

Cloud Computing : Project Deliverable 2

Group 6:

- Anantnaval Gaikwad
- Aniket Ghumed
- Rimjhim Jain
- Saumit Chinchkhandi
- Sudhanshu Dalvi

GITHUB ACCESS TO ALL MEMBERS

The screenshot shows the GitHub repository page for 'Group6_amazon_products_project'. The repository is public, created by 'anikethghumed23' last week, and has 5 commits. The README file contains the following content:

```
Group6_amazon_products_project

Project: Predictive Modeling for Best-Selling Products in E-commerce

Project Scope and Business Goal

Project Scope
The machine learning component of this project aims to address the specific problem of accurately predicting the best-selling products in the e-commerce industry. The focus is on developing a robust predictive model using advanced data analytics techniques. This aligns with the educational program goals of AWS Academy Cloud Foundations and Data Engineering, providing hands-on experience in applying machine learning concepts to real-world scenarios.
```

The screenshot shows the 'Who has access' section of the GitHub repository settings. It indicates that the repository is a PUBLIC REPOSITORY and has DIRECT ACCESS for 4 contributors. The list of contributors includes:

- Anantnaval Gaikwad (Collaborator)
- Rimjhim Jain (Collaborator)
- SaumitC (Collaborator)
- sddalv10 (Collaborator)

CREATE BUCKET

aws Services Search [Option+S]

Amazon S3 > Buckets > Create bucket

Create bucket Info

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

AWS Region

Copy settings from existing bucket - optional
Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Object Ownership Info
Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

ACLs disabled (recommended)
All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

ACLs enabled
Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership
Bucket owner enforced

Block Public Access settings for this bucket

aws Services Search [Option+S]

Successfully created bucket "amazon-product-data-bucket"
To upload files and folders, or to configure additional bucket settings, choose [View details](#).

Amazon S3 > Buckets

Account snapshot
Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

Buckets (1) Info
Buckets are containers for data stored in S3. [Learn more](#)

[View Storage Lens dashboard](#)

Name	AWS Region	Access	Creation date
amazon-product-data-bucket	US East (N. Virginia) us-east-1	Bucket and objects not public	November 19, 2023, 16:09:43 (UTC-05:00)

aws Services Search [Option+S]

Upload succeeded
[View details below](#).

Upload: status

The information below will no longer be available after you navigate away from this page.

Summary

Destination	Succeeded	Failed
s3://amazon-product-data-bucket	2 files, 358.5 MB (100.00%)	0 files, 0 B (0%)

Files and folders Configuration

Files and folders (2 Total, 358.5 MB)

Find by name	Name	Folder	Type	Size	Status	Error
	amazon_products.csv	amazon_products_dataset/	text/csv	358.5 MB	Success	-
	amazon_categories.csv	amazon_products_dataset/	text/csv	6.7 KB	Success	-

Data Exploration for Insight and Pre-processing

The following screenshots demonstrate the setup and execution of an AWS Glue crawler to extract data from an Amazon S3 bucket.

Screenshot 1: AWS Glue Crawler Configuration

The screenshot shows the AWS Glue console with the "amazon-products_crawler" configuration page. The crawler has been successfully created and is currently in a "READY" state. It is configured to use the "amazon_products_crawler" IAM role and the "amazon_database" database. The crawler runs every 52 seconds and has processed 1 table change and 0 partition changes.

Screenshot 2: Amazon S3 Bucket Objects

The screenshot shows the Amazon S3 console with the "amazon-product-data-bucket" bucket. A new folder named "query_location/" has been successfully created. The bucket contains three folders: "amazon_categories/", "amazon_products/", and "query_location/".

Screenshot 3: AWS Glue Crawler Execution

The screenshot shows the AWS Glue console with the "amazon-products_crawler" configuration page. The crawler is now starting, indicated by the message "Crawler successfully starting". The crawler has started and completed its run, processing 1 table change and 0 partition changes.

AWS Glue Services Search [Option+S] N. Virginia vocabs/user2753084@aghamed@uncc.edu @ 6291-9672-6182 Create table X

AWS Glue

- Getting started
- ETL jobs
- Visual ETL
- Notebooks
- Job run monitoring
- Data Catalog tables
- Data connections
- Workflows (orchestration)
- Data Catalog**
- Data Integration and ETL**
- Legacy pages**
- What's New Documentation AWS Marketplace
- Enable compact mode
- Enable new navigation

You can now create Apache Iceberg tables in the AWS Glue Data Catalog. To learn more, visit the documentation.

amazon_database

Last updated (UTC) November 19, 2023 at 21:54:34 Edit Delete

Database properties

Name	Description	Location	Created on (UTC)
amazon_database	-	-	November 19, 2023 at 21:51:17

Tables (1)

Last updated (UTC) November 19, 2023 at 21:54:35 Edit Delete Add tables using crawler Add table

Name	Database	Location	Classification	Deprecated	View data	Data quality
amazon_products_csv	amazon_database	s3://amazon-product-data-buck	CSV	-	Table data	View data quality

AWS Services Search [Option+S] N. Virginia vocabs/user2753084@aghamed@uncc.edu @ 6291-9672-6182 Get started X

Have you tried Amazon Athena for Apache Spark? With Athena for Spark, you can issue Spark SQL statements within Spark Python notebooks to create more complex insights. Read more about this in our blog. If you have tried Athena Spark, let us know how we can make the feature better by providing us your feedback here.

Amazon Athena > Query editor > Manage settings

Manage settings

Choose S3 data set

S3 buckets > amazon-product-data-bucket

Objects (1/3)

Key	Last modified	Size
amazon_categories	-	-
amazon_products	-	-
query_location	-	-

Cancel Choose

AWS Services Search [Option+S] N. Virginia vocabs/user2753084@aghamed@uncc.edu @ 6291-9672-6182 Get started X

Have you tried Amazon Athena for Apache Spark? With Athena for Spark, you can issue Spark SQL statements within Spark Python notebooks to create more complex insights. Read more about this in our blog. If you have tried Athena Spark, let us know how we can make the feature better by providing us your feedback here.

Amazon Athena > Query editor > Manage settings

Manage settings

Query result location and encryption

Location of query result - optional Enter an S3 prefix in the current region where the query result will be saved as an object.

s3://amazon-product-data-bucket/query_location View Browse S3

You can create and manage lifecycle rules for this bucket Use Amazon S3 Lifecycle rules to store your query results and metadata cost effectively or to delete them after a period of time. Learn more

Lifecycle configuration

Expected bucket owner - optional Specify the AWS account ID that you expect to be the owner of your query results output location bucket. Enter AWS account ID

Assign bucket owner full control over query results Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

Encrypt query results

Cancel Save

AWS Glue

Some crawlers successfully created
The following crawlers are now created: "amazon_categories", "organization-test", "amazon-products_crawler"

amazon_categories

Crawler properties

Name amazon_categories	IAM role LabRole	Database amazon_database	State READY
Description	Security configuration	Lake Formation configuration	Table prefix
Maximum table threshold	-	-	-

Advanced settings

Crawler runs (0)
The list of crawler runs for this crawler.

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
You don't have any crawler runs. Run crawler					

Last updated (UTC) November 19, 2023 at 23:12:54 | Run crawler | Edit | Delete

AWS Glue

Crawlers successfully starting
The following crawlers are now starting: "amazon_categories", "organization-test", "amazon-products_crawler"

amazon_categories

Crawler properties

Name amazon_categories	IAM role LabRole	Database amazon_database	State READY
Description	Security configuration	Lake Formation configuration	Table prefix
Maximum table threshold	-	-	-

Advanced settings

Crawler runs (1)
The list of crawler runs for this crawler.

Start time (UTC)	End time (UTC)	Current/last duration	Status	DPU hours	Table changes
November 19, 2023 at 23:15:00	-	01 s	Running	-	-

Last updated (UTC) November 19, 2023 at 23:12:54 | Run crawler | Edit | Delete

Amazon Athena

Query editor

Editor | Recent queries | Saved queries | Settings | Workgroup primary

Data

Data source: AwsDataCatalog
Database: amazon_database

Tables and views

Tables (2): amazon_categories, amazon_products
Views (0):

Query 1

SQL: Ln 1, Col 1

Run | Explain | Cancel | Clear | Create | Reuse query results up to 60 minutes ago

Query results

Results

No results

aws Services Search [Option+S] N. Virginia vodlays/user2753084@ghummed@unc.edu @ 6291-9672-6182

Typeahead suggestions are turned on by default. You can change this setting in query editor preferences.

Data

Query 1: `SELECT * FROM "amazon_database"."amazon_products" limit 10;`

Data source: AwsDataCatalog Database: amazon_database

Tables and views: Create

Tables (2): `amazon_categories`, `amazon_products`

Views (0)

SQL Ln 1, Col 1

Run again Explain Cancel Clear Create

Query results: Completed Time in queue: 98 ms Run time: 655 ms Data scanned: 1020.06 KB

Results (10)

#	asin	title	imgurl
1	B0C1B7C9J9	"Screen Protector Compatible with 2023 Honda Accord 12.3 Inch Touchscreen HD Clear Anti Scratch	Garmin GoPro Combo Mount
2	B07G2445M5	"Out-Front Mount for Garmin Edge	Garmin GoPro Combo Mount

Copy Download results

aws Services Search [Option+S] N. Virginia vodlays/user2753084@ghummed@unc.edu @ 6291-9672-6182

Tables (2): `amazon_categories`, `amazon_products`

Views (0)

SQL Ln 1, Col 1

Run again Explain Cancel Clear Create

Query results: Completed Time in queue: 98 ms Run time: 655 ms Data scanned: 1020.06 KB

Results (10)

#	asin	title	imgurl
1	B0C1B7C9J9	"Screen Protector Compatible with 2023 Honda Accord 12.3 Inch Touchscreen HD Clear Anti Scratch	Garmin GoPro Combo Mount
2	B07G2445M5	"Out-Front Mount for Garmin Edge	Dropproof AirTag Keychain Cover
3	B0BW453V3T	"2Pack Resin Airtag Case Holder with Airtag Keychain	Zumo 450
4	B0BYRT13QT	"Cameron Sino Battery for Garmin Zumo 400	Anti-Scratch Protective Case Air Tag Cover Card Keychain Tile Sticker Mate Cover Pendant for Shoulder Bag Backpack Wallet -D(insect)
5	B0BD7114BH	Anti-Scratch Protective Case Air Tag Cover Card Keychain Tile Sticker Mate Cover Pendant for Shoulder Bag Backpack Wallet -D(insect)	https://m.media-amazon.com/images/I/418LBE
6	B07KG18DC	Mobile Phone Holder for Brompton Folding Bike	https://m.media-amazon.com/images/I/31-ZB6
7	B09523R55T	"Case for Air Tag	Cute Cartoon Piggy Silicone Cover for AirTags L
8	B0B8GXHP91	"Abanen Quick Fit 22mm Leather Watch Band for Garmin Fenix 7/ Fenix 6/ Fenix 5/ EPIX 2	Fiber Pattern & Leather with Silicone Lining Soft
9	B0C2C56N81	"Screen Protector Case Compatible with Garmin Forerunner 965	Full Coverage Hard PC Plated Bumper Cover Pro
10	B0BHT6C294	"Tempered Screen Protector for Chevrolet Malibu Premier 2016-2023	LS/RS 2019-2023

Copy Download results

aws Services Search [Option+S] N. Virginia vodlays/user2753084@ghummed@unc.edu @ 6291-9672-6182

Data

Query 1: `SELECT * FROM "amazon_database"."amazon_categories" limit 10;`

Data source: AwsDataCatalog Database: amazon_database

Tables and views: Create

Tables (2): `amazon_categories`, `amazon_products`

Views (0)

SQL Ln 1, Col 1

Run again Explain Cancel Clear Create

Query results: Completed Time in queue: 92 ms Run time: 424 ms Data scanned: 6.67 KB

Results (10)

#	id	category_name
1		category_name
2	1	Beadling & Jewelry Making
3	2	Fabric Decorating

Copy Download results

aws Services Search [Option+S] N. Virginia v vocabls/user2753084=aghmed@unc.edu @ 6291-9672-6182

Tables (2) < 1 >

amazon_categories
amazon_products

Views (0) < 1 >

SQL Ln 1, Col 1

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed

Results (10)

Search rows

#	id	category_name
1		category_name
2	1	Beading & Jewelry Making
3	2	Fabric Decorating
4	3	Knitting & Crochet Supplies
5	4	Printmaking Supplies
6	5	Scrapbooking & Stamping Supplies
7	6	Sewing Products
8	7	Craft & Hobby Fabric
9	8	Needlework Supplies
10	9	*Arts

Time in queue: 92 ms Run time: 424 ms Data scanned: 6.67 KB

Copy Download results

aws Services Search [Option+S] N. Virginia v vocabls/user2753084=aghmed@unc.edu @ 6291-9672-6182

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings Workgroup primary

Athena now supports typeahead code suggestions to speed up SQL query development

Typeahead suggestions are turned on by default. You can change this setting in query editor preferences.

Edit preferences

Data < 1 >

Query 1:

```
1 SELECT ac.category_name, COUNT(*) AS total_products_sold
2 FROM amazon_products ap
3 JOIN amazon_categories ac ON ap.category_id = ac.id
4 GROUP BY ac.category_name
5 ORDER BY total_products_sold DESC
6 LIMIT 10;
```

Tables and views < 1 >

Filter tables and views

Tables (2) < 1 >

amazon_categories
amazon_products

Views (0) < 1 >

SQL Ln 7, Col 1

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed

Results (10)

Search rows

#	category_name	total_products_sold
1	Printmaking Supplies	68767
2	Girls' Clothing	24859
3	Boys' Clothing	20541
4	*Arts	18703
5	Men's Shoes	16500
6	Women's Clothing	15653
7	Men's Clothing	15079
8	Women's Handbags	14412
9	Girls' Jewelry	14264
10	Car Care	14099

Time in queue: 105 ms Run time: 1.774 sec Data scanned: 358.53 MB

Copy Download results

aws Services Search [Option+S] N. Virginia v vocabls/user2753084=aghmed@unc.edu @ 6291-9672-6182

Tables (2) < 1 >

amazon_categories
amazon_products

Views (0) < 1 >

SQL Ln 7, Col 1

Run again Explain Cancel Clear Create

Reuse query results up to 60 minutes ago

Query results Query stats

Completed

Results (10)

Search rows

#	category_name	total_products_sold
1	Printmaking Supplies	68767
2	Girls' Clothing	24859
3	Boys' Clothing	20541
4	*Arts	18703
5	Men's Shoes	16500
6	Women's Clothing	15653
7	Men's Clothing	15079
8	Women's Handbags	14412
9	Girls' Jewelry	14264
10	Car Care	14099

Time in queue: 105 ms Run time: 1.774 sec Data scanned: 358.53 MB

Athena now supports typeahead code suggestions to speed up SQL query development

```
1 SELECT ac.category_name, AVG(ap.price) AS avg_price
2 FROM amazon_products ap
3 JOIN amazon_categories ac ON ap.category_id = ac.id
4 GROUP BY ac.category_name
5 ORDER BY avg_price DESC;
```

Data source: AwsDataCatalog, Database: amazon_database

Tables and views: Tables (2)

Query results: Completed

Time in queue: 106 ms Run time: 2.121 sec Data scanned: 358.53 MB

Query results: Completed

Results (248)

#	category_name	avg_price
1	Computers & Tablets	602.2680041928688
2	Smart Home - Heating & Cooling	436.60887931034483
3	Computer Servers	400.1230541455161
4	Smart Home: Home Entertainment	226.553384297521
5	Computer External Components	180.07671527777782
6	Computer Monitors	178.93385335413302
7	Baby Gifts	176.4557894736845
8	Data Storage	173.27178166584386
9	Computers	164.0675935688536342
10	Camera & Photo	157.5771672902182
11	Virtual Reality Hardware & Accessories	151.38394557822892
12	Video Projectors	140.19007058823547
13	Baby Strollers & Accessories	136.68692049808408
14	Smart Home: Vacuums and Mops	134.58823529411768
15	Pregnancy & Maternity Products	133.0015435294115

Athena now supports typeahead code suggestions to speed up SQL query development

```
1 SELECT title, reviews, stars, category_name
2 FROM amazon_products ap
3 JOIN amazon_categories ac ON ap.category_id = ac.id
4 WHERE isbestseller = true
5 ORDER BY reviews DESC
6 LIMIT 10;
```

Data source: AwsDataCatalog, Database: amazon_database

Tables and views: Tables (2)

Query results: Completed

Time in queue: 98 ms Run time: 1.51 sec Data scanned: 358.53 MB

Results (10)

Amazon Athena > Query editor

Editor Recent queries Saved queries Settings Workgroup primary

Athena now supports typeahead code suggestions to speed up SQL query development. Typeahead suggestions are turned on by default. You can change this setting in query editor preferences. Edit preferences X

Data

Data source: AwsDataCatalog Database: amazon_database

Tables and views: Create ▾

Tables (2) Completed

- amazon_categories
- amazon_products

Views (0)

Query 1

```

1 SELECT
2   CASE
3     WHEN price <= 50 THEN '0-50'
4     WHEN price <= 100 THEN '51-100'
5     WHEN price <= 200 THEN '101-200'
6     ELSE '200+'
7   END AS price_range,
8   COUNT(*) AS products_count
9 FROM amazon_products
10 GROUP BY
11   CASE
12     WHEN price <= 50 THEN '0-50'
13     WHEN price <= 100 THEN '51-100'
14     WHEN price <= 200 THEN '101-200'
15     ELSE '200+'

```

SQL Ln 17, Col 22

Run again Explain Cancel Clear Create Reuse query results up to 60 minutes ago

Query results **Query stats**

Completed Time in queue: 98 ms Run time: 1.51 sec Data scanned: 358.53 MB

Results (10)

#	title	re
1	essence Lash Princess False Lash Effect Mascara Gluten & Cruelty Free	34
2	"LuxClub 4 PC Sheet Bed Sheets Deep Pockets 18" Eco Friendly Wrinkle Free Kids Fitted Sheets Machine Washable Hotel Bedding Silky Soft - White Twin"	1
3	DualShock 4 Wireless Controller for PlayStation 4 - Jet Black	14
4	Handcraft Frankincense Essential Oil - 100% Pure & Natural - Premium Therapeutic Grade with Premium Glass Dropper - Huge 4 fl. Oz	9
5	PlayStation DualSense Wireless Controller - Midnight Black	91
6	Meta Quest 2 — Advanced All-in-One Virtual Reality Headset — 128 GB	81
7	Orzly Hard EVA Case for Nintendo Switch & Switch OLED Console - Durable Anti-Shock Zipper Storage Hard Case for Nintendo Switch Console and Accessories	71
8	Hot Firming Lotion Sweat Enhancer - Skin Tightening Cream for Stomach Fat and Cellulite - Sweat Cream for Better Workout Results - Long Lasting Moisturizing Pre and Post Workout Massage Lotion	71
9	DELOMO Upgrade Pet Grooming Gloves Cat Brushes Gloves for Gentle Shedding - Efficient Pets Hair Remover Mittens - Dog Washing Gloves for Long and Short Hair Dogs & Cats & Horses - 1 Pair (Blue)	71
10	Nintendo Switch Pro Controller	65

Copy **Download results**

< 1 > ⌂

Query 2

```

1 SELECT
2   CASE
3     WHEN price <= 50 THEN '0-50'
4     WHEN price <= 100 THEN '51-100'
5     WHEN price <= 200 THEN '101-200'
6     ELSE '200+'
7   END AS price_range,
8   COUNT(*) AS products_count
9 FROM amazon_products
10 GROUP BY
11   CASE
12     WHEN price <= 50 THEN '0-50'
13     WHEN price <= 100 THEN '51-100'
14     WHEN price <= 200 THEN '101-200'
15     ELSE '200+'

```

SQL Ln 17, Col 22

Run again Explain Cancel Clear Create Reuse query results up to 60 minutes ago

Query results **Query stats**

Completed Time in queue: 99 ms Run time: 1.15 sec Data scanned: 358.52 MB

Results (4)

#	price_range	products_count
1	0-50	944808
2	101-200	40162
3	200+	373199
4	51-100	68169

Copy **Download results**

< 1 > ⌂

Data

Query 1 | X | Query 2 | X

```
1 | SELECT * FROM "amazon_database"."avg_price_by_category_view" limit 10;
```

Data source: AwsDataCatalog
Database: amazon_database

Tables and views Create @

Tables (2) < 1 >

- amazon_categories
- amazon_products

Views (4) < 1 >

- avg_price_by_category_view
- best_products_most_reviews_view
- price_range_distribution_view
- total_products_sold_view

SQL Ln 1, Col 1 Run again Explain Cancel Clear Create

Query results | Query stats

Completed Time in queue: 104 ms Run time: 1.987 sec Data scanned: 358.53 MB

Results (10)

#	category_name	avg_price
1	Computers & Tablets	602.2680041928688
2	Smart Home - Heating & Cooling	436.60887931034483
3	Computer Servers	400.1230541455161

Copy Download results

File Edit Data Insert Sheets Objects Search PUBLISH NEW LOOK

Data

Dataset: amazon_categories + ADD CHANGE VISUAL TYPE

Search fields Q

+ CALCULATED FIELD

category_name id

GROUP/COLOR category_name

VALUE id (Count)

SMALL MULTIPLES Add a dimension

Visuals

Pie chart

Sheet 1

Count of id by Category_name

SHOWING TOP 20 IN CATEGORY_NAME

Group By: category_name ▼
Size: id (Count) ▼

Properties

Category N... X

Value

Title: id (Count)

Sort

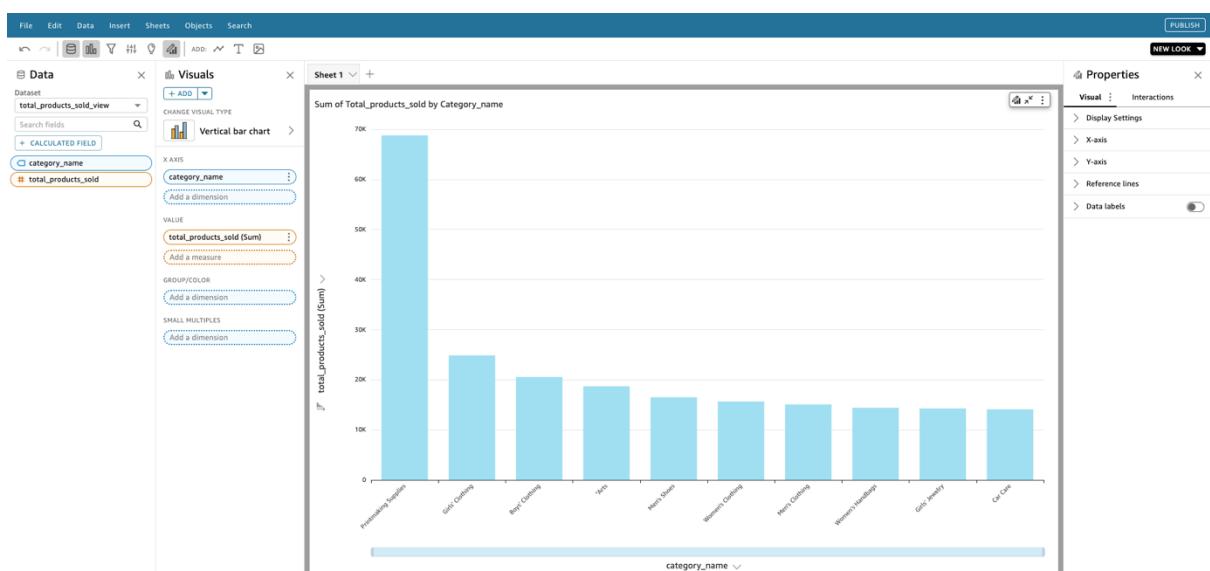
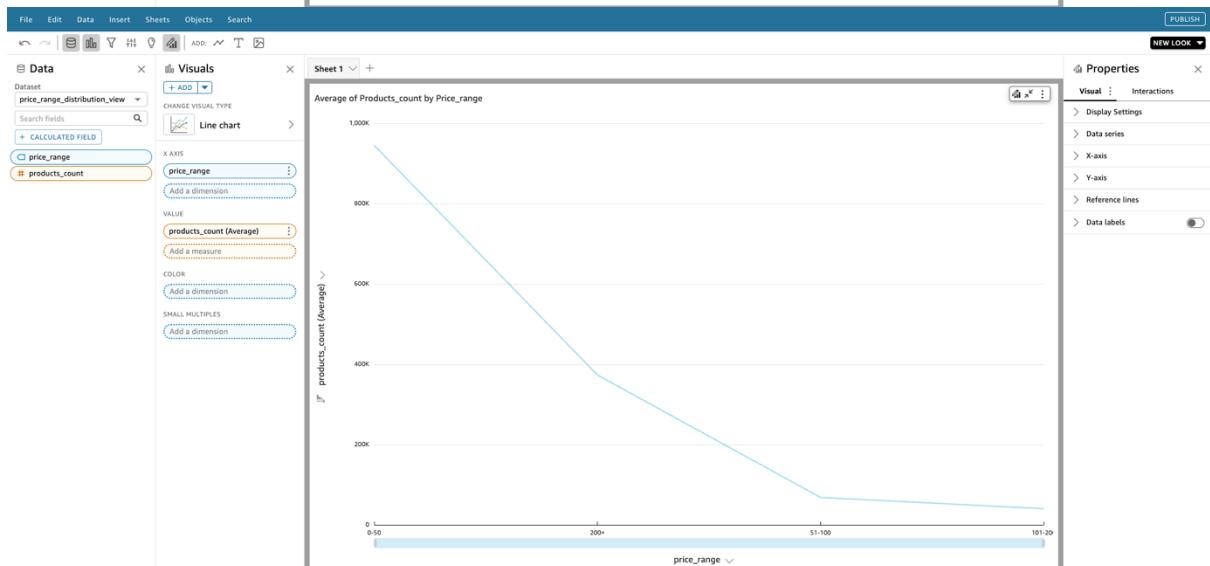
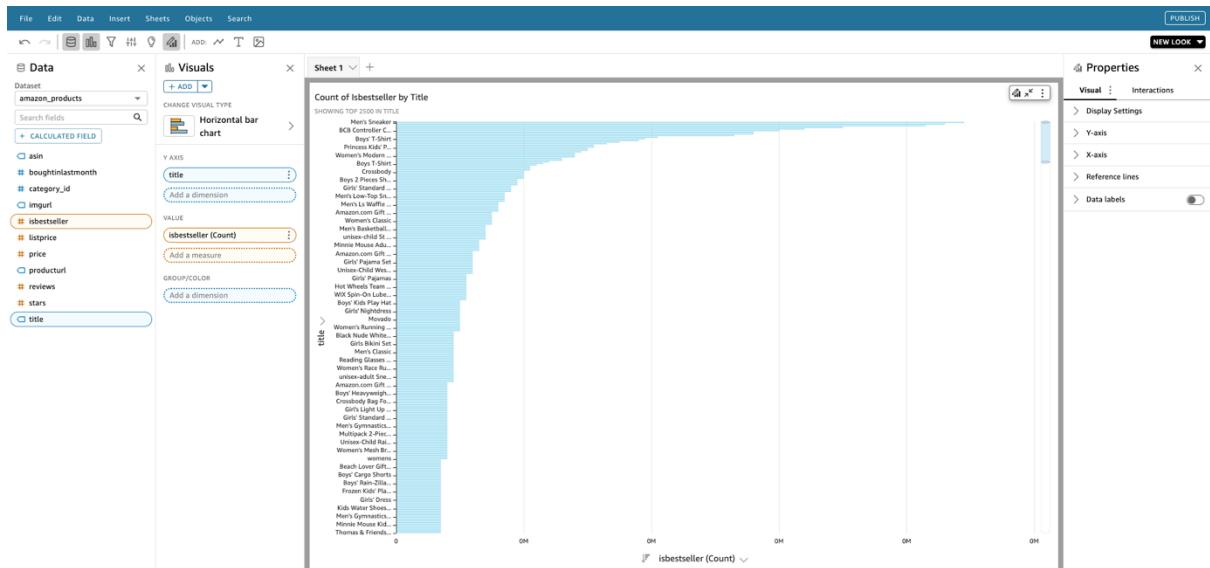
Group/Color

Legend (On)

Data labels (On)

Category Name Legend:

- *Arts
- *Foot
- *Heating
- *Hydraulics
- *Industrial...
- *Lights
- *Nintendo...
- *Nursery F...
- *Paint
- *Painting
- *PlayStation...
- *PlayStation...
- *PlayStation...
- *PlayStation...
- *Slot Cars
- *Sony PSP...
- *Test
- *Tirecycles
- *Other



AWS Glue ETL Job:

The screenshot displays three sequential views of an AWS Glue ETL job configuration for processing an Amazon products dataset.

Initial Configuration (Top):

- Job Name:** amazon_products
- Data Source:** Amazon S3 (s3://amazon-product-data-*)
- Data Format:** CSV
- Delimiter:** Comma (,)
- Escape character - optional:** None
- Quote character:** Double quote (")
- First line of source file contains column headers:** Checked
- Records in source files can span multiple lines:** Unchecked

Intermediate Transformation (Middle):

- Job Name:** amazon_products
- Transforms:**
 - Input: Amazon S3 (s3://amazon-product-data-*)
 - Transform - Filter: Filter_stars_4_5 (Key: stars, Value: ^[2-4](\d+)?[4,5]\$)
 - Transform - Aggregate: Aggregate (Fields to group by: title, Aggregation function: max(price))

Final Transformation (Bottom):

- Job Name:** amazon_products
- Transforms:**
 - Input: Filter_stars_4_5 (Filter - Transform)
 - Transform - Aggregate: Aggregate (Fields to group by: title, Aggregation function: max(price))

Data Preview:

Middle View Data Preview:

price	reviews	boughtinlastmonth	stars	category_id	isbn
224.99	0	0	4.3	104	Fal

Bottom View Data Preview:

title	max(price)
19 Degree Short Trip Exp andable 4 Wheeled Packi ng Case - Navy	895.0
26 Inch Luggage Large Su itcase With Spinner Whee ls Tsa Luggage Suitcases Traveler's Choice Hard Ca	189.99

AWS Glue

Getting started
ETL Jobs
 Visual ETL
 Notebooks
 Job run monitoring
 Data Catalog tables
 Data connections
 Workflows (orchestration)

▶ Data Catalog
 ▶ Data Integration and ETL
 ▶ Legacy pages

What's New Documentation AWS Marketplace

Enable compact mode Enable new navigation

amazon_products

Successfully updated job amazon_products. To run the job choose the Run Job button.

Last modified on 19/11/2023, 20:05:47 Try new UI Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Data preview (198) Filter sample dataset End session Previewing 2 of 2 fields

title max_price

19 Degree Short Trip Expandable 4 Wheeled Packing Case - Navy 895.0

26 Inch Luggage Large Suitcase With Spinner Wheel Is Tsa Luggage Suitcases Traveler's Choice Hard Ca 189.99

AWS Glue

Getting started
ETL Jobs
 Visual ETL
 Notebooks
 Job run monitoring
 Data Catalog tables
 Data connections
 Workflows (orchestration)

▶ Data Catalog
 ▶ Data Integration and ETL
 ▶ Legacy pages

What's New Documentation AWS Marketplace

Enable compact mode Enable new navigation

amazon_products

Successfully updated job amazon_products. To run the job choose the Run Job button.

Last modified on 19/11/2023, 20:09:15 Try new UI Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Data target properties - S3

Name Amazon S3
 Node parents Choose which nodes will provide inputs for this one. Choose one or more parent node
 Change Schema ApplyMapping - Transform

Format CSV
 Compression Type None
 S3 Target Location Choose an S3 location in the format s3://bucket/prefix/object/ with a trailing slash (/). /amazon-products-data-bucket View Browse S3

Data Catalog update options Info
 Choose how you want to update the Data Catalog table's schema and partitions. These options will only apply if the Data Catalog table is an S3 backed source.
 Do not update the Data Catalog
 Create a table in the Data Catalog and on subsequent runs, update the schema and add new partitions
 Create a table in the Data Catalog and on subsequent runs, keep existing schema and add new partitions

Partition keys - optional Add partition keys

AWS Glue

Getting started
ETL Jobs
 Visual ETL
 Notebooks
 Job run monitoring
 Data Catalog tables
 Data connections
 Workflows (orchestration)

▶ Data Catalog
 ▶ Data Integration and ETL
 ▶ Legacy pages

What's New Documentation AWS Marketplace

Enable compact mode Enable new navigation

amazon_products

Successfully updated job amazon_products. To run the job choose the Run Job button.

Last modified on 19/11/2023, 20:09:15 Try new UI Actions Save Run

Visual Script Job details Runs Data quality Schedules Version Control

Data target properties - S3

Name Amazon S3
 Node parents Choose which nodes will provide inputs for this one. Choose one or more parent node
 Change Schema ApplyMapping - Transform

Format CSV
 Compression Type None
 S3 Target Location Choose an S3 location in the format s3://bucket/prefix/object/ with a trailing slash (/). /job_transformations_data/ View Browse S3

Data Catalog update options Info
 Choose how you want to update the Data Catalog table's schema and partitions. These options will only apply if the Data Catalog table is an S3 backed source.
 Do not update the Data Catalog
 Create a table in the Data Catalog and on subsequent runs, update the schema and add new partitions
 Create a table in the Data Catalog and on subsequent runs, keep existing schema and add new partitions

Partition keys - optional Add a partition key

AWS Glue

Successfully started job
Successfully started job amazon_products. Navigate to Run details for more details.

amazon_products

Last modified on 19/11/2023, 20:11:00 Try new UI Actions Save Run

Data target properties - S3

Node parents
Choose which nodes will provide inputs for this one.
Choose one or more parent node

Change Schema
ApplyMapping - Transform

Format
CSV

Compression Type
None

S3 Target Location
Choose an S3 location in the format s3://bucket/prefix/object/ with a trailing slash (/).
s3://amazon-product-data- View Browse S3

Data Catalog update options
Choose how you want to update the Data Catalog table's schema and partitions. These options will only apply if the Data Catalog table is an S3 backed source.
 Do not update the Data Catalog
 Create a table in the Data Catalog and on subsequent runs, update the schema and add new partitions
 Create a table in the Data Catalog and on subsequent runs, keep existing schema and add new partitions

Partition keys - optional
Add partition keys.

Add a partition key

AWS Glue

Successfully started job
Successfully started job amazon_products. Navigate to Run details for more details.

amazon_products

Last modified on 19/11/2023, 20:11:00 Try new UI Actions Save Run

Job runs (1/1) Info

Last updated (UTC) November 20, 2023 at 01:11:20 View details Stop job run Table View Card View

Run status	Retries	Start time (UTC)	End time (UTC)	Duration	Capacity (DPU)	Worker type	Glue version
Running	0	2023/11/20 01:11:02	-	10 s	10 DPU	G.1X	4.0

Run details Input arguments (10) Continuous logs Run insights Metrics Spark UI Stop job run

Job name	Start time (UTC)	Glue version	Last modified on (UTC)
amazon_products	November 19, 2023 1:11:02 AM	4.0	November 19, 2023 1:11:06 AM
ID	End time (UTC)	Worker type	Log group name
jr_62ed2b07776f5fb1ebc6ea17606d9024b0eaab86d6fc15 70621eb6b2db02940	-	G.1X	/aws-glue/jobs
Run status	Start-up time	Max capacity	Number of workers
Running	0	10 DPU	10

AWS Glue

Successfully started job
Successfully started job amazon_products. Navigate to Run details for more details.

amazon_products

Last modified on 19/11/2023, 20:11:00 Try new UI Actions Save Run

Job runs (1/1) Info

Last updated (UTC) November 20, 2023 at 01:21:55 View details Stop job run Table View Card View

Run status	Retries	Start time (UTC)	End time (UTC)	Duration	Capacity (DPU)	Worker type	Glue version
Succeeded	0	2023/11/20 01:11:02	2023/11/20 01:15:21	2 m 11 s	10 DPU	G.1X	4.0

Run details Input arguments (10) Continuous logs Run insights Metrics Spark UI

Job name	Start time (UTC)	Glue version	Last modified on (UTC)
amazon_products	November 19, 2023 1:11:02 AM	4.0	November 19, 2023 1:15:21 AM
ID	End time (UTC)	Worker type	Log group name
jr_62ed2b07776f5fb1ebc6ea17606d9024b0eaab86d6fc15 November 19, 2023 1:15:21 AM 70621eb6b2db02940	-	G.1X	/aws-glue/jobs
Run status	Start-up time	Max capacity	Number of workers
Succeeded	7 seconds	10 DPU	10
Retry attempt number	Execution time	Execution class	Timeout