

Cloud Functions :

Cloud function name : **chicago-taxi-trip-etl**

URL : <https://us-central1-gke-elastic-394012.cloudfunctions.net/chicago-taxi-trip-etl>

Project ID : gke-elastic-394012

This API does all the ETL(Generates a Pyspark job in a Dataproc cluster with script in a gcs bucket: chicago-taxi-trip-dataset-01) directly from the public dataset of chicago taxi trip datasets and creates our processed table in our project.

Examples Request and Response :

Request :

POST <https://us-central1-gke-elastic-394012.cloudfunctions.net/chicago-taxi-trip-etl>

```
{
    "project_id" : "gke-elastic-394012",    # project id of the GCP and is not a mandatory field
and it is the default value
    "region" : "us-central1",              # region of the GCP and is not a mandatory field and it is
the default value
    "cluster_name" : "cluster-411a",        # Cluster of the Dataproc in GCP and is not a
mandatory field and it is the default value
    "bucket_name" : "chicago-taxi-trip-dataset-01",    #GCS Bucket name and is not a
mandatory field and it is the default value
    "script_path" : "ETL.py",              #Script name in the GCS Bucket and is not a mandatory
field and it is the default value
    "dataset_id" : "chicago_taxi_trips",   # Dataset id for the new processed table and is not
a mandatory field and it is the default value
    "output_table_name" : "sample_taxi_trips", # Table name for the new processed table
and is not a mandatory field and it is the default value

    "sample_pct" : "0.001"                 #Percentage Size of the public dataset table to be taken
for the demo/sample and is not a mandatory field and it is the default value
}
```

Response :

Ok , Status : 200 # This means a pyspark job has been initiated successfully if not there will be an error shown up.

Cloud function name : model-training-and-test

URL : <https://us-central1-gke-elastic-394012.cloudfunctions.net/model-training-and-test>

Project ID : gke-elastic-394012

This API does all the model training and evaluation from the processed dataset of chicago taxi trip dataset and generates Model evaluations and predictions and also saves the model in a GCS Bucket(chicago-taxi-trip-dataset-01) in our project.

Examples Request and Response :

Request :

POST <https://us-central1-gke-elastic-394012.cloudfunctions.net/model-training-and-test>

```
{
  "project_id" : "gke-elastic-394012", # project id of the GCP and is not a mandatory field
and it is the default value
  "dataset_id" : "chicago_taxi_trips", # Dataset ID of the new processed table and is not
a mandatory field and it is the default value
  "table_name" : "sample_taxi_trips", # Table name of the new processed table and is not
a mandatory field and it is the default value
}
```

"model" : "poly_rgr" # Model to be used for the training and evaluation and the options for this field are : 'grad_boost', 'svr', 'ann', 'poly_rgr' and is not a mandatory field and it is the default value

}

Response :

{

"Mean Absolute Error": "2.3684585397414315",
"Mean Squared Error": "345.6775536633942",
"R-squared Score": "-49.50672142916113",
"Root Mean Squared Error": "18.592405806226214",
"predictions": [
2.7524421506552095,
0.7449622621105275,
0.4115087314503131,
0.7642214016808654,
0.6733461906451901,
0.6573344760011903,
0.5054483062117173,
0.8552346738641384,
0.7396520317928549,
0.7870325012536793,
1.2683317945998174,
3.050225639075027,
1.0667639115553038,
0.856621055818887,
1.162192965118465,
0.756809783032522,
0.7022209867720952,

.....

]
}