



12

Visualizing Data with Amazon QuickSight

In *Chapter 11, Ad Hoc Queries with Amazon Athena*, we looked at how **Amazon Athena** enables data analysts to run ad hoc queries against data in the data lake using the power of SQL and Spark. And while this is an extremely powerful tool for querying large datasets, often, the quickest way to understand a summary of a dataset is to visualize the data in graphs and dashboards.

In this chapter, we will do a deeper dive into **Amazon QuickSight**, a **Business Intelligence (BI)** tool that enables the creation of rich visualizations that summarize data, with the ability to filter and drill down into datasets in numerous ways. In addition, QuickSight also enables the creation of formatted, multi-page reports, and brings advanced functionality, such as the ability to ask questions of data in natural language.

In smaller organizations, a data engineer may be tasked with setting up and configuring a BI tool that data consumers can use. Things may be different in larger organizations, where there may be a dedicated team to manage the BI system. However, it is still important for a data engineer to understand how these systems work, as these systems will consume data that the data engineer will have played a part in creating.

The purpose of BI tools is to enable users to quickly understand complex datasets by enabling the exploration of data *visually*. And while we will focus on Amazon QuickSight in this chapter, many of the con-

cepts in this chapter can be applied to other popular BI applications, such as Tableau, Microsoft Power BI, and Qlik.

Amazon QuickSight is a serverless BI solution and is fully managed by AWS. Organizations don't need to pay for any infrastructure or licensing costs, but rather pay a fixed amount per QuickSight user on a subscription basis based on which features are enabled.

In this chapter, we will cover the following topics:

- Representing data visually for maximum impact
- Understanding Amazon QuickSight's core concepts
- Ingesting and preparing data from a variety of sources
- Creating and sharing visuals with QuickSight analyses and dashboards
- Exploring QuickSight's advanced features
- Hands-on – creating a simple QuickSight visualization

Technical requirements

At the end of this chapter, you will get hands-on by creating a QuickSight visual from scratch. To complete the steps in the hands-on section, you will need the appropriate user permissions to sign up for a QuickSight subscription.

If you have administrator permissions for your AWS account, these permissions should be sufficient to sign up for a QuickSight subscription. If not, you will need to work with your IAM security team to create a custom policy. See the AWS documentation titled *IAM Policy Examples for Amazon QuickSight* and refer to the *All Access for Standard Edition* example policy as a reference.

At the time of writing, Amazon QuickSight includes a free trial subscription for 30 days for new QuickSight subscriptions. If you do not intend to use QuickSight past these 30 days, ensure that your user is

also granted the `quicksight:Unsubscribe` permission so that you can unsubscribe from QuickSight after completing the hands-on section.

Note that the *All Access for Standard Edition* example policy has a specific deny for the unsubscribe permission, so this may need to be modified based on your requirements. Work with your security team to implement a custom IAM policy.

You can find the code files of this chapter in the GitHub repository using the following link: <https://github.com/PacktPublishing/Data-Engineering-with-AWS-2nd-edition/tree/main/Chapter12>

Representing data visually for maximum impact

Data lakes are designed to capture large amounts of raw data and enable the processing of that data to draw out new insights that provide business value. The insights that are gained from a data lake can be represented in many ways, such as reports that summarize sales data and top sales items, **machine learning (ML)** models that can predict future trends, and visualizations and dashboards that effectively summarize data. Each of these ways of representing data offers different benefits, depending on the business purpose:

- If you're a data analyst who needs to report sales figures, profit margins, inventory levels, and other data for each category of product a company produces, you would probably want access to detailed tabular data. You would want the power of SQL to run powerful queries against the data to draw varied insights so that you can provide this data to different departments within the organization.
- If you're a logistics manager and are responsible for supplying all your retail stores with the correct amount of inventory, you would

want your data science team to develop an ML model that can predict inventory requirements for each store. The model could take in raw data from the data lake and predict how much inventory each store may require.

- If you're a sales manager for a specific product category, you need to have an updated view of sales for the products in your category at all times. You need to be able to determine which products are selling well, and which marketing campaigns are most effective. Seeing a visual representation of relevant data provides you with the most effective way to quickly understand the product and campaign's performance at a high level.

Having raw, granular data available to an organization is important, but when you need to make decisions quickly based on that data, having a visual representation of the data is critical.

It is not practical to identify trends or outliers in a dataset by examining a spreadsheet containing 10,000 rows. However, if you aggregate and summarize the data into a well-designed visual representation of the data, it becomes very easy to identify those trends and outliers.

Of course, you do need to be cautious when creating visualizations to ensure that you do not select a subset of data that is not representative of the full dataset. As Mark Twain (supposedly) said: "*There are three kinds of lies: lies, damned lies, and statistics*". It is important when designing a visual representation of data that you are careful not to misrepresent the data, and if you are viewing a visualization of data, that you are sure you can trust the author of that visualization.

Benefits of data visualization

A well-designed visual representation of data can reflect multiple different datasets in a single picture. It can do so in a way that enables the consumer of the visual to immediately gain insights that would take significant time and effort to gather from raw data.

There is a common fact often mentioned in articles on the internet that people can process images 60,000 times faster than they can process text. As it turns out, this is just an often-repeated claim with no evidence to back it up. However, while the number may be exaggerated, the basic claim that the human brain can process images quicker than text is without a doubt true.

And you don't need to look too far to validate this claim. For example, look at the rise of visual-based social media sites such as Instagram and Pinterest, or how people use emojis and animated GIFs to quickly and effectively communicate how they feel about something.

In the same way, we can use the power of visuals (images, graphs, word clouds, and many other types) to effectively communicate data from our data lake in a way that makes it easy for the consumer of the visual to quickly draw insights from the data.

Let's examine some common uses of visualizations that enable a user to quickly understand complex information.

Popular uses of data visualizations

Visualizations can be used to draw insights from many different types of data, in various ways. In this section, we will look at a few examples of some common types of visualizations to demonstrate the impact of a well-designed visual.

Trends over time

A common usage of analytic tools is to crunch through raw data to help surface trends, or changes in the data, over time. For example, we may want to understand how our spending on the AWS platform changes over time, as this can help identify areas where we need to focus on cost optimizations. A line graph can be a useful way to illustrate changes in data over a certain period.

The following diagram was created using a popular spreadsheet application and provides a visual of raw Amazon S3 spend per month, over 9 months, for a fictional company:

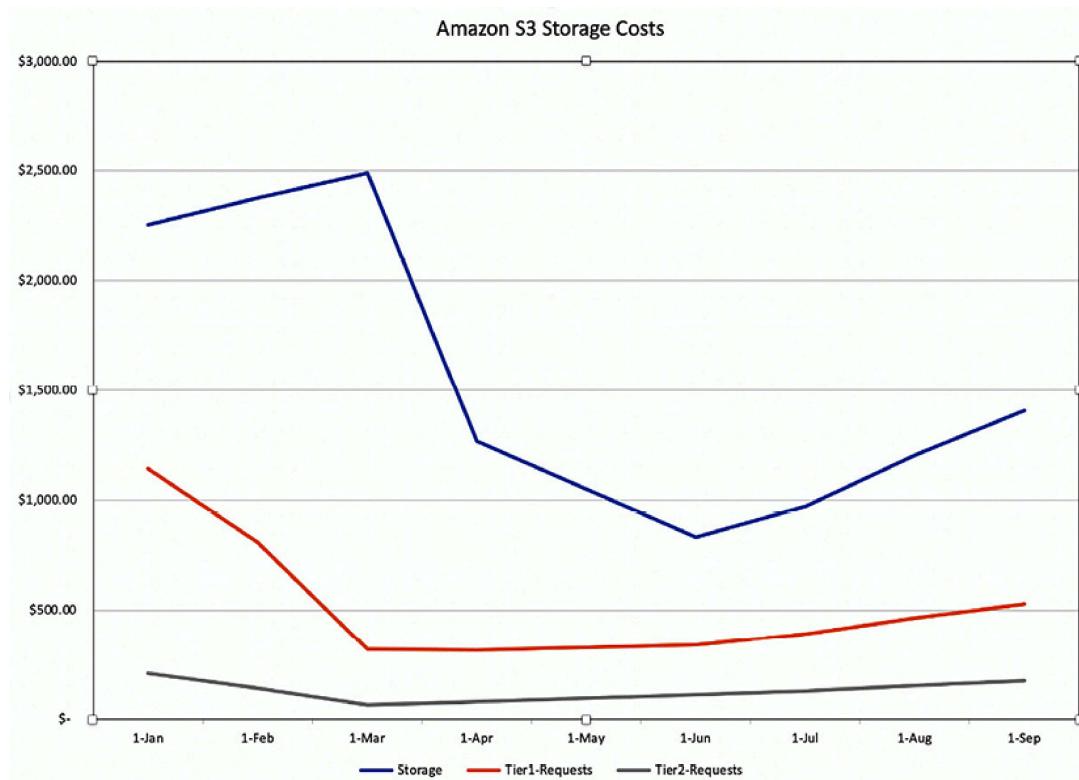


Figure 12.1: Line chart showing data over a certain period

In this visualization, we can see that our `Tier1-Requests` cost (middle line) significantly decreased from January to March. These costs are for API calls for operations such as `PUT`, `COPY`, `POST`, and `LIST`. Before February, we used to ingest a large number of small files, resulting in millions of `PUT` requests when writing these files to Amazon S3. After changing our transformation pipeline to write out fewer, larger files, this visualization clearly shows how those costs decreased.

In the visualization, we can also see that in March, our storage consumption (top line) significantly decreased. This makes sense as, during March, our fictional company had a project to implement Amazon S3 life cycle rules that deleted older versions of data from S3.

Showing summarized data over a certain period in a visual format makes it much easier to track and understand trends in our data, as well as to spot anomalies.

Data over a geographic area

In our first example, we looked at how we could graph trends over time, but another really useful visualization is to look at trends over a geographic area. There are many uses for this type of visualization, such as the following:

- Understanding the popularity of a certain product in different geographic regions
- Quickly visualizing hotspots for the spread of an infectious disease (such as flu outbreaks) in different geographic regions
- Visualizing the population sizes of different cities in different regions
- Showing differences in temperature in different geographic areas

These types of charts are often known as **geospatial charts**, although they go by many different names. The chart may also come in different formats, but a common format is to use circles of different sizes on the map, with the size of each circle representing the value of one of the columns in the dataset (the larger the value, the bigger the circle). Circles may also be different colors to represent different rows in the dataset.

For example, the following chart (created with Amazon QuickSight) uses city population data from <https://simplemaps.com/data/world-cities>. In this chart, we have filtered the data to show all cities with a population above 5 million people, and the size of each circle represents the relative population size. In the hands-on section of this chapter, you will use Amazon QuickSight to recreate this chart so that you can interact with the chart (filter for different values and so on):

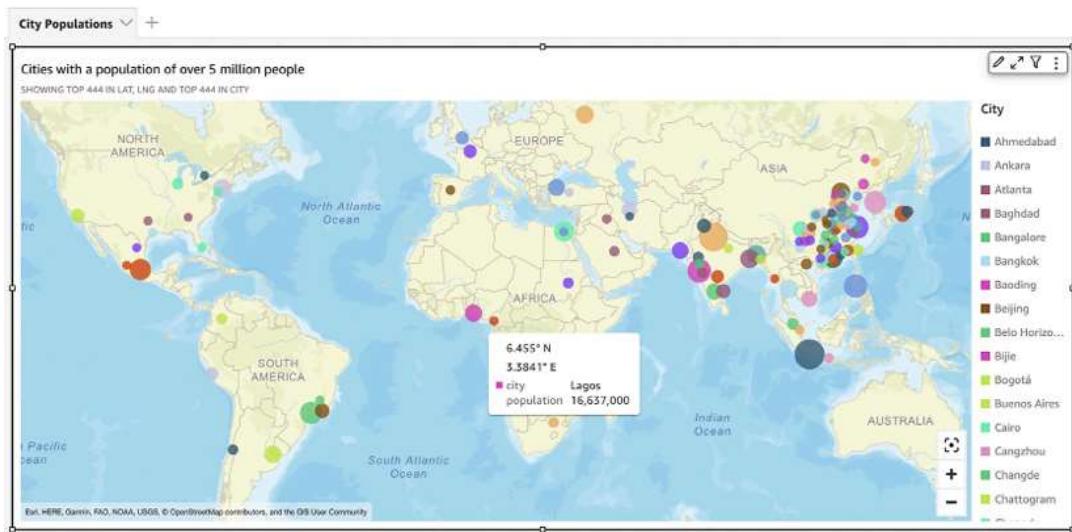


Figure 12.2: Map chart showing data by geographic region

The preceding map chart enables us to quickly understand which parts of the world have the most populated cities, and which parts are less populated. The same type of chart could be used to show the spread of disease, vaccination rates, poverty levels, water quality, or just about any other data that is associated with a specific location.

Heat maps to represent the intersection of data

Another common use of visualization tools is to understand the relationship between different sets of data. Often, we may have a gut feeling that there could be a correlation between two different datasets, but it is only when we explore the data more fully that we can understand those relationships.

As a very simple example, we would probably suspect that sales of ice cream, water, and other cold goods would be more popular in the summer months, and that the sales of coffee, hot chocolate, and soup would be more popular in the winter months:

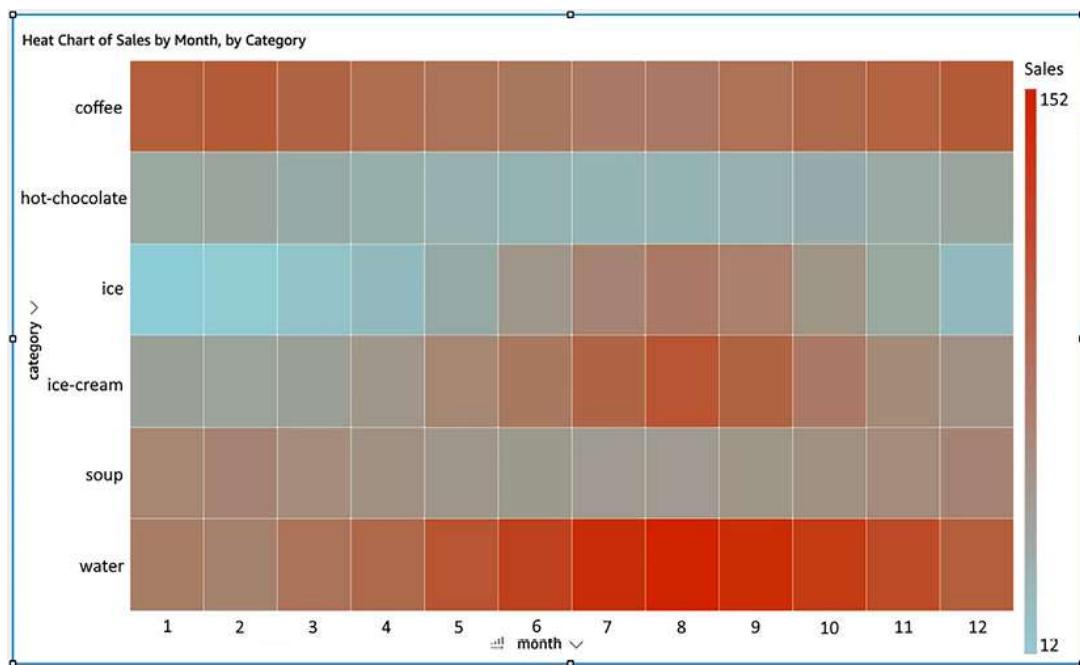


Figure 12.3: Heat map showing product sales by category and month

The preceding diagram shows a heat map that plots the relationship between sales in different categories, by month. The darker squares illustrate a higher sales value, while the lighter squares represent lower sales values.

As you can see, the sales of both coffee and water are strong throughout the year, but we can see that water has higher sales in the Northern Hemisphere summer (months 6 - 9), while coffee has higher sales in winter (months 11 - 2). Another insight we can gain quickly is that sales of ice are very low in the winter months and only have strong sales for a few summer months of the year (months 6 - 10). What other insights can you gain about sales of hot chocolate, ice cream, and soup by examining the heat map?

While this example may have been a fairly simple one, there are many other relationships between datasets that are not always as obvious, and heat maps can be useful to highlight these relationships visually.

We do not have sufficient space in this chapter to cover all the many varied types of charts that can be used to visually represent data, but

as we continue on to the other sections of this chapter, we will explore some other common chart types along the way. In the next section, we are going to dive deeper into Amazon QuickSight's core concepts.

Understanding Amazon QuickSight's core concepts

At its core, QuickSight lets us ingest data from a wide variety of sources, perform some filtering or other transformation tasks on the data, and then create dashboards with multiple types of visuals that can be easily shared with others, or highly formatted multi-page PDF reports.

The QuickSight service is fully managed by AWS, and there are no up-front costs for using the service. Instead, the service uses a pricing model of a monthly cost per user and offers both Standard and Enterprise editions. To include specific functionality, such as QuickSight Q (for making natural language queries of data), a higher price per user is charged. There is also an option for capacity pricing, where you pay for the number of sessions per month, or per year, instead of per user.

QuickSight also includes a powerful in-memory storage and computation engine to enable the best performance for working with a variety of data sources. In this section, we'll examine the differences between the Standard and Enterprise editions of QuickSight, and also do a deeper dive into SPICE, the in-memory storage and computation engine.

Standard versus Enterprise edition

The **Standard edition** of QuickSight is useful for those who are just starting to explore the power of a BI tool and enables users to create visualizations from a variety of sources. However, for larger organizations, the **Enterprise edition** of QuickSight provides several addi-

tional features that most large organizations would want to make use of.

The following is a subset of some of the additional functionality available in the Enterprise edition, but refer to the Amazon QuickSight pricing page for full details on the differences between the versions. If just getting started with exploring the functionality of a BI application, then it makes sense to start with the Standard edition, as you can upgrade to the Enterprise edition at any time.

However, the following features are only available in the Enterprise edition:

- Integration with **Active Directory (AD)** and the ability to use AD groups for the management of QuickSight resources
- The ability to embed dashboards into custom applications
- The ability to email reports to QuickSight users on a schedule
- Fine-grained access control over AWS data sources (such as S3 and Athena)
- Automatic insight generation using ML Insights
- Encryption of data at rest

Note that if you select to use the Enterprise edition, it is not possible to downgrade to the Standard edition. Therefore, when starting out with QuickSight it's recommended to start with the Standard edition, and then upgrade to the Enterprise edition as you have more users start using QuickSight, and if there are specific features from the Enterprise edition that you require.

While the cost per author for the Enterprise edition is more expensive, if you have a very large number of users that just need to access QuickSight as readers (i.e., do not create visualizations, but only view visualizations created by others), then you can take advantage of the **reader** pricing that is available in the Enterprise edition to reduce your costs overall.

With the Enterprise edition, there is a fixed monthly cost for users that have the author role, while users with the reader role are charged per session. Each session provides access to QuickSight dashboards for a user for up to 30 minutes after they have logged in. During this time, readers can fully interact with the dashboards (filtering data, doing drill-downs into data, and so on).

At the time of writing, a session costs \$0.30, and there is a maximum monthly cost of \$5 per reader (increased to \$10 per reader, if QuickSight Q is enabled), no matter how many sessions are used. In comparison, the Standard edition has a fixed cost (at the time of writing) of \$12 per user (if paying month-to-month) and all users have full read and author capabilities.

For example, if you have 50 users that need to access QuickSight, the Standard edition would cost \$600/month. However, with the Enterprise edition, if you had 5 authors and 45 readers, the cost would be \$345/month.

Refer to the QuickSight pricing page (<https://aws.amazon.com/quicksight/pricing/>) for the current pricing for your Region, as pricing may change occasionally. Let's now look at the in-memory storage and computation functionality available in QuickSight.

SPICE – the in-memory storage and computation engine for QuickSight

Like many other BI tools, Amazon QuickSight provides a storage engine for storing imported data and performing rapid calculations on that data. In QuickSight, **SPICE** is the acronym that's used to refer to this engine, and it stands for **Super-fast, Parallel, In-memory, Calculation Engine**. When you're creating a new dataset in QuickSight, you can select whether to perform direct queries of the dataset at the source, or whether you want to import data into SPICE.

If you choose to query the dataset directly from the source, then each time the visualization is accessed, QuickSight will make a connection to the data source (such as an Amazon RDS MySQL database) and query the data. This ensures that the dashboard always reflects the latest data. However, there is some latency in making the connection to the data source and retrieving data.

Alternatively, you can choose to import data into the SPICE engine. That way, when the visualization is accessed, QuickSight can read the data directly from SPICE, and this can significantly improve performance. For data that does not change constantly, it makes sense to import data into SPICE and query the data from there.

For example, if you have a dataset covering daily store sales, and the store sales are only updated at the close of business each evening (meaning the data is only updated daily), then it makes sense to import that into SPICE once a day. If you have 5,000 stores, and therefore 5,000 store managers that want to query the previous day's data for their store, having the data in SPICE reduces the number of direct queries of the database, and significantly improves query performance.

However, if creating a dashboard that reflects online sales, and sales are immediately updated in a centralized database, you may want to directly query that database. For example, you may create a dashboard for your marketing team that shows the top 10 products by sales over the past hour, and you may want to enable your marketing team to refresh the dashboard to get the latest data at any moment in time. In this case, you would want to use a direct query of the data source. This is because you always want the most up-to-date data, and the marketing team is relatively small, so there is not a massive load on the source database from direct queries.

You also have the option of scheduling a refresh of the data in SPICE so that QuickSight will regularly connect to the data source and retrieve

the latest data to store in SPICE. With both the Standard and Enterprise editions of QuickSight, you can schedule the refresh to be done daily, weekly, or monthly. With the Enterprise edition of QuickSight, however, you gain the additional option of performing incremental refreshes, and the ability to do an incremental refresh as often as every 15 minutes (or a full refresh every hour). You can also use an API call to trigger the refresh of SPICE data, enabling you to build an event-driven strategy for refreshing SPICE data. For more information, see the AWS blog post titled *Event-driven refresh of SPICE datasets in Amazon QuickSight* at <https://aws.amazon.com/blogs/big-data/event-driven-refresh-of-spice-datasets-in-amazon-quicksight/>.

NOTE

There is a 2-minute timeout for generating visuals in QuickSight. Therefore, if your direct query takes 2 minutes or longer to perform the query and generate the visualization, a timeout will occur. In these cases, you either need to improve the performance of the query (filtering data, only selecting specific columns, and so on) or you should import the data into SPICE. For more information on data source quotas and limitations, refer to the AWS documentation at:

<https://docs.aws.amazon.com/quicksight/latest/user/data-source-limits.html>.

If you're using a data source that charges for each query (such as Amazon Athena in on-demand mode, or Amazon Redshift Spectrum), importing the data into SPICE can help reduce costs. Storing the data in SPICE means you only pay for the query when the data is initially loaded, as well as for when the data is refreshed. With a direct query, you would pay for the query each time the visualization is accessed.

Managing SPICE capacity

Your account is granted 10 GB of SPICE storage for every paid user that has the author role (this would be every user in the Standard Edition, and users with the Author role in the Enterprise edition). SPICE storage is shared by all QuickSight users in an account and is on a per-region basis.

For example, if you have QuickSight Enterprise edition and you have 10 users with the Author role and 100 users with the Reader role, all in the Northern Virginia (us-east-1) Region, then your QuickSight account in us-east-1 would have 100 GB of SPICE storage available.

If additional SPICE storage is needed, you can purchase additional SPICE capacity. For example, if you needed 130 GB of total SPICE storage for the datasets you wanted to import, you could purchase an additional 30 GB of capacity each month. At the time of writing, additional SPICE capacity for the Enterprise edition is charged at \$0.38 per GB.

There are also limits on the size of a single dataset in SPICE. At the time of writing, datasets are limited to a maximum of 1 billion rows, or 1 TB, for QuickSight Enterprise edition. For the Standard edition, the limit is 25 million rows, or 25 GB of data. There are also other limits for each dataset (such as the number of columns and the length of column names), so ensure you refer to the latest QuickSight documentation for updated information on these limits (see

<https://docs.aws.amazon.com/quicksight/latest/user/data-source-limits.html>.

Now that we have reviewed the core Amazon QuickSight concepts, let's move on and review QuickSight's functionality for importing and preparing data.

Ingesting and preparing data from a variety of sources

Amazon QuickSight can use other AWS services as a source, as well as on-premises databases, imported files, and even some **Software as a Service (SaaS)** applications.

For example, you can easily connect to Oracle, Microsoft SQL Server, Postgres, and MySQL databases, either running as part of the Amazon RDS managed database service, or as instances running on Amazon EC2 or in your own data centers. You can also connect to data warehouse systems such as Amazon Redshift, Snowflake, and Teradata. Other AWS services are also supported as data sources, including Amazon S3, Amazon Athena, Amazon OpenSearch Service, Amazon Aurora, and AWS IoT Analytics.

In addition to these traditional data sources, QuickSight can also connect to various SaaS offerings, including ServiceNow, Jira, Adobe Analytics, Salesforce, GitHub, and Twitter.

Data stored in files, such as a Microsoft Excel Spreadsheet (XLSX files), JSON documents, and CSV files, can also be imported into QuickSight. These files can be directly uploaded through the QuickSight console, or they can be imported from Amazon S3.

The rich variety of potential data sources for QuickSight is shown in the following screenshot:

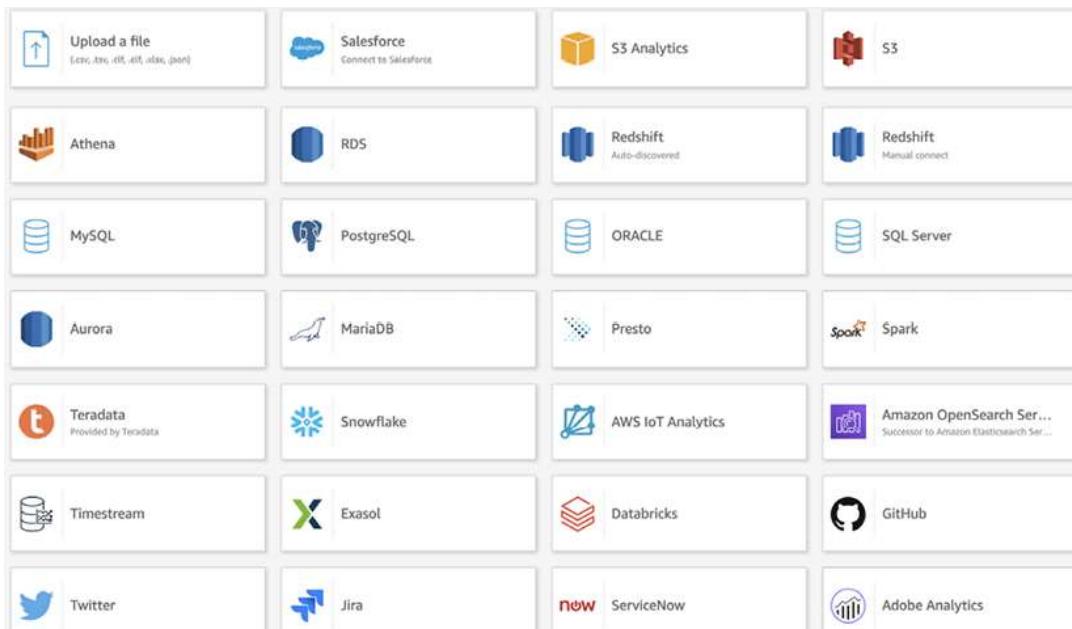


Figure 12.4: Data sources that can be imported into Amazon QuickSight

For the latest list of supported data sources, and to confirm information about supported versions, see the AWS QuickSight documentation topic *Supported Data Sources* at

<https://docs.aws.amazon.com/quicksight/latest/user/supported-data-sources.html>.

For data sources not directly supported, you can use other ingestion methods (such as those discussed in *Chapter 6, Ingesting Batch and Streaming Data*) to ingest data into your S3-based data lake. You can then create visualizations of that data by using the Amazon Athena data source integration to enable QuickSight to query the data directly in Amazon S3.

To learn more about managing your SPICE memory capacity, refer to the AWS documentation at:

<https://docs.aws.amazon.com/quicksight/latest/user/managing-spice-capacity.html>.

Once your data has been imported, you can use QuickSight to do some level of ETL on the datasets, which we will look at in the next section.

Preparing datasets in QuickSight versus performing ETL outside of QuickSight

QuickSight includes functionality for performing data transformations on imported data. For example, you can do the following:

- Join two different datasets
- Exclude specific fields
- Filter data
- Change the data type or name of a field
- Create a new calculated field

All of these data preparation tasks can be done using a simple visual interface.

If you select to join two different datasets, then you need to import the data into SPICE. However, if you're just working with a single data source, the transformations you specify will be applied when the data is read from the data source.

Ultimately, you need to decide whether you should perform data transformations and joins in QuickSight, or whether you should perform those transformations outside of QuickSight. For example, you could join two datasets, drop unneeded columns, change the data types and column names, and create new calculated fields using tools such as AWS Glue DataBrew or AWS Glue Studio.

There are several factors to consider when making this decision, including the following:

- If this dataset may be used outside of QuickSight, such as for queries using Amazon Athena, then it makes sense to perform the ETL with other tools before using the dataset in QuickSight.
- If the required transformations are relatively simple and the resulting dataset will only be used in QuickSight, then you may

choose to perform the transformation using QuickSight. This could include transforms such as adding additional calculated fields, changing the names or data types of a few columns, dropping a few columns, and so on.

The decision about where to perform data transformations can be complex, and it may not be an easy decision. However, an important factor to take into account is the controls that may be in place for formal data pipelines, versus those for more informal transformations (such as those performed by data analysts using tools such as Amazon QuickSight).

If you have strong governance controls around your formal data engineering pipelines (such as code reviews and change control), then you may choose to ensure that all the transformations are done within formal processes. However, you need to balance this against ensuring that you don't tie up your end user teams in formal processes that slow the business down.

Often, you need to balance the two sides – ensuring that your business teams have the flexibility to perform minor transformations using tools such as QuickSight, while also ensuring that new datasets or visualizations that business users may use to make important business decisions have the correct governance controls around them.

It is not always easy to find this balance, and there are no specific rules that apply universally when making this decision. Therefore, much thought needs to be given to this decision to find the right balance between enabling the business to make decisions quickly, without being constrained by overly formal processes for even minor data transformations.

The business ultimately needs to take the time required to put in place governance and controls that communicate the types of ad hoc data transformations that data analysts and others can perform. These poli-

cies should also make it clear as to when transformations need to be performed by data engineering teams using formal processes.

For more information on the types of transforms you can do in QuickSight, see the QuickSight documentation titled *Preparing data in Amazon QuickSight* at <https://docs.aws.amazon.com/quicksight/latest/user/preparing-data.html>.

Once you have your data ready in QuickSight (whether you did the transforms inside or outside of QuickSight), you are ready to start building your dashboards and reports, as covered in the next section.

Creating and sharing visuals with QuickSight analyses and dashboards

Once a dataset has been imported (and optionally transformed), you can create visualizations of this data using **QuickSight analyses**. This is the tool that is used by QuickSight authors to create new dashboards, with these dashboards containing one or more visualizations that can be shared with others in the business.

When you create a new analysis/dashboard, you choose one or more datasets to include in the analysis (up to a maximum of 50 datasets per dashboard). Each analysis consists of one or more sheets (or tabs, much like browser tabs) that display a group of visualizations. You can have up to 20 sheets (tabs) per dashboard, and each sheet can have up to 30 visualizations.

Once you have created an analysis (consisting of multiple visuals, optionally across multiple sheets), you can choose to publish the analysis as a dashboard. When you're publishing a dashboard, you can select various parameters related to how readers can interact with the dashboard, including the following:

- If they can apply their own ad-hoc filters to the data in the dashboard
- If they can download data in the dashboard as a CSV file
- If they can perform drill-down and drill-up actions (when supported in a dashboard)
- If they can sort the data

Once the dashboard has been published, you can select who to share the dashboard with. You can either share the dashboard with everyone in the account (providing them with read access to the dashboard) or you can select specific users and groups to share with.

By default, when you create a new analysis, the analysis contains a single sheet, with a single empty visualization that is set to a type of **AutoGraph**, as seen in the following screenshot.

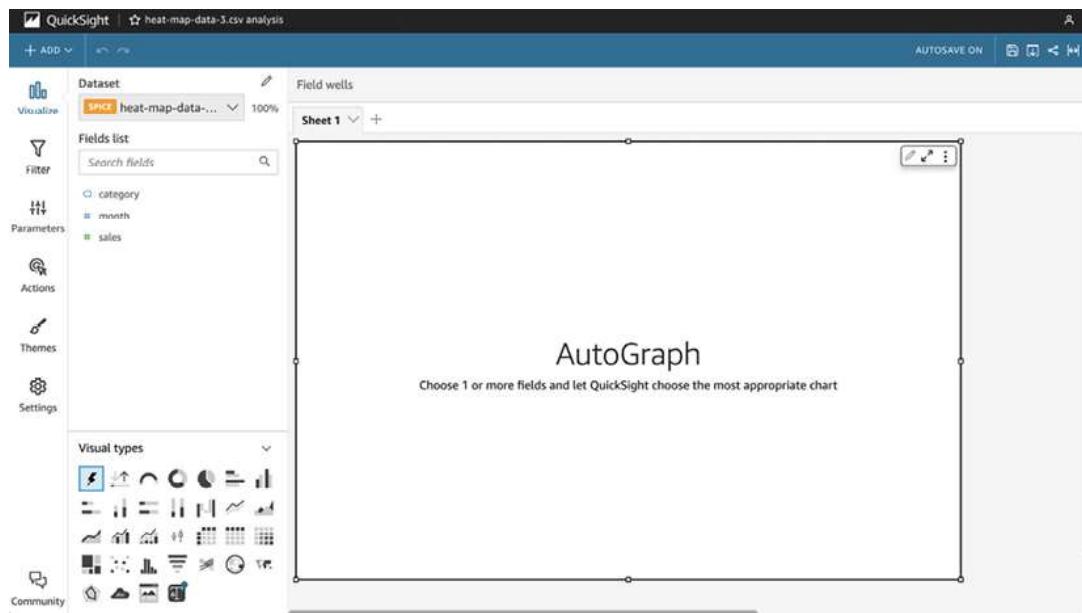


Figure 12.5: New analysis screen in Amazon QuickSight

QuickSight supports many different types of visualizations (as can be seen in the **Visual types** section of the preceding screenshot). Let's dive deeper into some of these visual types.

Visual types in Amazon QuickSight

In this section, we will discuss several **data visualization types** supported by Amazon QuickSight. There are many different types of visualizations that are supported, and in this section, we will cover some of the most popular ones, but you can review the full list of visualizations in the *Amazon QuickSight documentation*

(<https://docs.aws.amazon.com/quicksight/latest/user/working-with-visual-types.html>).

AutoGraph for automatic graphing

While this is not an actual type of visual, you can select AutoGraph as a visual type to let QuickSight automatically choose the visual type for you. Based on the number of fields you select, and the data type of each field that is selected, QuickSight automatically uses the most appropriate visual type for your data. This is often a good way to start exploring your data if you're unsure of the specific type of graph you want to use.

Line, geospatial, and heat maps

Earlier in this chapter, we discussed three common types of visualizations:

- **Line charts:** Displays data as a series of data points and is often used to plot data over a certain period
- **Geospatial charts:** Displays data points overlayed on a map, combining geospatial data with other data
- **Heat maps:** Displays data in a chart with values represented by darker or lighter colors

All three of these types of charts (and variations of these charts) are supported by Amazon QuickSight, and can be used to create rich visualizations from many different data sources.

Bar charts

Bar charts are a common visualization type, and QuickSight supports multiple types of bar charts. For example, you can have a simple bar chart showing a single value for a dimension (such as *sales per region*) or a multi-measure bar chart that shows multiple measures for a dimension (such as *sales goal* and *achieved sales per region*).

There are also additional bar chart types that are supported, such as stacked bar charts and clustered bar charts. Bar charts can be displayed horizontally or vertically.

Key performance indicators

A **Key performance indicator (KPI)** is often used to show progress against a specific goal. For example, you may have a goal of achieving a specific amount of revenue in a quarter.

A KPI visual could display the current revenue as a percentage of the target revenue in a visual. A dashboard showing this KPI (or multiple KPIs) can help management keep track of how the business is performing based on several key metrics.

In QuickSight, a KPI displays a comparison of two values and includes a progress bar indicating the percentage difference between the values:

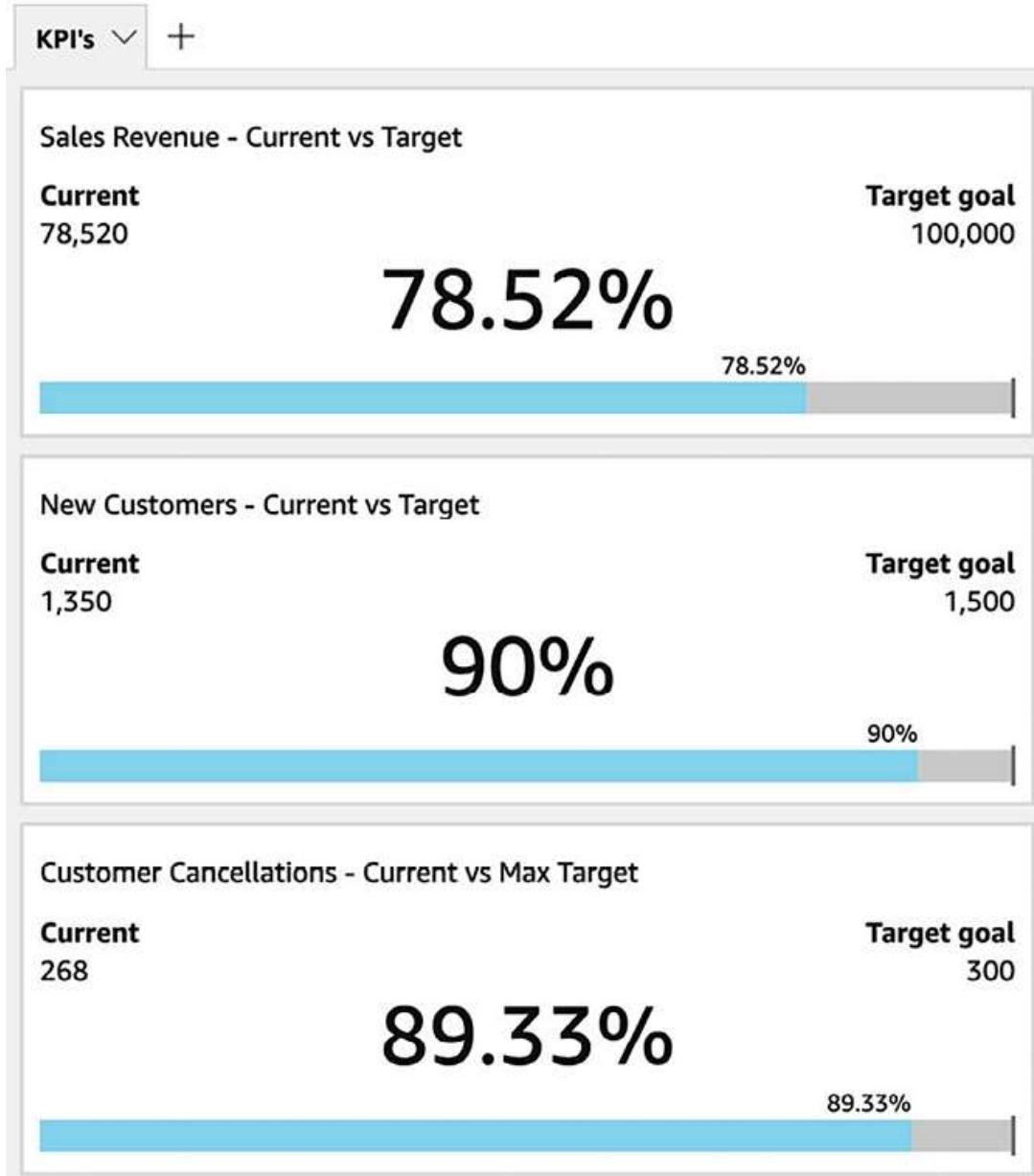


Figure 12.6: Dashboard with KPI visuals

In the preceding screenshot, a sales manager can quickly view how their organization is performing against several key metrics. This chart shows that revenue is nearly at 80% of the target, new customers are at 90% of the target, and that the team is within 11% of the target maximum customer cancellations for that period.

Tables as visuals

There may be use cases where you want to display the raw data of a table on a dashboard, without converting the data into a specific

visual.

QuickSight supports displaying tables directly within an analysis/dashboard and supports up to 200 columns in the visual. However, directly displaying raw table data should ideally only be done with small tables, where you display just a limited amount of raw data.

Custom visual types

QuickSight lets you include several **custom visuals** within a dashboard, including the following:

- Custom images (such as a company or product logo)
- Custom videos
- An online form
- An embedded web page

These visual types help you customize and personalize your dashboards. For example, you may want to embed your company logo on a visual, or include a video that provides a guide for working with a specific dashboard.

Note that when you embed custom content in an analysis/dashboard, you need to specify the HTTP URL of the resource. Also, while QuickSight does include functionality for emailing dashboards to users, embedded custom visual types (pictures, videos, forms, and web pages) will not be displayed in the email copy of a dashboard.

There are also other limitations to using embedded content. For example, the web content needs to support opening the content in an iframe; otherwise, the content may not appear in QuickSight. When you're looking to embed content into a QuickSight analysis/dashboard, you should look for content that has an embeddable URL (which is often available when you choose to share content).

Other visual types

There are many other types of charts that are supported in QuickSight, and new types are added over time. These include the following common chart types:

- Pie charts
- Box plots
- Donut charts
- Gauge charts
- Histograms
- Pivot tables
- Sankey diagrams
- Tree maps
- Waterfall charts
- Word clouds

As we have discussed in this section, QuickSight lets us create many different types of visuals, then publish and share those visuals as dashboards. However, QuickSight also includes advanced functionality that can automatically reveal new insights in your data and lets you embed dashboards into custom applications, as we will see in the next section.

Understanding QuickSight's advanced features

The **Enterprise edition** of Amazon QuickSight includes advanced features that can help you draw out additional insights from your data, ask questions of your data using natural language, and enable you to widely share your data by embedding dashboards into applications. We will review some of these features next.

Amazon QuickSight ML Insights

QuickSight ML Insights uses the power of ML algorithms to automatically uncover insights and trends, forecast future data points, and identify anomalies in your data.

All of these ML Insights functionalities can easily be added to an analysis/dashboard without the author needing to have any ML experience or any real understanding of the underlying ML algorithms. However, for those who are interested in the underlying ML algorithms used by QuickSight, Amazon provides comprehensive documentation on this topic.

Review the Amazon QuickSight documentation titled *Understanding the ML algorithm used by Amazon QuickSight* for more information: <https://docs.aws.amazon.com/quicksight/latest/user/concept-of-ml-algorithms.html>.

However, to make use of ML Insights, there are specific requirements for your data, such as having at least one metric and one category dimension. For ML forecasting, the more historical data you have, the better. For example, if you want to forecast based on daily data, you need at least 38 daily data points, or to forecast on quarters, you need at least 35 quarterly data points. The full details on the data requirements are documented in the Amazon QuickSight documentation titled *Dataset requirements for using ML Insights with Amazon QuickSight*:

<https://docs.aws.amazon.com/quicksight/latest/user/ml-data-set-requirements.html>.

Let's examine some of the different types of ML Insights in more detail.

Amazon QuickSight autonarratives

Autonarratives provide natural language insights into your data, providing you with an easy-to-read summary of what is displayed in a vis-

ual. Effectively, autonarratives enable you to provide a plainly stated summary of your data, as the following autonarrative examples show:

- Year-to-date revenue *decreased* by 4.6% from \$906,123 to \$864,441 compared to the same period last year. We are at 89.3% achievement for the YTD goal and 77.9% achievement for the annual goal.
- Daily revenue for *Accessories / Cell Phone Covers* on September 3, 2021 was higher than expected at \$3,461.21.

You can add a variety of autonarratives to an analysis, such as bottom-ranked items, growth rate, anomaly detection, top movers, and many others. For the full list of available autonarratives, see the Amazon QuickSight documentation titled *Insights that include autonarratives*:

<https://docs.aws.amazon.com/quicksight/latest/user/auto-narratives.html>.

ML-powered anomaly detection

Amazon QuickSight can perform **anomaly detection** (also sometimes referred to as outlier detection) across millions of metrics contained in your data, identify non-obvious trends, and highlight outliers in the data. These types of insights are difficult to draw out of data without using the power of modern ML algorithms.

You can add an autonarrative widget to an analysis and specify the type as being anomalous. Then, you can configure several settings related to how QuickSight detects outliers in the data and can set a schedule for when outliers are calculated (ranging from once an hour to once a month). You can also configure QuickSight to analyze the top items that contributed to the anomaly.

Once an anomaly has been detected, you can choose to **explore the anomalies** on the insight. This opens a screen where you can change various settings related to anomaly detection, enabling you to explore different types of anomalies in the dataset.

ML-powered forecasting

Amazon QuickSight can use the power of ML algorithms to provide reliable forecasts against your data. When you create a visual that uses a date field and contains up to three metrics, you can select an option in the widget to add a forecast of future values.

QuickSight will automatically analyze historical data using an ML model and graph out future predicted values for each metric. You can also configure the forecast properties by setting items such as forecast length (how many future periods to forecast and how much historical data to analyze).

The ML model that's used by QuickSight for forecasting automatically excludes data that it identifies as outliers and automatically fills in any missing values. For example, if you had a short spike in sales due to a promotion, QuickSight could exclude that spike when calculating the forecast. Or, if there were a few days where historical data was missing, QuickSight could automatically determine likely values for the missing period.

It is important to remember that the QuickSight ML Insight features (including autonarratives, anomaly detection, and forecasting) are available in the Enterprise edition of QuickSight only, and will not be available if you only have a Standard edition subscription.

In this section, we looked at how QuickSight enables you to draw out powerful new insights from your data using ML. In the next section, we will look at another advanced feature available in QuickSight, and that is the ability to ask questions of your data using natural language.

Amazon QuickSight Q for natural language queries

ML approaches have enabled new abilities in many products, and one of those is the ability to query data with free-form text that uses everyday language. A number of popular BI products offer functionality for **natural language queries (NLQ)**, and this includes Amazon QuickSight.

With the Amazon QuickSight Q feature, a user could type in a query such as “*Show me the top 3 product categories by revenue for 2023*”, and QuickSight would create a graph displaying that information. However, a QuickSight author does first need to do some work to set up topics for QuickSight Q in order to enable the functionality.

QuickSight Q is a feature that is only available with the QuickSight Enterprise edition, and there is an additional cost to enable this functionality. At the time of writing, there is a \$250 charge per month for each account that has QuickSight Q enabled, and authors enabled to use Q have an extra charge of \$10/month, and the reader maximum cost increases from \$5 to \$10 for readers with Q.

Generative BI dashboarding authoring capabilities

In September 2023, AWS announced the preview of new generative BI authoring capabilities as part of Amazon QuickSight Q. The three new capabilities in preview that were part of the announcement provide new functionality for the following:

- Building visualizations by specifying what you want to see using natural language
- Creating complex calculations rapidly, by specifying the expected outcome of the calculation using natural language
- Refining and tweaking visualizations using natural language prompts

These new capabilities are powered by **Amazon Bedrock**, a service that offers a variety of **Large Language Models (LLMs)**, suited for different tasks. For more information on this new functionality, see the QuickSight documentation at <https://docs.aws.amazon.com/quicksight/latest/user/generative-bi-author-experience.html>.

Let's now take a deeper look at **QuickSight Q Topics**, the key concept employed to use QuickSight Q against a dataset.

QuickSight Q Topics

Q Topics are a collection of one (or more) datasets that are used to represent a business subject area that users can ask questions about. Q Topics need to be created by a QuickSight author, and the author then needs to configure the topic to best enable natural language queries.

The easiest way to create and configure a topic, is to create a topic based on an existing analysis/dashboard. When you create a topic using this method, **QuickSight's Automated Data Preparation** automatically configures the topic using the power of ML, including automation of the following:

- **Selection of fields:** With Automated Data Preparation, QuickSight examines the existing analysis to understand which fields are most commonly used in your dashboard, and includes those fields in the topic, and excludes fields that are not commonly used. In addition, it also automatically includes any calculated fields that you created in the analysis.
- **Naming of fields:** It is common to use acronyms and shortened versions of field names in a database or data warehouse, such as naming a column *DOB* for *Date Of Birth*. With Automated Data Preparation, QuickSight can attempt to automatically rename fields to a more natural name that someone is likely to use in a query. In addition, it can automatically add additional synonyms for each

field, such as *Birth Date* for the *Date of Birth* field, or *salesperson* and *account representative* for the *SALES_REP* field.

- **Field formatting:** With Automated Data Preparation, QuickSight can automatically determine an appropriate format for a field when using that field in an answer. For example, with a field that shows *sales revenue* value, QuickSight can automatically assign a currency format to that field so that results using this field display the value with a dollar prefix, for example.

When using Automated Data Preparation, authors are strongly encouraged to review the automated configuration that QuickSight applies and edit it further as needed. For example, based on the knowledge that an author has of the business, they may choose to add additional synonyms for certain fields.

If an author creates a new topic directly (i.e., by not creating a topic based on an existing dashboard), then the author needs to perform the above tasks manually to select fields, rename fields, add synonyms, and apply a format to relevant fields.

Once a topic has been created, and your users start querying using natural language prompts, it is important to review how the topic is performing, so that adjustments can be made as needed. Let's take a closer look at how to fine-tune your topics.

Fine-tuning your QuickSight Q Topics

QuickSight Q includes functionality that lets you review the performance of the topics you have created, so that you can fine-tune the topic to further improve performance.

Once your topic has been available to your users for a while, you can review the topic to see information such as the queries that your users have run, whether QuickSight Q was able to provide a result, where users needed to provide more information, etc. This enables you to

fine-tune the field names, synonyms, and field formats, in order to improve the results that QuickSight Q provides.

For example, you may find that users typed in a query and no result was available, but you may find that they used a specific term (such as *rep* for *sales representative*) that had not been configured as a synonym, and you can then add a new synonym so that future similar queries will return an appropriate result.

It is important that you regularly review your topics, to understand how users are interacting with the topic, so that you can refine the topic settings to improve results. When a user provides feedback on how well Q responded to a question, this is recorded and available on the topic's **Summary** and **User Activity** tabs.

In the **Summary** tab, you can view metrics such as the number of questions asked over time, as well as a distribution of questions that received positive, negative, or no feedback. In the **User activity** tab, you can see a list of questions that were asked, as well as positive or negative feedback and comments. For more details on monitoring topic feedback, see the Amazon QuickSight documentation at <https://docs.aws.amazon.com/quicksight/latest/user/quicksight-q-topics-performance.html>.

In the next section, we will look at another popular feature of the Enterprise edition of QuickSight, a feature that enables you to easily distribute your published dashboards more widely.

Amazon QuickSight embedded dashboards

For use cases where you don't want your users to have to log in to QuickSight via the AWS Management Console or QuickSight portal, you can embed QuickSight directly into your applications or website.

You can embed either the full console experience (including authoring tools for creating new analyses and managing datasets) or embed published dashboards only. Embedded dashboards have the full interactive capabilities that they do in the console, which means that users can filter and sort data, and even drill down into data (so long as the author enabled those levels of interactivity when they published the dashboard).

Embedding for registered QuickSight users

QuickSight supports several authentication methods, including AD SAML 2.0, as well as SSO using AWS Identity Center (or other identity providers such as Okta, Auth0, and PingOne).

As such, your users can authenticate with your existing website or HTML-based application using one of the supported authentication methods and, using that identity, map to an existing QuickSight user. If that user has not accessed QuickSight before, a new QuickSight user will be created for the user.

You can elect to either embed the full console experience or only embed dashboards. Users will be able to open any dashboards that their QuickSight user has been given access to.

With the QuickSight embedding experience, you can optionally customize the display theme using your branding. This enables the embedded QuickSight objects to appear as a direct part of your application, rather than looking like an embedded external application. However, even when you have a customized theme, the embedded QuickSight application does display a **Powered by QuickSight** label.

Embedding for unauthenticated users

For use cases where your users do not authenticate with your website or application, you still have the option of embedding QuickSight dashboards for anonymous user access.

To enable anonymous access, you need to purchase reader session capacity pricing. This offers a set number of QuickSight sessions per month, or per year (depending on your plan), and these sessions can be consumed by anonymous users. The bonus of purchasing an annual plan for QuickSight sessions is that the **Powered by QuickSight** label can be removed from embedded resources.

An example use case for this functionality is for a local government health department that wants to share the latest information on a virus outbreak with their community. The health department could embed an Amazon QuickSight dashboard into its website that is linked to the latest data on the spread of the virus.

Users accessing the website could interact with the dashboard, filtering data for their specific location, sorting data, or even downloading a CSV version of the data for their additional analysis. These users would not need to log into the health department website to access the dashboard, and the health department could use an annual plan for reader session capacity. For more information on pricing for reader session plans, see the Amazon QuickSight pricing page:

<https://aws.amazon.com/quicksight/pricing/>

When embedding dashboards for unauthenticated users, you need to be very aware of what data you make available, and that you do not accidentally expose company confidential data via an authenticated embedding.

Let's now look at another advanced feature in Amazon QuickSight: the ability to generate multi-page formatted reports.

Generating multi-page formatted reports

The Amazon QuickSight Paginated Reports functionality (launched in November 2022) enables authors to create highly formatted, multi-page PDF reports. This extends Amazon QuickSight functionality beyond just creating rich visualizations in dashboards, to enabling report generation as well.

With this functionality, QuickSight authors are able to create reports where they specify attributes such as page size, length, orientation (portrait or landscape), as well as the arrangement of images, charts, and tables. These reports can easily be printed, or distributed via email as PDF attachments. A report created with this functionality can generate up to 1,000 pages in a PDF, and reports can be scheduled to run using QuickSight's scheduling mechanism.

Note that there is an extra cost to enable paginated reports in QuickSight. At the time of writing, the cost for 500 reports units a month (with a report unit being 100 pages long, or 100 MB in size) was \$500 per month. To learn more about paginated reports, see the QuickSight documentation titled *Working with paginated reports in Amazon QuickSight* at

<https://docs.aws.amazon.com/quicksight/latest/user/working-with-reports.html>.

Having learned more about QuickSight's advanced functionality, let's get hands-on by creating a QuickSight visualization.

Hands-on – creating a simple QuickSight visualization

Earlier in this chapter, we discussed how data can be represented over a geographic area. We used the example of data containing information on the population of world cities, and how we could use that to easily visualize how large cities are geographically distributed. The example visual in *Figure 12.2* showed cities with a population of over 5 million people, displayed on top of a map of the world.

For the hands-on section of this chapter, we are going to recreate that visual using Amazon QuickSight.

Setting up a new QuickSight account and loading a dataset

Before we start creating a new dashboard, we need to download a sample dataset of world city populations. We will use the basic dataset available from <https://simplemaps.com/>, which is freely distributed under the *Creative Commons Attribution 4.0* license (<https://creativecommons.org/licenses/by/4.0/>):

1. Use the following link to download the **basic** dataset from simplemaps.com: <https://simplemaps.com/data/world-cities>. If the file downloaded is a ZIP file, make sure to extract the actual city data CSV file.
2. Log into the **AWS Management Console** and use the top search bar to search for, and open, the **QuickSight** service.
3. If you have not used QuickSight before in this account, you will be prompted with a **Sign up for QuickSight** button. Click the button to start the signup process.
4. The default page opens to the QuickSight Enterprise edition. For this exercise, only the Standard edition is needed, so click on the **Sign up for Standard Edition here** link at the very bottom of the page, as per the screenshot in *Figure 12.7*

Edition	<input checked="" type="radio"/> Enterprise	<input type="radio"/> Enterprise + Q Learn more
Team trial for 30 days (4 authors)*	FREE	FREE
Author per month (yearly)**	\$18	\$28
Author per month (monthly)**	\$24	\$34
Readers (pay-per-Session)	\$0.30 / session (max \$5)****	\$0.30 / session (max \$10)****
Additional SPICE per month	\$0.38 per GB	\$0.38 per GB
QuickSight Q regional fee	N/A	\$250 / mo / region
Natural language query with QuickSight Q	N/A	INCLUDED
Single Sign On with SAML or OpenID Connect	✓	✓
Connect to spreadsheets, databases & business apps	✓	✓
Access data in Private VPCs	✓	✓
Row-level security for dashboards	✓	✓
Secure data encryption at rest	✓	✓
Connect to your Active Directory	✓	✓
Use Active Directory groups***	✓	✓
Send email reports	✓	✓
Embed QuickSight	✓	✓
Capacity-based pricing	✓	✓
Supported regions	Learn more	Learn more

* Trial authors are auto-converted to month-to-month subscription upon trial expiry

** Each additional author includes 10GB of SPICE capacity

*** Active Directory groups are available in accounts connected to Active Directory

**** Sessions of 30-minute duration. Total charges for each reader are capped at \$5 per month. [Conditions apply.](#)

QuickSight Standard Sign-up

[Sign up for Standard Edition here.](#)



Figure 12.7: Setting up a new QuickSight account

- For **Authentication method**, select **Use IAM federated identities only**, and then select your preferred **AWS Region**. Under **Account info**, provide a unique name for your **QuickSight account** (such as `data-engineering-<initials>`) along with a **Notification email address** that can be used to send QuickSight notifications to you. Leave all other settings as-is and click **Finish**:

Create your QuickSight account

Standard

[Back](#)

Authentication method

Use IAM federated identities & QuickSight-managed users

Authenticate with single sign-on (SAML or OpenID Connect), AWS IAM credentials, or QuickSight credentials

Use IAM federated identities only

Authenticate with single sign-on (SAML or OpenID Connect) or AWS IAM credentials

QuickSight region

Select a region



US East (Ohio)



Account info

QuickSight account name

You will need this for you and others to sign in



data-i

Notification email address

For QuickSight to send important notifications

gare

.l.com

Figure 12.8: Configuring a new QuickSight account

6. After a while, you should receive a pop-up message confirming that you have signed up for Amazon QuickSight. Click on the **Go to Amazon QuickSight link**, and then click through the welcome screens, which provide an overview of Amazon QuickSight's functionality.
7. From the left-hand side menu, click on **Datasets** to go to the dataset management screen. On this screen, you will see several pre-loaded sample datasets:

Name	Owner	Last Modified	⋮
Sales Pipeline	SPICE	Me	a few seconds ago
Web and Social Media Analytics	SPICE	Me	a few seconds ago
Business Review	SPICE	Me	a few seconds ago
People Overview	SPICE	Me	a few seconds ago

Figure 12.9: Pre-loaded datasets for a new QuickSight account

8. Click on **New dataset** to create a new dataset. On the new dataset screen, click on **Upload a file**.
9. When you're prompted to provide the file to upload, navigate to where you downloaded the World Cities data from *simplymaps.com* (in Step 1 of this exercise) and upload the `worldcities.csv` file.
10. Once the file has been uploaded, you will be presented with a popup to confirm the file upload settings. Click on **Next**.
11. On the next screen, click on **Visualize**. This will open a new analyses screen where you can create your analysis/dashboard based on the World Cities dataset.

Now that we have subscribed to QuickSight, downloaded our World Cities dataset, and uploaded the dataset to QuickSight, we are ready to create our first visual.

Creating a new analysis

We are now on the analysis authoring page for QuickSight. Using this interface, we can build out new analyses consisting of multiple visualizations and, optionally, containing multiple sheets (tabs). Then, we can publish our analysis as a dashboard that can be consumed by QuickSight readers.

Initially, we will receive a pop-up dialog that enables us to select a layout (with a default of **Tiled**), and to select a resolution to optimize the display. Leave the defaults and click on **Create**.

The following screenshot shows the analysis workspace after importing our `worldcities.csv` dataset and accepting the defaults:

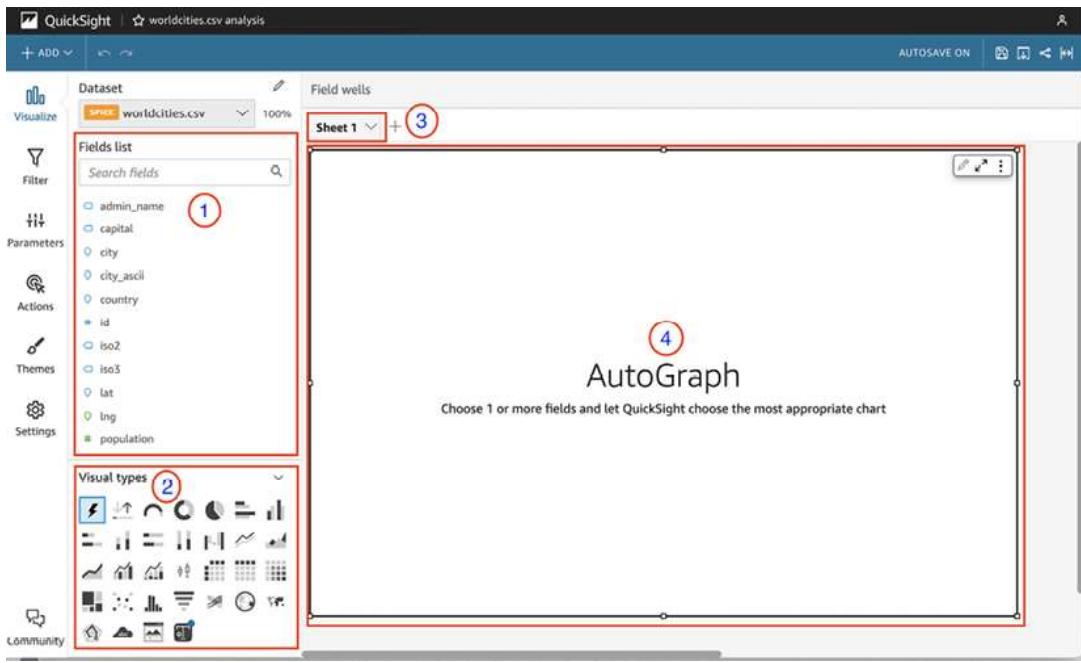


Figure 12.10: The different parts of a new QuickSight analysis

In this screenshot, we can see the following components of the analysis workspace. Note that the numbers in the following list correspond to the component numbers shown in the preceding screenshot:

- A list of fields in our selected dataset (`worldcities.csv`).
- A list of different types of charts that we can use in our visuals (bar, pie, heat map, and so on).
- The sheet bar, which shows us our current sheet (**Sheet 1**). Clicking the + sign would enable us to create additional sheets (much like tabs in a browser). We can also rename sheets.
- The visual display area. Once we select a chart type and add some fields to the visual, the chart will be displayed here. Notice that the size of the visual area can be dragged to be larger or smaller, and we can click on + **Add** in the top menu bar if we want to add additional visuals to this sheet.

To create our map of the world showing cities with populations greater than 3 million people, perform the following steps:

1. Click on the **AutoGraph** box, and then in the **Visual types** box, find and select the **Points on map** visual type.
2. From **Fields list**, drag **lat** into the **Geospatial** field well (at the top of the visual-designer workspace), and then drag **lng** into the same **Geospatial** field well. Make sure that you drag **lng** either above or below **lat**; otherwise, you will end up replacing the existing **lat** field.
3. Drag **population** into the **Size** field well and drag **city** into the **Color** field well.

Your visual designer should look as follows at this point:



Figure 12.11: Creating a new Points on Map visual

At this point, our visual is displaying population data for all 41,000 cities in the dataset. However, for our use case, we only want to display data for cities that have a population of above 3 million people. Perform the following steps to filter the data to just cities with a population above a certain size.

4. From the left-hand side QuickSight menu, click on **Filter**, and then click **Add Filter** (as shown in the following screenshot):

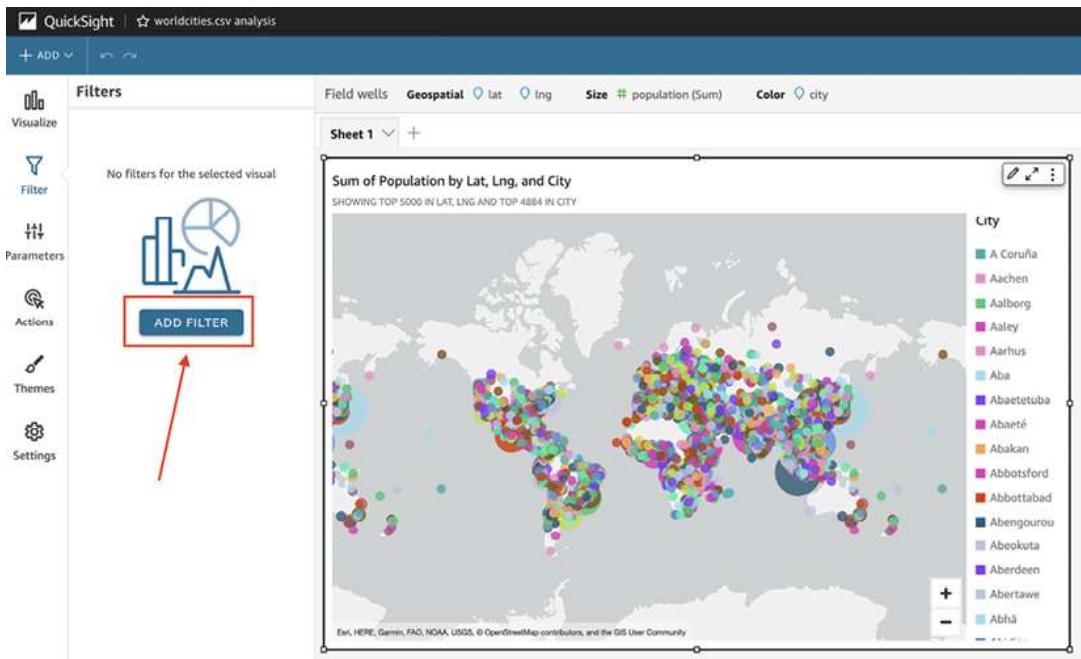


Figure 12.12: Configuring a filter for a visual

5. In the pop-up that shows the list of fields, click on the **population** field. This displays a filters list with **population** showing as the only filter.
6. From the filters list, click on **population**. Change the **Equals** drop-down to **Greater than or equal to** and enter a value of **3000000** (3 million), as shown in the following screenshot. Then, click on **APPLY**:

< Edit filter

X

Applied to

Only this visual

population



:

Equals - none

Aggregation

Sum

Filter condition

Greater than or equal to

Use parameters

Minimum value

3000000

Null options

Include nulls

OR

ADD FILTER CONDITION ▾

Note: There are limitations on how you
can group filters.

[Learn more](#)

APPLY

DELETE FILTER

Figure 12.13: Editing the filter for a visual

Our visual now displays only those cities that have a population of 3 million people or more. Note how you can position the mouse

over a city to get a popup of the city's name, along with its latitude, longitude, and population details.

You can also modify the following aspects of the visual:

1. Drag the corners of the visual to increase the size of the visual.
 2. Experiment with the visual by changing the filter on population size (for example, change the filter to 5 million people).
 3. Zoom in and out on the map to size it to display just the parts of the map you want to show.
-
7. Double-click on the title of the visual to change the title (to something such as Cities with a population of over 5 million people).
 8. Click on the **Pencil Icon** (at the top right of the visual) and change the **Base map to Streets**.
 9. Click the **down arrow** next to the title of the sheet (by default, Sheet 1) and rename the sheet (for example, changing the name to City Populations).

The completed visual now looks as follows:

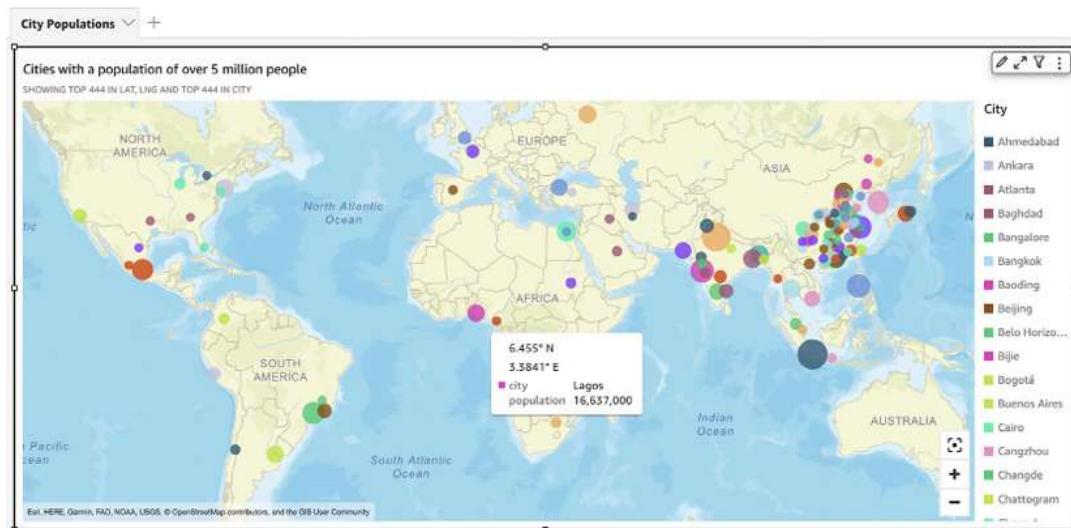


Figure 12.14: A completed visual showing cities with a population of over 5 million people

Let's now look at how to publish our new visual.

Publishing our visual as a dashboard

We can now publish our dashboard to make it available to QuickSight readers.

To publish our analysis, click on the **SHARE** icon in the top menu bar, and then select **Publish dashboard**.



Figure 12.15: The share icon on a QuickSight analysis

On the **Publish a dashboard** pop-up, enter a dashboard name into the **Publish new dashboard as** field, such as `World Cities by Population`, then click **Publish dashboard**.

Once we have published our analysis, other users in our QuickSight account can access the visual through the QuickSight dashboards tab.

In the hands-on section of this chapter, you signed up for a new QuickSight account and imported a new file-based dataset that contained information on world cities. This included geospatial data (latitude and longitude), as well as the size of the population of the city. Then, you created a new visual based on this data, filtering the data to only show cities with a population of 5 million or more people.

IMPORTANT – AVOIDING FUTURE QUICKSIGHT SUBSCRIPTION COSTS

If you do not intend to use QuickSight after the initial 30-day subscription, ensure that you unsubscribe from QuickSight to avoid future subscription charges. For more information, see the AWS documentation titled *Deleting your Amazon QuickSight subscription and closing the account* (<https://docs.aws.amazon.com/quicksight/latest/user/closing-account.html>).

Summary

In this chapter, you learned more about the Amazon QuickSight service, a **BI** tool that is used to create and share rich visualizations of data.

We discussed the power of visually representing data, and then explored core Amazon QuickSight concepts. We looked at how various data sources can be used with QuickSight, how data can optionally be imported into the SPICE storage engine, and how you can perform some data preparation tasks using QuickSight.

We then did a deeper dive into the concepts of analyses (where new visuals are authored) and dashboards (published analyses that can be shared with data consumers). As part of this, we also examined some of the common types of visualizations available in QuickSight.

We then looked at some of the advanced features available in QuickSight. This included ML Insights (which uses ML to detect outliers in data and forecast future data trends), QuickSight Q (which enables the use of natural language queries to create visualizations), as well as embedded dashboards (which enable you to embed either the full QuickSight console, or specific dashboards, directly into your websites and applications) and paginated reports.

We wrapped up this chapter with a hands-on section that took you through the steps of configuring QuickSight within your AWS account and creating and customizing a new visualization.

In the next chapter, we will do a deeper dive into some of the many AWS ML and Artificial Intelligence services that are available. We will also review how these services can be used to draw new insights and context out of existing structured and unstructured datasets.

Learn more on Discord

To join the Discord community for this book – where you can share feedback, ask questions to the author, and learn about new releases – follow the QR code below:

<https://discord.gg/9s5mHNyECd>

