CARTEL

Superset: Experience Brief
September 18 2019

Table of Contents

- Current Product Audit

- Personas

- Recommendations

CARTEL

Current Product Audit

Current Product Audit

We evaluated the current Superset product to get an understanding of the existing functionality, design issues, and areas for improvement.

We reviewed the product across the following areas:

- Navigation / UI
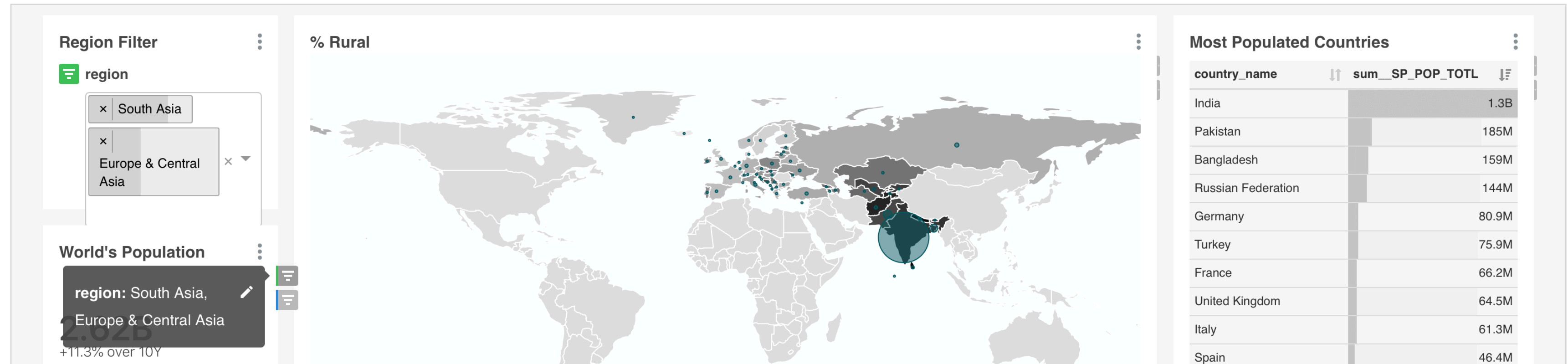- Dashboards
- Explore
- SQL Labs

CARTEL

Navigation / UI

- Order of items in navigation doesn't map to most common usage

- No clear sense of orientation or wayfinding; inconsistent application of on state for current product section and lack of clear hierarchy

- Object organization is inconsistent between charts, dashboards, and SQL objects/tools

- Object lists are data dense but lacking some useful metrics/information around usage

- Object lists visually simplistic and miss opportunity to guide users with thumbnails

- No clear navigation to the Explore tool outside of the "New" flow

- Unclear what "Manage" contains, and has items that could be better organized elsewhere

- Sources misses opportunity to better explain data organization and options, as well as serve as a place to find available information
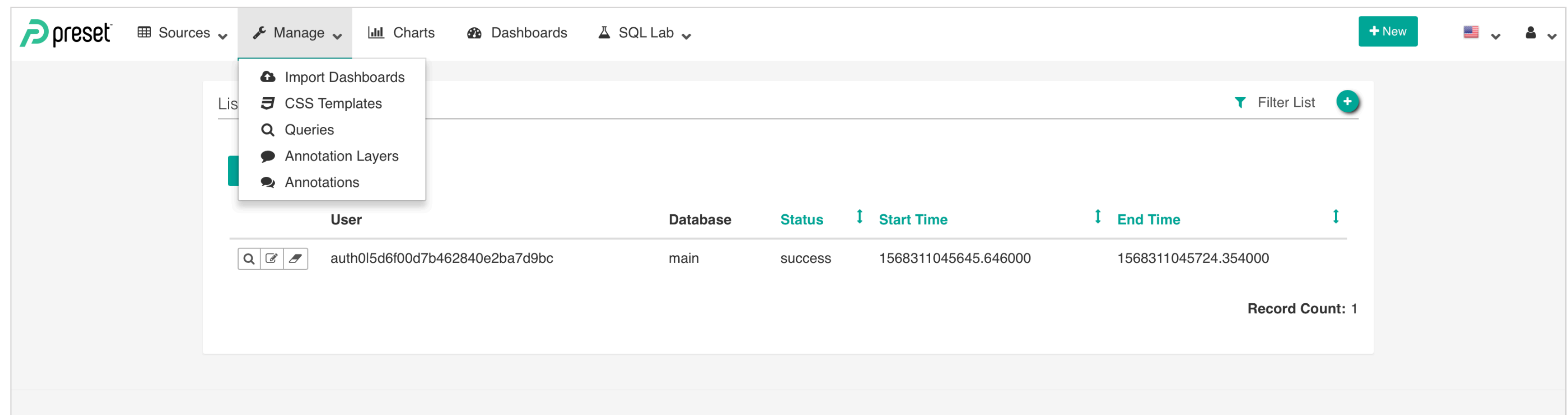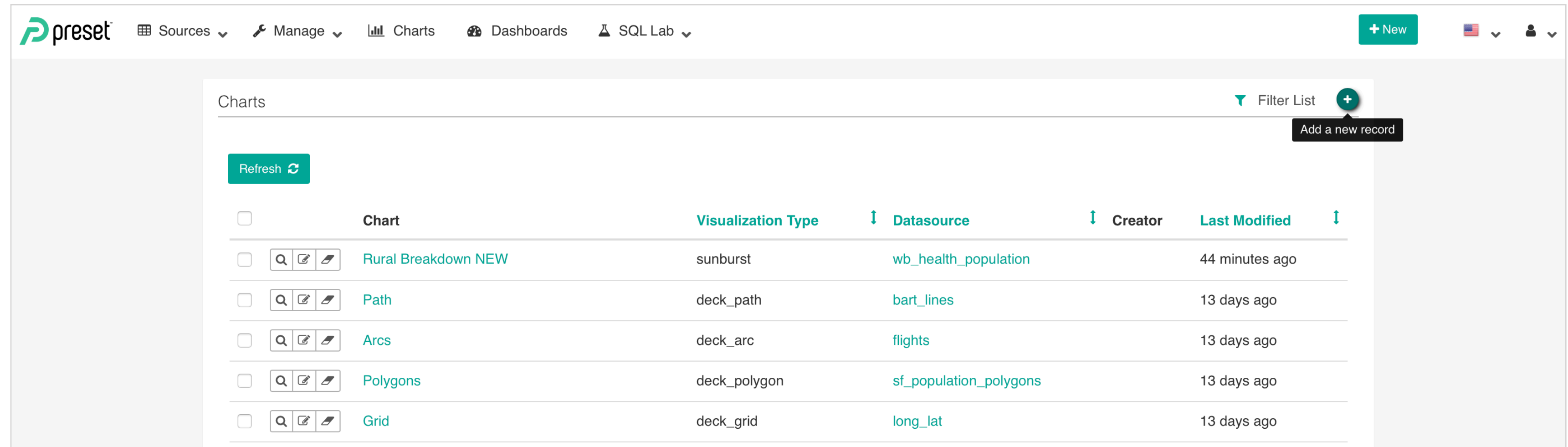
Dashboards

- Use of underlying grid creates an attractive uniform appearance for dashboards, regardless of how content is customized

- Sharing options are limited and somewhat hidden

- Filter tabs require hover to understand where filters are applied

- Lack of distinction between filter panels vs data panels requires thought to understand what is happening

- Actions button/dropdown has a somewhat non-intuitive approach

- "Edit Chart" flows leave users in an unexpected place and require re-navigating to the dashboard

- Save interaction behaves slightly oddly, options could be grouped more intuitively

- Resize behavior should snap to columns while resizing

- Tab interaction is somewhat unexpected, moving multiple charts into a tab is laborious

- Chart addition would benefit from thumbnails, possibly additional filters and option to not show charts that are already on the dashboard

Explore

- Hierarchy and layout could be improved to better reflect structure and importance of information and actions
- Controls are inconsistent with other parts of the product
- Query settings are difficult to understand and use
- Query panel shows a large amount of settings, many of which are advanced. Could be hidden until asked for to better support simple queries while allowing access to more complicated settings
- Lack of query setting consistency across visualizations
- Datasource and viz take up a lot of room and force query options lower
- Difficult to understand what data is available
- "Customize" is not an intuitive label for visualization options, location may be problematic as well
- Customize controls are all presented at once, which can be overwhelming
- Left panel is not collapsible if more real estate is desired
- Too many visualizations types shown at once, and not grouped effectively. Most common charts could be called out more prominently while less common available on request/reduced in prominence
  Undo/history for changes could be improved to allow users to explore with less risk of work loss
- Time settings are very manual, ideally would be automatically chosen.
- Advanced settings should be hidden

SQL Lab

- Query tabs become difficult to manage

- Saving is manual and history/undo inconsistent, leading to fear of losing information (and high likelihood)

- Database selection, data table display, and query writing interaction is very confusing; difficult to understand what data scope is particularly after changing database/schema/table

- Table preview contains useful information, but limited. Missing opportunity for greater understanding of the data

- Tabs for Results, Query history, Preview, etc. create a false equivalency between contents, becomes unwieldy after multiple previews

- Missing opportunity for more visualization options than simple tables

- SQL typeahead/auto-complete could be smarter, contain better information

- SQL resize control could be more prominent

Personas

Personas

Personas are archetypes of the different types of people who will use the product.

We have identified the following user personas, based on feedback from the Superset community and reviewing existing research. Personas do not represent individual roles, but serve as composites for similar use cases.

Primary Personas:
- Data Engineer
- Data Scientist
- Explorer
- Consumer

CARTEL

Data Engineer

"There's a huge demand within the organization for data from our systems and applications, but making that data accessible to those that want it is a big challenge. Speed is a big concern, and making sure the right data is not only available to the tools people use but also fast to query takes up a lot of my time. I need a solution that allows me to focus on the data infrastructure and spend less time helping massage and query the data."

Overview

The Data Engineer is often the initial touchpoint between Superset and an organization attempting to use it. Within an organization they are responsible for data governance and security, building and managing data infrastructure, and ensuring data reliability and speed. Downstream users may enlist the Data Engineer to help in querying, both for their knowledge of the underlying data and for help getting the data into a format they can use.

Part of maintaining the data infrastructure involves building and managing ETL workflows to make the data accessible and fast. The Data Engineer may spend a significant amount of time responding to requests to speed up queries via summarization and other approaches (e.g. data cubes).

The Data Engineer is often a bottleneck, and is looking for a solution that will enable Data Scientists and other downstream users to analyze data with a minimum of assistance. The easier it is for those users to get data into the system and manipulate it by themselves, the more time they can spend on building and supporting the data infrastructure.

Usage Attributes

Attribute	Rating (out of 5)
Data Ingestion	●●●●●
Data Preparation/Modeling	●●●●●
SQL Knowledge	●●●●●
Ad-Hoc Exploration	●●○○○
Structured Exploration	○○○○○
Chart/Dashboard Creation	○○○○○
Chart/Dashboard Consumption	○○○○○
Collaboration/Sharing	●○○○○
Frequency of Use	●○○○○

Goals

- Promote exploration of all the data within the organization
- Enable basic users to analyze data with a minimum of support
- Allow broader access to browse and explore data to less technical users
- Focus on speed and accessibility of data
- Reduce need for custom ETL workflows for query speed improvements

Challenges

- May be the only one who understands the data, so ends up as a bottleneck in the organization
- Solutions like Looker/Tableau are too expensive
- Prefers open-source both philosophically and for the ability to customize and control
- Heterogeneous data sources make it difficult to provide optimal datasets
- People in the organization often don't know what data is available
- Providing broad access to the data is the only way to scale their work

CARTEL

Data Scientist

"Our business runs on data, but I'm one of the few people in the company who knows how to write complex queries in SQL. I'm constantly inundated with questions, which makes it difficult to get my own work done. I need a tool that allows me to quickly query the data for myself, but also enables other users to answer their questions on their own. With more time, I'd also like to be able to explore the data to understand what's available and gain insights I didn't know to ask for."

Overview

Data Scientists are the people in an organization closest to the data. They are fluent in SQL and are generally in the business of mining the data to provide information to stakeholders at all levels across the business. They may be involved in the preparation and ingestion of the data into Superset, and also for enriching/enhancing the data via semantic modeling. Additionally they may be responsible for building and training models with the data, often in tools outside of Superset.

Data Scientists will use tools like Superset when possible, but often use a host of tools outside as well depending on which is best suited. They are comfortable in more complex solutions such as R, and will sometimes access Superset purely through the API.

Data Scientists frequently become bottlenecks when less technical users are unable to answer their questions themselves. They may spend significant time preparing queries for other users, or teaching them how to use the software. This reduces time they could spend accomplishing their own work. If they had more time, they'd like to be able to explore the data and look for insights that they didn't know to ask for.

Usage Attributes

Attribute	Rating
Data Ingestion	●○○○○
Data Preparation/Modeling	●●●○○
SQL Knowledge	●●●●●
Ad-Hoc Exploration	●●●●●
Structured Exploration	●●●●●
Chart/Dashboard Creation	●●●●○
Chart/Dashboard Consumption	●●●○○
Collaboration/Sharing	●●●○○
Frequency of Use	●●●●●

Goals

- Write custom queries in both ad-hoc and structured scenarios to answer business questions
- Visualize query results in the most consumable manner
- Create, save, and organize queries for later retrieval and use
- Create charts and dashboards for their own use as well as for other users
- Enable less technical users to self-serve their own queries
- Understand "data lineage" and trust the results
- Understand what data is available to them

Challenges

- Bottleneck for other users; frequently stuck in a role as a data concierge
- Not always clear if data is really accurate of truth
- Organizing and managing queries and query history is difficult
- Need to be able to move between SQL and visualization UI/controls fluidly
- System speed has significant impact on the ability to get work done

CARTEL

Explorer

"I need access to the data so I can answer business questions, but I don't know SQL well enough to always get what I need. I know what I want the answer to look like, but I'm often not sure how to use the tools to get it in the shape I want it. If the data isn't there or not in the right format, I have to go bug people to help me get it the way I want it. I use other tools like Tableau and Excel, but I'd ideally like to have one tool that can do most of what I want."

Overview

The Explorer persona typically has a primary role that isn't data-focused, but needs to be able to do data analysis to support their primary function. They may need to report on performance or demographics to other parts of the company, or use data analysis to understand their own efficacy and direct future work. Explorers are often PMs or Engineers, and may have a decent understanding of the data but are not highly fluent in SQL or have lots of experience with querying data directly. They are familiar with tools like Excel and Tableau, and generally select the tool that seems most appropriate to arrive at what they are looking to accomplish.

Explorers usually have a specific task they are trying to accomplish or question they need answered. They may do some light ad-hoc exploration, but generally usage is more task-driven. Explorers often have an idea of what they want their query results to look like, but struggle to make the tool they are using produce that result.

Explorers frequently collaborate with Data Scientists and Data Engineers to get access to the right data and for help with queries. They are often tasked with creating reports and dashboards for other personas to consume.

Usage Attributes

Attribute	Rating
Data Ingestion	○○○○○
Data Preparation/Modeling	○○○○○
SQL Knowledge	●●○○○
Ad-Hoc Exploration	●●○○○
Structured Exploration	●●●●●
Chart/Dashboard Creation	●●●●●
Chart/Dashboard Consumption	●●●●○
Collaboration/Sharing	●●●●●
Frequency of Use	●●●●●

Goals

- Quickly and easily create charts and add them to dashboards
- Customize data displays to best convey insights
- Share insights with other parts of the business
- Be as self-reliant as possible when constructing queries and accessing data
- Reduce the set of tools they need to use to accomplish their tasks
- Organize and manage their queries for easy retrieval and use

Challenges

- Data they need may not be available or in the right format
- Query builders can be unintuitive or unfamiliar
- Lack of SQL knowledge means they must rely on what is available in the UI
- Multiple tools to analyze data make it difficult to know which is the right one for their task

CARTEL

Consumer

"Our organization is increasingly data-driven, and I need accurate and up-to-date metrics about the most important aspects of my role. I don't have time to sift through data; I want it on demand and in an easy to consume format. I have to trust the data and tell at a glance if something needs my attention. I'd prefer not to have to log into another tool, but I will if the value is high enough."

Overview

Consumers rely on data to make decisions about the business, but are generally consumers of the data rather than creators. The usually need data in the aggregate, but may want to be able to drill in for deeper insights. While they may not do much data discovery and preparation themselves, many have a technical background and are familiar with data concepts (particularly in smaller companies and startups).

Most of the content Consumers use is created by Data Scientists and Explorers. A primary use case is to have a well defined data set upon which they want to report on a regular basis (e.g. Monthly Sales by Channel). Such reports are typically consumed in the form of a chart/visualization on a dashboard, but in some cases may be delivered outside of a product in the form of an email/pdf report. Consumers sometimes have ad-hoc questions they need answered, but will generally delegate the investigation of the question to an Explorer or Data Scientist and receive the answer in the form of one or more visualizations.

Consumers want charts and dashboards presented in a clean, easy to consume (and share) format. Dashboard appearance, customization, and performance are key features to attract and support this persona. Because many executives tend to fall into this persona, supporting their needs can help to build critical buy-in to continue using a particular product.

Usage Attributes

Attribute	Rating
Data Ingestion	○ ○ ○ ○ ○
Data Preparation/Modeling	○ ○ ○ ○ ○
SQL Knowledge	○ ○ ○ ○ ○
Ad-Hoc Exploration	○ ○ ○ ○ ○
Structured Exploration	● ○ ○ ○ ○
Chart/Dashboard Creation	○ ○ ○ ○ ○
Chart/Dashboard Consumption	● ● ● ● ●
Collaboration/Sharing	● ● ● ○ ○
Frequency of Use	● ● ○ ○ ○

Goals

- Consume regular reports and visualizations
- Share reports and dashboards with others
- Delegate ad-hoc questions and receive timely and understandable results
- Be able to do light data exploration when questions arise

Challenges

- Multiple products to log in to
- Charts/dashboard performance can affect usefulness
- Trusting the data is difficult without understanding data provenance
- Exploration tools are fairly difficult to use, and no time or inclination to learn them

CARTEL

Usage Matrix

	DATA ENGINEER	DATA SCIENTIST	EXPLORER	CONSUMER
Data Ingestion	●●●●●	●○○○○	○○○○○	○○○○○
Data Preparation/Modeling	●●●●●	●●●○○	○○○○○	○○○○○
SQL Knowledge	●●●●●	●●●●●	●●○○○	○○○○○
Ad-Hoc Exploration	●●○○○	●●●●●	●●○○○	○○○○○
Structured Exploration	○○○○○	●●●●●	●●●●●	●○○○○
Chart/Dashboard Creation	○○○○○	●●●●●	●●●●●	○○○○○
Chart/Dashboard Consumption	○○○○○	●●●○○	●●●●○	●●●●●
Collaboration/Sharing	●○○○○	●●●○○	●●●●●	●●●○○
Frequency of Use	●○○○○	●●●●●	●●●●●	●●○○○

Open Questions

The following questions are fundamental to determining the direction of Superset, as they directly impact the conceptual and interaction models behind the new design.

1 Should the Explore view and SQL view come closer together or stay separate?

2 What is the object model as it relates to queries and reports/charts?

3 How do we want workspaces and groups to behave?

CARTEL

Recommendations

Overall

- Create a consistent visual language and set of interactions/controls, and apply them consistently throughout the product

- Refine workflows throughout the product to improve cohesion and align to user goals and common usage patterns

- Consider user onboarding and improve first-time experiences configuring and using the product

- Focus on improving the overall cohesion and usability around end-to-end exploration, object creation, and insight consumption

- Consider global search to make it easier to find objects from a variety of characteristics (e.g. name, sql query, created by)

CARTEL

Navigation

- Organize navigation around most common usage patterns

- Improve orientation and wayfinding by implementing clear, consistent on-states, page headers, and hierarchy

- Include Explore tool and SQL Lab as top-level items to give them a place to live in the architecture as well as to improve consistency

- Consider object groupings (e.g. workspaces) and their effect on the navigation

- Relocate Manage out of primary navigation, consider renaming, and rethink what product sections it contains

- Potentially move language selection a level deeper

CARTEL

Object Management

- Consider reducing object management overhead and user conceptual load by merging saved queries with charts; could potentially be one object type produced by different tools
- Improve listing views of objects with better filters and object location methods
- Improve display of objects in listing views with more visual layout and presentation (e.g. cards)
- Maintain data-dense table view of objects as a secondary view option, and improve layout and controls to facilitate additional object location methods and bulk manipulation use cases
- Facilitate improved user understanding of object relevance and importance with additional information around object use and popularity in their organization (e.g. frequency of use, etc.)
- Consider higher-level organization concepts like workspaces to improve object organization that aligns to complex organizations and facilitates collaboration

CARTEL

Dashboards

- Organize dashboard actions in a more discoverable and intuitive way
- Make sharing more prominent and accessible, and improve the options available as well as their presentation
- Redesign modal windows to better organize information and options intuitively (e.g. Save, etc.)
- Improve resize behavior hinting to help users better understand what the result will be
- Better differentiate between dashboard filters and content panels
- Improve interaction and display of filters to better facilitate use and clear understanding of filter effects
- Improve tab implementation and provide bulk actions to make it easier to take advantage of grouping
- Bolster data confidence with improved access to underlying data and display of relevant information
- Enhance chart addition with better object display and additional filters (e.g. hiding charts already on dashboard, adding thumbnails, etc.)
- Refine common workflows to align with user expectations (e.g. chart editing, save, etc)
- Consider templates or other accelerators to achieve effective layouts with a minimum of work

CARTEL

Explore

- Improve page hierarchy and layout to better reflect object structure, importance of information, and actions
- Choose a direction for the query controls and create a more intuitive approach to data exploration
- Rethink semantics to improve label comprehension and usability
- Reduce the number of controls shown by default, while maintaining access to advanced functionality
- Increase consistency of control layout across visualization types as much as possible
- Automatically apply appropriate settings where possible to reduce user workload (e.g. time settings)
- Provide improved access to data available in the datasource to better understand what is queryable
- Consider enhancing datasource display to provide more insight into the data itself (e.g. distributions of values, data cleanliness, etc.)
- Decouple datasource selection, querying, and visualization controls to divide user task load into discrete sections and reduce confusion

CARTEL

Explore (cont.)

- Reduce the number of visualizations presented by default to common types with more advanced options available on request to reduce user cognitive load
- Allow selection of visualization types not renderable with the current query but provide clear instructions of how to manipulate the query to match the viz
- Group visualization controls with the visualization itself and align more closely with publishing tasks
- Provide access to the data table returned by the query at all times while creating visualizations to better help users structure their query and map the data
- Reduce risk and encourage exploration by improving history and undo options

CARTEL

SQL Lab

- Improve page hierarchy and layout to better reflect object structure, importance of information, and actions
- Remove query tabs and consider alternate methods of enabling multiple lines of inquiry
- Improve access to saving and explore additional methods to prevent work loss (e.g. autosaving during session without permanent object creation)
- Provide better access to query history as well as undo controls to reduce risk of work loss and encourage exploration
- Rationalize system of datasource selection and data display to make it clearer what is possible to use in your query vs. what is shown on the left
- Consider enhancing datasource display to provide more insight into the data itself (e.g. distributions of values, data cleanliness, etc.
- Explore improving sql helpers/auto-complete
- Add ability to visualize data in additional ways, make controls and presentation consistent with Explore view
- Consider merging SQL Lab with Explore view, or improving flow between tools

CARTEL

CARTEL

Thank you.