



Trends in Obesity Rates by Age Group in the USA

Husky-group 3

Ananya Kakani

Jui Ying Lee

Sandeep Samsani

Bhavana Deshetty

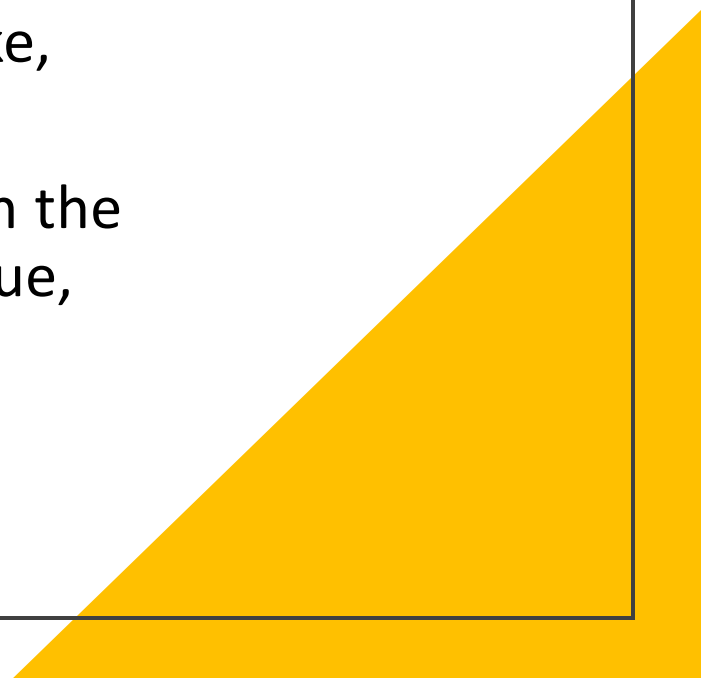


Outline

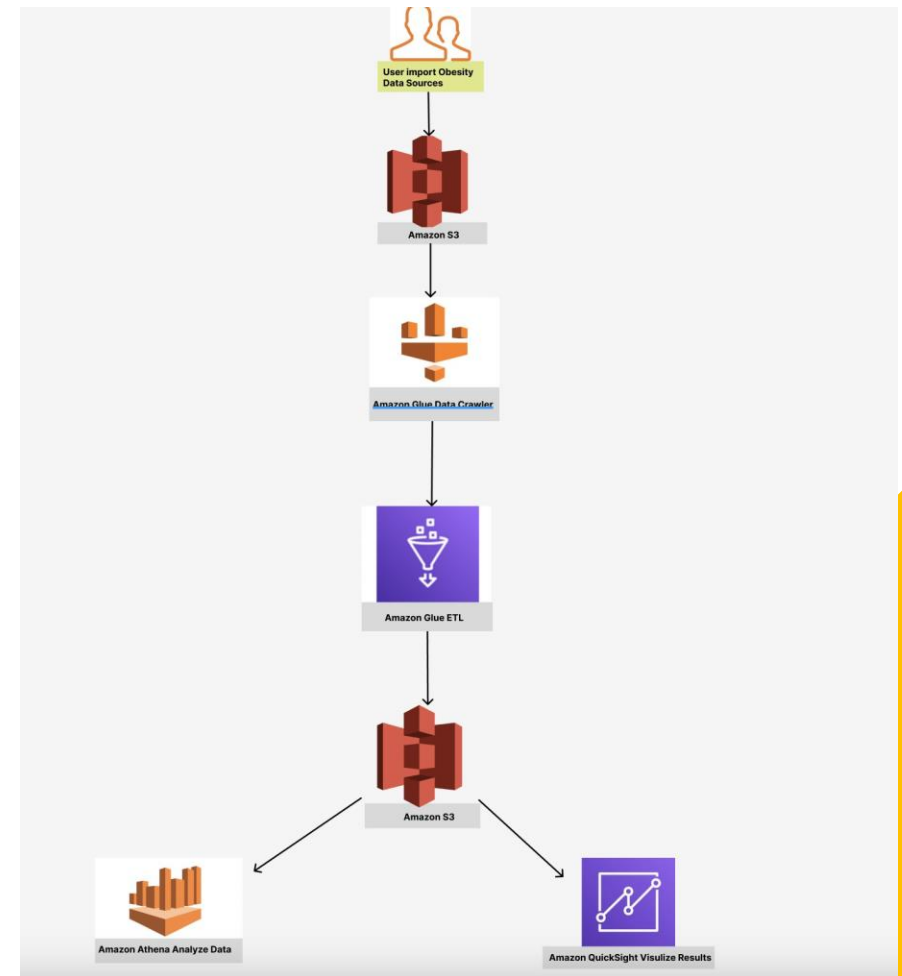
- Introduction
- AWS Architecture
- Data
- AWS Services
- Results
- Conclusion
- Live Demo
- Q&A




Introduction

- Obesity is a major public health problem in the United States, affecting millions of people of all ages. Obesity is a risk factor for many chronic diseases, including heart disease, stroke, type 2 diabetes, and some types of cancer.
 - This project aims to analyze obesity rates by age group in the USA using AWS services. The project will use AWS S3, Glue, Athena, and QuickSight to retrieve, clean, analyze, and visualize the data
- 
- A large yellow triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

AWS Architecture



Data

- Data retrieval: The data will be retrieved from the CDC NHANES website and stored in an AWS S3 bucket.
 - Data cleaning and preparation: After cleaned the data locally, AWS Glue will be used to clean and prepare the data for analysis.
 - Data analysis: AWS Athena will be used to analyze the data and calculate obesity rates by age group.
 - Data visualization: AWS QuickSight will be used to visualize the results of the data analysis.
- 
- A large yellow triangle is positioned in the bottom right corner of the slide, pointing towards the top right.

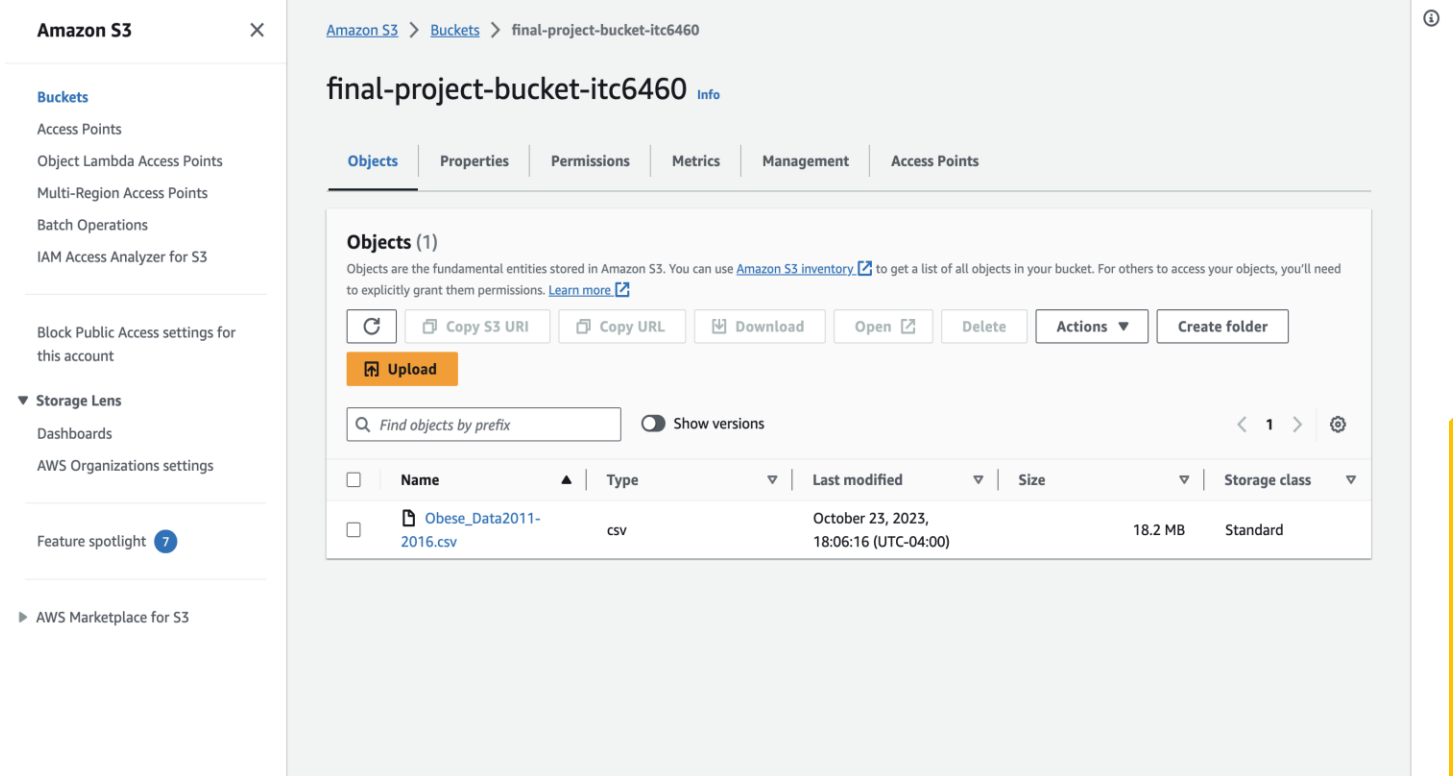
Data

- Generally, we've cleaned the data in our local system, and later uploaded to the AWS S3 bucket. Afterwards, AWS Glue is used to create a data catalog and transform the data into a format that is compatible with AWS Athena.
- AWS Athena will be used to query the data and analyze the trends in obesity rates by age group.



Implementation : AWS S3

- Step 1: Create the S3 bucket and upload the dataset
- In this step, we created a new s3 bucket (No.1) to upload the dataset after cleaned successfully
- We also need to create the second S3 bucket successfully to store the analyzed data



The screenshot displays the AWS S3 console interface. On the left, the 'Amazon S3' sidebar is visible with a 'Buckets' section. The main content area shows the details for a bucket named 'final-project-bucket-itc6460'. The 'Objects' tab is selected, showing a list of objects. One object is listed: 'Obese_Data2011-2016.csv', which is a CSV file, 18.2 MB in size, and stored in the 'Standard' storage class. The object was last modified on October 23, 2023, at 18:06:16 (UTC-04:00). The console also shows options to upload, copy, download, and delete objects, as well as a search bar and a 'Show versions' toggle.

Amazon S3

Buckets

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

Amazon S3 > Buckets > final-project-bucket-itc6460

final-project-bucket-itc6460 Info

Objects Properties Permissions Metrics Management Access Points

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Copy S3 URI Copy URL Download Open Delete Actions Create folder

Upload

Find objects by prefix Show versions

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	Obese_Data2011-2016.csv	csv	October 23, 2023, 18:06:16 (UTC-04:00)	18.2 MB	Standard

1



AWS Glue

- Step 1: Create an IAM role
- Step 2: Add permissions for Glue Access

The screenshot displays the AWS IAM console interface. On the left is a navigation sidebar for 'Identity and Access Management (IAM)' with sections for 'Access management' (User groups, Users, Roles, Policies, Identity providers, Account settings) and 'Access reports' (Access analyzer, Archive rules, Analyzers, Settings, Credential report, Organization activity, Service control policies (SCPs)). The main content area shows the details for the role 'AWSGlueServiceRole-FinalProj'. The 'Summary' tab is active, displaying the creation date (October 21, 2023, 19:40 (UTC-04:00)), last activity (58 minutes ago), ARN (arn:aws:iam::182198172686:role/service-role/AWSGlueServiceRole-FinalProj), and maximum session duration (1 hour). Below the summary, the 'Permissions' tab is selected, showing a list of 'Permissions policies (6)'. The policies listed are AmazonAthenaFullAccess, AmazonS3FullAccess, and AWSGlueConsoleFullAccess, all of which are AWS managed and attached to the role. The bottom of the console shows a footer with 'CloudShell', 'Feedback', and copyright information for Amazon Web Services, Inc. or its affiliates.

Identity and Access Management (IAM)

Search IAM

Dashboard

▼ Access management

- User groups
- Users
- Roles**
- Policies
- Identity providers
- Account settings

▼ Access reports

- Access analyzer
- Archive rules
- Analyzers
- Settings
- Credential report
- Organization activity
- Service control policies (SCPs)

Related consoles

- CloudShell
- Feedback

AWSGlueServiceRole-FinalProj Info

Delete

Summary Edit

Creation date
October 21, 2023, 19:40 (UTC-04:00)

ARN
arn:aws:iam::182198172686:role/service-role/AWSGlueServiceRole-FinalProj

Last activity
58 minutes ago

Maximum session duration
1 hour

Permissions | Trust relationships | Tags | Access Advisor | Revoke sessions

Permissions policies (6) Info

You can attach up to 10 managed policies.

Search

Filter by Type
All types

<input type="checkbox"/>	Policy name	Type	Attached entities
<input type="checkbox"/>	AmazonAthenaFullAccess	AWS managed	2
<input type="checkbox"/>	AmazonS3FullAccess	AWS managed	3
<input type="checkbox"/>	AWSGlueConsoleFullAccess	AWS managed	2

© 2023, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Glue

- In Glue service, crawlers will scan the S3 bucket for data and create a table with the help of Glue data catalog.
- We can click on the “Crawlers” and then click “Create tables with crawler”. After entering the essential details, select the s3 bucket we created in the previous step.

The screenshot displays the AWS Glue console interface. On the left is a navigation sidebar with categories like 'Getting started', 'Data Catalog', and 'Data Integration and ETL'. The 'Crawlers' link is highlighted. The main panel shows the configuration for the 'Obesity_Crawler'. It includes fields for Name, Description, IAM role (AWSGlueServiceRole-FinalProj), Database (obesity_usa_before), State (READY), and various configurations like Security, Lake Formation, and Table prefix. Below this is a section for 'Crawler runs' showing a single completed run on October 23, 2023, with a status of 'Completed' and a duration of 40 seconds. The footer contains copyright information for Amazon Web Services and links to Privacy, Terms, and Cookie preferences.

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
October 23, 2023 at 22:14:01	October 23, 2023 at 22:14:41	40 s	Completed	0.081	1 table change, 0 partition changes

AWS Glue ETL

- Step 1: Create and run a Glue ETL job
- Step 2, click "Create Job," input essential details like name and description. Select tables and databases generated by the crawler. Then, choose and add desired transform steps.

The screenshot displays the AWS Glue console interface for a job named 'Obesity_etl'. The left sidebar shows the navigation menu with categories like 'Getting started', 'ETL jobs', 'Data Catalog', and 'Data Integration and ETL'. The main panel is in the 'Visual' tab, showing a workflow diagram with two nodes: 'Data source - Data Catalog AWS Glue Data Catalog' and 'Transform - Change Sch... Change Schema'. The 'Data source' node is selected, and its properties are shown on the right: Name 'AWS Glue Data Catalog', Database 'obesity_usa_before', and Table 'final_project_bucket_itc6460'. A 'Data preview' section at the bottom provides instructions on how to start a data preview session. The bottom of the console shows the footer with copyright information and links to CloudShell, Feedback, Privacy, Terms, and Cookie preferences.

AWS Glue ETL

- Step 3: Run the ETL Jobs and get the results

The screenshot displays the AWS Glue console interface. On the left is a navigation sidebar with categories like 'Getting started', 'ETL jobs', 'Data Catalog', and 'Data Integration and ETL'. The main panel shows the 'Job Run' details for a specific job. The job name is 'Obesity_etl' and its status is 'Succeeded'. A table below provides various metrics such as start time, end time, execution time, and worker details.

AWS Glue × [AWS Glue](#) > [Monitoring](#) > Job run ⓘ

Job Run - jr_7c64167136866406acc36da5a6f839d234815fd68ab43cb25584f605d189b1df

Run details [Info](#) Rewind job bookmark ↺

jr_7c64167136866406acc36da5a6f839d234815fd68ab43cb25584f605d189b1df

Job name	Id	Run status	Glue version
Obesity_etl	jr_7c64167136866406acc36da5a6f839d234815fd68ab43cb25584f605d189b1df	✔ Succeeded	4.0
Retry attempt number	Start time	End time	Start-up time
Initial run	October 23, 2023 6:22:31 PM	October 23, 2023 6:23:35 PM	9 seconds
Execution time	Last modified on	Trigger name	Security configuration
54 seconds	October 23, 2023 6:23:35 PM	-	-
Timeout	Max capacity	Number of workers	Worker type
2880 minutes	10 DPUs	10	G.1X
Execution class	Log group name	Cloudwatch logs	Performance and debugging recommendations
Standard	/aws-glue/jobs	<ul style="list-style-type: none">All logsOutput logsError logs	<ul style="list-style-type: none">View in CloudWatch

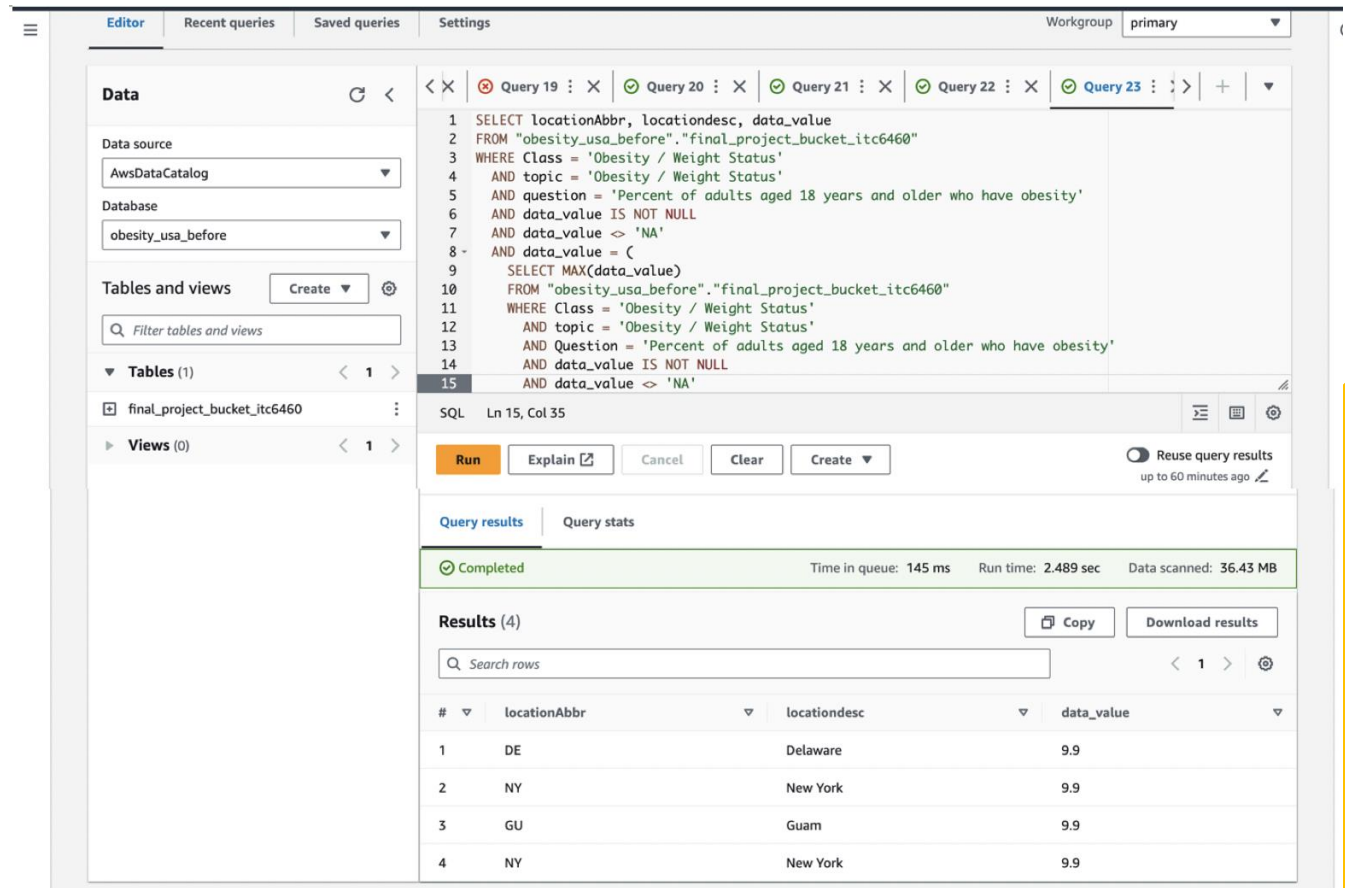
☒ Enable compact mode

[CloudShell](#) [Feedback](#)

© 2023, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

AWS Athena

- In this query, we're trying to find the states that have the highest data value based on their obesity.



The screenshot displays the AWS Athena console interface. On the left, the 'Data' panel shows the 'Data source' as 'AwsDataCatalog' and the 'Database' as 'obesity_usa_before'. Below this, a list of tables includes 'final_project_bucket_itc6460'. The main editor area contains a SQL query (Query 23) that selects location abbreviations and descriptions, filtering for obesity-related data and finding the maximum value for each location. The query is as follows:

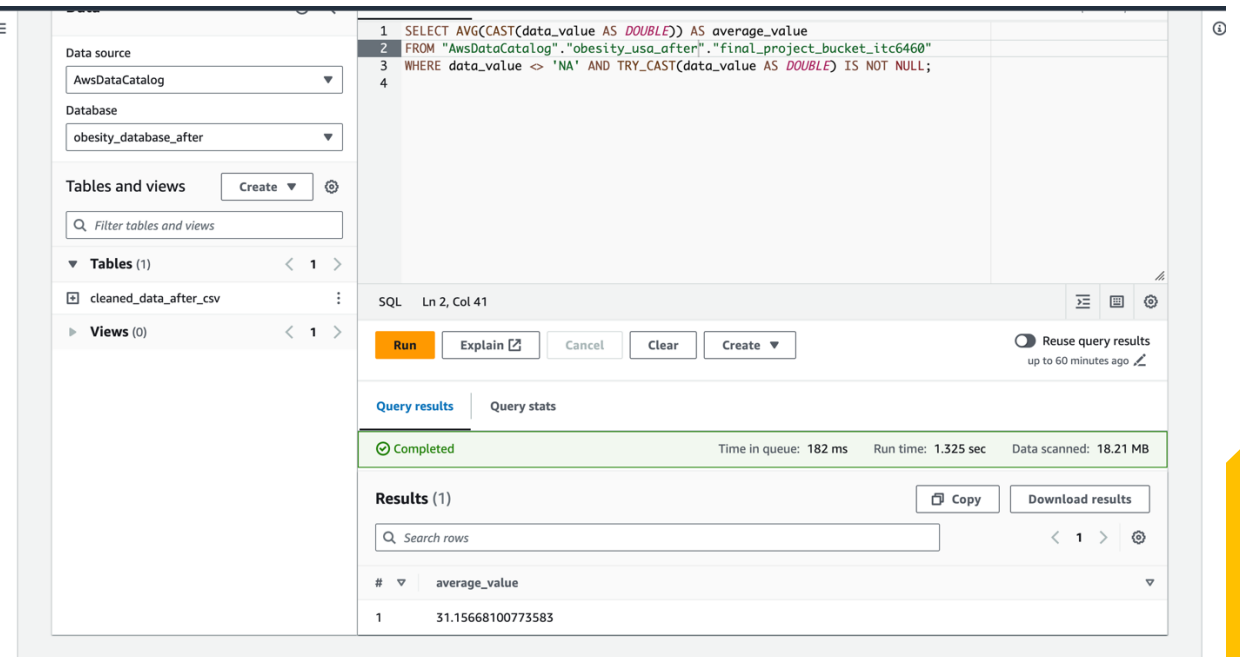
```
1 SELECT locationAbbr, locationdesc, data_value
2 FROM "obesity_usa_before"."final_project_bucket_itc6460"
3 WHERE Class = 'Obesity / Weight Status'
4 AND topic = 'Obesity / Weight Status'
5 AND question = 'Percent of adults aged 18 years and older who have obesity'
6 AND data_value IS NOT NULL
7 AND data_value <> 'NA'
8 AND data_value = (
9     SELECT MAX(data_value)
10    FROM "obesity_usa_before"."final_project_bucket_itc6460"
11   WHERE Class = 'Obesity / Weight Status'
12     AND topic = 'Obesity / Weight Status'
13     AND Question = 'Percent of adults aged 18 years and older who have obesity'
14     AND data_value IS NOT NULL
15     AND data_value <> 'NA'
```

Below the query editor, the 'Run' button is highlighted. The 'Query results' tab shows the query is 'Completed' with a run time of 2.489 seconds and 36.43 MB of data scanned. The results are displayed in a table with 4 rows:

#	locationAbbr	locationdesc	data_value
1	DE	Delaware	9.9
2	NY	New York	9.9
3	GU	Guam	9.9
4	NY	New York	9.9

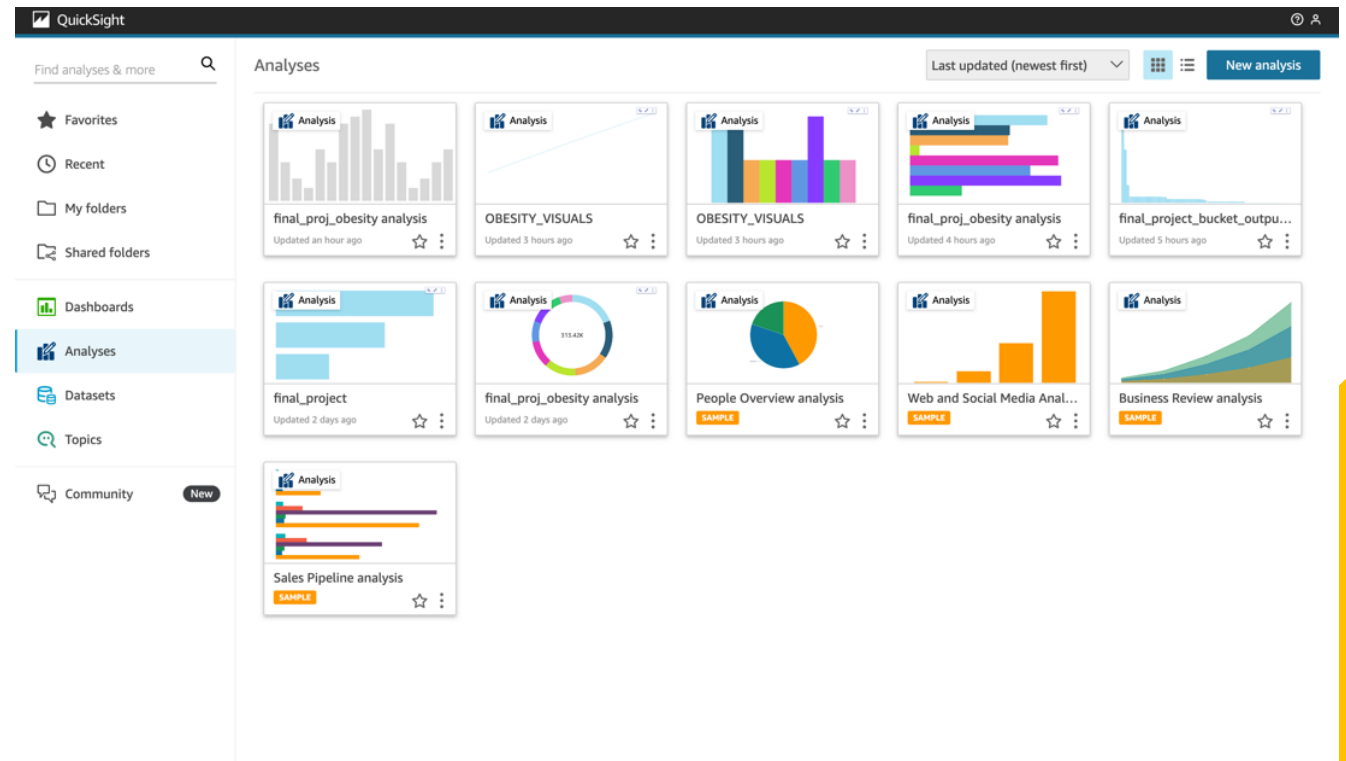
AWS Athena

- We effectively queried and obtained the average data value for college graduates, categorized by gender. This data provides quantitative insights, encompassing obesity rates, sample sizes, and confidence levels.

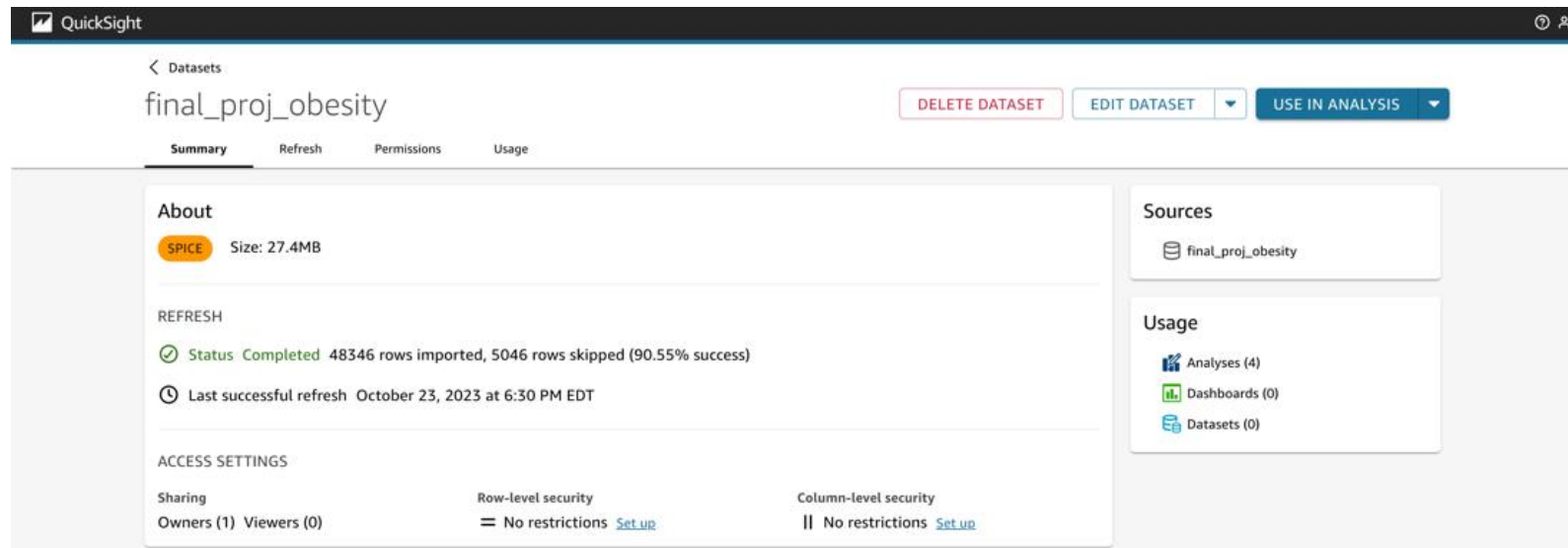


AWS QuickSight

Step 1: Create a dashboard to visualize our transformed data.



AWS QuickSight

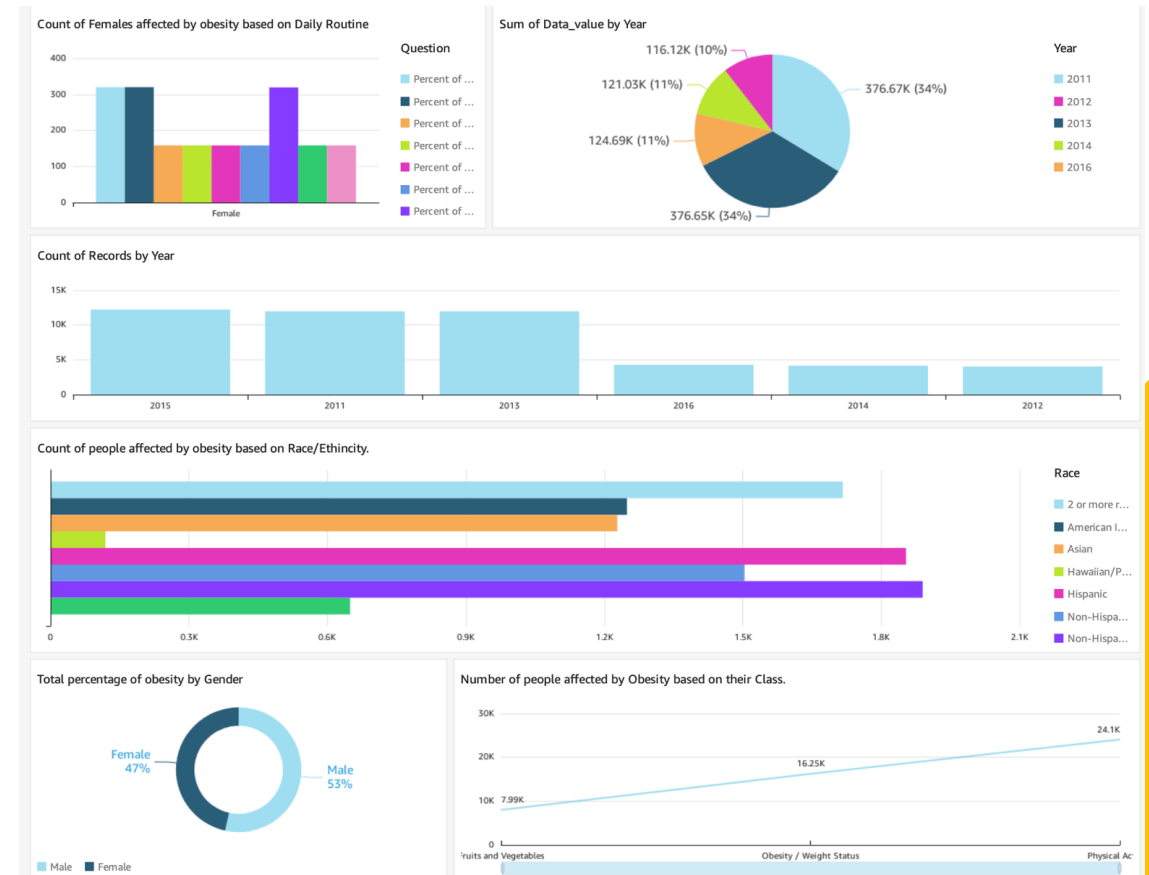


Step 2: Once the analysis is complete, we simply choose the desired data in the "Visualizations" section and save it. We employ Quicksight for data presentation.

AWS QuickSight Result

Step 3: We've successfully created a diverse set of visualizations, exportable in PDF format. Our visuals include line charts, bar charts, pie charts, and tables.

Utilizing Quicksight, we've enhanced interactivity with features like filters and actions, making our visualizations more user-friendly and interactive.



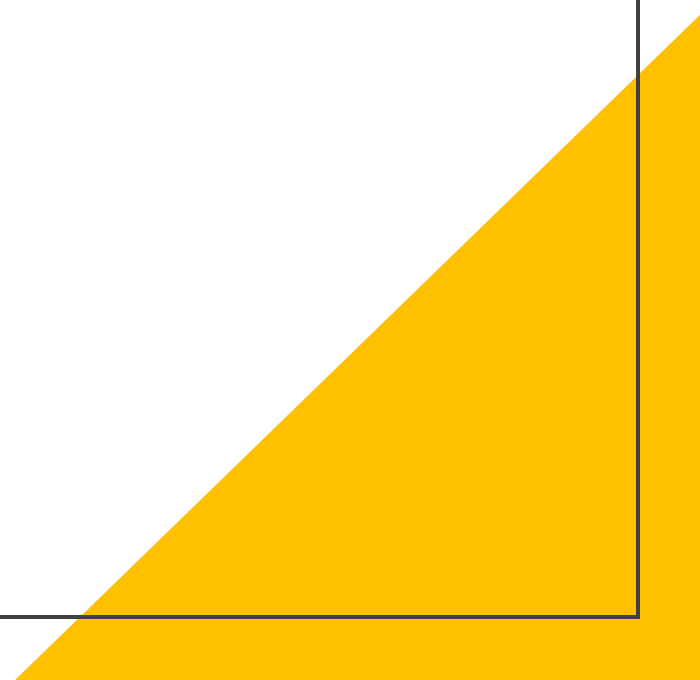
Conclusion

Future Directions:

- **In-Depth Demographic Analysis:** Expanding our analysis to encompass influential factors like race, ethnicity, and socioeconomic status for a more nuanced understanding of obesity trends.
- **Predictive Modeling:** Constructing models to anticipate obesity risk enables proactive healthcare initiatives and personalized interventions.

By embracing this multifaceted approach, we aim to deepen our understanding of the causes and consequences of obesity, enabling more effective, precisely targeted interventions. This will ultimately advance public health outcomes and address one of the most pressing health challenges of our time.

LIVE DEMO



THANK YOU!

