**Extract Customers Sentiment in a smart way using AWS Comprehend**

Why do Sentiment Analysis?

In simple terms Sentiment Analysis is leveraging Data Analytics to assess how customers feel about a product or service. Apart from the obvious, Sentiment Analysis can give valuable insights about all the aspects of business like what the customer feels about the quality oof product, the package, the way it was delivered, the staff who delivered it, the technical support he or she received for the product, and the list goes on. It is a smart way for businesses to gain information about what went well with the product and what went wrong. It can be a powerful tool to stay ahead of competitors.

What is AWS Comprehend?

[Amazon Comprehend](https://docs.aws.amazon.com/comprehend/latest/dg/what-is.html) is a natural language processing (NLP) service provided by Amazon Web Services (AWS) that uses machine learning to uncover insights and relationships in text. Comprehend provides a number of features useful to businesses and users working with unstructured text data. It uses natural language processing (NLP) to extract insights about the content of documents. Amazon Comprehend processes any text file in UTF-8 format, image files (JPG, PNG, or TIFF), and semi-structured documents (PDF or Word files).

Business Case

Consider a business scenario where Ms. Alice is the Product Manager of XYZ Company. Ms. Alice needs to assess the opinion of customers about product PQR which they launched recently. Company XYZ has enormous data related to the customer feedback about product P but they are not labelled as positive comments, negative comments or neutral ones. So, following the traditional method of creating a Machine Learning model to predict the sentiments is not a feasible solution. Also, XYZ doesn’t have a mature Data Science team which can do the Machine Learning model creation by writing the code and developing the model. Ms. Alice needs to fetch the customer sentiments without much delay or else there is a chance of losing the customers to their competitor. What can Ms. Alice do to handle this situation in a cost-effective way?

Challenges with traditional method of predicting sentiments

Close analysis of the above situation shows the fact that Ms. Alice is facing 2 major challenges in estimating the customer sentiments towards product P.

1.No labelled data: Traditional methods of Sentiment Analysis rely on the creation of a predictive model, which consumes enormous data labelled for each class of output. The output classes of a simple Sentiment Analysis Model are Positive, Negative and Neutral. For building a predictive model there must be sufficient amount of labelled data from each class. This arises the need of going through the process of Labelling data needs dedicated human resources to do the job. Data labelling is a tedious and time-consuming process. Possibility of human error while labelling huge data cannot be eliminated.

2.Skilled Human Resource: Creating a Machine Learning model requires enormous data, and mature data analysis environment. Machine Learning Engineers put a lot of effort from data gathering, data pre-processing till model creation for creating a predictive model. Sentiment Analysis model also goes through all these phases.

What are the benefits that AWS Comprehend can offer?

1.No need of labelled data: AWS Comprehend is powered by enormous data from a wide variety of domains and so users can make Sentiment analysis on the data without going through the hassles of data labelling. Sentiment Analysis can also be done using Unsupervised Learning methods where the data points need not be labelled. But research shows that Unsupervised learning gives back a much lesser accuracy of prediction compared to the Supervised mode where data points are labelled.

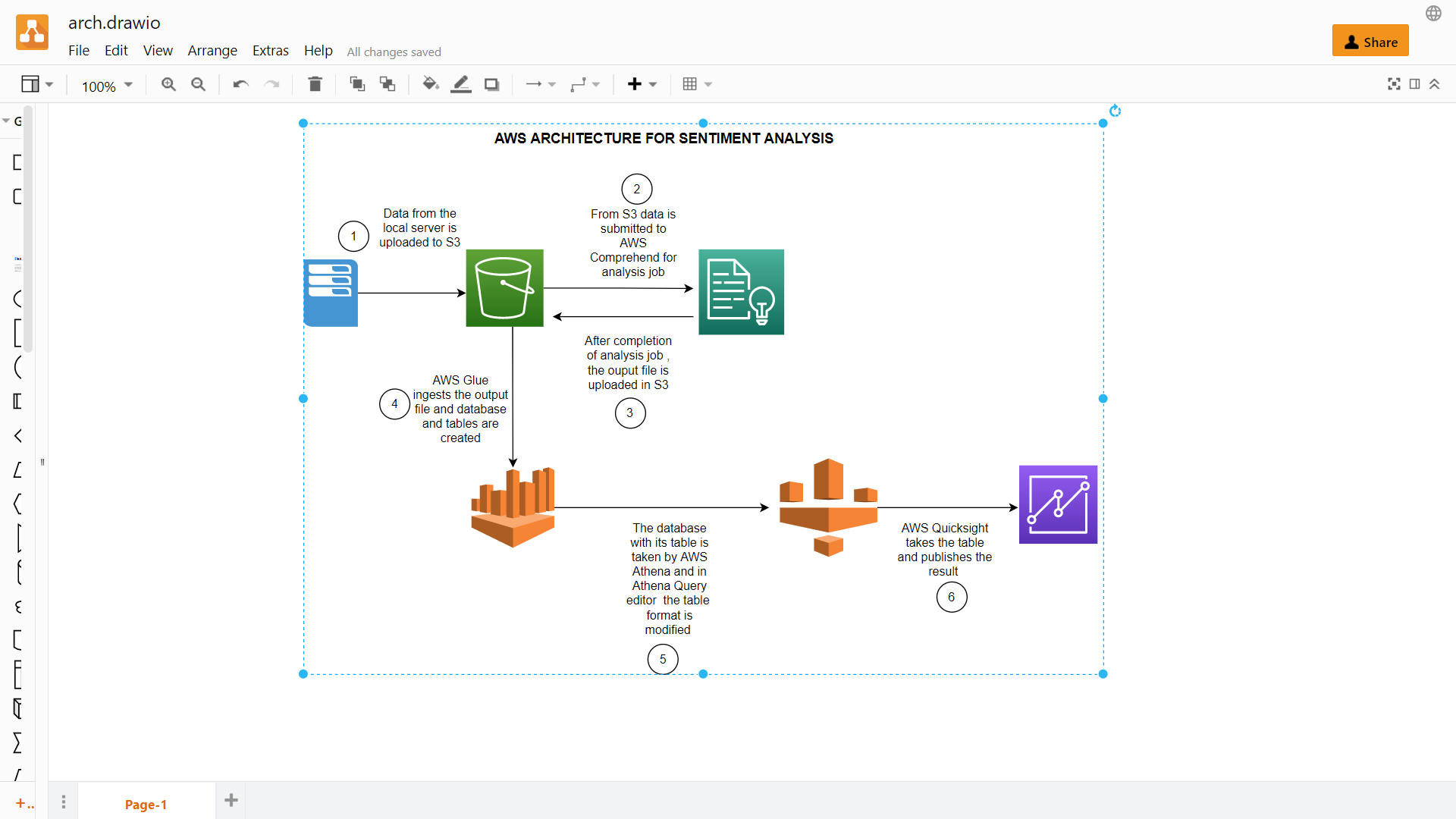
2.Quick Results. The results are available within hours as opposed to days or even month of going through all the steps of Model creation.

3. Ease of Development and low maintenance: There is minimal or almost no code to be written for all the data analysis done through AWS Comprehend. This eliminates the need for having skilled Machine Learning Engineers to do the job and thus eases development efforts and helps in low maintenance efforts.

4.High Accuracy: Compared to the training data that a company can have for model creation; AWS Comprehend is trained on enormous data and that too from different domains. So, analysis done by AWS Comprehend will be naturally of higher accuracy.

The Architecture

The diagram below depicts architecture for Sentiment Analysis on a sample dataset. The dataset is Google Playstore App review dataset and can be downloaded from [here](https://www.kaggle.com/prakharrathi25/google-play-store-reviews). The data contains over 12000 reviews of different app store applications by real users. Only the column with textual review has to be passed to AWS Comprehend for analysis.

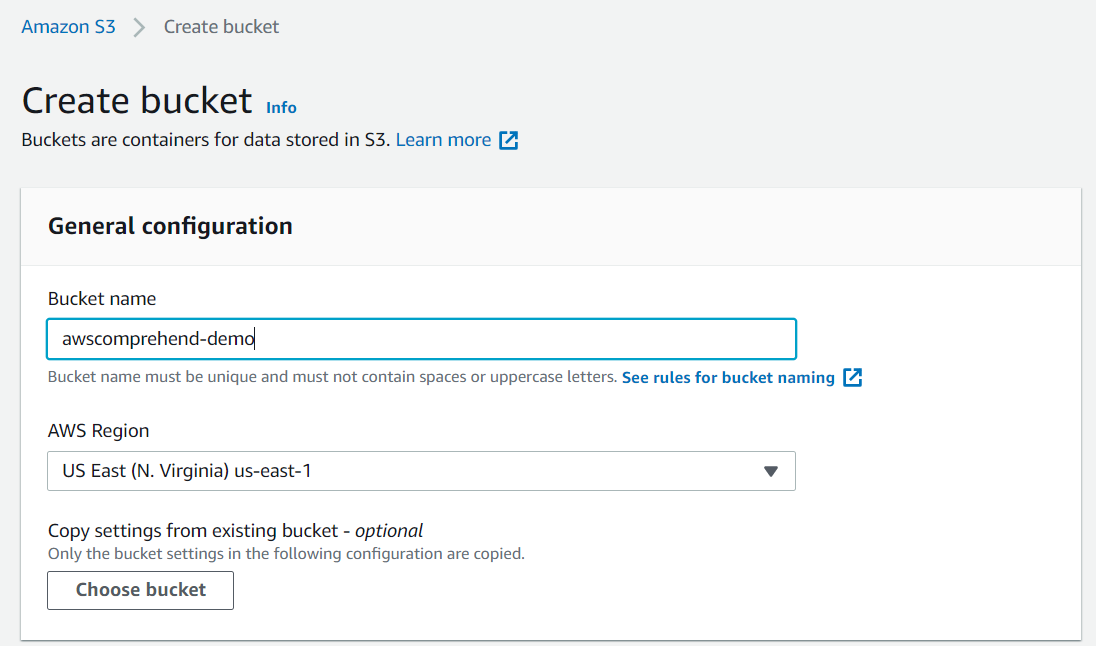


Steps in Sentiment Analysis

Step 1: All the other columns from the dataset are dropped and a new csv file containing only the ‘content ‘column is created. [Jupyter Notebook](https://jupyter.org/) can be used to handle this.

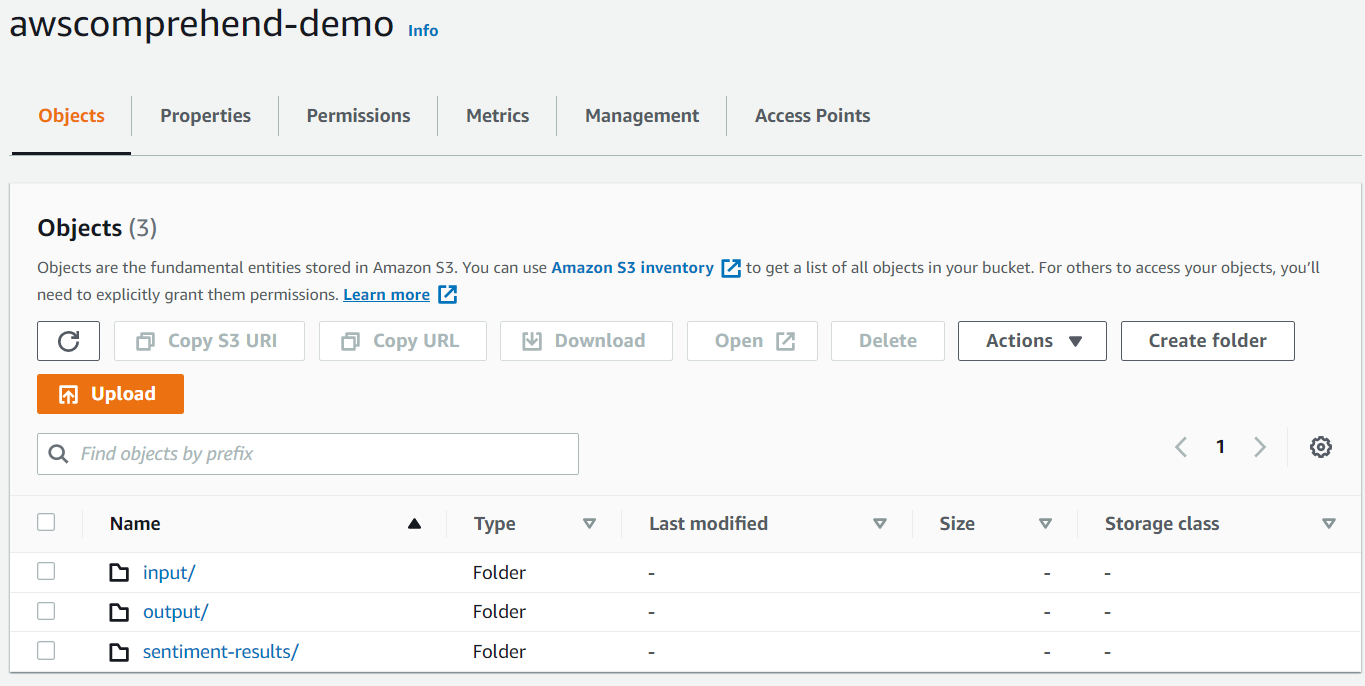


Step 2 : reviews\_content .csv is the new file from step 1 and is uploaded into an [S3 bucket](https://aws.amazon.com/s3/) .

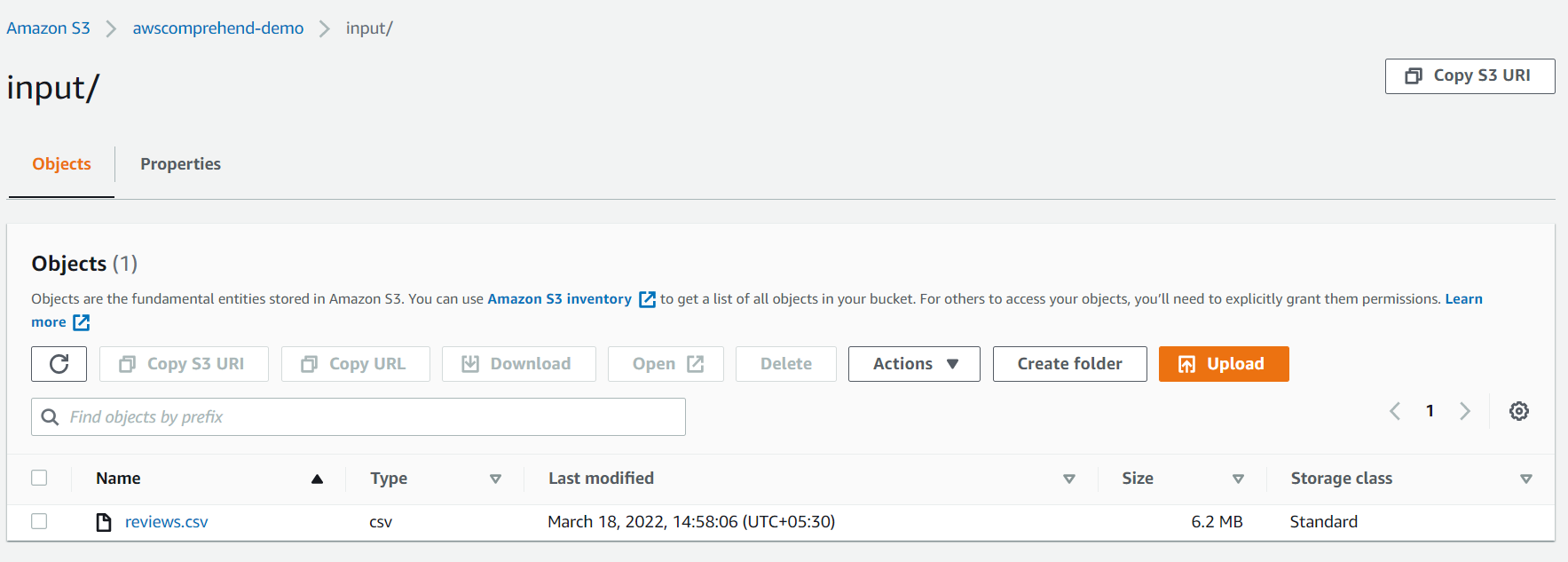


All the other details are kept as default values while creating the bucket. The Bucket name specified must be a unique one.

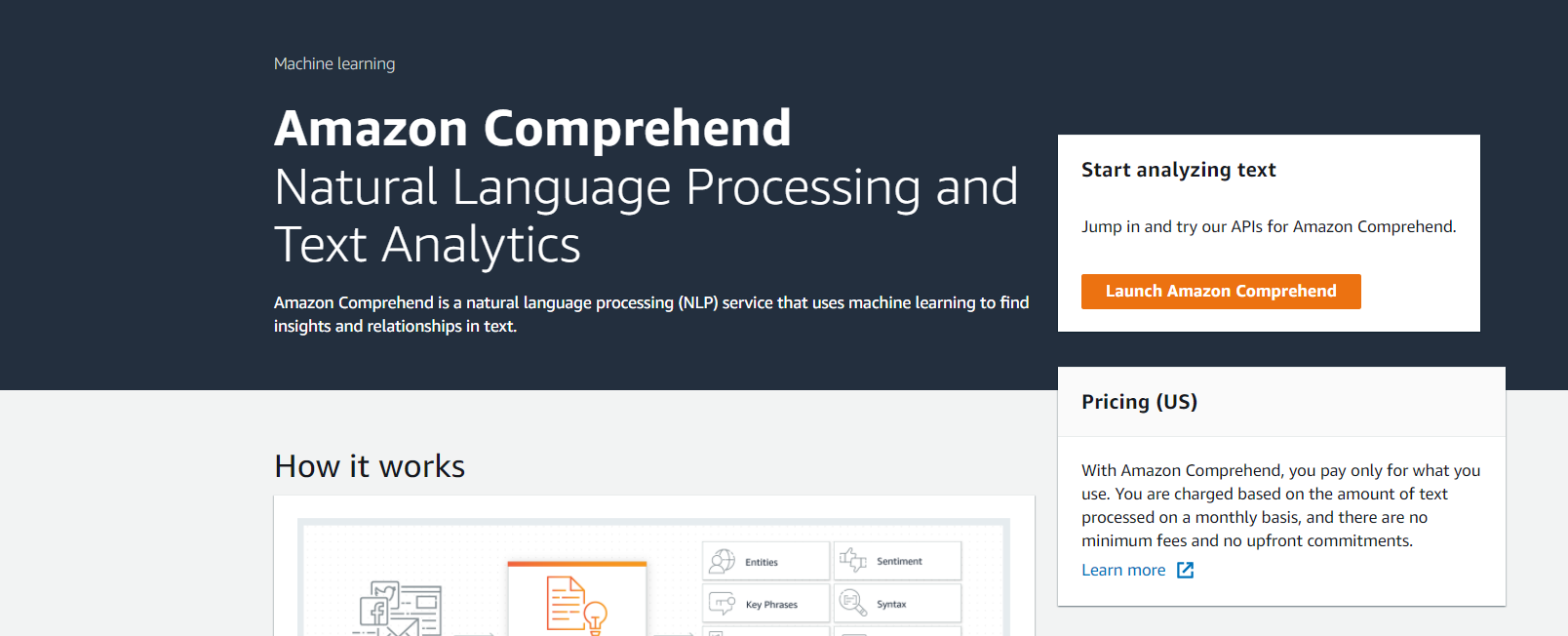
Step3: Upload the csv file in S3 bucket created. Inside the S3 bucket awscomprehend-demo create 3 folders input/, output/ and sentiment-results/.



Step 4 : Upload the reviews\_content .csv file in input/ folder.

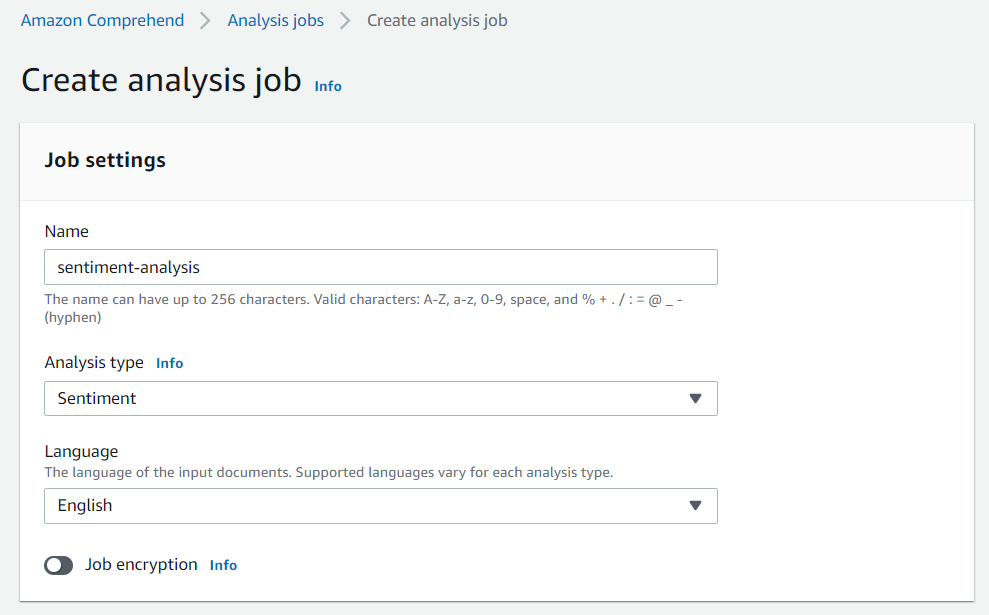


Step 4: Open AWS Comprehend from console and click Launch Amazon Comprehend .Make sure AWS Comprehend is launched in same region as the region in which S3 bucket is created.

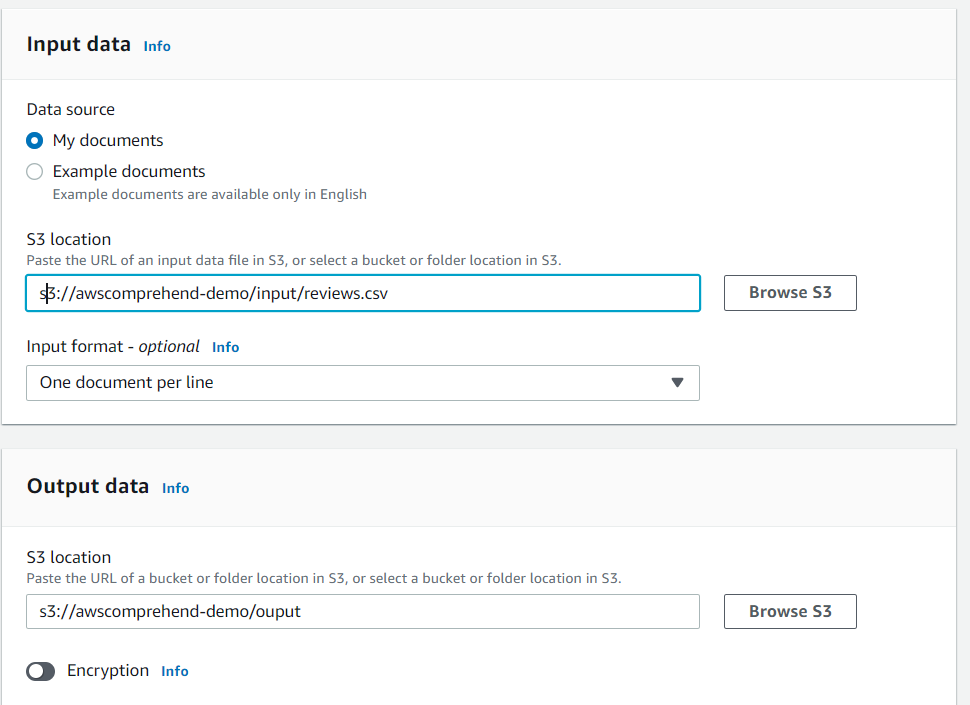


Step 5:Click on Analysis Job from left pane and Click on Create job button.

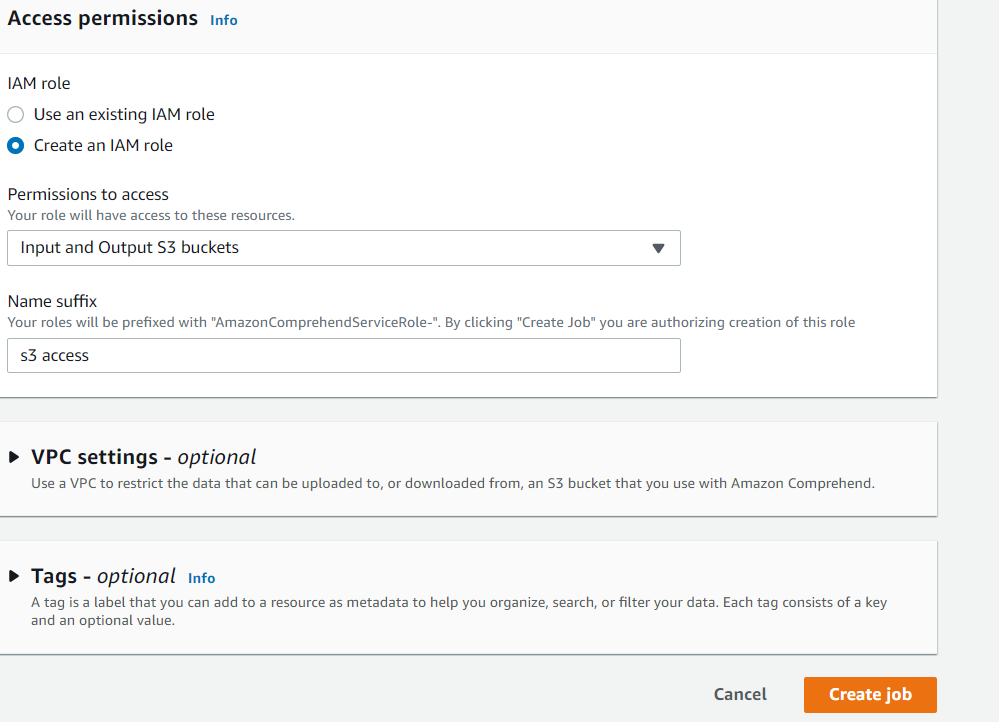
Step 6: Enter Name ,Analysis type and Language for the Analysis job as sentiment-analysis, Sentiment and English for creating an Analysis Job.



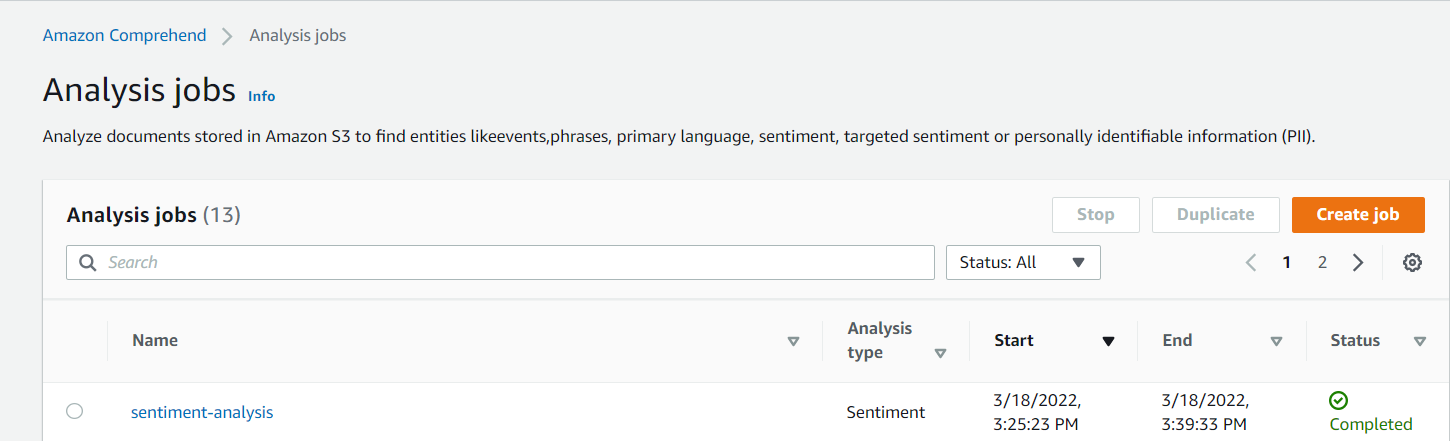
Step 7: Specify the input and output locations in S3 from which the Analysis job takes input and write the output.



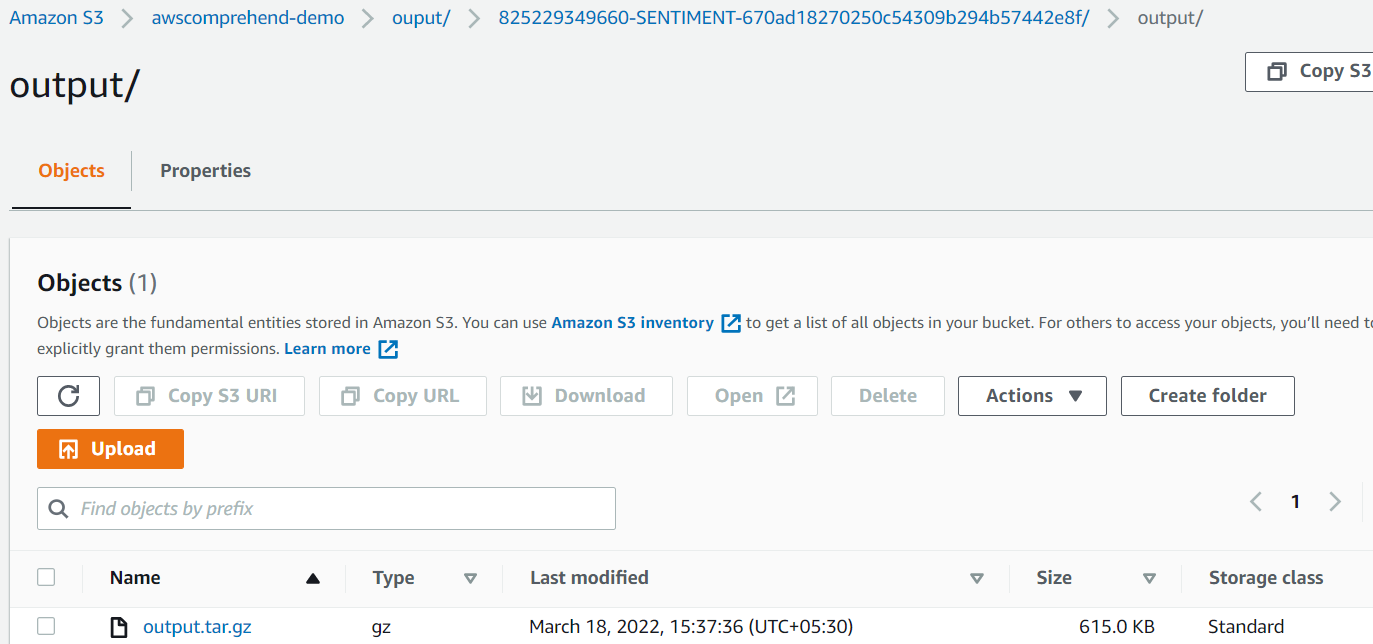
Step 8:Specify the access permissions for the role by Create a role or Use an existing role. Once all these steps are done click Create job button. It will take a few minutes to create the job and then AWS Comprehend automatically starts executing it.



Step 9: Once the job is completed ,the status of the job changes to Completed.

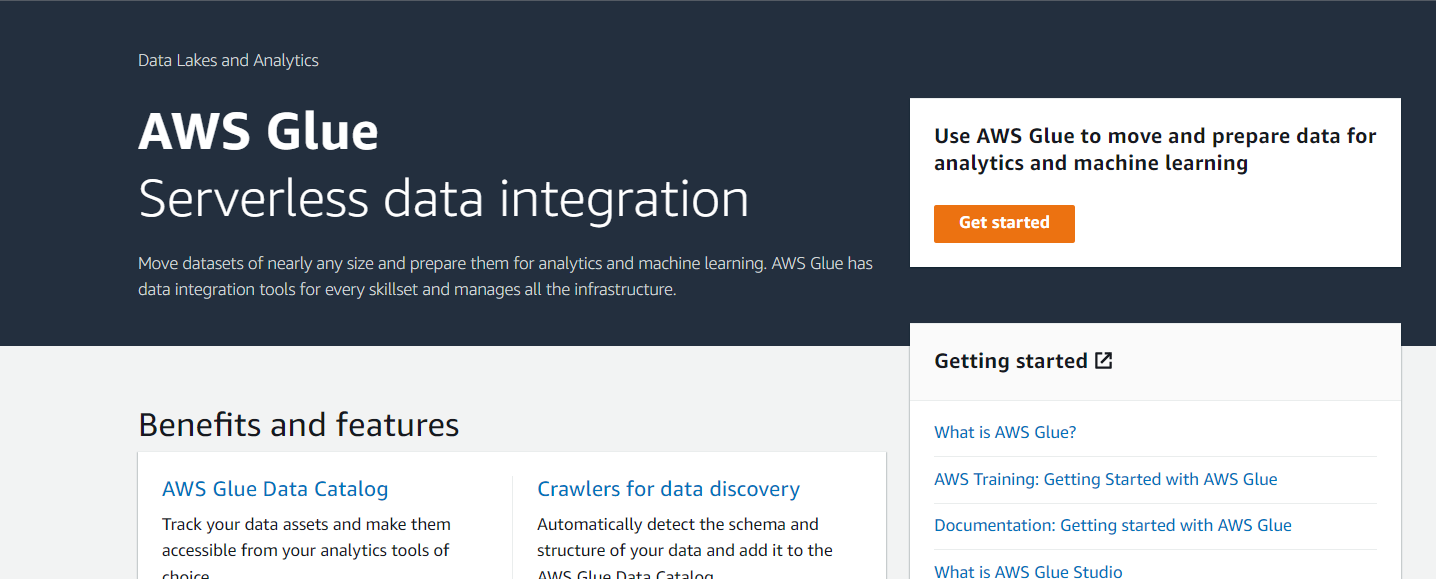


Step 10: We can see the oupt.tar.gz file inside the output/ folder in the S3 bucket location we specified. Download this file and unzip it in local machine.



Step 11: Upload the unzipped file into the sentiment-results folder inside the S3 bucket.

Step 12: Open [AWS Glue](https://aws.amazon.com/glue/) console.



Step 13: A Glue crawler is created with the following details:

Crawler name: comprehend-crawler

Crawler source type: Data stores

Repeat crawls of S3 data stores: Crawl all folders

Choose a data store: S3

Include path: s3://awscomprehend-demo/sentiment-results/

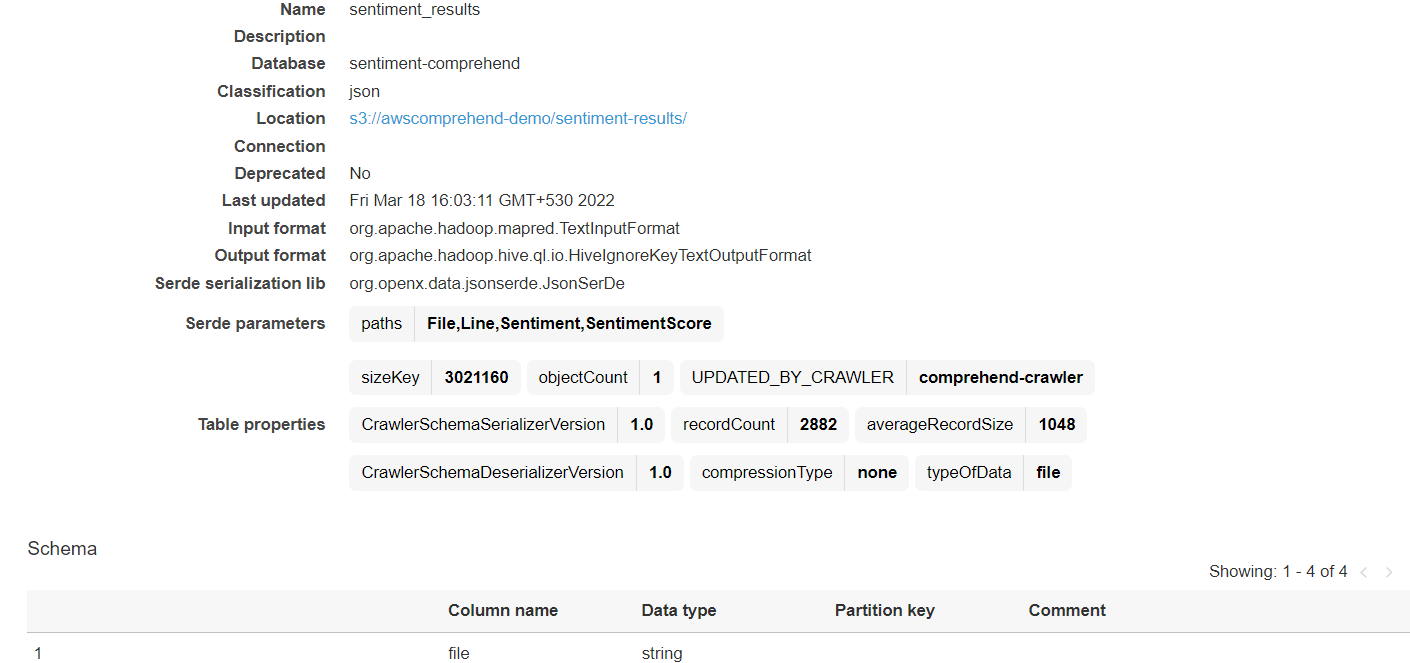
Add another data store: No

Iam role:<Create one or use an existing one>

Frequency: Run on demand

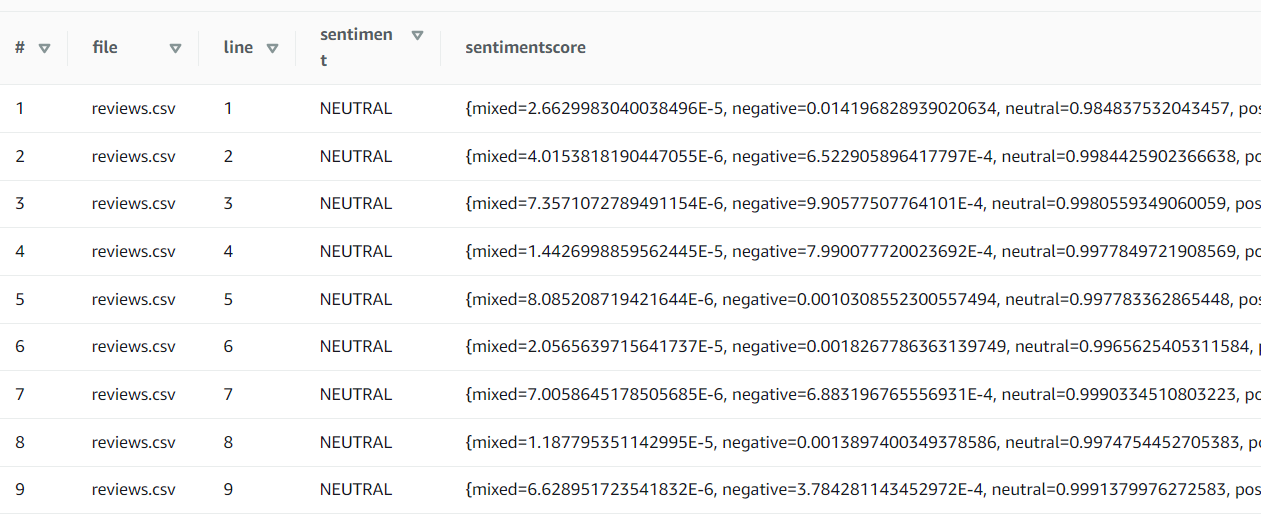
Database name: sentiment-comprehend

Click Finish to complete creating the crawler. Select the crawler comprehend-crawler and click Run crawler. Once the crawler job is complete we can see the tables by clicking the Tables option from left pane of window.

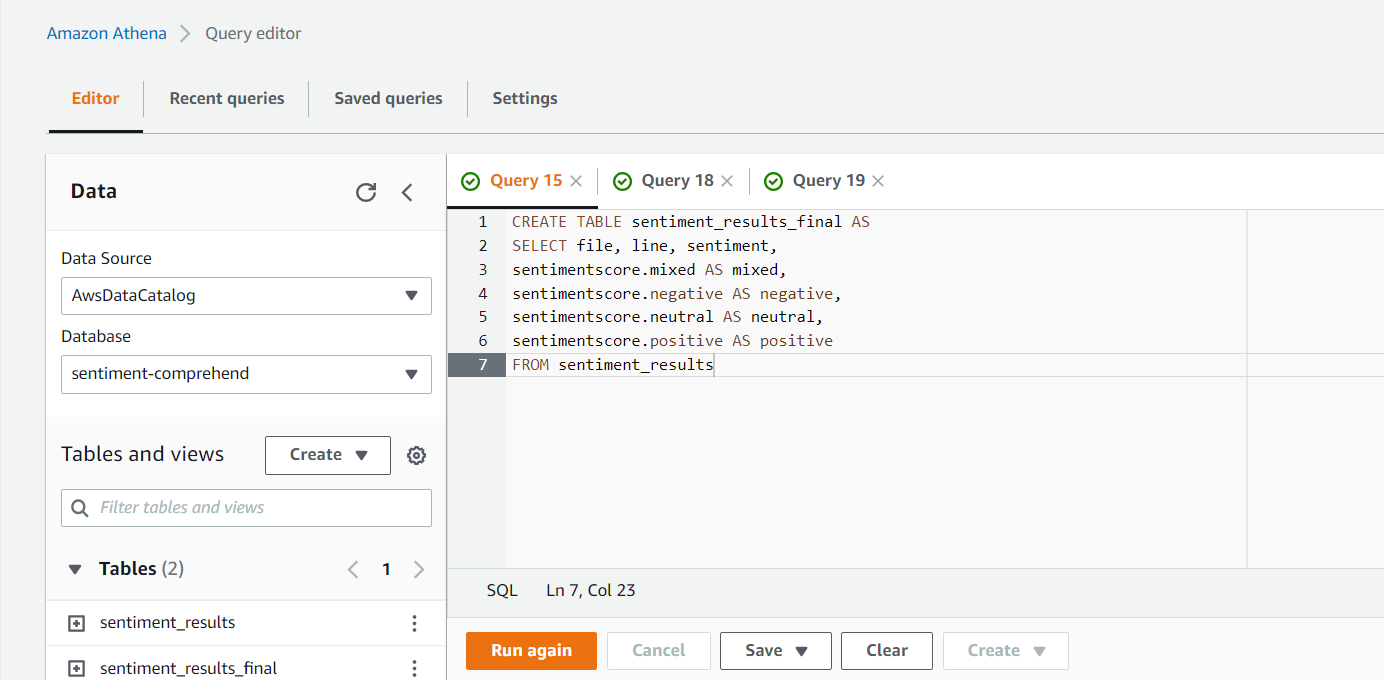


Step 14 : Open [Amazon Athena](https://aws.amazon.com/athena/) console and select the Database and Table name as sentiment-comprehend and sentiment\_results.

Step 15:Right click on the table name and select preview. We can see the sentiment score as a nested json document.



Step 16: To unnest the json document, enter the following query in Athena Query Editor and click Run.

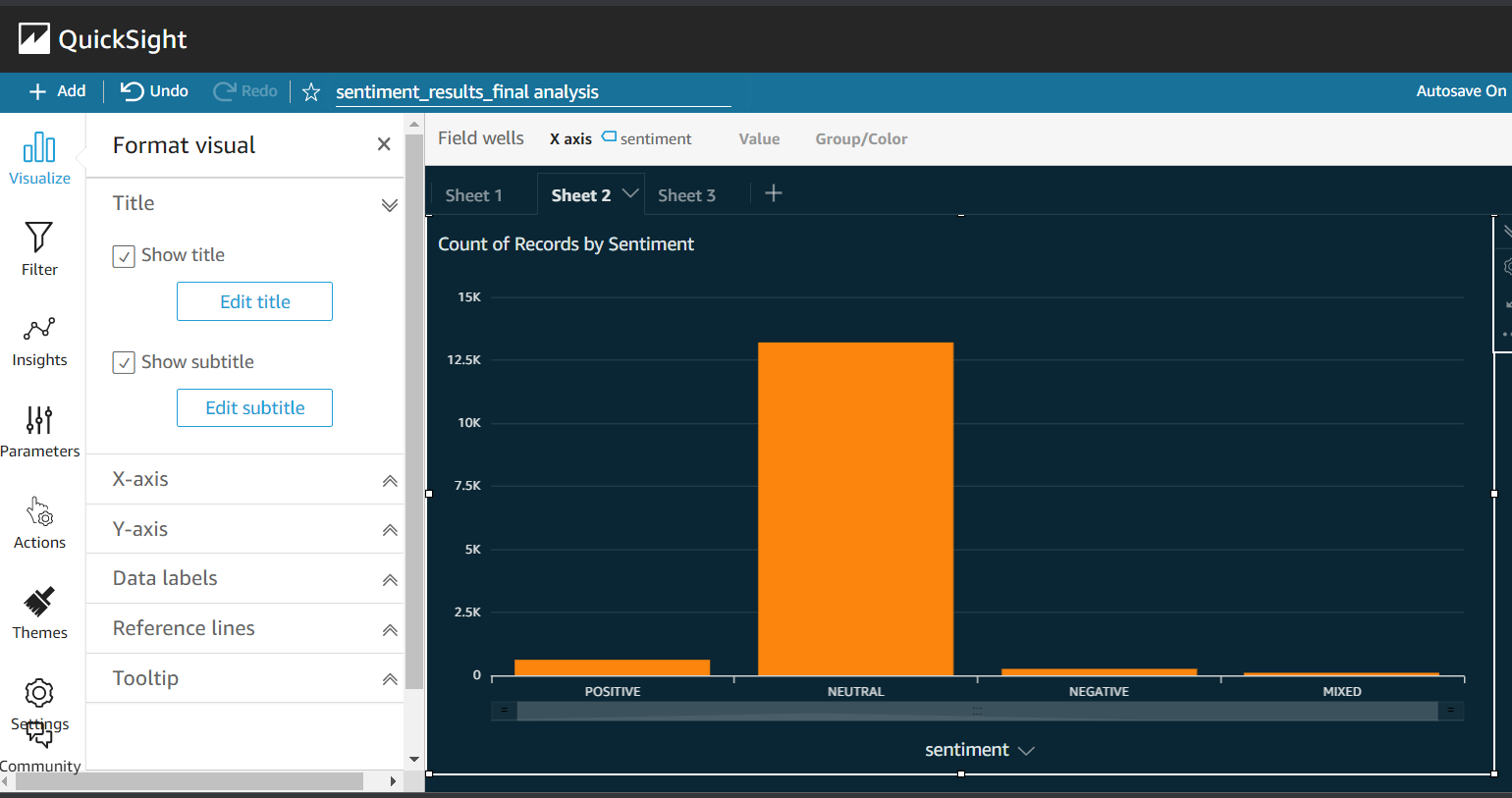


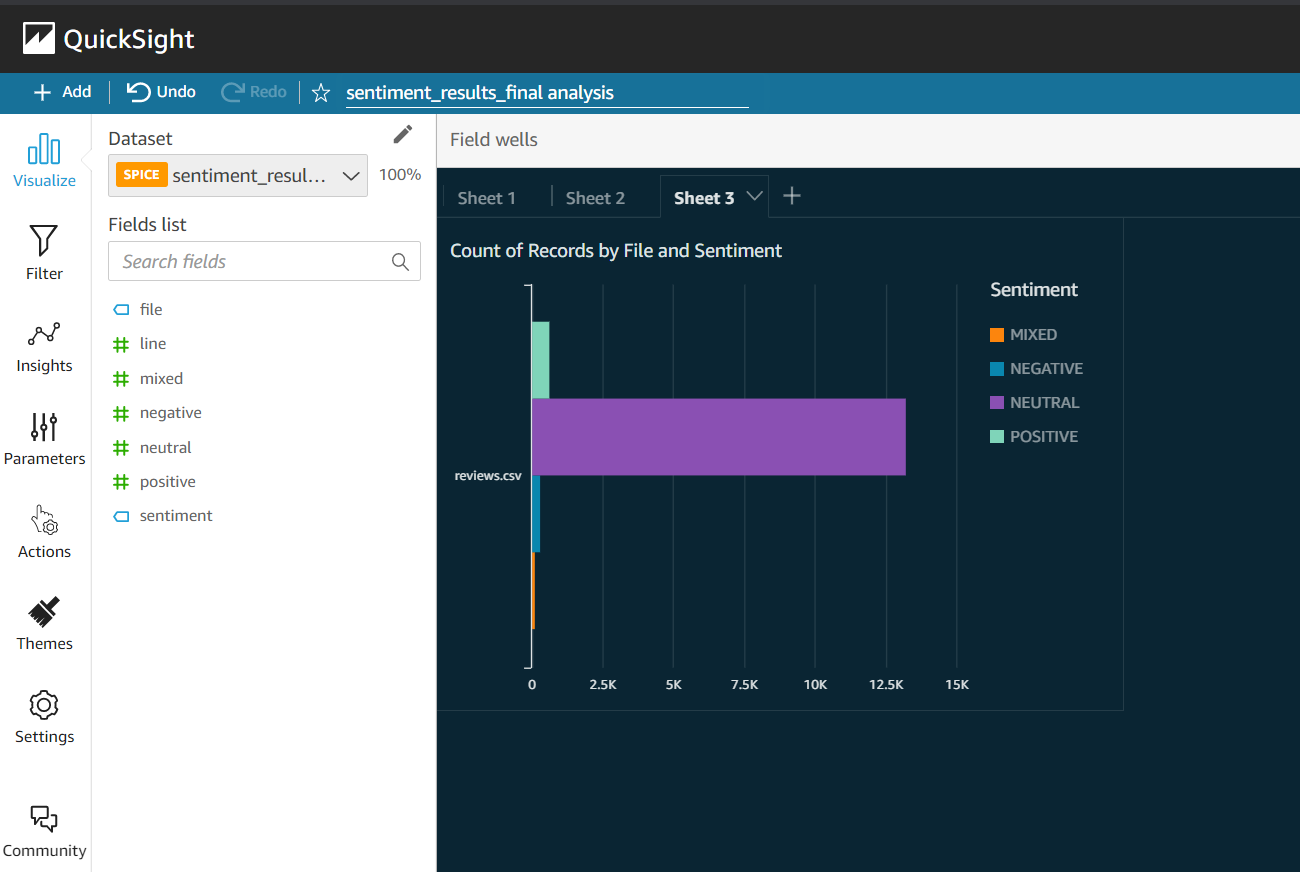
Step 16: Open [AWS Quicksight](https://aws.amazon.com/quicksight/) console. Click New analysis. Select New Dataset .  **enter the following details**

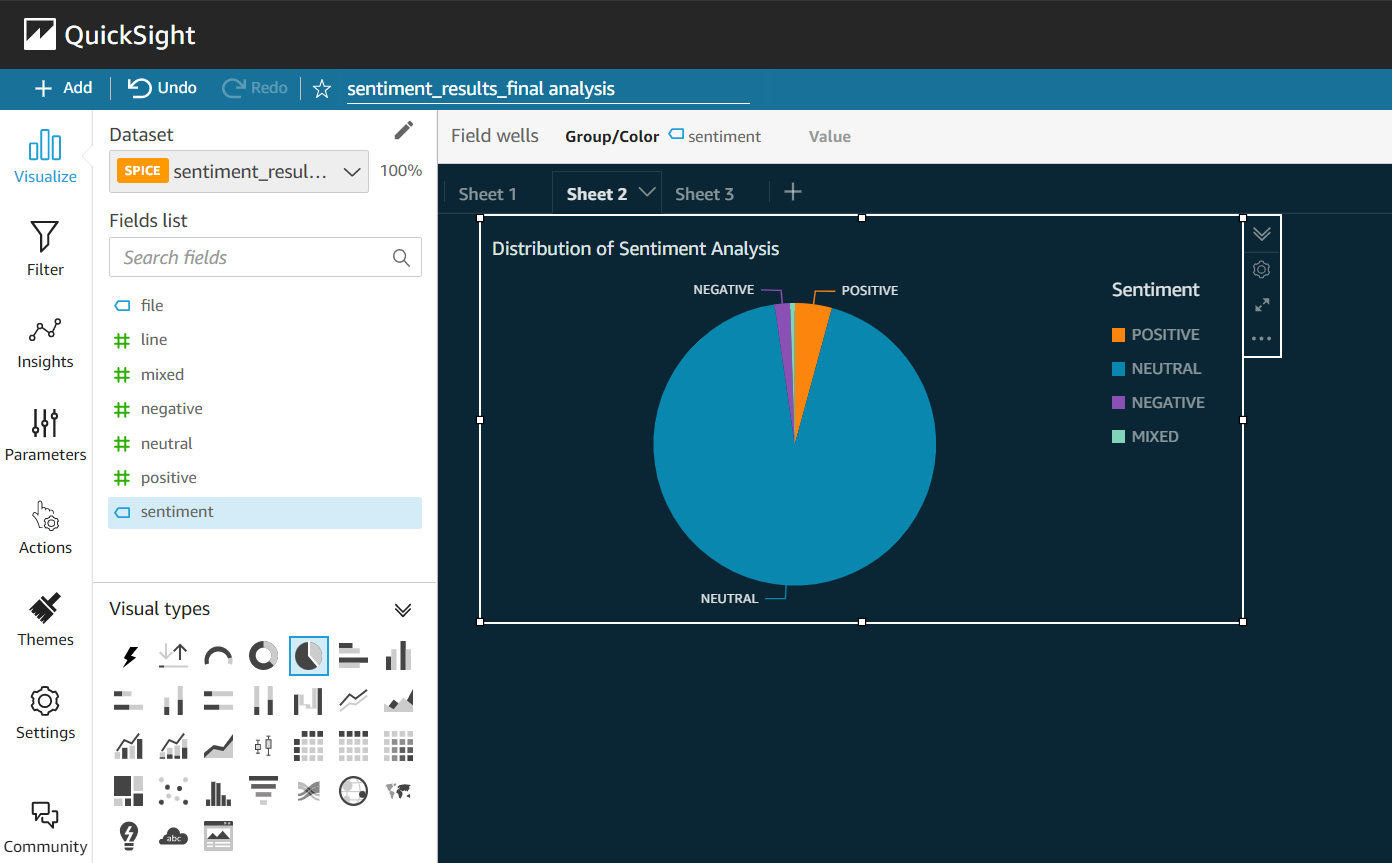
**Data source name:** **comprehend-sentiment-demo**

**Choose database: sentiment-comprehend**

**Select sentiment\_results\_final from the list of tables. Click on Visualize. We can select the desired chart type for visualizing the results.**







Result Analysis

From the charts it can be easily seen that majority of comments were neutral, followed by positive comments. That is, people are neither happy, nor disappointed with the product or service. Customers having Neutral Sentiment points to the fact that a customer churn may happen in case the competitor is giving a discount or offer.

Who should use Sentiment Analysis?

Any industry whose final product or service reaches the customer must use Sentiment Analysis. Irrespective of their size, a small, medium or large company can leverage Sentiment Analysis for improving customer satisfaction in an intelligent way. AWS Comprehend provides a smart way of analysing sentiment with minimum investment. From a profitability perspective, Sentiment Analysis using AWS Comprehend gives insights as to whether the company is moving into a profitable position or not and also whether all the efforts done by the other departments like Marketing, Operations and Logistics are bearing fruit and which are the areas where rectifications are to be made. Sentiment score can be incorporated as a KPI (Key Performance Indicator) that demonstrates how effectively the company is achieving key business objectives. Tracking what customers are saying about the company and its products, gives an accurate view of the brand reputation. By staying abreast of customer satisfaction and brand reputation, the company can get ahead of any issues that may harm its business.