
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

FWM 318

How to build location-aware workflows to detect data anomalies

Matt Nightingale (he/him)
Startup Solutions Architect
AWS

Drishti Arora (she/her)
Cloud Application Architect
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Defining and identifying location anomalies
- Solving location anomalies
- re:Invent Pizza Store example use-case

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Defining location data anomalies



- Completeness
- Accuracy
- Fit for Purpose

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

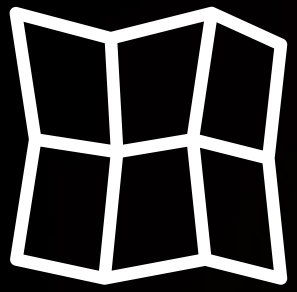
Anomalous location data

- Completeness
- Accuracy
- Fit for Purpose

UID	Business	Address	City	State	ZipCode	Latitude	Longitude
1	“Wayne Brady”	"3355 Las Vegas Blvd S"	Las Vegas	NV	89101	36.121451	-115.1696112
2	"The UPS Store"	”null"	Las Vegas	NV	89117	36.159901	-115.278969
3	"Hobby Lobby"	"4955 S Fort Apache Rd"	Las Vegas	NV	0	36.098997	-115.299264
4	"Glam Sass Makeup Artist"	”3651 S Lindell Rd Suite H"	Topaz	NV	89148	37.374431	-121.856157
5	“Steak ‘n Shake”	“Rt 374”	Las Vegas	NV	89183	36.011406	-115.175787
6	“/usr/lib”	“520 E Freemont St, Ste 200”	Las Vegas	NV	89101	36.169116	-115.139599

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Techniques to identify location data anomalies



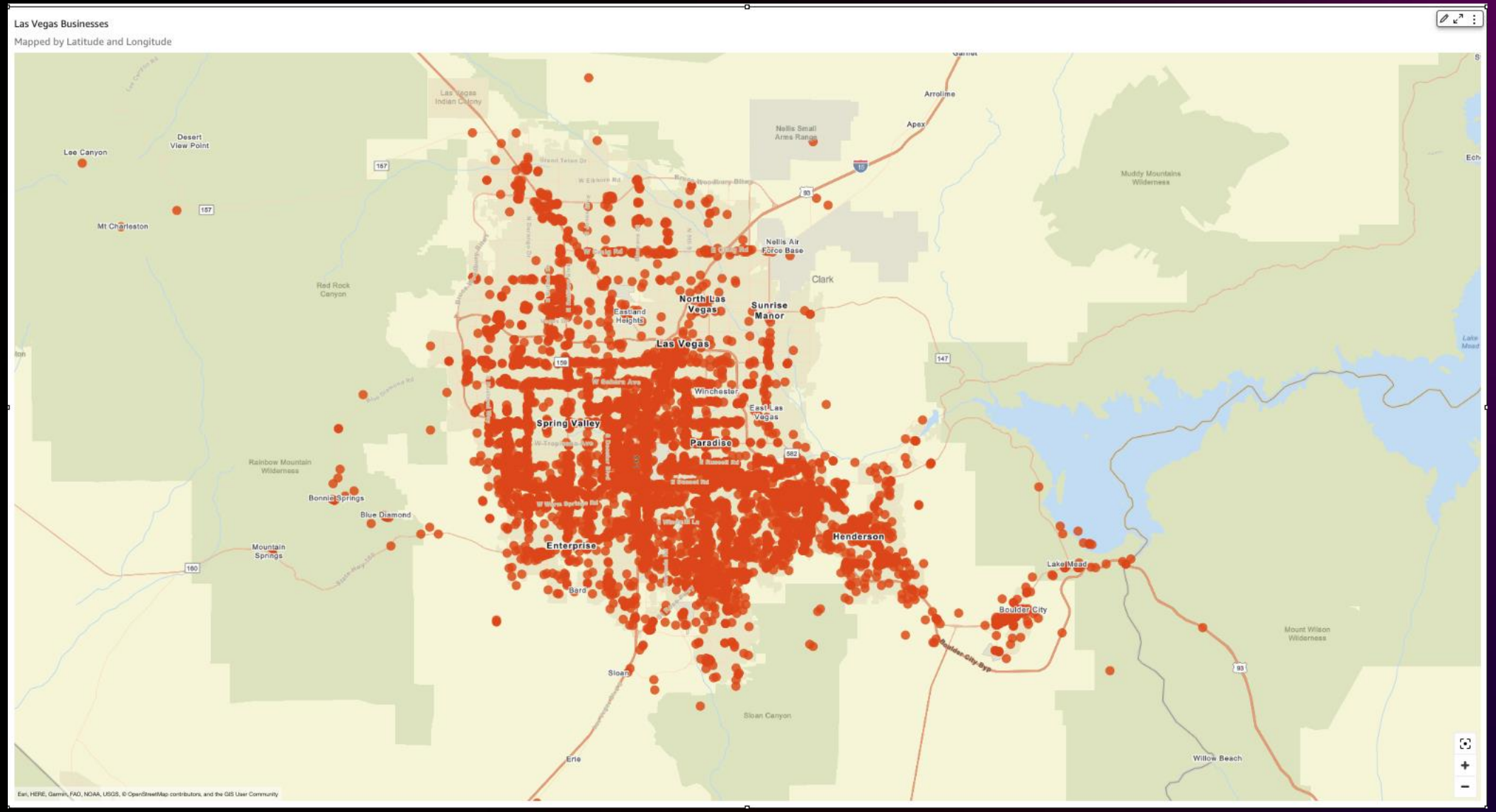




- Mapping
- Data sampling
- Data analysis

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Mapping anomalies with Amazon QuickSight

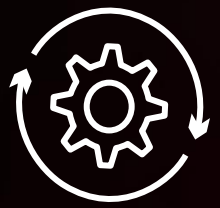


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Defining a sample size

Population size

Margin of error



Yamane's Formula

Sample size

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Sample sizes

	Size of Population					
Margin of Error	200	500	1,000	2,500	5,000	>5,000
+/- 10%	65	81	88	93	94	96
+/- 7.5%	92	127	146	160	165	171
+/- 5%	132	217	278	333	357	384
+/- 3%	169	341	516	748	880	1,067

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The AWS analytics portfolio

Data, visualization, engagement, & machine learning

Data Exchange | QuickSight | Pinpoint | SageMaker | Comprehend | Lex | Polly | Rekognition | Translate

+ many more

Analytics

Redshift | EMR (Spark & Hadoop) | AWS Glue DataBrew | AWS Glue (Spark & Python) | Athena | Kinesis Data Analytics

Data lake infrastructure & management

S3/Glacier | Lake Formation | AWS Glue

Data movement

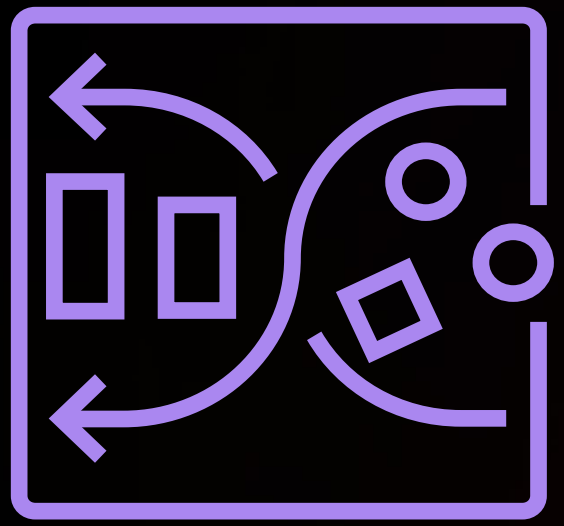
Database Migration Service | Snowball | Snowmobile | Kinesis Data Streams | Kinesis Data Firehose | Managed Streaming for Apache Kafka

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue DataBrew

Flag anomalies with built-in Glue DataBrew recipes



Visual interface to clean and normalize data without needing to write code

350+ built-in transformations without writing code

Flag outliers and missing values, deduplicate data

Data Deduplication

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue Jobs

Transform your dataset to become anomaly aware with Glue Jobs



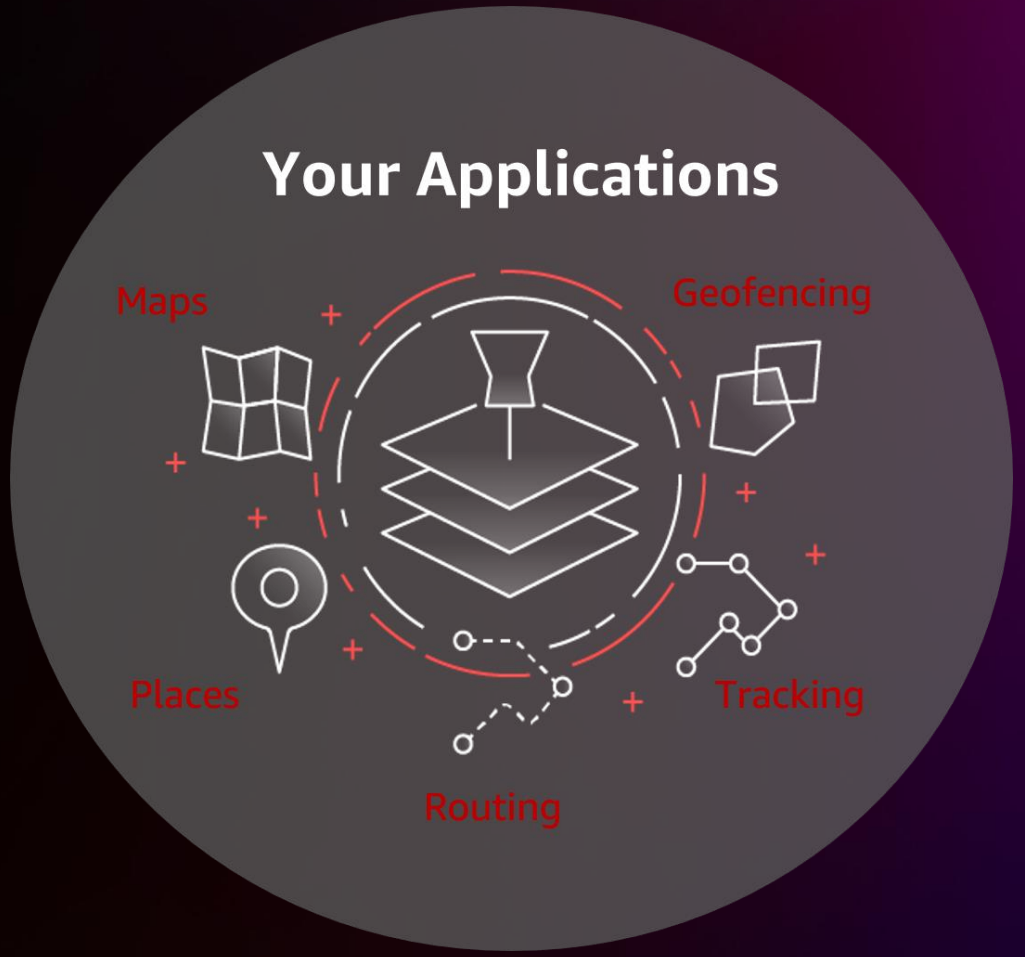
- Serverless, scalable execution model
- Support for Spark, Streaming ETL, Scala and Python Shell
- Automate jobs with event-based triggers
- Complete data integration

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Solving location data anomalies with Amazon Location Service



Amazon Location Service

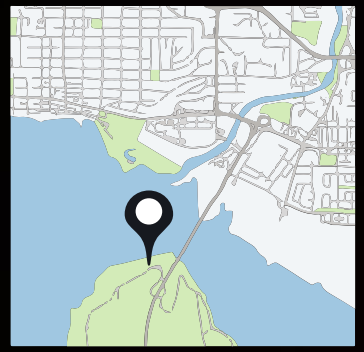


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

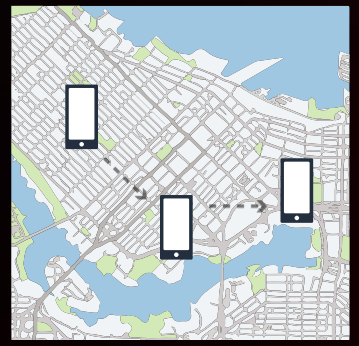
Amazon Location Service features



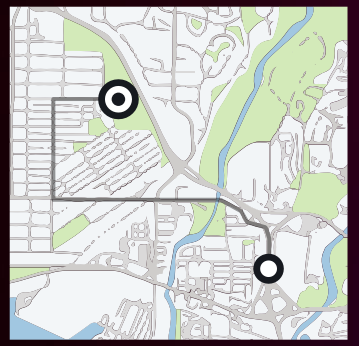
Maps



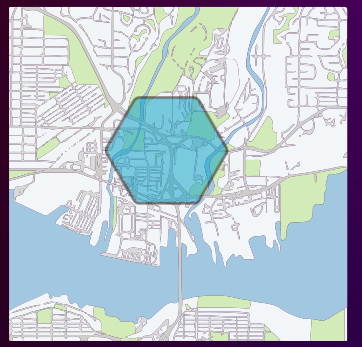
Places



Trackers



Routes



Geofences

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Geocoding with Amazon Location Service

```
place_id = location.search_place_index_for_text(
                    IndexName=location_index,
                    FilterBBox=[-115.3347, 35.5252,
                                -114.5145, 36.2631],
                    #Bounding Box for Clark County, NV
                    Text="Southern Highlands Golf Club")
```

```
{'Place': {'AddressNumber': '1', 'Country': 'USA', 'Geometry':
{'Point': [-115.19361, 35.97211]}, 'Interpolated': False, 'Label':
'Southern Highlands Golf Club, 1 Robert Trent Jones Ln, Las Vegas,
NV 89141-6038, United States', 'Municipality': 'Las Vegas',
'Neighborhood': 'Southern Highlands', 'PostalCode': '89141-6038',
'Region': 'Nevada', 'Street': 'Robert Trent Jones Ln', 'SubRegion':
'Clark', 'TimeZone': {'Name': 'America/Los_Angeles', 'Offset': -
25200}}
```

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Reverse-Geocoding with Amazon Location Service

```
place_id = location.search_place_index_for_position(
                        IndexName=location_index,
                        Position=[-115.1935488, 35.9723318])
```

```
{'Place': {'AddressNumber': '1', 'Country': 'USA', 'Geometry':
{'Point': [-115.19361, 35.97211]}, 'Interpolated': False,
'Label': '1 Robert Trent Jones Ln, Las Vegas, NV 89141-6038,
United States', 'Municipality': 'Las Vegas', 'Neighborhood':
'Southern Highlands', 'PostalCode': '89141-6038', 'Region':
'Nevada', 'Street': 'Robert Trent Jones Ln', 'SubRegion':
'Clark', 'TimeZone': {'Name': 'America/Los_Angeles', 'Offset': -
25200}}
```

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Place Suggestion with Amazon Location Service

```
suggestion = location.search_place_index_for_suggestions(
    FilterBBox=[-115.3347, 35.5252,
                -114.5145, 36.2631],
    #Bounding Box for Clark County, NV
    IndexName=location_index,
    Language='en',
    MaxResults=2,
    Text='Southern Highlands'
)
```

```
{'Text': 'Southern Highlands Golf Club, 1 Robert Trent Jones Ln, Las Vegas, NV 89141-6038, United States'}
```

aws

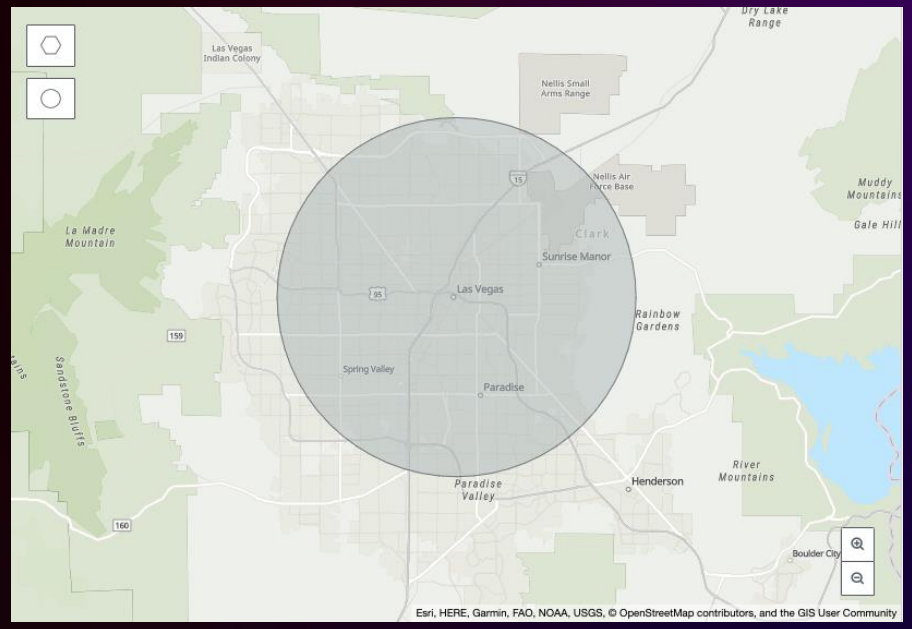
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Address anomaly use-case

Pizza Store Order Processing:

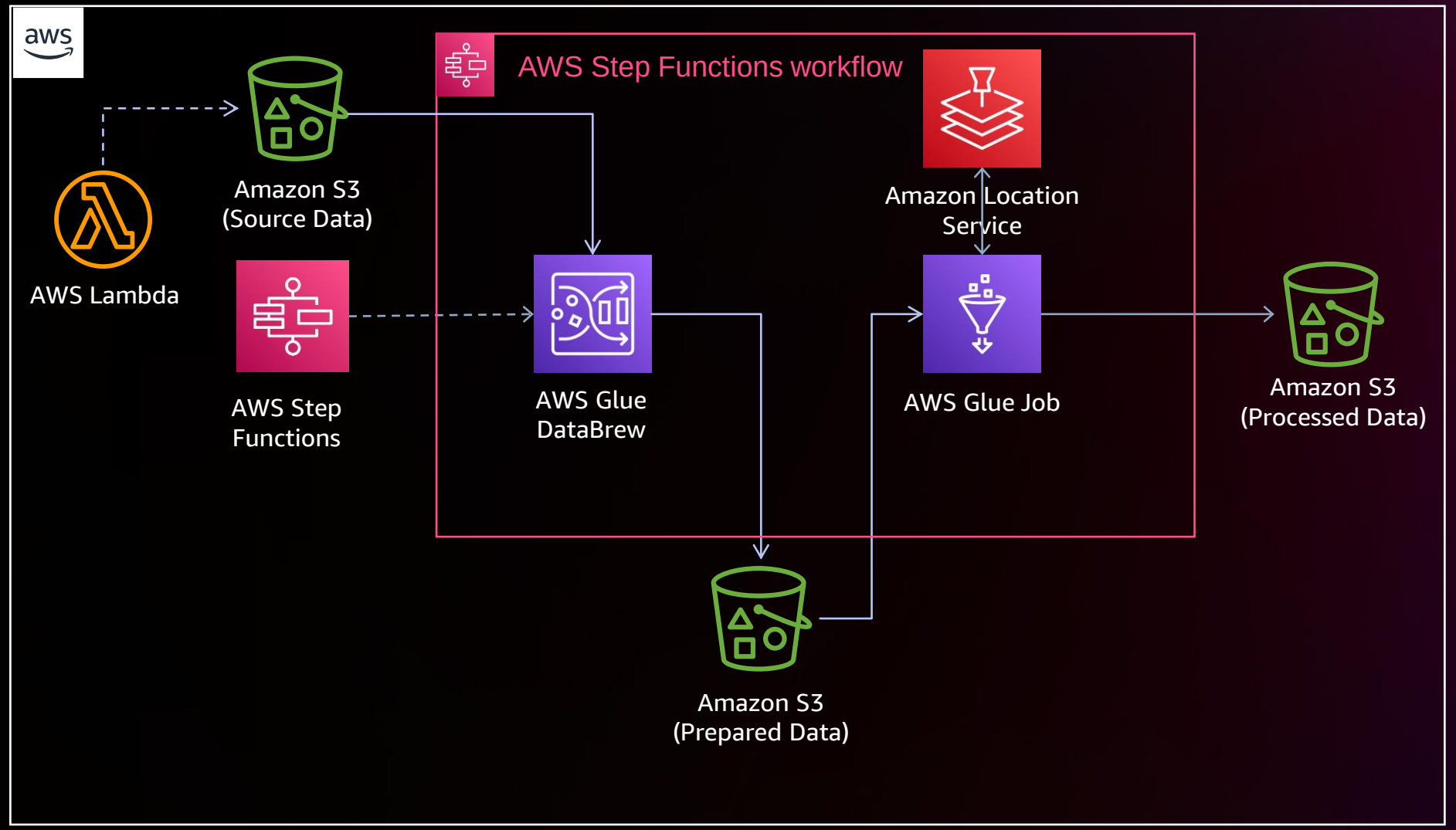
- re:Invent Pizza Company has a store location at position X in Las Vegas
- re:Invent Pizza needs to provide delivery to its customers and ensure the pizza arrives hot and on-time to a happy customer at the correct location
- re:Invent pizza has a database of customer addresses which needs to be checked for anomalies





© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Anomaly detection workflow



Scan for GitHub Repo!

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Matt Nightingale
nghtm@amazon.com

Drishti Arora
adrishti@amazon.com

Connect on LinkedIn!

Connect on LinkedIn!



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.