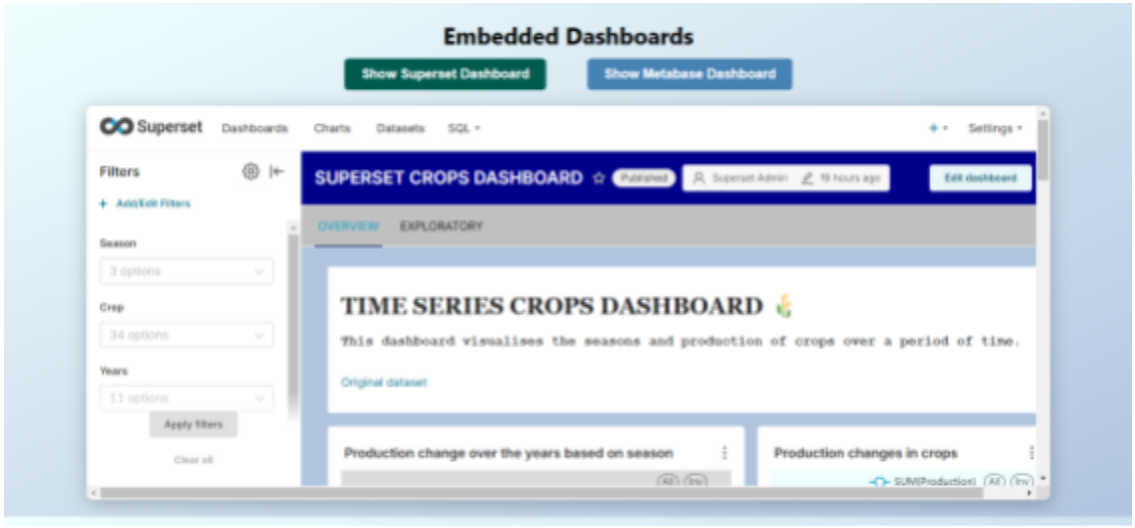

Fig. 32: Embedded Superset Dashboard

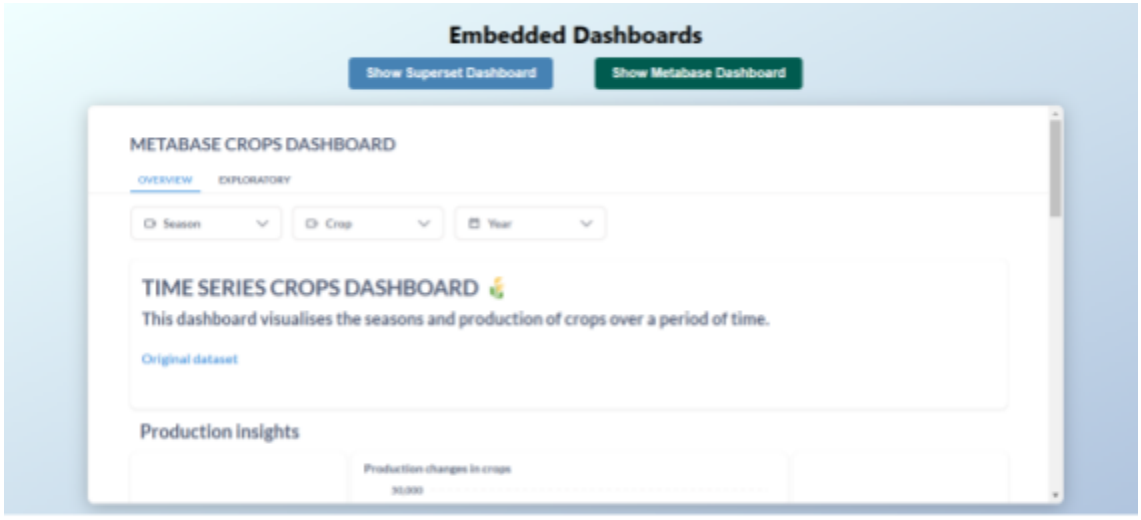

Fig. 33: Embedded Metabase Dashboard

6. Comparison of the two BI tools

FEATURE	METABASE	SUPERSET
Core functionalities: ● Data connection ● Query Builder ● Visualization ● Map charts	Supports 15 different SQL databases or engines	Supports 39 different SQL databases or engines
	Point-and-click query builder	SQL lab for advanced querying
	Basic charts and dashboard, about 20 default charts	Very broad range of visualization types, over 60 default charts
	Basic support for simple maps	Advanced mapping capabilities; customizable map charts
Flexibility	Limited flexibility in terms of customization and advanced visualizations	Highly flexible with extensive customization options for charts and dashboards
Scalability	For small-medium sized business	For large organizations
Usability: ● User capability ● Setup	Focuses heavily on empowering non-technical users.	Users might require technical expertise and some SQL knowledge.
	Faster, simple and straightforward setup	Takes longer to setup Steeper learning curve
Unique offerings: ● User interface ● Automatic insights ● Versatility ● Collaboration	Intuitive and simple user interface	Comprehensive user interface
	Provides automatic insights	Lacks automatic insights
	Strong focus on simplicity	Supports a broad range of visualizations and advanced analytics.
	Accessible to all users irrespective of their technical expertise. Easily share dashboards.	Limits collaboration to technical users. Sharing of dashboards is complex
Predictive Analytics	Offers basic analytics options, advanced predictive analytics features are limited.	Can integrate with machine learning models, making it better suited for predictive analytics

Advanced features	Limited	More sophisticated
Support for HTML embedding	Does not support embedding HTML content directly within dashboards.	Allows for embedding HTML content in dashboards, offering greater flexibility in design and interactivity.
CSS customizations	Limited customization; Predefined themes	Extensive customization available; Has live CSS editor
Pros	● User friendly, easy to setup ● Ideal for non-technical users ● Strong community support ● Quick drill -through analyses, one-click-to-create charts are available.	● Highly customizable and flexible for technical users. ● Wide range of data visualization options. ● Filters can be easily added to gain targeted insights. ● Helpful if you want only some people querying your data.
Cons	● Limited advanced features ● Customization Constraints ● May not satisfy the needs of technical users. ● The Chart collection is not vast.	● Resource-intensive ● Duplicating charts or dashboards is hard ● Complex setup and learning process. ● You need to know the kind of visualizations required beforehand and not on an adhoc basis.

7. Conclusion, Final Recommendations and Use cases

When deciding between Metabase and Superset, it's essential to consider the specific needs of the organization, how advanced the visualization requirements are, budget and more. Both tools provide strong data visualization and analysis capabilities, but they cater to different user capacities and use cases.

Metabase is ideal for teams with non-technical users who need easy access to data and straightforward reporting with no extensive training. It supports basic data analysis, interactive dashboards, quick drill-through analyses, and simple chart

creation. Overall, Metabase is a great choice for those who prefer a user-friendly interface that facilitates quick data exploration and reporting.

Superset is more suited for organizations with a more technical user base, that require advanced SQL querying, advanced analytics, customization, and a broader range of visualization options. Although it has a steeper learning curve, the advantages of customization and scalability are substantial for larger or more complex data environments.

Metabase use cases

1. **E-commerce:** *Sales performance analysis*

A sales team leverages Metabase to analyze customer behavior, monitor monthly sales performance across different regions, and optimize product recommendations. The user-friendly query builder allows team members to create reports quickly, without using any technical SQL skills, helping them identify trends and make informed decisions based on real-time data.

2. **Finance:** *Financial performance analysis*

A finance team uses Metabase to create a dashboard that conducts risk analysis, tracks portfolio performance, monitors market trends, and identifies investment opportunities. Non-technical team members can easily filter and visualize data to assess the effectiveness of their financial strategies.

Superset use cases

1. **Customer Support Analytics:**

A customer support team makes use of Superset to track metrics like inbound calls, resolutions, customer satisfaction scores, and support team performance. With advanced SQL querying, complex visualizations and customizations, the team can easily monitor trends and identify areas for improvement, leading to enhanced service quality and faster response times.

2. **Healthcare Quality Metrics:**

A healthcare organization employs Superset to visualize patient outcomes, analyze treatment patterns, readmission rates, and treatment effectiveness. With its ability to create detailed reports and visualizations, Superset helps healthcare providers monitor performance indicators, facilitating informed decision-making to improve patient care and operational efficiency.

In summary, the choice ultimately depends on the team's technical expertise, the complexity needs, and the level of customization required.

8. Advanced capabilities

Metabase

- Intuitive dashboards can be created with filters and multiple visualizations, allowing users to gain targeted insights.
- Metabase offers alerting features that notify users of changes in key metrics, helping teams stay informed.
- Users can configure scheduled email reports that automatically send insights to stakeholders at specified intervals, providing timely updates without manual intervention.
- Users can integrate Metabase reports and dashboards into applications or websites, facilitating wider access to insights throughout the organization.
- A single query is enough to join multiple datasets, allowing for a comprehensive analysis across related data sources.
- It supports the ability to layer charts, enabling users to overlay one chart on top of another. This highlights correlations or comparisons directly on a single visualization, enhancing interpretability.

Superset

- The advanced SQL Lab allows users to write, test, and visualize complex SQL queries, enabling deeper data exploration.
- Users can build highly customizable dashboards featuring complex visualizations, charts, and filters organized to cater to specific analysis needs.
- Superset supports role-based access control, enabling organizations to implement granular permissions and maintain data security.
- Superset can connect to real-time data sources, enabling live updates to dashboards and visualizations, which is essential for time-sensitive analytics.
- With the SQL Lab, users can write complex queries to join multiple datasets, allowing for intricate analyses leveraging multiple data sources.
- Users can use one chart to drive another, allowing for interactive and drill-down capabilities.

9. Appendix

GitHub Repository

- https://github.com/lolly-pop-pb/IDEAS-PROJECT

Dashboard Links:

- http://localhost:8080/superset/dashboard/p/9o5PRVpPXOV/
- http://localhost:3000/public/dashboard/f4e7831a-c383-42ea-a9ba-e3bf24fa0d91

Resources

- https://superset.apache.org/
- https://www.metabase.com/

- https://www.metabase.com/lp/metabase-vs-superset
- https://preset.io/blog/superset-vs-metabase/
- https://www.metabase.com/docs/latest/installation-and-operation/running-metabase-on-docker
- https://preset.io/blog/customizing-superset-dashboards-with-css/
- https://docs.preset.io/docs/css-mini-guide-1
- https://medium.com/@tusharkantanayak/how-to-customize-superset-charts-using-css-external-css-36591e8c7d57
- https://www.w3.org/wiki/CSS/Properties/color/keywords
- https://www.youtube.com/watch?v=yrrLxpz9J8c
- https://www.youtube.com/watch?v=vBSaUNxLE0c
- https://www.youtube.com/watch?v=EQm7Kk1D1Cw
- https://www.youtube.com/watch?v=dnT2Xj52mnk&t=1s
- https://www.youtube.com/watch?v=aAdxrX2VwcY
- https://www.youtube.com/watch?v=r6NOwBv5r-Y
- https://www.youtube.com/watch?v=wALCw0F8e9M
- https://www.youtube.com/watch?v=mAnp_yugoWM