

Feature Engineering

Feature challenges

- Hard to share and reuse
- Hard to reliably serve in production with low latency
- skew in feature values between training and serving



Fully managed Feature Store

A rich feature repository to serve, share, and re-use ML features.



Share and reuse ML features across use cases

Centralized feature repository with easy APIs to search and discover features, fetch them for training/serving, and manage permissions.



Serve ML features at scale with low latency

Offload the operational overhead of handling infrastructure for low-latency scalable feature serving.



Alleviate training-serving skew

Compute feature values once; re-use for training and serving. Track and monitor for drift and other quality issues.



Use batch and streaming feature ingestion

Ingest features efficiently in large batches or in real time as data streams in. ETA for streaming ingestion is Q3 2021.

Feature Store

Features [+ CREATE ENTITY TYPE](#) [VIEW INGESTION JOBS](#) [VIEW BATCH SERVING JOBS](#)

Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

Region: us-central1 (Iowa) [?](#)

Filter Enter a property name

Feature	Entity type	Featurestore	Description
budget_id	budget	hello_world	entity id column
newspaper_budget	budget	hello_world	—
radio_budget	budget	hello_world	—
sales_total	budget	hello_world	—
tv_budget	budget	hello_world	—

- Feature Store is a top-level container for features and their values
- Permitted users can add and share their features
- Users can define features and ingest values from various sources

Entity type

Features [+ CREATE ENTITY TYPE](#) [VIEW INGESTION JOBS](#) [VIEW BATCH SERVING JOBS](#)

Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

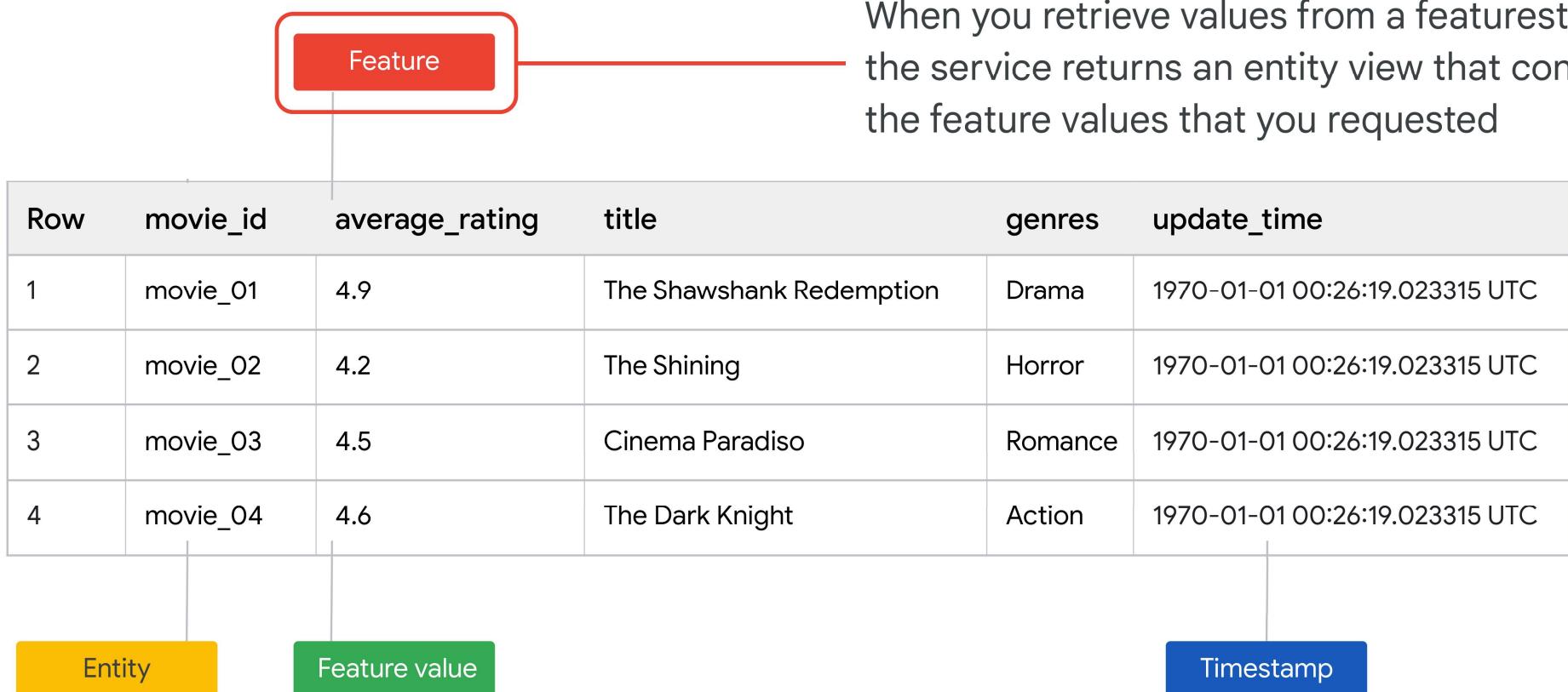
Region: us-central1 (Iowa) [?](#)

Filter Enter a property name

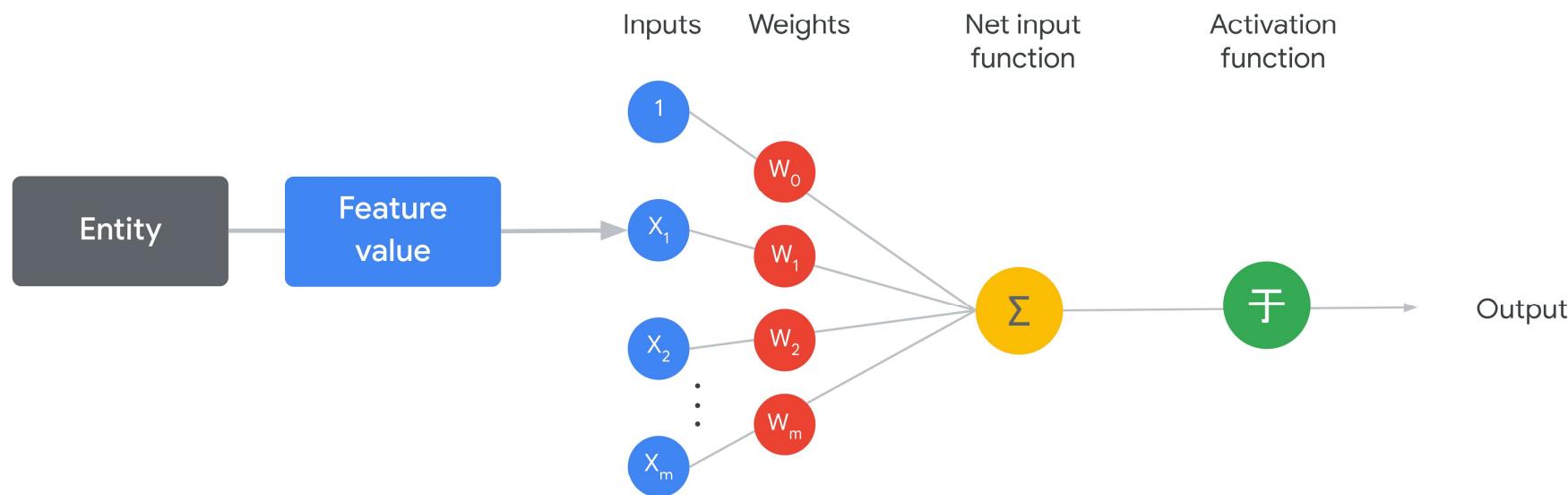
Feature	Entity type	Featurestore	Description
budget_id	budget	hello_world	entity id column
newspaper_budget	budget	hello_world	—
radio_budget	budget	hello_world	—
sales_total	budget	hello_world	—
tv_budget	budget	hello_world	—

- An *entity type* is a collection of semantically related features
- You define your own entity types based on the concepts that are relevant to your use case

Entity view

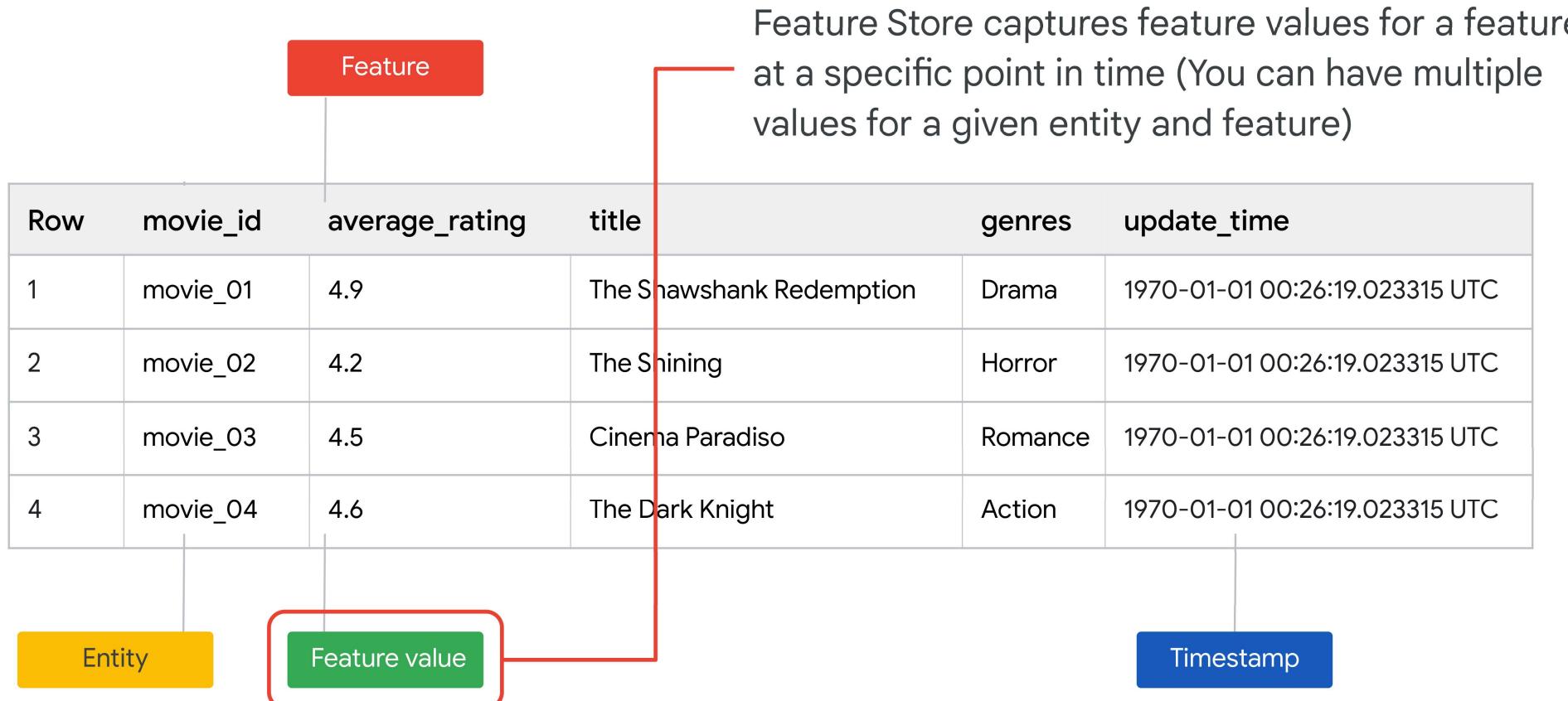


A feature describes some entity



Examples: Age of user, price of product, category of web page

Feature value



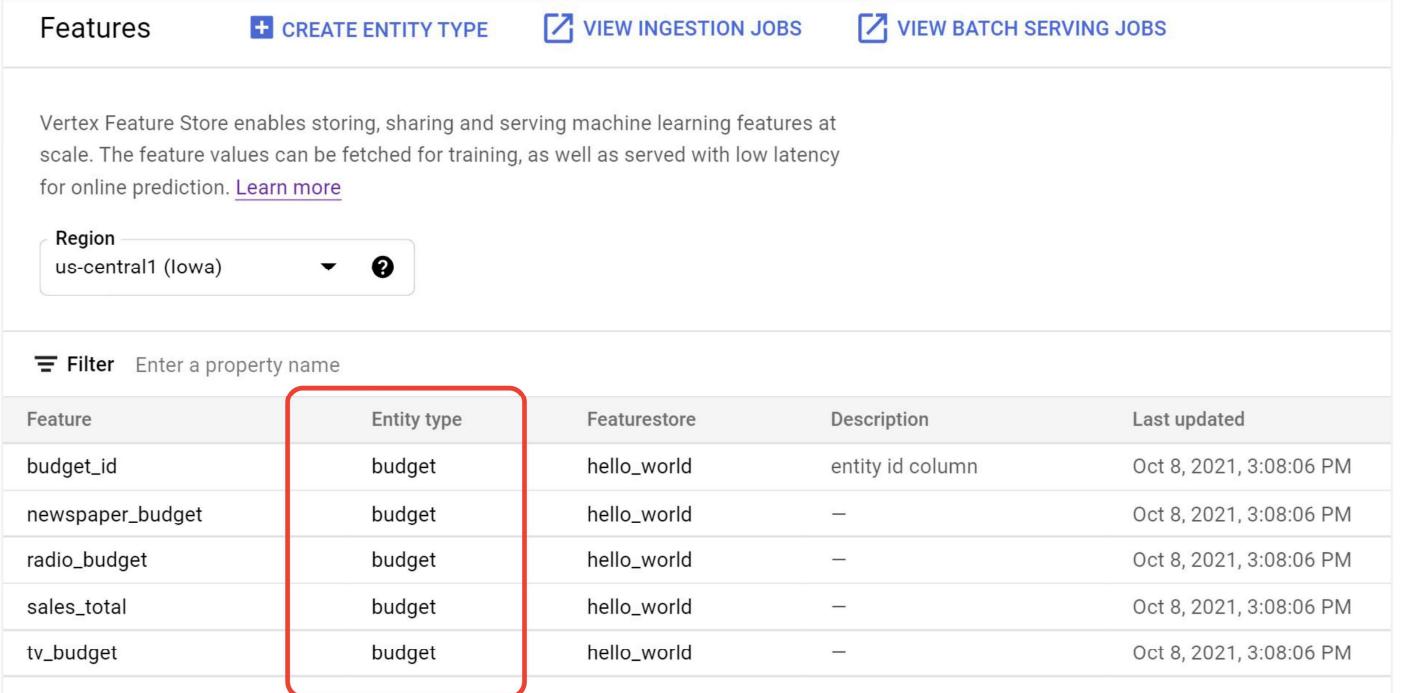
Timestamp

The diagram illustrates a movie dataset with various annotations:

- Feature:** A red callout points to the "Feature" column header.
- Entity:** A yellow callout points to the "movie_id" column header.
- Feature value:** A green callout points to the "average_rating" column header.
- Timestamp:** A blue callout points to the "update_time" column header, which is also highlighted with a red border.
- Text:** A descriptive text box states: "The timestamp column indicates when the feature values were generated."

Row	movie_id	average_rating	title	genres	update_time
1	movie_01	4.9	The Shawshank Redemption	Drama	1970-01-01 00:26:19.023315 UTC
2	movie_02	4.2	The Shining	Horror	1970-01-01 00:26:19.023315 UTC
3	movie_03	4.5	Cinema Paradiso	Romance	1970-01-01 00:26:19.023315 UTC
4	movie_04	4.6	The Dark Knight	Action	1970-01-01 00:26:19.023315 UTC

In Vertex AI Feature Store



The screenshot shows the Vertex AI Feature Store interface. At the top, there are navigation links: 'Features' (selected), '+ CREATE ENTITY TYPE', 'VIEW INGESTION JOBS', and 'VIEW BATCH SERVING JOBS'. Below this, a descriptive text states: 'Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction.' A 'Learn more' link is provided. A 'Region' dropdown is set to 'us-central1 (Iowa)'. A 'Filter' input field is present. The main table lists six features: 'budget_id', 'newspaper_budget', 'radio_budget', 'sales_total', and 'tv_budget'. All features have the same 'Entity type' value: 'budget'. The 'Featurestore' column shows 'hello_world' for all entries. The 'Description' column contains 'entity id column' for the first entry and '-' for the others. The 'Last updated' column shows 'Oct 8, 2021, 3:08:06 PM' for all entries.

Feature	Entity type	Featurestore	Description	Last updated
budget_id	budget	hello_world	entity id column	Oct 8, 2021, 3:08:06 PM
newspaper_budget	budget	hello_world	-	Oct 8, 2021, 3:08:06 PM
radio_budget	budget	hello_world	-	Oct 8, 2021, 3:08:06 PM
sales_total	budget	hello_world	-	Oct 8, 2021, 3:08:06 PM
tv_budget	budget	hello_world	-	Oct 8, 2021, 3:08:06 PM

One entity type
(budget)

In Vertex AI Feature Store

Features [+ CREATE ENTITY TYPE](#) [VIEW INGESTION JOBS](#) [VIEW BATCH SERVING JOBS](#)

Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

Region: us-central1 (Iowa) [?](#)

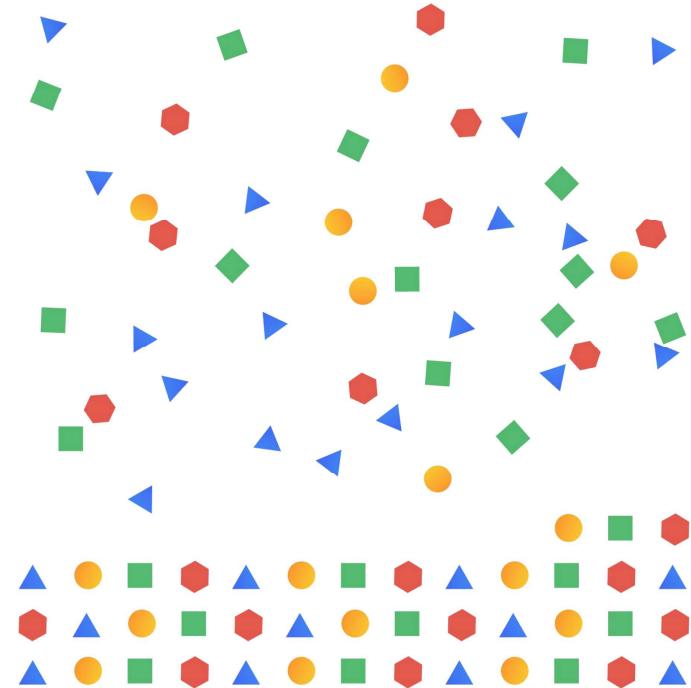
Filter Enter a property name

Feature	Entity type	Featurestore	Description	Last updated
budget_id	budget	hello_world	entity id column	Oct 8, 2021, 3:08:06 PM
newspaper_budget	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM
radio_budget	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM
sales_total	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM
tv_budget	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM

Many values

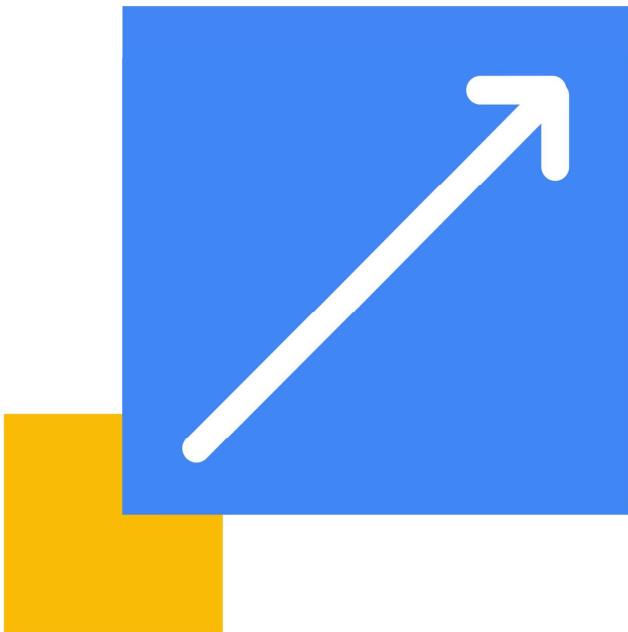
Feature ingestion

- *Feature ingestion* is the process of importing feature values computed by feature engineering jobs into a featurestore
- Entity type and features must be defined in the featurestore
- Batch ingestion means you can do a bulk ingestion of values into a featurestore



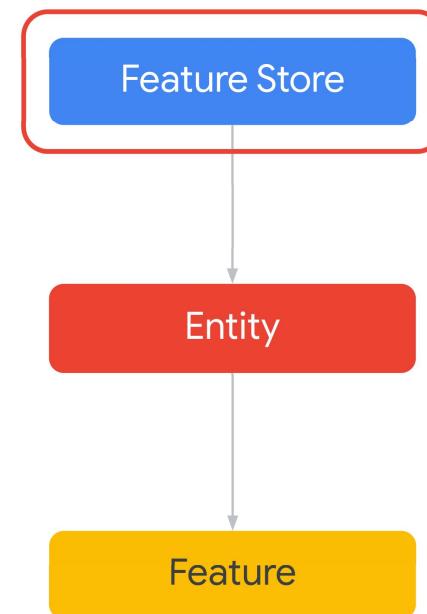
Feature serving

- *Feature serving* is the process of exporting stored feature values for training or inference
- Feature Store offers two methods for serving features:
 - *Batch serving* is for high throughput and serving large volumes of data for offline processing
 - *Online serving* is for low-latency data retrieval of small batches of data for real-time processing (like for online predictions)



Feature Store *creation*

Step 1. Create a featurestore



 Features  CREATE ENTITY TYPE  VIEW INGESTION JOBS  VIEW BATCH SERVING JOBS

Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

Region: us-central1 (Iowa)  

 Filter Enter a property name

Feature	Entity type	Featurestore	Description	Last updated
				

You don't have any features in this region yet
Get started by creating a featurestore. [Learn more](#)

 CREATE FEATURESTORE

Create featurestore

Name *

hello_world



The name of your featurestore

Region

us-central1 (Iowa)



The region containing your featurestore

Number of online serving nodes *

1

The number of nodes to allocate for your featurestore

Encryption

Use a customer-managed encryption key (CMEK)

▲ SHOW LESS

CREATE

CANCEL

 Features  CREATE ENTITY TYPE  VIEW INGESTION JOBS  VIEW BATCH SERVING JOBS

 Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

 Region: us-central1 (Iowa)  

 Filter: Enter a property name

Feature	Entity type	Featurestore	Description	Last u
-	-	<u>Featurestore</u> <u>hello_world</u> 	-	-

The screenshot shows the 'hello_world' model details page in the Vertex AI interface. The top navigation bar includes a back arrow, the model name 'hello_world', and an 'EDIT CONFIGURATION' button. Below the navigation is a horizontal menu with three tabs: 'PROPERTIES' (highlighted with a red box), 'ENTITY TYPES', and 'METRICS'. On the left, a sidebar lists various configuration sections with corresponding icons. The main content area displays 'Basic info' and 'Online serving configuration' sections.

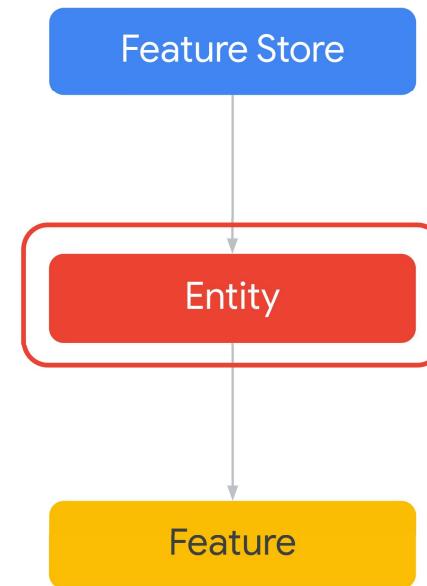
Basic info

Status	Active
Created	Oct 6, 2021, 10:38:32 PM
Last updated	Oct 6, 2021, 10:38:32 PM
Region	us-central1
Encryption type	Google-managed key

Online serving configuration

Nodes per cluster	1
-------------------	---

Step 2. Create an entity type



 Vertex AI

Features

+ CREATE ENTITY TYPE

 [Dashboard](#)

 [Datasets](#)

 [Features](#)

 [Labeling tasks](#)

 [Workbench](#)

VIEW INGESTION JOBS **VIEW BATCH SERVING JOBS**

Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

Region: us-central1 (Iowa) ▾ ?

Filter: Enter a property name

Create entity type

Entity types group and contain related features. For example, a “movies” entity type might contain features like “title” and “genre”. [Learn more](#)

Region

us-central1 (Iowa) ▼ ?

Featurestore *

hello_world

Entity type name *

budget_id

Must start with a letter or underscore. Can use letters, numbers, and underscores.

Description

Optional text description of the entity type

Feature monitoring PREVIEW

Feature monitoring PREVIEW

Provides descriptive statistics and distribution shapes. Enables feature monitoring for all features in the entity type. You can also edit feature monitoring at the feature level, which will override this setting.

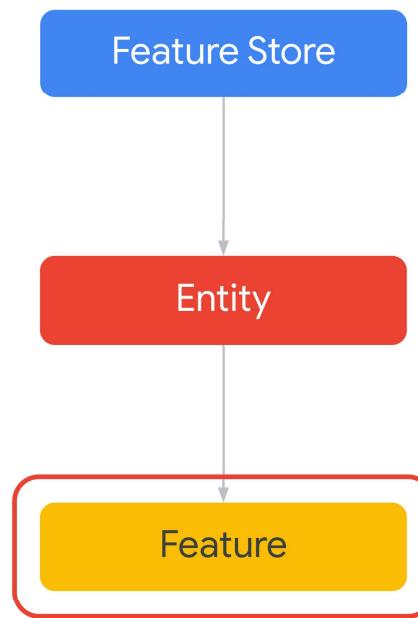
Disabled

Monitoring time interval

1 days

CREATE CANCEL

Step 3. Add features



 Vertex AI

[← budget](#)  INGEST VALUES  EDIT INFO  DELETE

	FEATURES	ENTITY TYPE PROPERTIES
 Dashboard		
 Datasets		
 Features	Features	
 Labeling tasks	A feature is a measurable attribute of an entity type. After you add features in your entity type, you can then associate your features with values stored in BigQuery or Cloud Storage. Learn more	
 Workbench	ADD FEATURES	

Add features

A feature is a measurable attribute of an entity type. After you add features in your entity type, you can then associate your features with values stored in BigQuery or Cloud Storage. [Learn more](#)

Feature names can include lowercase letters, numbers and underscores and must start with a lowercase letter or underscore

Feature name *	Value type *	Description	Override monitoring values	Feature monitoring	Interval *
Feature name 1 *	Value type 1 *	Description 1	<input type="checkbox"/> Override	<input checked="" type="checkbox"/> Disabled	Interval 1 1 days

[+ ADD ANOTHER FEATURE](#)

[SAVE](#) [CANCEL](#)

Features

[+ CREATE ENTITY TYPE](#)[VIEW INGESTION JOBS](#)[VIEW BATCH SERVING JOBS](#)

Vertex Feature Store enables storing, sharing and serving machine learning features at scale. The feature values can be fetched for training, as well as served with low latency for online prediction. [Learn more](#)

Region

us-central1 (Iowa)

**Filter** Enter a property name

Feature	Entity type	Featurestore	Description	Last updated	Created
budget_id	budget	hello_world	entity id column	Oct 8, 2021, 3:08:06 PM	Oct 8, 2021, 3:08:06 PM
newspaper_budget	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM	Oct 8, 2021, 3:08:06 PM
radio_budget	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM	Oct 8, 2021, 3:08:06 PM
sales_total	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM	Oct 8, 2021, 3:08:06 PM
tv_budget	budget	hello_world	—	Oct 8, 2021, 3:08:06 PM	Oct 8, 2021, 3:08:06 PM

Ingestion jobs

The screenshot shows the Vertex AI interface with the 'Features' menu item selected. The main content area is titled 'Ingestion jobs' and displays a message about importing feature data from BigQuery or Cloud Storage. A table lists one completed ingestion job, which is highlighted with a red border.

Ingestion jobs import feature data from BigQuery or Cloud Storage so it can be used in a featurestore. Before you import data, you need to define the corresponding entity type and features. [Learn more](#)

Ingestion job	Entity type	Featurestore	Elapsed time	Entities	Feature values	Last updated
3745873525550874624	budget	hello_world	5 min 33 sec	1,200	6,000	Oct 14, 2021, 11:01:47 PM

Vertex AI

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments
- Marketplace

Ingestion job details

FEATURES PROPERTIES

This ingestion job finished on October 14, 2021 at 11:01:47 PM GMT-7

Status	Finished
Job ID	3745873525550874624
Created	Oct 14, 2021, 10:56:13 PM
Elapsed time	5 min 33 sec
Region	us-central1
Workers	1
Data source	bq://cloud-training-demos.xyz_team_dataset.tidyadvertising_1_string_int
Entity type	<u>budget</u>
Featurestore	<u>hello_world</u>
Ingested entities	1,200

Format of input data for batch ingestion

Data for ingestion should have the following columns:

- Entity_id: The ID of ingested entity
- Timestamp: The timestamp at which the feature was generated/computed (this timestamp is critical for correct feature serving)
- Feature columns that match the destination feature name

Example:

user_id	timestamp	Feature 1	Feature 2	Feature 3
88	1552521600	true	177	"abc"
88	1552521600	true	666	"xyz"
96	1569542400	false	235	"foobar"

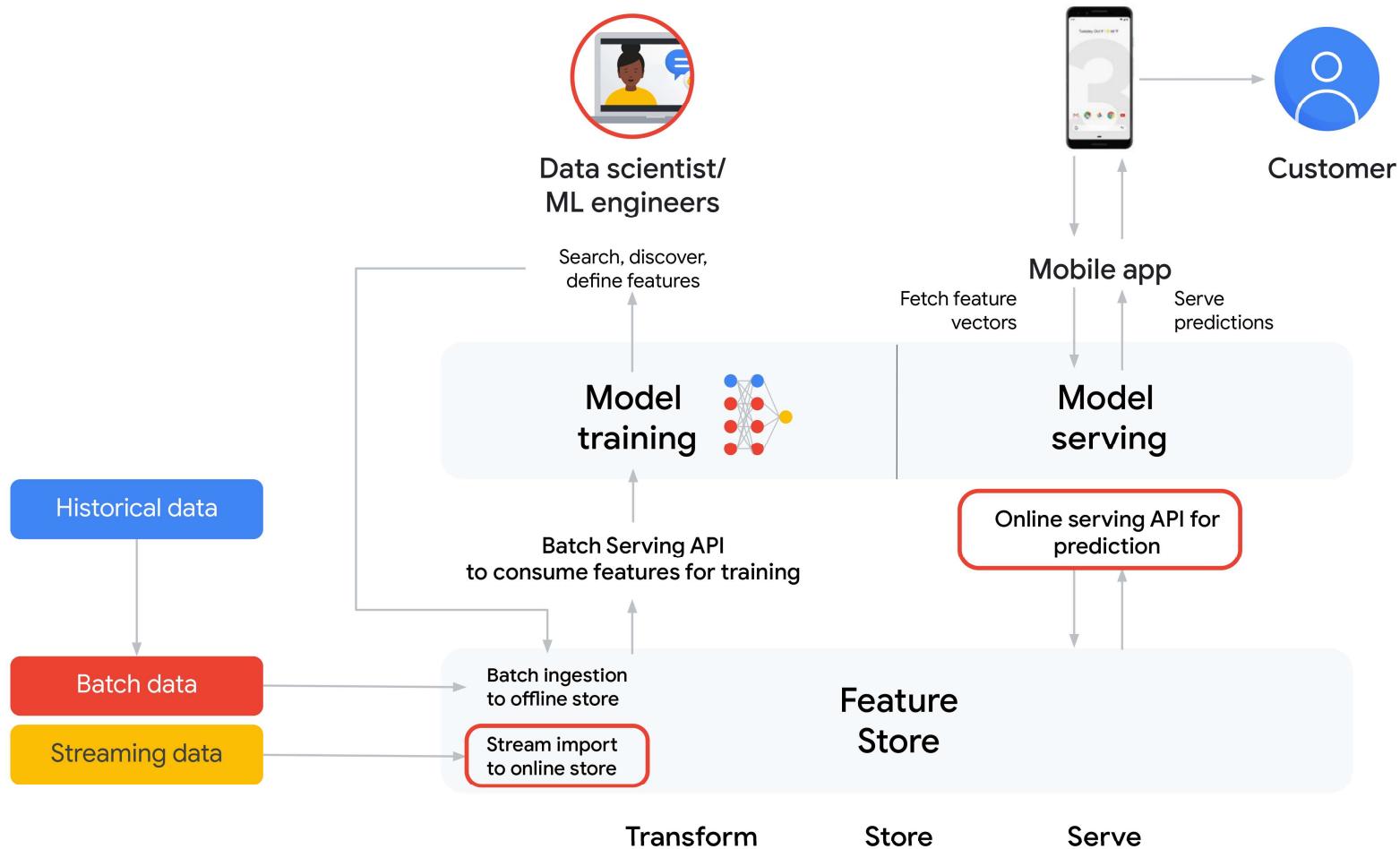
Feature serving:

Batch and online methods

Feature serving

- *Feature serving* is the process of exporting stored feature values for training or inference
- Feature Store offers two methods for serving features: batch and online





Feature serving

Vertex AI

Batch serving jobs

REFRESH

Region
us-central1 (Iowa)

Batch serving job Featurestore Elapsed time Entity types Feature values Last updated Started Status

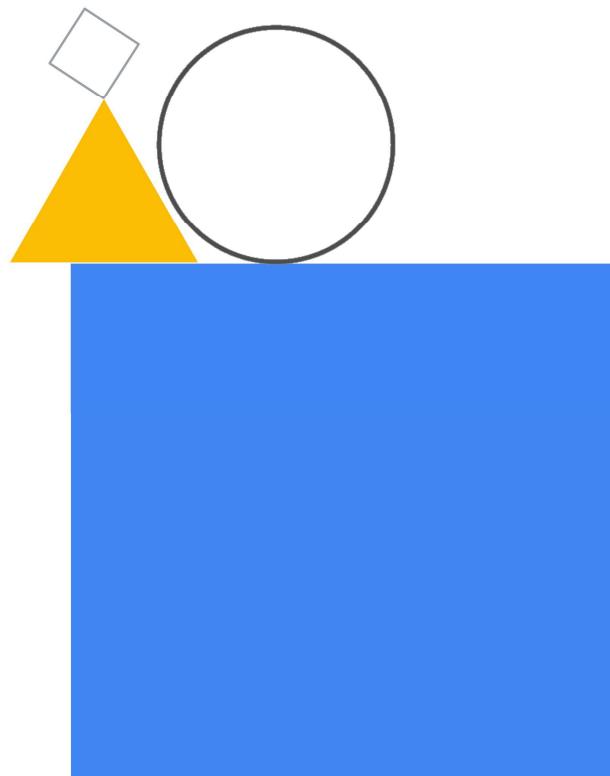


You don't have any batch serving jobs in this project yet
You can create a batch serving job through the Feature Store API

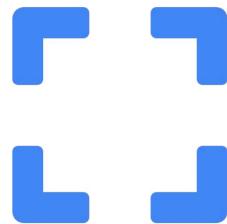
Show debug pane

- Dashboard
- Datasets
- Features
- Labeling tasks
- Workbench
- Pipelines
- Training
- Experiments
- Marketplace

Raw Data to Features



What raw data do we need to collect to predict the price of a house?



Lot size
Number of rooms

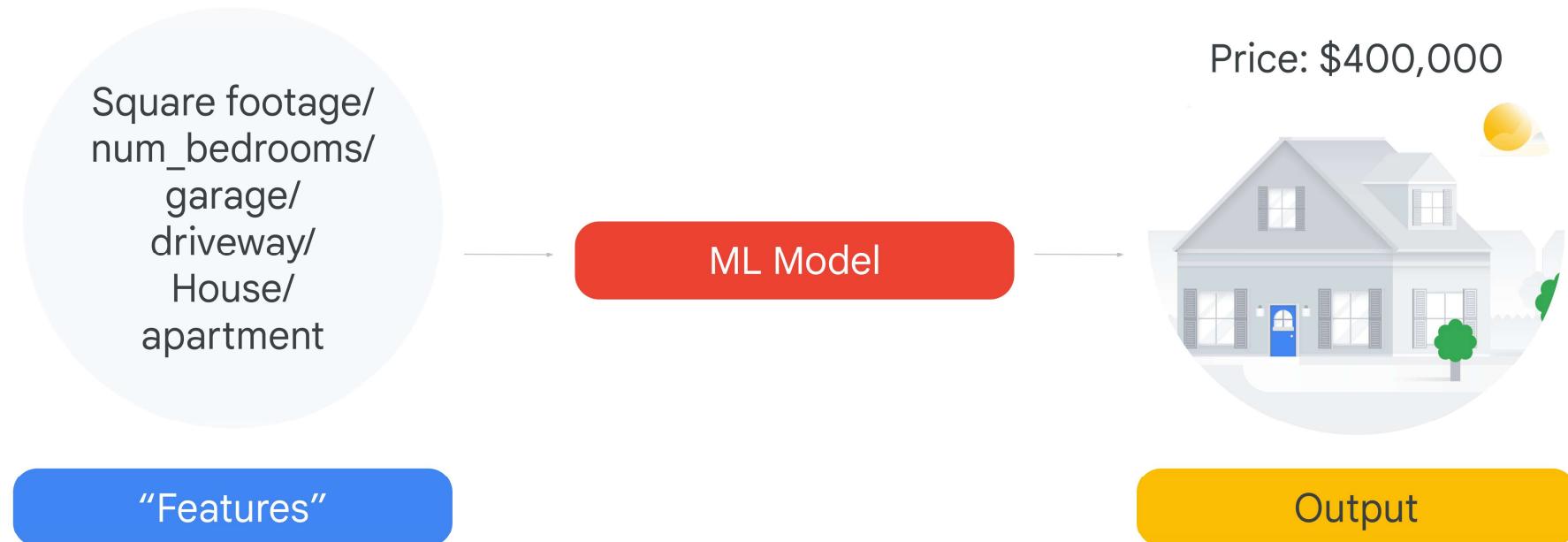


Historical
sale price



Location, location,
location

Predict property value from historical data



Raw data must be mapped into numerical feature vectors

```
0 : {  
    house_info : {  
        ✓ num_rooms: 6  
        ✓ num_bedrooms: 3  
        ✓ street_name: "Main Street"  
        ✓ num_basement_rooms: -1  
        ...  
    }  
}
```



```
[  
    6.0,  
    1.0,  
    0.0,  
    0.0,  
    0.0,  
    9.321,  
    -2.20,  
    1.01,  
    0.0,  
    ....  
]
```

What makes a good feature?



— 1 —

— 2 —

— 3 —

— 4 —

— 5 —

Be related to the
objective

Be known at
prediction-time

Be numeric with
meaningful
magnitude

Have enough
examples

Bring human
insight to
problem

Categorical variables should be one-hot encoded

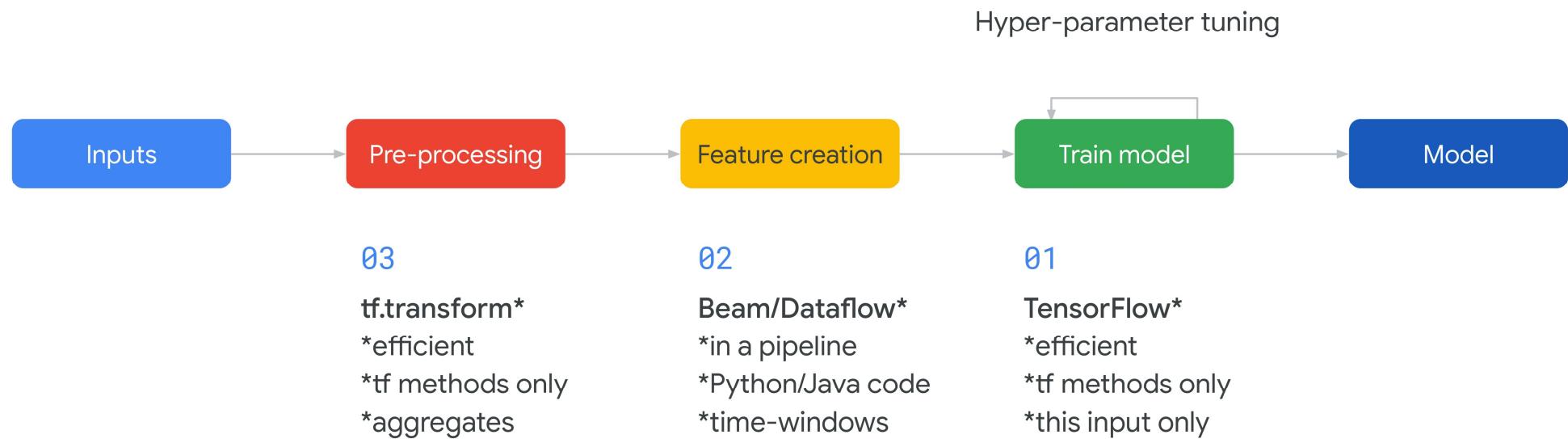
```
{  
  "transactionId": 42,  
  "name": "Ice Cream",  
  "price": 2.50,  
  "tags": ["cold", "dessert"],  
  "servedBy": {  
    "employeeId": 72365,  
    "waitTime": 1.4,  
    "customerRating": 4},  
  "storeLocation": {  
    "latitude": 35.3,  
    "longitude": -98.7}  
},
```

[..., 4, ...]

(OR)

[..., 0, 0, 0, 1, 0, ...]

There are three possible places to do feature engineering, each of which has its pros and cons



Thanks