

# Adidas Sales Analysis

By Santanu Adhikary



**Do follow me on LinkedIn - <https://www.linkedin.com/in/santanu-adhikary-dev/>**

# **Contents**

1.Introduction -----	03
a) Amazon Redshift	
b) Amazon S3	
c)Quick Sight Tool	
2.Business Requirements -----	04
3.Overview of the project -----	05
4.Implementation -----	06
5.Visualisation -----	34
6.Conclusion -----	40

# Introduction

Regardless of whether you are a small local business or a national corporation, tracking your retail sales is incredibly important. The valuable activity lets you identify potential sales patterns, close more deals in much less time, and make better and informed sales forecasts. However, it is seen that most retail sales managers struggle to keep a tap on their sales performance. This, in turn, impacts their retail productivity.

The amount of data gathered by retail businesses keeps growing at an alarmingly high rate. Sales managers are unable to ensure the adequate utilization and analysis of data, contributing to the issue of data silos. Besides, it results in the most common challenges for retailers: a lack of consistency in the implementation of the sales process, limited line-of-sight for managers, and reps spending too much time not selling.

It is crucial to find a technology solution that can effectively handle all the generated data and use it to unlock meaningful insights into sales performance for the retail industry. This is where **AWS Redshift** comes into the picture. In today's blog, we will understand how using Amazon Redshift can help to significantly **improve sales performance** and tracking.

## **A). Amazon Redshift:**

Amazon Redshift is a fully managed data warehouse service provided by Amazon Web Services (AWS). It's designed for analyzing large datasets and offers scalability, high performance, and SQL compatibility. Redshift uses a distributed architecture with massively parallel processing (MPP), making it efficient for complex queries. Data is stored in a columnar format, which enhances query speed. Whether you're running business intelligence reports or analyzing historical data, Amazon Redshift is a powerful choice for data analytics and reporting.

## **B). Amazon S3 Storage:**

Amazon S3 (Simple Storage Service) is an object storage service provided by Amazon Web Services (AWS). It offers industry-leading scalability, data availability, security, and performance. Here are the key points about Amazon S3:

- Scalability: Amazon S3 allows you to store and retrieve any amount of data from anywhere on the web. It's designed to handle both small and large-scale applications.
- Object Storage: Instead of traditional file storage, Amazon S3 stores data as objects within resources called buckets. These objects can be photos, audio files, videos, or any other type of data.

- Durability and Availability: Amazon S3 provides 99.999999999% durability (meaning your data is highly resistant to loss) and 99.99% availability of objects over a given year.

## C). Quick Sight:

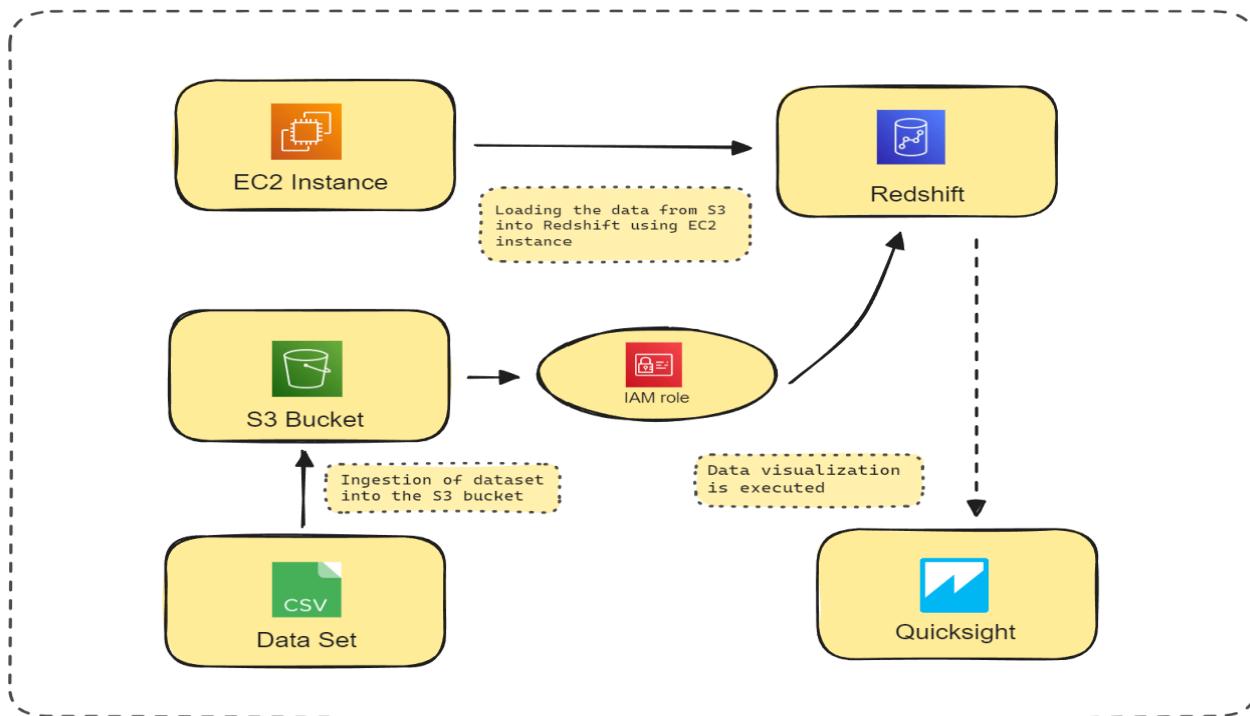
Amazon Quick Sight is a cloud-native, serverless business intelligence (BI) service provided by Amazon Web Services (AWS). Here are the key points about Quick Sight:

- Unified BI at Hyperscale: Quick Sight powers data-driven organizations with unified BI capabilities. It allows users to meet varying analytic needs from the same source of truth through:
- **Modern Interactive Dashboards:** Create interactive visualizations and explore data.
- **Paginated Reports:** Generate critical operational reports and dashboards.
  - **Embedded Analytics:** Embed dashboards or natural language query capabilities in applications.

## Business Requirements

1. Import sales data into Amazon Redshift.
2. Create SQL queries to aggregate sales data by product, region, and time period.
3. Identify top-selling products by calculating total sales or units sold.
4. Determine highest revenue-generating regions by analyzing sales revenue per region.
5. Analyze trends over time by plotting sales data over different time periods (e.g., daily, monthly, quarterly).
6. Visualize the sales data using charts and graphs (e.g., bar charts, line graphs) to provide insights into sales performance.

## Overview of the Project



1	2	3	4	5	6
Fetching data from kaggle. And Loading the data to S3 bucket.	Creating IAM role for S3 Access from Redshift	Deploying a Redshift Cluster	Deploy EC2 instance ,connect to Redshift, Load the data from S3 to Redshift .	performing data analysis using Redshift Query Editor	Loading the data to Quick sight and identifying trends and patterns through visualization.

## Implementation of Project

## Step 1: Open AWS console

The screenshot shows the AWS Console Home page. On the left, there's a sidebar with 'Recently visited' services: S3, IAM, Amazon Redshift, AWS Glue, RDS, and EC2. Below this is a 'Welcome to AWS' section with a 'Getting started with AWS' link and a rocket icon. To the right, there's an 'AWS Health' section showing 0 open issues and 0 scheduled changes over the past 7 days. Further right is an 'Applications' section which is currently empty, with a 'Create application' button. At the bottom of the page, there are links for CloudShell, Feedback, and various AWS terms like Privacy, Terms, and Cookie preferences.

## Step 2: Create VPC

The screenshot shows the VPC dashboard. On the left, there's a sidebar with options like Your VPCs, Subnets, Route tables, Internet gateways, Egress-only Internet gateways, DHCP option sets, Elastic IPs, Managed prefix lists, Endpoints, Endpoint services, NAT gateways, Peering connections, and Security. The main area is titled 'Resources by Region' and shows various Amazon VPC resources: VPCs (0), Subnets (0), Route Tables (0), Internet Gateways (0), Egress-only Internet Gateways (0), NAT Gateways (0), VPC Peering Connections (0), Network ACLs (0), Security Groups (0), Customer Gateways (0), and DHCP option sets (1). A pink box highlights the 'VPCs' card. On the right, there are sections for 'Service Health' (with a link to view complete service health details), 'Settings' (with links for Zones and Console Experiments), 'Additional Information' (with links for VPC Documentation, All VPC Resources, Forums, and Report an Issue), and 'AWS Network Manager' (with a brief description and a 'Get started with Network Manager' link). The footer includes links for CloudShell, Feedback, and various AWS terms.

Open VPC and create default VPC and add subnets

The screenshot shows the AWS VPC dashboard. On the left, there's a sidebar with options like 'VPC dashboard', 'EC2 Global View', 'Virtual private cloud' (with 'Your VPCs' selected), 'Subnets', 'Route tables', etc. The main area is titled 'Your VPCs' with a search bar. A message says 'No VPCs found in this Region'. On the right, there's an 'Actions' dropdown menu with several options: 'Create VPC' (highlighted with a yellow box), 'Create default VPC', 'Create flow log', 'Edit VPC settings', 'Edit CIDRs', 'Manage middlebox routes', 'Manage tags', and 'Delete VPC'.

This screenshot shows the 'Create default VPC' wizard. The path in the top navigation bar is 'VPC > Your VPCs > Create default VPC'. The main content area is titled 'Default VPC' and contains a brief description: 'A default VPC enables you to launch Amazon EC2 resources without having to create and configure your own VPC and subnets. We'll create a default VPC with a default subnet in each Availability Zone, an internet gateway, and a route table with a route to the internet gateway.' At the bottom right of this section is a 'Create default VPC' button, which is also highlighted with a yellow box. There are 'Cancel' and 'Next Step' buttons below it.

The screenshot shows the AWS VPC dashboard. On the left sidebar, under 'Virtual private cloud', 'Your VPCs' is selected. A table titled 'Your VPCs (1/1)' lists one entry: 'MyVPC' with VPC ID 'vpc-0d7dd00aa70582c12'. The row for 'MyVPC' is highlighted with a pink border. The main panel displays the details for 'vpc-0d7dd00aa70582c12 / MyVPC'. The 'Details' tab is selected, showing the following configuration:

VPC ID	State	DNS hostnames	DNS resolution
vpc-0d7dd00aa70582c12	Available	Enabled	Enabled
Tenancy	DHCP option set	Main route table	Main network ACL
Default	dopt-00ce19470d19a4ce5	-	-
Default VPC	IPv4 CIDR	IPv6 pool	IPv6 CIDR (Network border group)
Yes	172.31.0.0/16	-	-
	Route 53 Resolver DNS	Owner ID	

At the bottom of the main panel, there are links for CloudShell, Feedback, and the footer contains copyright information: © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences.

Screenshot of the AWS VPC Subnets 'Create subnet' page.

**VPC ID:** vpc-003a627b3c7ad7116 (MyVPC) (highlighted with a red box)

**Associated VPC CIDRs:** IPv4 CIDR: 10.0.0.0/16

**Subnet settings:**

- Subnet 1 of 3**
- Subnet name:** Public1A (highlighted with a red box)
- Availability Zone:** Asia Pacific (Mumbai) / ap-south-1a (highlighted with a red box)
- IPv4 VPC CIDR block:** 10.0.0.0/16

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Screenshot of the AWS VPC Subnets 'Create subnet' page, showing additional configuration options.

**Subnet settings:**

- Subnet 1 of 3**
- Subnet name:** Public1A
- Availability Zone:** Asia Pacific (Mumbai) / ap-south-1a
- IPv4 VPC CIDR block:** 10.0.0.0/16
- IPv4 subnet CIDR block:** 10.0.1.0/24 (highlighted with a red box)
- Tags - optional:** Key: Name Value: Public1A

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

**Subnet 2 of 3**

Subnet name  
Create a tag with a key of 'Name' and a value that you specify.  
**Public1B**

The name can be up to 256 characters long.

Availability Zone [Info](#)  
Asia Pacific (Mumbai) / ap-south-1b

IPv4 VPC CIDR block [Info](#)  
Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.  
**10.0.0.0/16**

IPv4 subnet CIDR block  
**10.0.2.0/24** 256 IPs

▼ Tags - optional

Key	Value - optional
<input type="text" value="Name"/>	<input type="text" value="Public1B"/> <a href="#">X</a> <a href="#">Remove</a>

Add new tag  
You can add 49 more tags.  
[Remove](#)

**Subnet 3 of 3**

Subnet name  
**Public1C**

Availability Zone [Info](#)  
Choose the zone which you want to use, or let Amazon choose one for you.  
Asia Pacific (Mumbai) / ap-south-1c

IPv4 VPC CIDR block [Info](#)  
Choose the VPC's IPv4 CIDR block for the subnet. The subnet's IPv4 CIDR must lie within this block.  
**10.0.0.0/16**

IPv4 subnet CIDR block  
**10.0.3.0/24** 256 IPs

▼ Tags - optional

Key	Value - optional
<input type="text" value="Name"/>	<input type="text" value="Public1C"/> <a href="#">X</a> <a href="#">Remove</a>

Add new tag  
You can add 49 more tags.  
[Remove](#)

[Add new subnet](#)

[Cancel](#) [Create subnet](#)

Name	Subnet ID	State	VPC	IPv4 CIDR
Public1A	subnet-01469a62a04a41375	Available	vpc-003a627b5c7ad7116   MyV...	10.0.1.0/24
Public1B	subnet-0c646b4bb238d9f1a	Available	vpc-003a627b5c7ad7116   MyV...	10.0.2.0/24
Public1C	subnet-0ed23317e3f42bf52	Available	vpc-003a627b5c7ad7116   MyV...	10.0.3.0/24

### Step3: Creating Security Groups

There is a default security group we need to change the inbound rules to give access to redshift from S3

Inbound rules

Security group rule ID	Type	Protocol	Port range	Source	Description - optional
-	Redshift	TCP	5439	A... 0.0.0.0/0	Delete
-	SSH	TCP	22	A... 0.0.0.0/0	Delete

Add rule

⚠ Rules with source of 0.0.0.0/0 or ::/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.

Cancel Preview changes Save rules

The screenshot shows the AWS VPC Security Groups page. On the left, the navigation menu includes options like Egress-only internet gateways, DHCP option sets, Elastic IPs, Managed prefix lists, Endpoints, Endpoint services, NAT gateways, Peering connections, Security groups (which is selected), DNS firewall, Rule groups, Domain lists, Network Firewall, Firewalls, Firewall policies, Network Firewall rule groups, and TLS inspection. The main pane displays 'Security Groups (1/1)' with a table showing one entry: Name (sg-0367ad96be6e4cd1b), Security group ID (sg-0367ad96be6e4cd1b), Security group name (default), and VPC ID (vpc-0dc452b3495). Below this is the 'Inbound rules (1)' section, which shows a single rule: Security group rule (sgr-0a4254bd960c1f332), IP version (-), Type (All traffic), and Protocol (All).

#### Step 4 : Open IAM Dashboard to create role for Redshift to access the S3

It is used to securely access the AWS services and resources. It helps to enhance the security by ensuring that only authorized individuals or system can interact with our AWS.

The screenshot shows the AWS IAM Dashboard. The left sidebar has sections for Identity and Access Management (IAM) (selected), Dashboard, Access management (User groups, Users, Roles, Policies, Identity providers, Account settings), Access reports (Access Analyzer, External access, Unused access, Analyzer settings, Credential report), and What's new. The main pane shows 'IAM resources' with counts: User groups (0), Users (0), Roles (17), Policies (1), and Identity providers (0). Below this is a 'What's new' section with a list of recent updates, including: IAM Access Analyzer now simplifies inspecting unused access to guide you toward least privilege, IAM Access Analyzer introduces custom policy checks powered by automated reasoning, Announcing AWS IAM Identity Center APIs for visibility into workforce access to AWS, and New organization-wide IAM condition keys to restrict AWS service-to-service requests.

The screenshot shows the AWS IAM Roles page. On the left, there's a navigation sidebar with options like Dashboard, Access management, Access reports, and Access analyzer. The main area displays a list of 17 IAM roles, each with a checkbox, a role name, and a description of the trusted entity. A yellow box highlights the 'Create role' button at the top right of the list.

Role Name	Trusted Entity
AWSServiceRoleForAutoScaling	AWS Service: autoscaling (Service-Linked Role)
AWSServiceRoleForECS	AWS Service: ecs (Service-Linked Role)
AWSServiceRoleForElasticLoadBalancing	AWS Service: elasticloadbalancing (Service-Linked Role)
AWSServiceRoleForGlobalAccelerator	AWS Service: globalaccelerator (Service-Linked Role)
AWSServiceRoleForOrganizations	AWS Service: organizations (Service-Linked Role)
AWSServiceRoleForRDS	AWS Service: rds (Service-Linked Role)
AWSServiceRoleForRedshift	AWS Service: redshift (Service-Linked Role)
AWSServiceRoleForSSO	AWS Service: sso (Service-Linked Role)
AWSServiceRoleForSupport	AWS Service: support (Service-Linked Role)
AWSServiceRoleForTrustedAdvisor	AWS Service: trustedadvisor (Service-Linked Role)

This screenshot shows the 'Select trusted entity' step of the 'Create role' wizard. It includes a sidebar with steps: Step 1 (Select trusted entity), Step 2 (Add permissions), and Step 3 (Name, review, and create). The main area has three tabs: 'Trusted entity type', 'Use case', and 'Permissions'. The 'Trusted entity type' tab is active, showing four options: 'AWS service' (selected), 'AWS account', 'Web Identity', and 'Custom trust policy'. The 'Use case' tab shows a dropdown menu with 'Redshift' selected. The 'Permissions' tab is partially visible at the bottom. A yellow box highlights the 'Redshift' option in the use case dropdown.

Screenshot of the AWS IAM 'Create role' wizard, Step 2: Add permissions.

The 'Permissions policies' section shows a list of AWS managed policies. The policy **AmazonS3FullAccess** is selected and highlighted with a red box. The 'Next' button at the bottom right is also highlighted with a red box.

Policy name	Type	Description
<a href="#">AmazonDMSRedshiftINSSRole</a>	AWS managed	Provides access to manage S3 settings for...
<b>AmazonS3FullAccess</b>	AWS managed	Provides full access to all buckets via the ...
<a href="#">AmazonS3ObjectLambdaExecutionRolePolicy</a>	AWS managed	Provides AWS Lambda functions permis...
<a href="#">AmazonS3OutpostsFullAccess</a>	AWS managed	Provides full access to Amazon S3 on Out...
<a href="#">AmazonS3ReadOnlyAccess</a>	AWS managed	Provides read only access to Amazon S3 ...
<a href="#">AWSBackupServiceRolePolicyForS3Backup</a>	AWS managed	Policy containing permissions necessary f...
<a href="#">AWSBackupServiceRolePolicyForS3Restore</a>	AWS managed	Policy containing permissions necessary f...
<a href="#">QuickSightAccessForS3StorageManagementAnalytic...</a>	AWS managed	Policy used by QuickSight team to access ...

**Set permissions boundary - optional**

Cancel Previous Next

Screenshot of the AWS IAM 'Create role' wizard, Step 3: Name, review, and create.

The 'Role details' section shows the role name **Success** and a description explaining it allows Redshift clusters to call AWS services on behalf of the user.

The 'Step 1: Select trusted entities' section shows the trust policy JSON:

```

1: {
2:   "Version": "2012-10-17",
3:   "Statement": [
4:     {
5:       "Effect": "Allow",
6:       "Action": "sts:AssumeRole",
7:       "Principal": "*"
8:     }
9:   ],
10:   "Service": "redshift.amazonaws.com"
11: }
12: }
13: }
14: }
15: }
16: }

```

The 'Step 2: Add permissions' section shows the attached policy **AmazonS3FullAccess**.

The 'Step 3: Add tags' section shows a note about adding optional tags. The 'Create role' button at the bottom right is highlighted with a red box.

The screenshot shows the AWS Identity and Access Management (IAM) service interface. In the top right corner, there is a message: "Role S3Access created." Below this, the "Roles" section displays a list of 18 existing roles, each with a role name, trusted entities, and last activity. A new role, "S3Access," is listed at the bottom of the table. The "Create role" button is visible in the top right of the table header. On the left sidebar, under the "Access management" section, the "Roles" option is selected. At the bottom of the page, there are sections for "Access AWS from your non AWS workloads" and "Temporary credentials".

## Step 5: Amazon S3 Bucket creation

The screenshot shows the Amazon S3 service interface. The main heading is "Amazon S3" with the subtext "Store and retrieve any amount of data from anywhere". Below this, a paragraph explains that Amazon S3 is an object storage service. A large "Create a bucket" button is highlighted with a pink box. To the right, there is a "Pricing" section with information about no minimum fees and a link to the AWS Simple Monthly Calculator. At the bottom, there is a "How it works" section featuring a video thumbnail titled "Introduction to Amazon S3". The navigation bar at the top includes "CloudShell" and "Feedback" on the left, and "Global" and "role-IIHT-LAB/lab-session @ 569945402693-mm1" on the right.

The screenshot shows the 'Create bucket' wizard in the AWS S3 console. The 'General configuration' section is active, displaying fields for 'AWS Region' (set to 'Asia Pacific (Mumbai) ap-south-1') and 'Bucket name' (set to 'intrim-bucket'). The 'Object Ownership' section is also visible below it.

**General configuration**

AWS Region: Asia Pacific (Mumbai) ap-south-1

Bucket name: intrim-bucket

Object Ownership

The screenshot shows the 'Default encryption' section of the 'Create bucket' wizard. It includes options for 'Encryption type' (selected: 'Server-side encryption with Amazon S3 managed keys (SSE-S3)'), 'Bucket Key' (selected: 'Enable'), and an 'Advanced settings' button. A note at the bottom states: 'After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.'

**Default encryption**

Encryption type: Server-side encryption with Amazon S3 managed keys (SSE-S3)

Bucket Key: Enable

Advanced settings

After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

The screenshot shows the AWS S3 console with a green success message at the top: "Successfully created bucket 'intrim-bucket'. To upload files and folders, or to configure additional bucket settings, choose View details." Below this, there's an "Account snapshot" section with metrics like Total storage (Pending), Object count (Pending), and Average object size (Pending). A note says you can enable advanced metrics in the "default-account-dashboard" configuration. Below the snapshot, there are tabs for "General purpose buckets" and "Directory buckets", with "General purpose buckets" selected. It shows one bucket named "intrim-bucket" with the following details: Name: intrim-bucket, AWS Region: Asia Pacific (Mumbai) ap-south-1, Access: Objects can be public, Creation date: March 21, 2024, 16:06:53 (UTC+05:30). There are buttons for Copy ARN, Empty, Delete, and Create bucket.

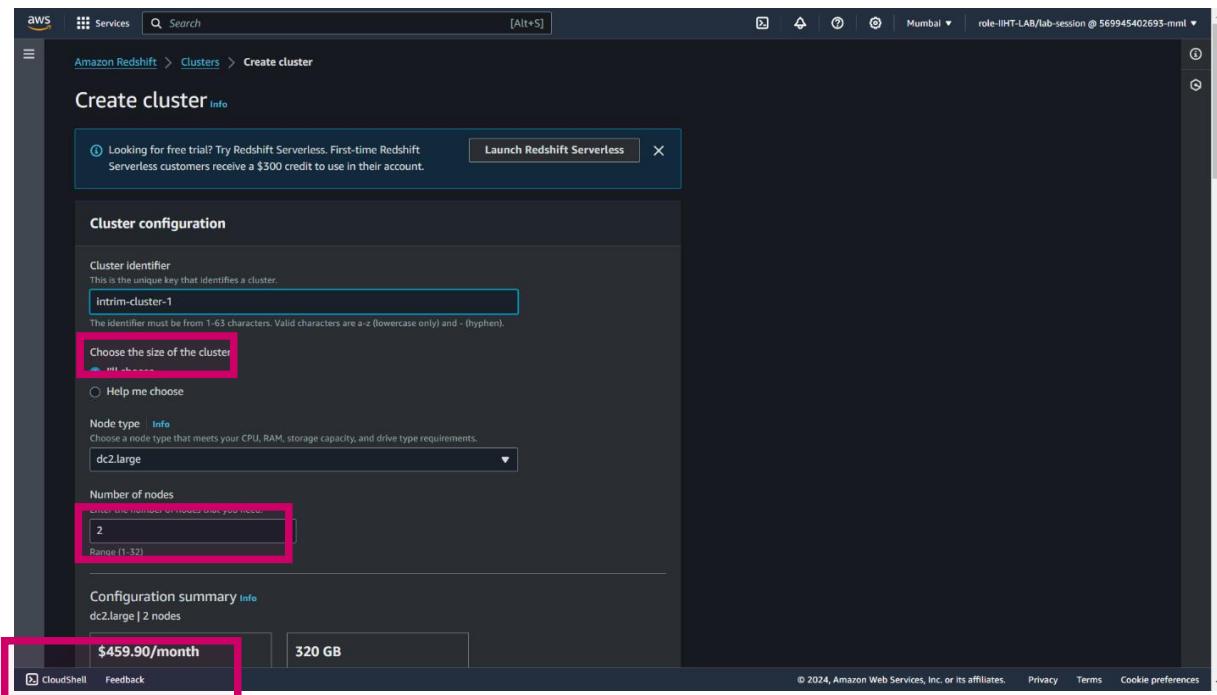
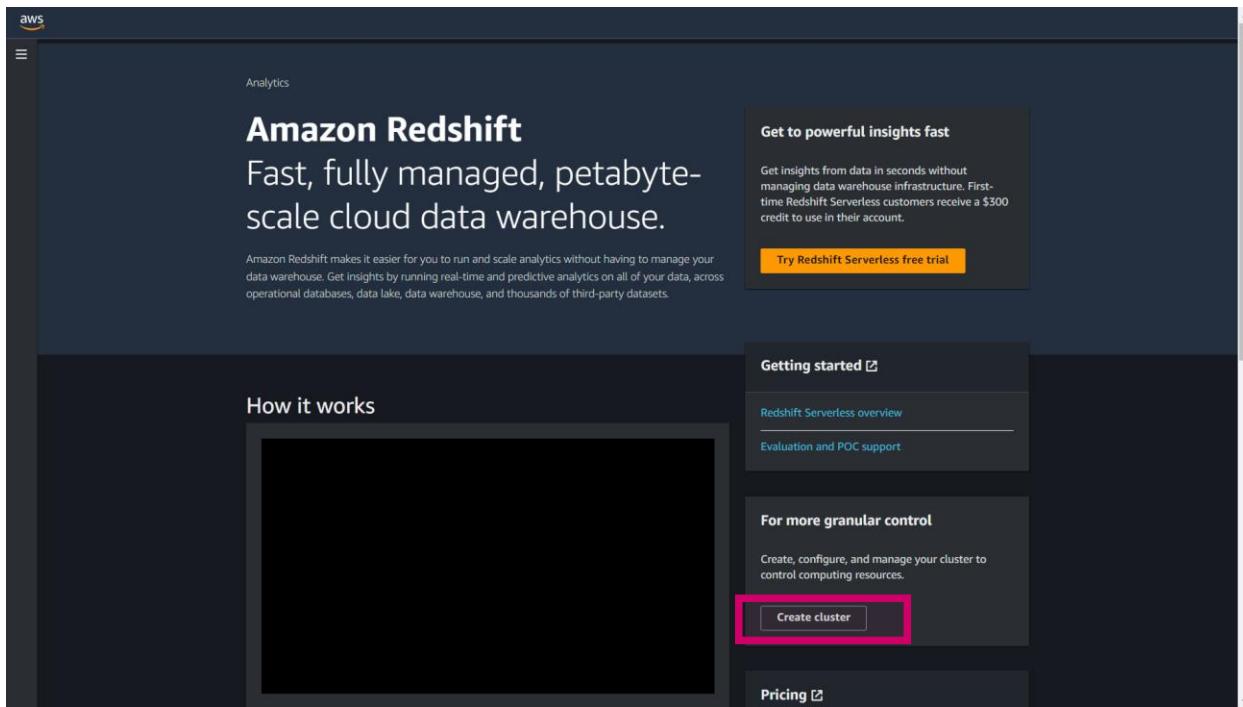
## Step6: Uploading .csv file into S3 bucket

The screenshot shows the AWS S3 console for the "intrim-bucket". The left navigation bar shows "Amazon S3 > Buckets > intrim-bucket". The main page displays the "Objects (0)" section with a message: "No objects. You don't have any objects in this bucket." Below this is a large "Upload" button. Above the upload button, there are several actions: Copy, Copy S3 URI, Copy URL, Download, Open, Delete, Actions, Create folder, and Upload. A note below the actions says: "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." The bottom of the screen includes standard AWS footer links: CloudShell, Feedback, © 2024, Amazon Web Services, Inc. or its affiliates., Privacy, Terms, and Cookie preferences.

The screenshot shows the AWS S3 'Upload' interface. At the top, the navigation bar includes 'Services', a search bar, and a global context switcher for 'role-IIHT-LAB/lab-session @ 569945402693-mm1'. Below the navigation is a breadcrumb trail: 'Amazon S3 > Buckets > intrim-bucket > Upload'. The main area is titled 'Upload' with an 'Info' link. A note at the top says: 'Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. Learn more [?]'. Below this is a large dashed box with the placeholder text 'Drag and drop files and folders you want to upload here, or choose Add files or Add folder.' A table titled 'Files and folders (1 Total, 1.1 MB)' lists one item: 'Adidas US Sales Datasets.csv' (text/csv). Buttons for 'Remove', 'Add files', and 'Add folder' are available above the table. The 'Destination' section shows 's3://intrim-bucket'. The bottom navigation bar includes 'CloudShell', 'Feedback', and copyright information: '© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences'.

The screenshot shows the AWS S3 'Upload success' summary page. At the top, a green banner displays a checkmark icon and the text 'Upload succeeded' followed by 'View details below.' Below the banner is a message: 'The information below will no longer be available after you navigate away from this page.' The main section is titled 'Summary' and shows the destination 's3://intrim-bucket' with a breakdown: 'Succeeded' (1 file, 1.1 MB (100.00%)) and 'Failed' (0 files, 0 B (0%)). Below this is a navigation bar with 'Files and folders' (selected) and 'Configuration'. The 'Files and folders' section lists one item: 'Adidas US S...' (text/csv, 1.1 MB, Status: Succeeded). The bottom navigation bar includes 'CloudShell', 'Feedback', and copyright information: '© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences'.

## Step 7: Creating Cluster



## Step 8: Creating Cluster Subnet Group

**Cluster subnet group details**

**Name**  
You can't modify the name after your subnet group has been created.  
**intrim-subnet-1**

**Description**  
subnet

**Add subnets**

**VPC**  
Choose the VPC that contains the subnets that you want to include in your cluster subnet  
**vpc-0d7dd0a0a70582c12**

**Add all the subnets for this VPC**

**Availability Zone**      **Subnet**  
Choose an Availability Zone      Subnet

**Subnets in this cluster subnet group (1)**

Availability Zone	Subnet ID	CIDR block	IPv6 CIDR block	Action
ap-south-1a	subnet-07e203...	172.31.32.0/20	-	<b>Remove</b>

**Create cluster subnet group**

**Create cluster**

Specify the availability zone to create the cluster in. Otherwise, Amazon Redshift chooses an availability zone for you.  
**No preference**

**Enhanced VPC routing**  
Enabling this option routes network traffic between your cluster and data repositories through a VPC, instead of through the Internet. [Learn more about getting started cluster in VPC](#)

Turn off  
 Turn on

**Publicly accessible**

Turn on Publicly accessible  
Allow public connections to Amazon Redshift.

**Elastic IP address**  
Select the Elastic IP address for connecting to the cluster.

**It can take about ten minutes for the setting to change and connections to succeed.**

**Database configurations** Info

**Maintenance** Info

**Monitoring**

**Backup**

**Create cluster**

The screenshot shows the AWS Redshift console under the 'Clusters' section. A green success message at the top states: 'intrim-cluster-1 has been successfully created.' Below this, there are three main sections: 'Query data using Redshift query editor', 'Work with your client tools', and 'Choose your JDBC or ODBC driver'. In the 'Clusters' table below, one cluster named 'intrim-cluster-1' is listed, showing it's 'Available' in 'ap-south-1a' with 'dc2.large | 2 nodes | 320 GB' storage. The entire cluster row is highlighted with a pink box.

## Step 9 : Creating an EC2 instance

The screenshot shows the AWS EC2 Dashboard. On the left sidebar, under 'Instances', the 'Launch Instances' option is selected. The main area displays 'Resources' and 'Launch instance' sections. The 'Launch instance' section contains a large 'Launch instance' button, which is highlighted with a pink box. Other buttons like 'Migrate a server' and 'View all AWS Free Tier offers' are also visible. The right side of the screen shows 'EC2 Free Tier' information and 'Account attributes'.

The screenshot shows the 'Launch an instance' wizard in the AWS EC2 console. The current step is 'Name and tags'. A pink box highlights the 'Name' field, which contains 'project-redshift-client'. Below it is a section titled 'Application and OS Images (Amazon Machine Image)'. A search bar is present, followed by a 'Quick Start' menu with options like Amazon Linux, macOS, AWS, and Ubuntu, with Ubuntu selected. A pink box also highlights the 'Ubuntu' button in this menu. To the right, the 'Summary' pane shows the instance configuration: 1 instance, Canonical, Ubuntu, 22.04 LTS AMI, t2.micro instance type, redshiftsg security group, and 1 volume(s) - 8 GiB storage. A tooltip for the free tier is visible.

The screenshot shows the 'Launch an instance' wizard in the AWS EC2 console. The current step is 'Instance type'. A pink box highlights the 'linuxKey' key pair selection in the 'Key pair (login)' section. Below it is a 'Network settings' section with Network and Subnet options. To the right, the 'Summary' pane shows the instance configuration: 1 instance, Canonical, Ubuntu, 22.04 LTS AMI, t2.micro instance type, redshiftsg security group, and 1 volume(s) - 8 GiB storage. A tooltip for the free tier is visible.

The screenshot shows two side-by-side AWS EC2 instance creation wizards.

**Left Wizard (Network settings):**

- Security Group:** The "Select existing security group" button is highlighted with a red box. A dropdown menu shows "redshiftsg sg-0582302ab92d09061" selected, also highlighted with a red box.
- Launch Instance Button:** The "Launch instance" button at the bottom right of the summary panel is highlighted with a yellow box.

**Right Wizard (Configure storage):**

- Root Volume:** A configuration panel shows "1x 8 GiB gp2" selected for the root volume, with a note: "Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage".
- Launch Instance Button:** The "Launch instance" button at the bottom right of the summary panel is highlighted with a yellow box.

**Common Elements:**

- Summary Panel:** Both wizards include a "Summary" panel with instance details like "Number of instances: 1", "Software Image (AMI)", "Virtual server type (instance type): t2.micro", and a "Free tier" callout.
- Bottom Buttons:** Both wizards have "Cancel", "Launch instance", and "Review commands" buttons at the bottom.
- Header:** Both screens show the AWS logo, services (S3, EC2, Console Home), search bar, and location (Mumbai).

The screenshot shows the AWS EC2 'Launch an instance' success page. At the top, there's a green success banner stating 'Successfully initiated launch of instance (i-01a7383c38de32c44)'. Below it, a 'Next Steps' section provides links to various EC2 features. A search bar at the top right says 'What would you like to do next with this instance, for example "create alarm" or "create backup"'.

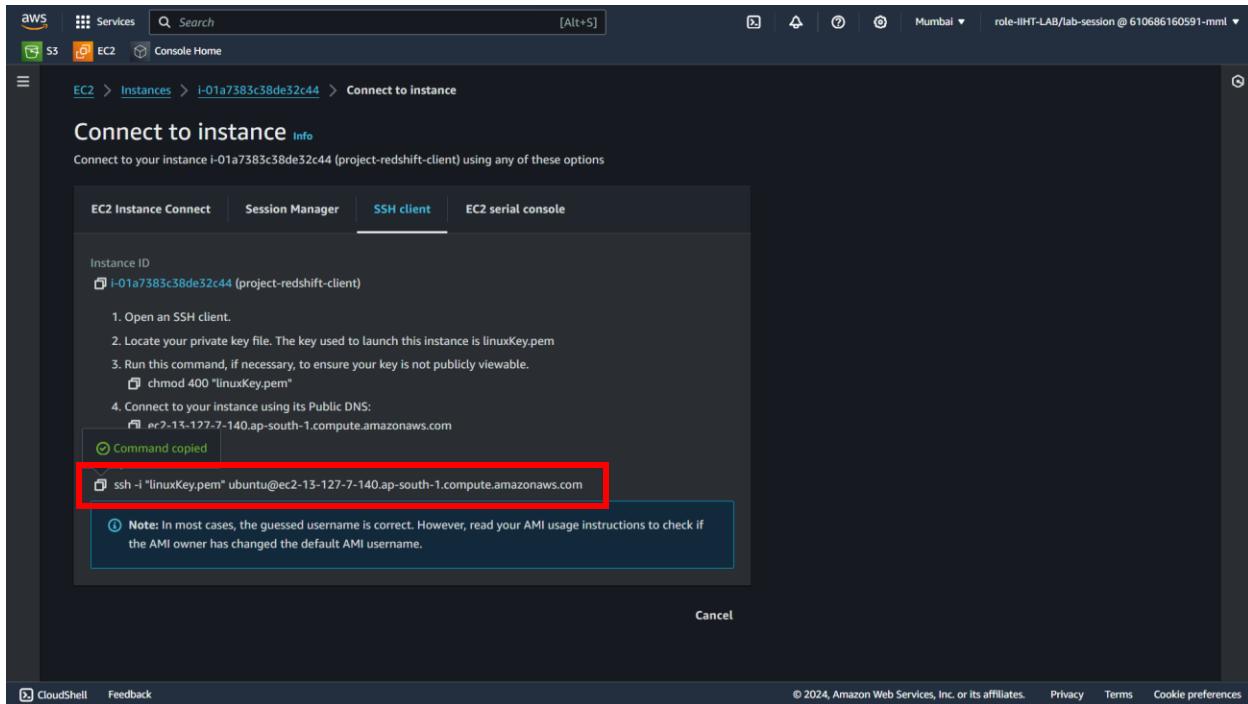
Create billing and free tier usage alerts	Connect to your instance	Connect an RDS database	Create EBS snapshot policy
To manage costs and avoid surprise bills, set up email notifications for billing and free tier usage thresholds. <a href="#">Create billing alerts</a>	Once your instance is running, log into it from your local computer. <a href="#">Connect to instance</a> <a href="#">Learn more</a>	Configure the connection between an EC2 instance and a database to allow traffic flow between them. <a href="#">Connect an RDS database</a> <a href="#">Create a new RDS database</a> <a href="#">Learn more</a>	Create a policy that automates the creation, retention, and deletion of EBS snapshots <a href="#">Create EBS snapshot policy</a>

Below this, there are four more sections: 'Manage detailed monitoring', 'Create Load Balancer', 'Create AWS budget', and 'Manage CloudWatch alarms'. The bottom of the page includes standard AWS footer links: CloudShell, Feedback, © 2024, Amazon Web Services, Inc. or its affiliates., Privacy, Terms, and Cookie preferences.

The screenshot shows the AWS EC2 Instances page. The left sidebar has categories like Instances, Images, Elastic Block Store, and Network & Security. The main area displays a table of instances. One row is highlighted with a red box:

<input type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
<input type="checkbox"/>	project-redshift...	i-01a7383c38de32c44	Running	t2.micro	Initializing	<a href="#">View alarms</a>	ap-south-1a	ec2-13-

Below the table, a modal window titled 'Select an instance' is open, showing the same instance details. The bottom of the page includes standard AWS footer links: CloudShell, Feedback, © 2024, Amazon Web Services, Inc. or its affiliates., Privacy, Terms, and Cookie preferences.



## Step 10 : Open terminal and run the commands to create a table in Ec2 instance.

These are the commands to run the table in Ec2 instance.

- “sudo apt-get update
- sudo apt-get install -y postgresql-client && psql --version
- psql -h redshift-cluster-1.cjlluv3z3ar1.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d dev -W”

Syntax : pql -h <hostname> -p <port> -U <user> -d <database> -W

```

PS C:\Users\2320846> cd downloads
PS C:\Users\2320846\downloads> ssh -i "linuxKey.pem" ubuntu@ec2-13-127-7-140.ap-south-1.compute.amazonaws.com
The authenticity of host 'ec2-13-127-7-140.ap-south-1.compute.amazonaws.com (13.127.7.140)' can't be established.
ED25519 key fingerprint is SHA256:iAINGLoqfOLVshcfpgNaja3pJ4SzbJHmNFFV2FVPo0.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'ec2-13-127-7-140.ap-south-1.compute.amazonaws.com' (ED25519) to the list of known hosts.
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.5.0-1014-aws x86_64)

 * Documentation: https://help.ubuntu.com
 * Management: https://landscape.canonical.com
 * Support: https://ubuntu.com/pro

 System information as of Thu Mar 21 14:42:13 UTC 2024

 System load: 0.0      Processes:          98
 Usage of /: 28.4% of 7.57GB   Users logged in:     0
 Memory usage: 19%           IPv4 address for eth0: 10.0.1.38
 Swap usage:  0%

Expanded Security Maintenance for Applications is not enabled.

0 updates can be applied immediately.

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

The list of available updates is more than a week old.
To check for new updates run: sudo apt update

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

ubuntu@ip-10-0-1-38:~$
```

- Connecting to the Redshift cluster from ec2 through postgreSQL client

```

ubuntu@ip-10-0-1-38:~$ psql -h redshift-cluster-1.c4hapfhrjfor.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d dev -W
Password:
psql (14.11 (Ubuntu 14.11-0ubuntu0.22.04.1), server 8.0.2)
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384, bits: 256, compression: off)
Type "help" for help.
```

- create database for our use

```

dev=# create database project ;
CREATE DATABASE
dev=#

```

- Again we connect to Redshift and login to the created database
- Create a table

```

dev@ip-10-0-1-38:~$ psql -h redshift-cluster-1.c4hapfhjfor.ap-south-1.redshift.amazonaws.com -p 5439 -U awsuser -d project -W
Password:
psql (14.11 (Ubuntu 14.11-0ubuntu0.22.04.1), server 8.0.2)
SSL connection (protocol: TLSv1.2, cipher: ECDHE-RSA-AES256-GCM-SHA384, bits: 256, compression: off)
Type "help" for help.

project=# \d
Did not find any relations.
project=#
CREATE TABLE public.sales_data (
    retailer character varying(255) ENCODE lzo,
    retailer_id integer ENCODE az64,
    Invoice_date character varying(255) ENCODE lzo,
    region
        character varying(255) ENCODE lzo,
    state character varying(255) ENCODE lzo,
    city character varying(255) ENCODE lzo,
    product character varying(255) ENCODE lzo,
    price_per_unit numeric(10, 2) ENCODE az64,
    units_sold integer ENCODE az64,
    total_sales numeric(10, 2) ENCODE az64,
    operating_profit numeric(10, 2) ENCODE az64,
    operating_margin numeric(5, 2) ENCODE az64,
    sales_method character varying(255) ENCODE lzo
) DISTSTYLE AUTO;
CREATE TABLE
project=#

```

## Step 11: Load the data from S3 to Redshift

To load the data we need S3 object URI

Amazon S3 > Buckets > redshift-project-4526 > Adidas US Sales Datasets.csv

**Properties**    Permissions    Versions

**Object overview**

Owner: e7531b78a3b63d9a9820b24960d752aafeeb1c776eff27916236165e3f6dd  
be

AWS Region: Asia Pacific (Mumbai) ap-south-1

Last modified: March 21, 2024, 10:04:29 (UTC+05:30)

Size: 1006.0 KB

Type: CSV

Key: Adidas US Sales Datasets.csv

S3 URI copied: s3://redshift-project-4526/Adidas US Sales Datasets.csv

Amazon Resource Name (ARN): arn:aws:s3:::redshift-project-4526/Adidas US Sales Datasets.csv

Entity tag (Etag): b61bd3dfcfba59d9fcf83e63df837d6

Object URL: https://redshift-project-4526.s3.ap-south-1.amazonaws.com/Adidas+US+Sales+Datasets.csv

We need the IAM S3 Role ARN that we created earlier

The screenshot shows the AWS IAM Roles page. A search bar at the top has 's3' typed into it, resulting in one match. The search result, 's3-full-access', is highlighted with a pink border. Below the search results, there's a section titled 'Roles Anywhere' with three options: 'Access AWS from your non AWS workloads', 'X.509 Standard', and 'Temporary credentials'. At the bottom of the page, there are links for CloudShell, Feedback, and a copyright notice.

The screenshot shows the detailed view of the 's3-full-access' role. The ARN of the role, 'arn:aws:iam::610686160591:role/s3-full-access', is highlighted with a pink border. Below the ARN, there's a 'Summary' section with information like creation date (March 19, 2024) and last activity (42 minutes ago). Under the 'Permissions' tab, there's a table showing a single attached policy, 'AmazonS3FullAccess', which is an AWS managed policy. The table includes columns for Policy name, Type, Attached entities, and a status indicator.

Then use the following command to load the data

```

project=# 
COPY project.public.sales_data FROM 's3://redshift-project-4526/Adidas US Sales Datasets.csv' IAM_ROLE 'arn:aws:iam::610686160591:role/s3-full-access' FORMAT AS CSV DELIMITER ',' QUOTE '\"' IGNOREHEADER 1 REGION AS 'ap-south-1' ;

INFO: Load into table 'sales_data' completed, 9648 record(s) loaded successfully.
COPY
project=# select * from sales_data limit 10 ;

```

COPY <tablename> FROM <S3 bucket URI> IAM\_ROLE <IAM role A> FORMAT AS CSV  
DELIMITER<file delimiter> REGION AS <cluster region>

retailer	retailer_id	invoice_date	region	state	city	product	price_per_unit	units_sold	total_sales	operating_profit	operating_margin	sales_method
Foot Locker	1185732	01-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1200	600000.00	300000.00	50.00	In-store
Foot Locker	1185732	02-01-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	1000	500000.00	150000.00	30.00	In-store
Foot Locker	1185732	03-01-2020	Northeast	New York	New York	Women's Street Footwear	40.00	1000	400000.00	140000.00	35.00	In-store
Foot Locker	1185732	04-01-2020	Northeast	New York	New York	Women's Athletic Footwear	45.00	850	382500.00	133875.00	35.00	In-store
Foot Locker	1185732	05-01-2020	Northeast	New York	New York	Women's Apparel	60.00	900	540000.00	150000.00	30.00	In-store
Foot Locker	1185732	06-01-2020	Northeast	New York	New York	Women's Apparel	50.00	1000	500000.00	125000.00	25.00	In-store
Foot Locker	1185732	07-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1250	625000.00	312500.00	50.00	In-store
Foot Locker	1185732	08-01-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	900	450000.00	135000.00	30.00	Outlet
Foot Locker	1185732	21-01-2020	Northeast	New York	New York	Women's Street Footwear	40.00	950	380000.00	133000.00	35.00	Outlet
Foot Locker	1185732	22-01-2020	Northeast	New York	New York	Women's Athletic Footwear	45.00	825	371250.00	129930.00	35.00	Outlet
Foot Locker	1185732	23-01-2020	Northeast	New York	New York	Men's Apparel	60.00	900	540000.00	162000.00	30.00	Outlet
Foot Locker	1185732	24-01-2020	Northeast	New York	New York	Women's Apparel	50.00	1000	500000.00	125000.00	25.00	Outlet
Foot Locker	1185732	25-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1220	610000.00	305000.00	50.00	Outlet
Foot Locker	1185732	26-01-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	950	450000.00	135000.00	30.00	Outlet
Foot Locker	1185732	27-01-2020	Northeast	New York	New York	Women's Street Footwear	40.00	950	380000.00	133000.00	35.00	Outlet
Foot Locker	1185732	28-01-2020	Northeast	New York	New York	Women's Athletic Footwear	45.00	800	360000.00	126000.00	35.00	Outlet
Foot Locker	1185732	29-01-2020	Northeast	New York	New York	Men's Apparel	60.00	850	510000.00	153000.00	30.00	Outlet
Foot Locker	1185732	30-01-2020	Northeast	New York	New York	Women's Apparel	50.00	950	475000.00	118750.00	25.00	Outlet
Foot Locker	1185732	31-01-2020	Northeast	New York	New York	Men's Street Footwear	50.00	1200	600000.00	300000.00	50.00	Outlet
Foot Locker	1185732	01-02-2020	Northeast	New York	New York	Men's Athletic Footwear	50.00	900	450000.00	135000.00	30.00	Outlet
(28 rows)												

Database schema :

	Field	Type	NL	CMP
A	retailer	character varying(50)	NULL	lzo
#	retailerid	integer	NULL	az64
A	invoice_date	character varying(50)	NULL	lzo
A	region	character varying(40)	NULL	lzo
A	state	character varying(40)	NULL	lzo
A	city	character varying(40)	NULL	lzo
A	product	character varying(100)	NULL	lzo
#	priceperunit	numeric(10,2)	NULL	az64
#	units_sold	integer	NULL	az64
#	total_sales	numeric(12,2)	NULL	az64
#	operatingprofit	numeric(12,2)	NULL	az64
#	operatingmargin	numeric(5,2)	NULL	az64
A	salesmethod	character varying(50)	NULL	lzo

## Step 12: Data Cleaning

Looking at the data we found some data mismatch , so we have to clean the data first before performing any analysis on the data .

- First we have to change the invoice\_date colum from character variable to date format

```
-- changing the invoice_date from character varying(50) to date format

-- step 1 : Adding new column
ALTER TABLE sales_data
ADD COLUMN order_date DATE;

-- Step 2: Update the new column with the converted values
UPDATE sales_data
SET order_date = TO_DATE(invoice_date, 'DD-MM-YYYY');

-- Step 3: Drop the old column
ALTER TABLE sales_data
DROP COLUMN INVOICE_DATE;

SELECT * FROM SALES_DATA;
```

Now the schema looking like this

	Field	Type	NL	CMP
A	retailer	character varying(50)	NULL	lzo
#	retailerid	integer	NULL	az64
A	region	character varying(40)	NULL	lzo
A	state	character varying(40)	NULL	lzo
A	city	character varying(40)	NULL	lzo
A	product	character varying(100)	NULL	lzo
#	priceperunit	numeric(10,2)	NULL	az64
#	units_sold	integer	NULL	az64
#	total_sales	numeric(12,2)	NULL	az64
#	operatingprofit	numeric(12,2)	NULL	az64
#	operatingmargin	numeric(5,2)	NULL	az64
A	salesmethod	character varying(50)	NULL	lzo
#	order_date	date	NULL	az64

- We observed that in the dataset the some column's data was mismatching , these columns are

- Total\_Sales column
- Operating profit column

```
select priceperunit , units_sold , total_sales , operatingprofit , operatingmargin from sales_data ;
```

	priceperunit	units_sold	total_sales	operatingprofit	operatingmargin
□	50	1200	600000	300000	50
□	50	1000	500000	150000	30
□	40	1000	400000	140000	35
□	45	850	382500	133875	35
□	60	900	540000	162000	30
□	50	1000	500000	125000	25
□	50	1250	625000	312500	50

So we perform some more data cleaning for

Total\_Sales :

```
25 -- Data cleaning for total sales column
26
27 -- ADDING A NEW COL
28 ALTER TABLE sales_data
29 ADD COLUMN TOTAL_COST INT;
30
31
32 -- MANUALLY MULTIPLY UNITS SOLD * PRICE PER UNIT
33 UPDATE sales_data
34 SET TOTAL_COST = PRICEPERUNIT * UNITS SOLD;
35
36
37 --DROP OLD COLUMN
38 ALTER TABLE SALES_DATA
39 DROP COLUMN TOTAL_SALES;
40
41
42 -- RENAME COLUMN TO THE ORIGINAL NAME
43 ALTER TABLE SALES_DATA
44 RENAME TOTAL_COST TO TOTAL_SALES;
45
```

	priceperunit	units_sold	total_sales	operatingprofit	operatingmargin
□	50	1200	60000	300000	50
□	50	1000	50000	150000	30
□	40	1000	40000	140000	35
□	45	850	38250	133875	35
□	60	900	54000	162000	30
□	50	1000	50000	125000	25

Operating profit :

```

52  -- Data cleaning for operating profit column
53
54
55  -- add NEW PROFIT col
56  ALTER TABLE sales_data
57  ADD COLUMN new_profit INT;
58
59  -- MANUALLY CALCULATING THE PROFIT
60  UPDATE sales_data
61  SET new_profit = ( operatingmargin * total_sales ) / 100 ;
62
63  --DROP OLD COLUMN OPERATING_MARGIN
64  ALTER TABLE SALES_DATA
65  DROP COLUMN operatingprofit;
66
67  -- RENAME THE NEW COL TO THE ORIGINAL COLUMN NAME
68  ALTER TABLE SALES_DATA
69  RENAME new_profit TO operating_profit;
70

```

	priceperunit	units_sold	total_sales	operating_profit	operatingmargin
□	50	1200	60000	30000	50
□	50	1000	50000	15000	30
□	40	1000	40000	14000	35
□	45	850	38250	13387	35
□	60	900	54000	16200	30
□	50	1000	50000	12500	25

Now we will move to Redshift cluster to run the queries

Note: We can directly run queries from the instance but for better visuals we used Redshift's query editor v2

### Step 13: Running SQL queries to analyze the sales trends

SQL queries to aggregate sales data by product, region, and time period.

```

1
2  -- aggregate sales data by product
3  SELECT
4    product,
5    SUM(total_sales) AS total_sales
6  FROM sales_data
7    GROUP BY product
8    Order by total_sales DESC;
9
10

```

Result 1 (6)

	product	total_sales
□	Men's Street Footwear	27680769
□	Women's Apparel	23870985
□	Men's Athletic Footwear	20577180
□	Women's Street Footwear	17201563
□	Men's Apparel	16520632
□	Women's Athletic Footwear	14315521

## By Region

```

7   --aggregate sales data by region
8 ↘ SELECT
9     product, region ,
10    SUM(total_sales) AS total_sales
11 ↘ FROM sales_data
12   GROUP BY region , product
13   Order by region;
14

```

Result 1 (30)

	product	region	total_sales
□	Men's Apparel	Midwest	2223786
□	Women's Apparel	Midwest	3453008
□	Men's Street Footwear	Midwest	4707360
□	Men's Athletic Footwear	Midwest	2619289
□	Women's Street Footwear	Midwest	1997448
□	Women's Athletic Footwear	Midwest	1673543
□	Men's Street Footwear	Northeast	6841324
□	Men's Athletic Footwear	Northeast	3895862
□	Women's Street Footwear	Northeast	3152823
□	Women's Athletic Footwear	Northeast	2668013
□	Men's Apparel	Northeast	3475037
□	Women's Apparel	Northeast	5045208
□	Men's Apparel	South	2811194
□	Women's Apparel	South	4224937
□	Men's Street Footwear	South	4048261
□	Men's Athletic Footwear	South	3647045
□	Women's Street Footwear	South	3242822
□	Women's Athletic Footwear	South	2629097
□	Men's Apparel	Southeast	3183237
□	Women's Apparel	Southeast	4109786
□	Men's Street Footwear	Southeast	4693836
□	Men's Athletic Footwear	Southeast	3653645
□	Women's Street Footwear	Southeast	3059884
□	Women's Athletic Footwear	Southeast	2674048
□	Men's Apparel	West	4827378
□	Women's Apparel	West	7038046
□	Men's Street Footwear	West	7389988
□	Men's Athletic Footwear	West	6761339

## By Time period

```

18 -- aggregate sales data by time period
19 ∵ SELECT
20   product ,
21   TO_CHAR(order_date::DATE,'MONTH' ) as month,
22   sum(total_sales) AS total_sales
23 ∵ FROM sales_data
24   GROUP BY product , month , REGION
25   ORDER BY total_sales desc ;
26

```

Virtual Servers in the Cloud 2

Result 1 (100)

product	month	total_sales
Women's Apparel	MAY	903900
Men's Street Footwear	JULY	894762
Men's Street Footwear	MAY	875565
Women's Apparel	JANUARY	870171
Men's Street Footwear	DECEMBER	865236
Women's Apparel	SEPTEMBER	845374
Men's Athletic Footwear	JANUARY	827784
Men's Athletic Footwear	MAY	798378
Men's Athletic Footwear	AUGUST	796036
Women's Street Footwear	SEPTEMBER	777403
Men's Street Footwear	AUGUST	775264
Men's Street Footwear	APRIL	771903
Women's Apparel	APRIL	757109
Men's Athletic Footwear	SEPTEMBER	753147
Men's Street Footwear	NOVEMBER	734340
Men's Street Footwear	JANUARY	733706
Men's Street Footwear	SEPTEMBER	732989
Women's Apparel	JUNE	727454
Men's Athletic Footwear	JULY	720404
Women's Apparel	JULY	718496
Men's Street Footwear	FEBRUARY	717585
Men's Athletic Footwear	FEBRUARY	717550
Women's Street Footwear	JANUARY	711685
Men's Street Footwear	JANUARY	699308
Women's Apparel	AUGUST	676628
Men's Street Footwear	JULY	674139
Women's Apparel	FEBRUARY	650157
Men's Street Footwear	JUNE	642840
Men's Street Footwear	MARCH	633749
Women's Apparel	DECEMBER	630685
Women's Street Footwear	MAY	627273
Women's Apparel	OCTOBER	623236
Men's Athletic Footwear	APRIL	620850
Men's Street Footwear	DECEMBER	620123
Men's Street Footwear	AUGUST	619224
Men's Street Footwear	SEPTEMBER	618857
Women's Street Footwear	FEBRUARY	617181
Women's Street Footwear	JULY	604788
Men's Apparel	MAY	601740
Women's Street Footwear	APRIL	595061
Men's Apparel	JULY	576730
Men's Street Footwear	REINC	5667140

Result 2 (100)

product	month	total_sales
Women's Street Footwear	JULY	600100
Men's Apparel	MAY	601740
Women's Street Footwear	APRIL	595061
Men's Apparel	JULY	578730
Men's Street Footwear	JUNE	566219
Men's Athletic Footwear	AUGUST	561319
Women's Athletic Footwear	MAY	559951
Men's Street Footwear	AUGUST	558716
Men's Street Footwear	MAY	557409
Men's Street Footwear	DECEMBER	552861
Men's Street Footwear	JULY	551740
Men's Athletic Footwear	JUNE	549573
Women's Athletic Footwear	JANUARY	543866
Men's Apparel	JANUARY	541701
Women's Street Footwear	JUNE	539924
Women's Street Footwear	AUGUST	538454
Women's Apparel	NOVEMBER	538436
Women's Apparel	DECEMBER	532298
Women's Apparel	JULY	525885
Men's Street Footwear	MARCH	524016
Women's Apparel	JULY	522292
Men's Apparel	SEPTEMBER	513620
Men's Street Footwear	OCTOBER	511351
Men's Street Footwear	OCTOBER	508353
Men's Street Footwear	APRIL	506990
Men's Street Footwear	JULY	503231
Women's Athletic Footwear	SEPTEMBER	501163
Men's Apparel	FEBRUARY	497137
Women's Apparel	NOVEMBER	495827
Women's Athletic Footwear	JULY	494946
Women's Apparel	SEPTEMBER	494112
Men's Athletic Footwear	DECEMBER	492676
Men's Athletic Footwear	MARCH	487809
Men's Athletic Footwear	SEPTEMBER	485515
Women's Apparel	APRIL	485057
Men's Street Footwear	AUGUST	484315
Women's Athletic Footwear	FEBRUARY	483639
Women's Street Footwear	OCTOBER	483456
Men's Athletic Footwear	OCTOBER	482089
Men's Athletic Footwear	AUGUST	480367
Men's Apparel	AUGUST	478597
Women's Street Footwear	MARCH	474928

Identify top-selling products by calculating total sales or units sold.

## By total sales

```
28 -- Identify top-selling products by calculating total sales
29
30 ✓ select
31   ✓ product
32   ✓ from (
33   ✓   select
34     ✓ product ,
35     ✓ sum(total_sales ) as total_sales
36   ✓ from sales_data
37   ✓ group by product
38   ✓ order by total_sales DESC
39 )
40 ;
41
```

	product
□	Men's Street Footwear
□	Women's Apparel
□	Men's Athletic Footwear
□	Women's Street Footwear
□	Men's Apparel
□	Women's Athletic Footwear

## By units sold

```
44 -- Identify top-selling products by calculating units sold.
45
46 select
47   product as top_selling_products
48   from (
49   ✓   select
50     ✓ product ,
51     ✓ sum(units_sold) as units_sold
52   ✓ from sales_data
53   ✓ group by product
54   ✓ order by units_sold DESC
55 )
56 ;
```

	top_selling_products
□	Men's Street Footwear
□	Men's Athletic Footwear
□	Women's Apparel
□	Women's Street Footwear
□	Women's Athletic Footwear
□	Men's Apparel

Determine highest revenue-generating regions by analyzing sales revenue per region.

```

55
56    -- Determine the highest revenue generating regions by analyzing sales revenue per region
57
58    select
59        region ,
60        sum(total_sales ) as total_sales
61    from sales_data
62        group by region
63        order by total_sales  desc ;
64
65
66
67
68
69

```

	region	total_sales
1	West	36436157
2	Northeast	25078267
3	Southeast	21374436
4	South	20603356
5	Midwest	16674434

The screenshot shows the AWS Redshift Query Editor v2 interface. The main area displays a SQL query:

```

1  select count(*) from sales_data;
2
3  ALTER TABLE sales_data
4  ADD COLUMN order_date DATE;
5
6  -- Step 2: Update the new column with the converted values
7  UPDATE sales_data

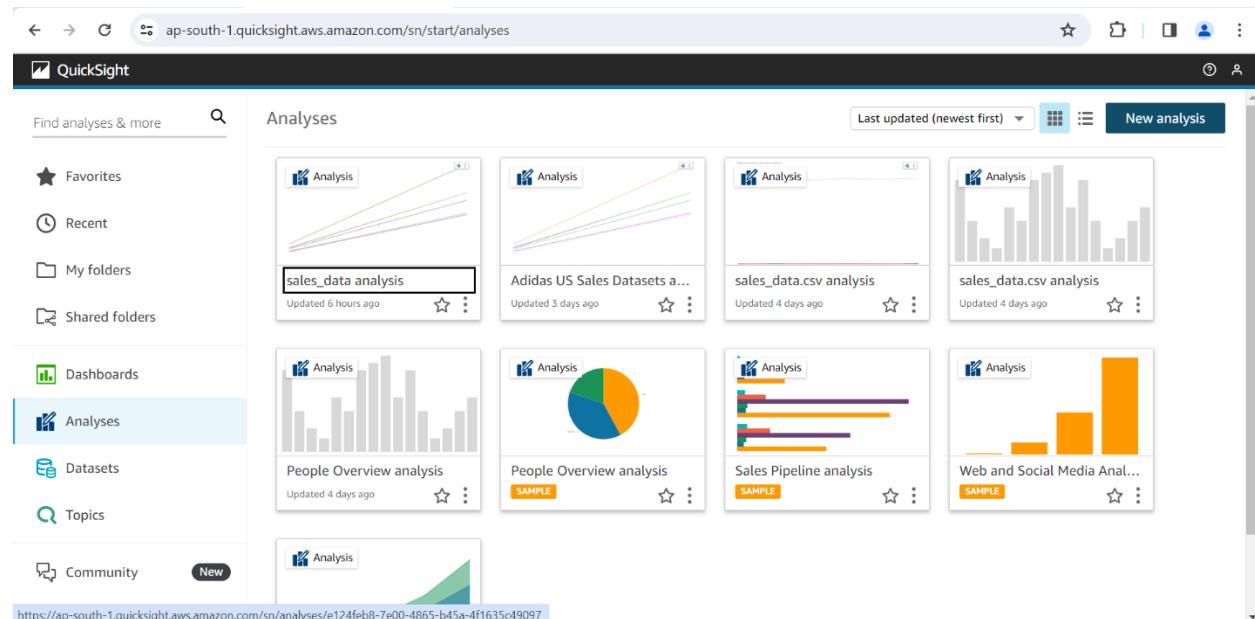
```

The sidebar on the left shows the database structure for the cluster "redshift-cluster-1". Under the "Tables" section, the "sales\_data" table is highlighted with a red box.

# Visualization

For dashboarding we will be using AWS QuickSight Service

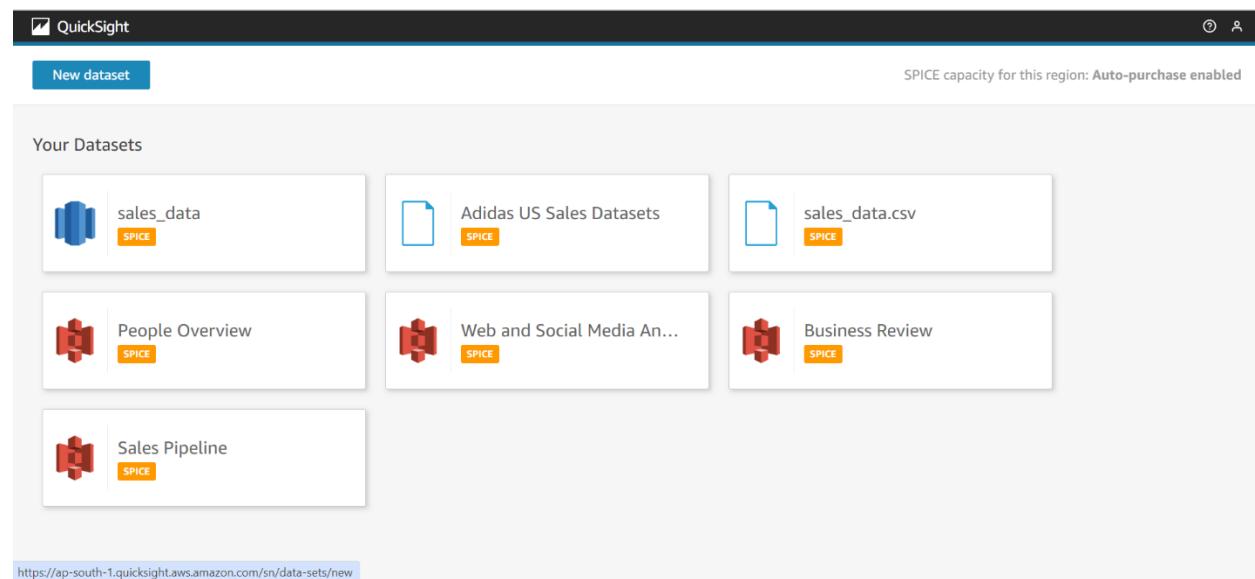
We first start importing data to QuickSight from Redshift Database



The screenshot shows the AWS QuickSight interface. On the left, there's a sidebar with navigation links: Favorites, Recent, My folders, Shared folders, Dashboards, Analyses (which is selected), Datasets, Topics, and Community. The main area is titled "Analyses" and displays several cards for different analyses:

- sales\_data analysis** (Updated 6 hours ago)
- Adidas US Sales Datasets a...** (Updated 3 days ago)
- sales\_data.csv analysis** (Updated 4 days ago)
- sales\_data.csv analysis** (Updated 4 days ago)
- People Overview analysis** (Updated 4 days ago)
- People Overview analysis** (SAMPLE)
- Sales Pipeline analysis** (SAMPLE)
- Web and Social Media Anal...** (SAMPLE)

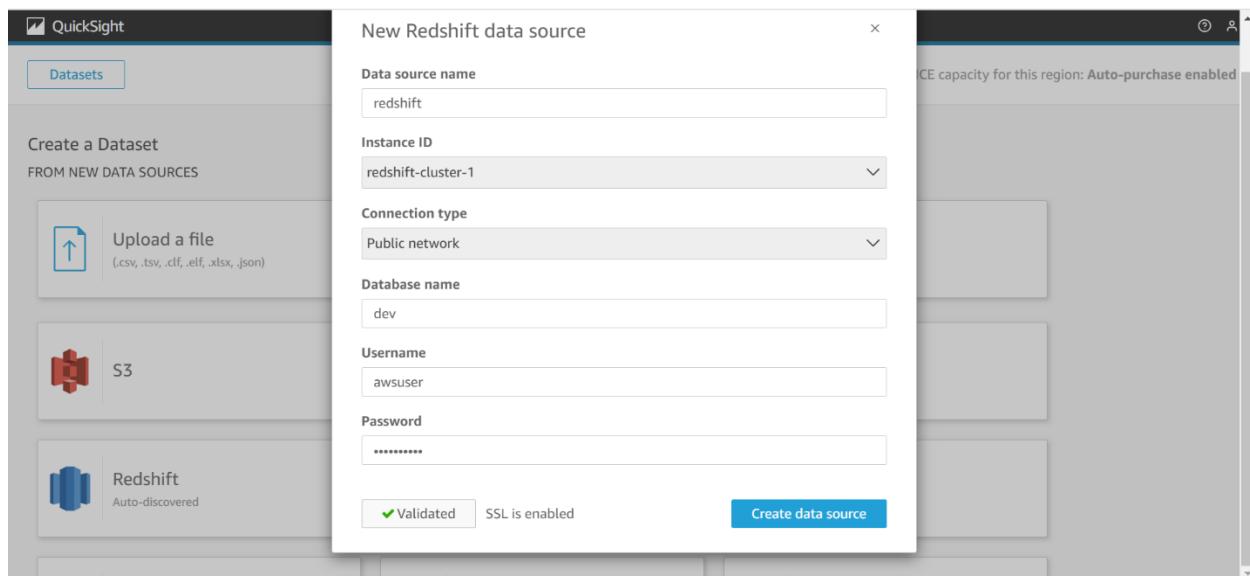
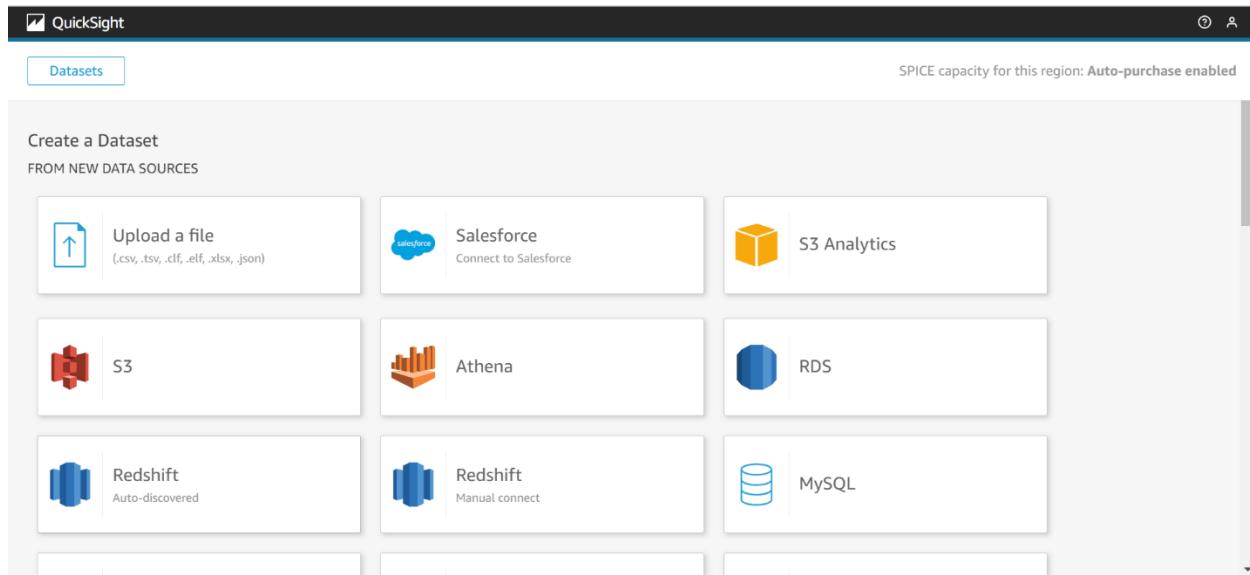
At the bottom of the page, the URL is shown: <https://ap-south-1.quicksight.aws.amazon.com/sn/analyses/e124feb8-7e00-4865-b45a-4f1635c49097>

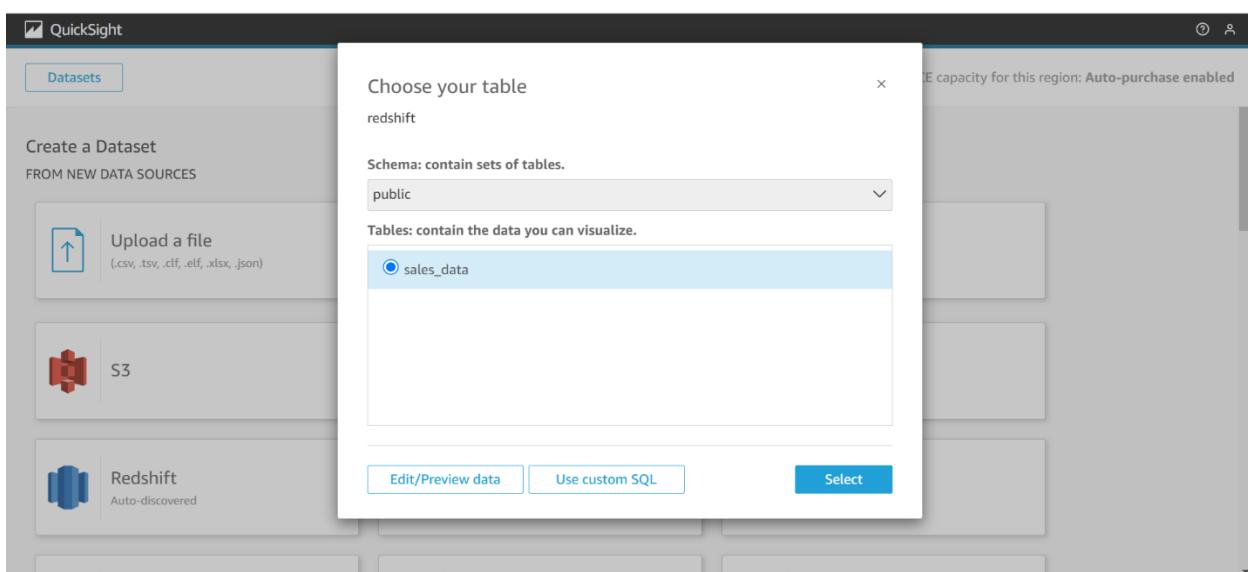
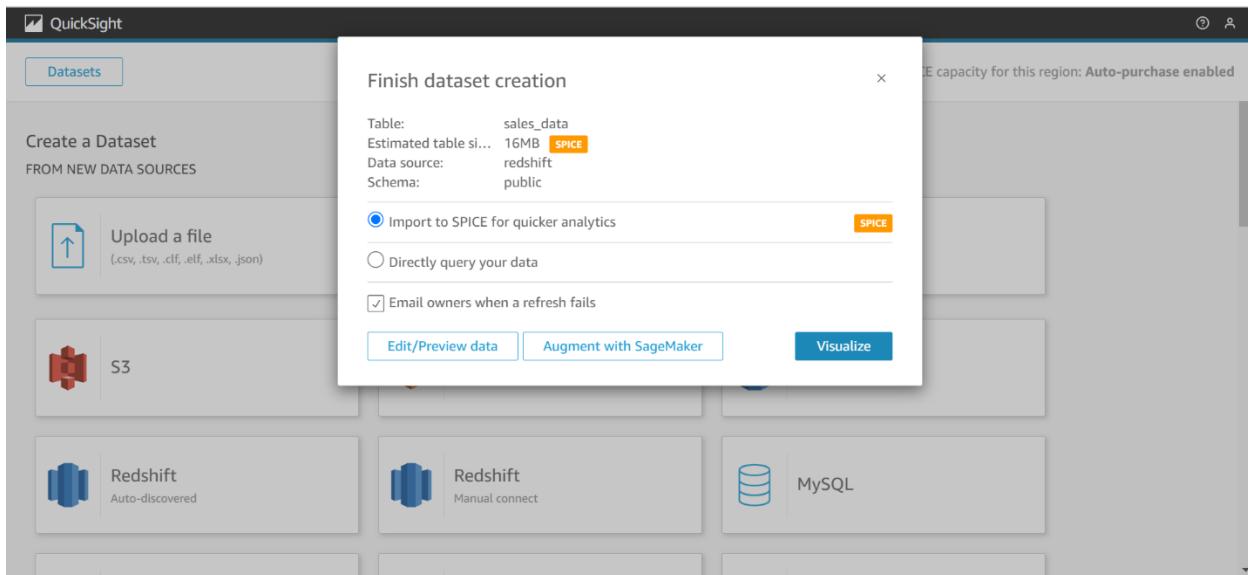


The screenshot shows the AWS QuickSight interface. At the top, there's a "New dataset" button and a note about SPICE capacity: "SPICE capacity for this region: Auto-purchase enabled". The main area is titled "Your Datasets" and lists six datasets:

- sales\_data** (SPICE)
- Adidas US Sales Datasets** (SPICE)
- sales\_data.csv** (SPICE)
- People Overview** (SPICE)
- Web and Social Media An...** (SPICE)
- Sales Pipeline** (SPICE)

At the bottom of the page, the URL is shown: <https://ap-south-1.quicksight.aws.amazon.com/sn/data-sets/new>





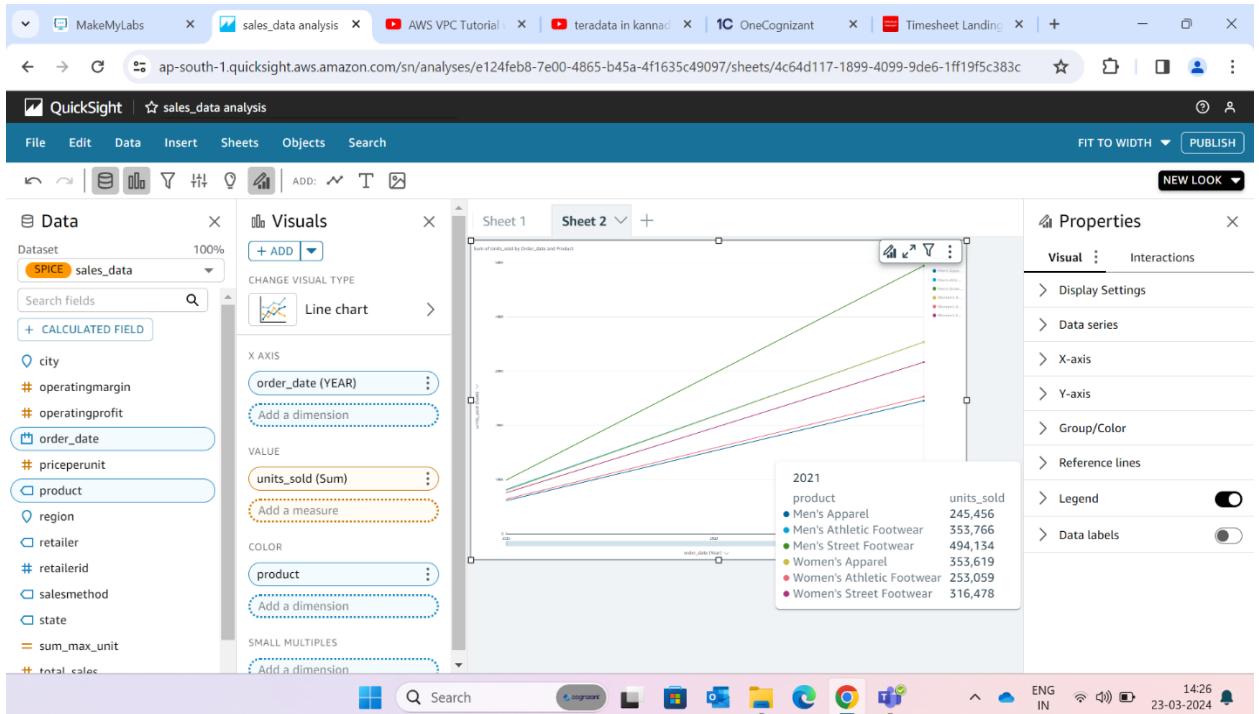
The screenshot shows the QuickSight interface with the following details:

- Top Bar:** QuickSight logo, project name "sales\_data analysis", and a help icon.
- Header:** File, Edit, Data, Insert, Sheets, Objects, Search.
- Data Panel:** Dataset dropdown set to "SPICE sales\_data" (100% complete), search fields input, and a "+ CALCULATED FIELD" button.
- Visuals Panel:** "CHANGE VISUAL TYPE" dropdown, followed by a grid of 4x6 visual icons including bar charts, pie charts, line graphs, and other data visualization types.
- Sheet 1:** A blank canvas area with a placeholder text "AutoGraph" and instructions "Add 1 or more fields to build a visual."
- Message Pop-up:** "Import complete: 100% success" and "9,648 rows were imported to SPICE", with a note that "0 rows were skipped".

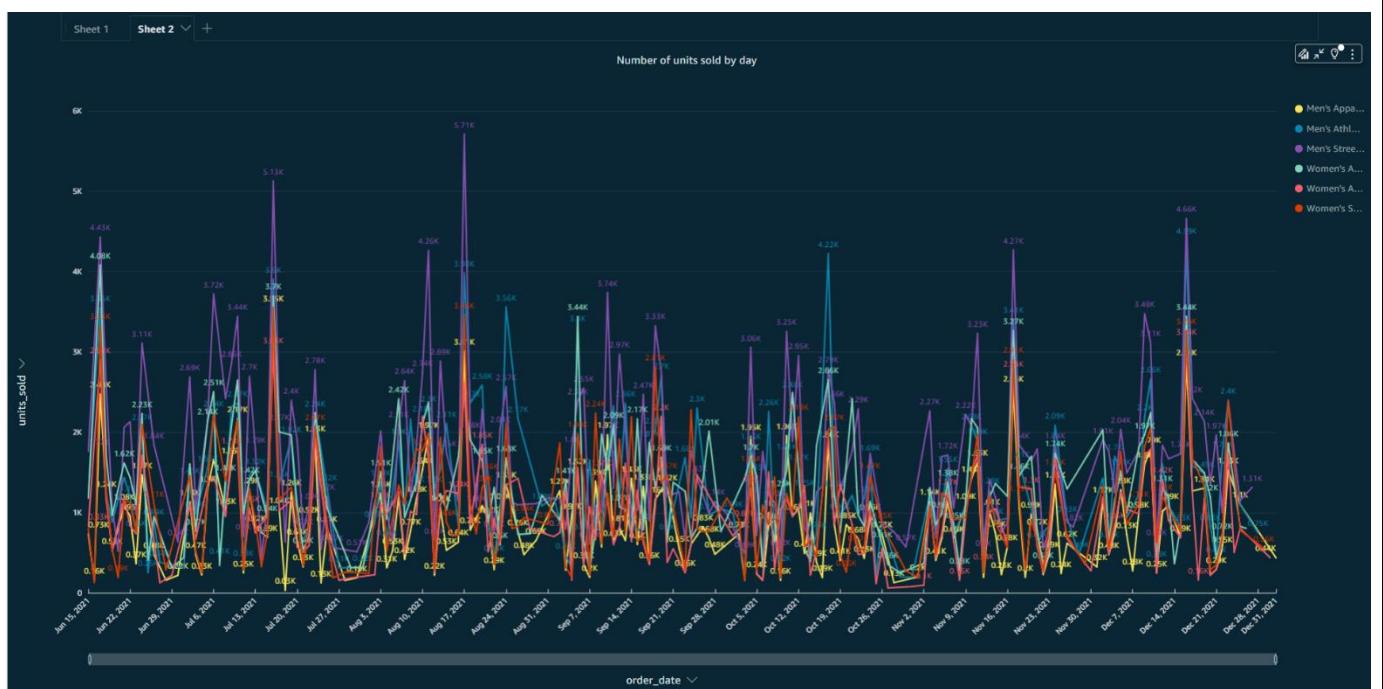
The screenshot shows the Amazon QuickSight interface with three visualizations:

- Line chart:** Shows the "Sum of units\_sold by Order\_date and Product". The Y-axis ranges from 0 to 50K. The X-axis shows dates from 2021-01-01 to 2023-01-01. Multiple lines represent different products, showing a significant spike in sales around January 2022.
- Bar chart:** Shows the "Sum of Total\_sales by Region and Product". The Y-axis ranges from 0 to 8M. The X-axis shows regions: West, Southeast, South, Northeast, and Midwest. For each region, there are five bars representing different products.
- Stacked Bar chart:** Shows the "Sum of Total\_sales by Product". The Y-axis ranges from 0 to 50M. The X-axis shows products: Men's Street Footwear, Men's Casual Footwear, Men's Dress Footwear, Men's Athletic Footwear, and Women's Footwear. The bars are stacked to show the total sales per product category.

The left sidebar shows the dataset "SPICE sales\_data" and various calculated fields like "order\_date", "product", and "region". The right sidebar contains properties for the visualizations, including "Display Settings", "Data series", "X-axis", "Y-axis", "Group/Color", "Reference lines", "Legend", and "Data labels".



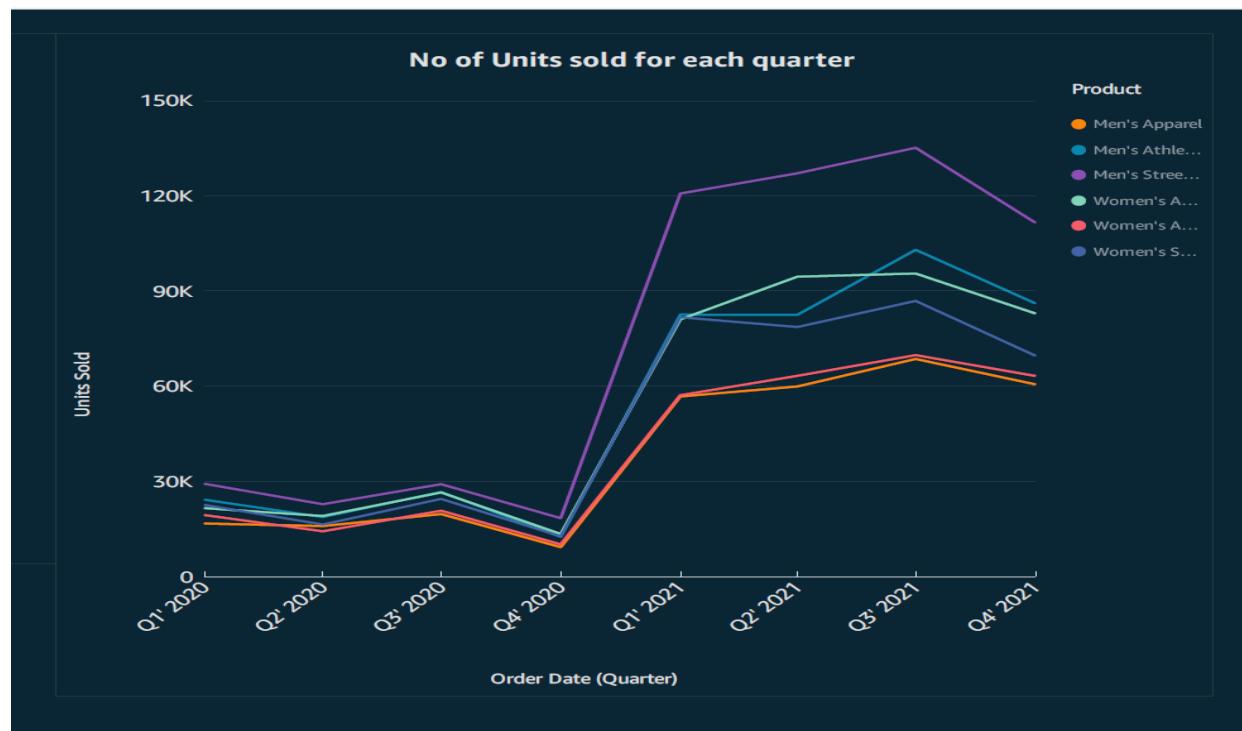
## Plotting no of units sold for each day



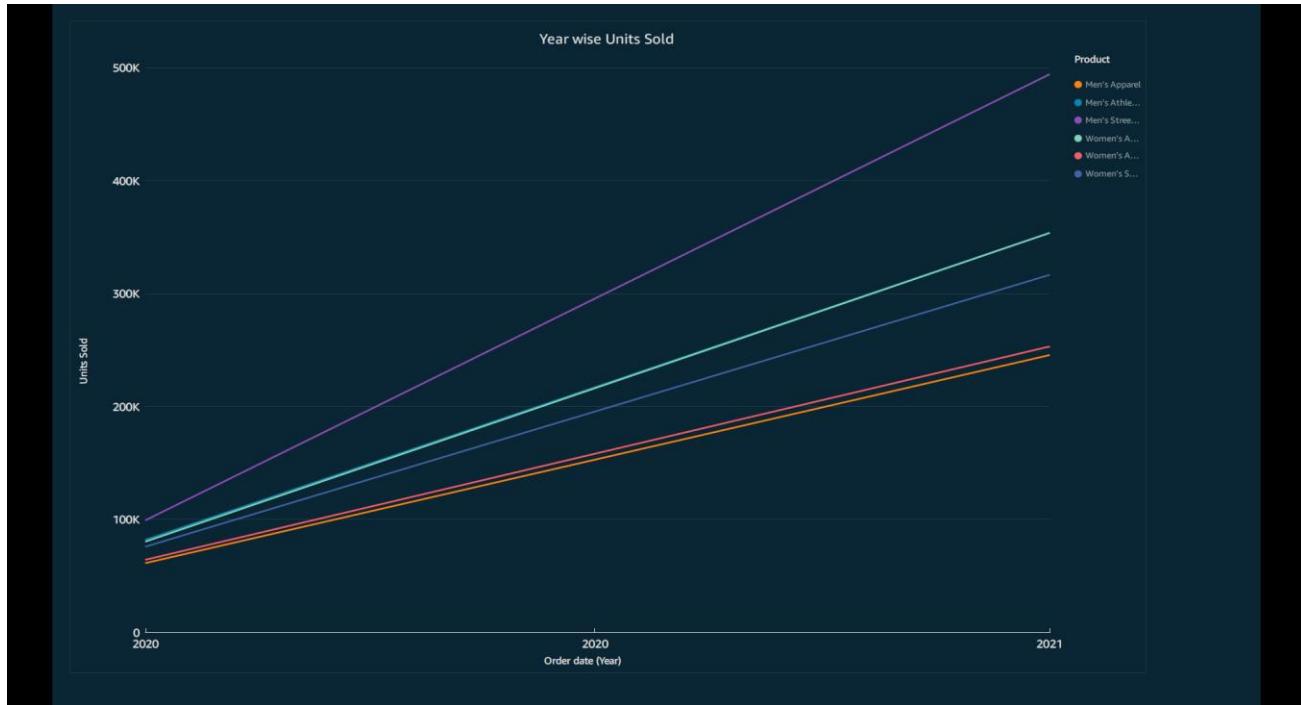
## Plotting no of units sold for each month



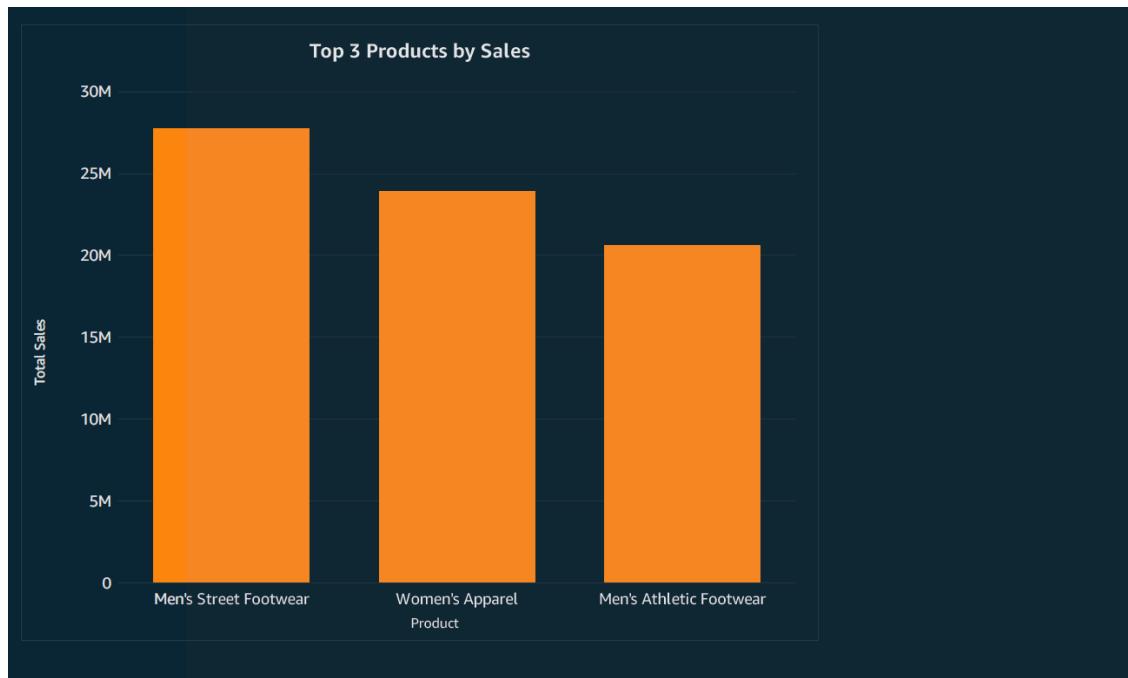
## Quarterly wise units sold



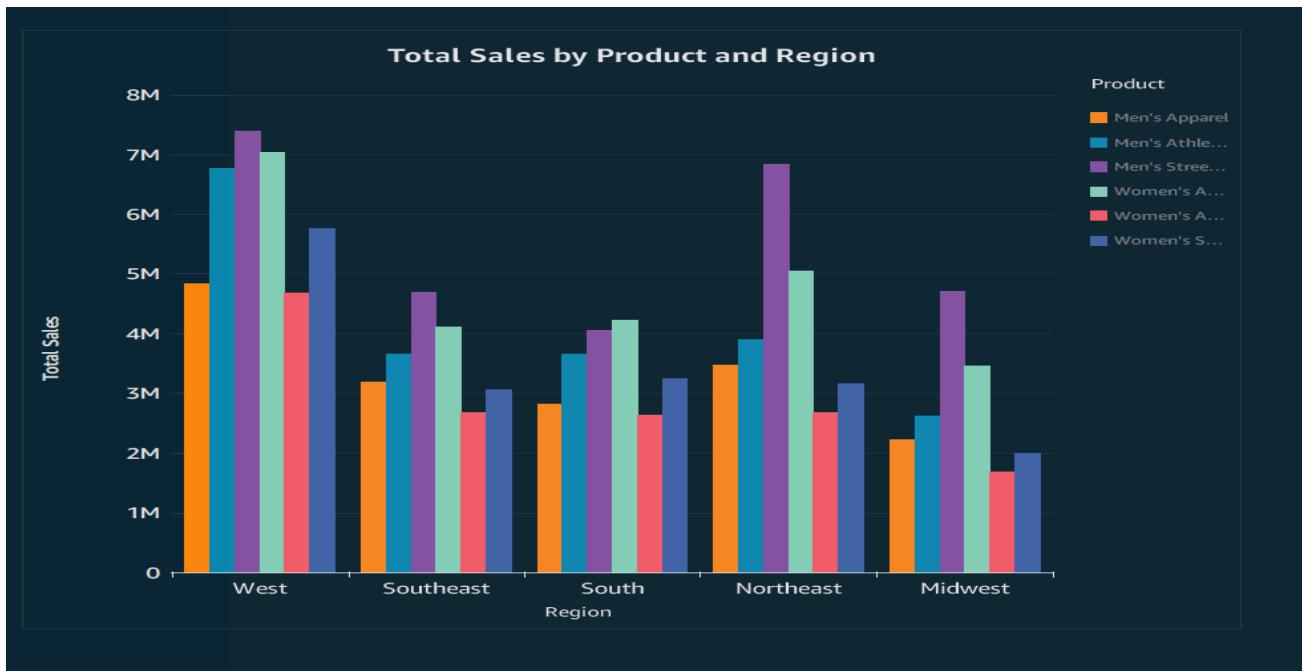
## Year wise units sold



## Top 3 products by sales



## Region wise product sales



# Conclusion

1. Data Warehousing with Amazon Redshift: By leveraging Amazon Redshift, we successfully centralized and managed the sales data. Redshift's scalability and performance allowed us to handle large volumes of information efficiently. We transformed raw data into a structured format, enabling complex queries and aggregations. The data warehouse served as the backbone for our analysis, providing a solid foundation for further insights.
2. Visual Insights with Amazon Quick Sight: Amazon Quick Sight played a pivotal role in turning data into actionable insights. We created interactive dashboards and visualizations that allowed stakeholders to explore sales trends, product performance, and customer behavior. Quick Sight's user-friendly interface empowered business users to make informed decisions without relying on technical expertise. Whether tracking revenue, monitoring inventory, or identifying growth opportunities, Quick Sight provided a responsive platform for data exploration.

“Armed with Redshift and Quick Sight, the sales team can gain valuable insights. They can identify top-performing products, can analyze regional sales patterns, and optimize inventory management. By embracing AWS tools, they can elevate their sales analysis capabilities, driving smarter business decisions and enhancing overall performance”.