

Assignment 7: optional catchup assignment 2 - VERTEX AI - for midterm and quiz - this will catch up midterm.


1) Build an AutoML Forecasting Model with Vertex AI






Reference: <https://codelabs.developers.google.com/codelabs/automl-forecasting-with-vertex-ai#0>


Objectives:


- Create a managed dataset
- Import data from a Google Cloud Storage Bucket
- Update the column metadata for appropriate use with AutoML
- Train a model using options such as a budget and optimization objective
- Make online batch predictions

Review Data

 Google Cloud Platform Select a project





Iowa Liquor Retail Sales

Iowa Department of Commerce

Liquor sales in Iowa since 2012, by store, by item, and by day

[VIEW DATASET](#)

OVERVIEW

SAMPLES

Overview

This dataset contains every wholesale purchase of liquor in the State of Iowa by retailers for sale to individuals since January 1, 2012. The State of Iowa controls the wholesale distribution of liquor intended for retail sale, which means this dataset offers a complete view of retail liquor sales in the entire state. The dataset contains every wholesale order of liquor by all grocery stores, liquor stores, convenience stores, etc., with details about the store and location, the exact liquor brand and size, and the number of bottles ordered.

Additional details

Type: [Datasets](#)
Category: [Analytics](#), [Financial services](#), [Machine learning](#)
Dataset source: [Iowa Alcoholic Beverages Division](#)
Cloud service: BigQuery
Expected update frequency: Monthly

Google Cloud Platform CMPE260 Search products and resources

FEATURES & INFO SHORTCUT DISABLE EDITOR TABS

Explorer + ADD DATA

sales

SCHEMA DETAILS PREVIEW

Table schema

Field name	Type	Mode	Policy Tags	Description
invoice_and_item_number	STRING	NULLABLE		Concatenated invoice and line number associated with the liquor order. This provides a unique identifier for the individual liquor products included in the store order.
date	DATE	NULLABLE		Date of order
store_number	STRING	NULLABLE		Unique number assigned to the store who ordered the liquor.
store_name	STRING	NULLABLE		Name of store who ordered the liquor.
address	STRING	NULLABLE		Address of store who ordered the liquor.
city	STRING	NULLABLE		City where the store who ordered the liquor is located
zip_code	STRING	NULLABLE		Zip code where the store who ordered the liquor is located
store_location	STRING	NULLABLE		Location of store who ordered the liquor. The Address, City, State and Zip Code are geocoded to provide geographic coordinates. Accuracy of geocoding is dependent on how well the address is interpreted and the completeness of the reference data used.
county_number	STRING	NULLABLE		Iowa county number for the county where store who ordered the liquor is located
county	STRING	NULLABLE		County where the store who ordered the liquor is located

PERSONAL HISTORY PROJECT HISTORY SAVED QUERIES

Import Data

Navigate to Vertex AI Datasets

Google Cloud Platform CMPE260 Search products and resources

Vertex AI

Datasets + CREATE REFRESH

Managed datasets contain data used to train a machine learning model. [Learn more](#)

Region us-central1 (Iowa)

Filter Enter a property name

Name	ID	Status	Region	Type	Items	Last updated	Labels
No results to display							

You haven't created any datasets in this project yet

CREATE DATASET

Create Dataset

Google Cloud Platform

CMPE260

Search products and resources

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

Create dataset

iowa_daily

Can use up to 128 characters.

Select a data type and objective

First select the type of data your dataset will contain. Then select an objective, which is the outcome that you want to achieve with the trained model. [Learn more about model types](#)

IMAGE

TABULAR

TEXT

VIDEO

Regression/classification

Predict a target column's value. Supports tables with hundreds of columns and millions of rows.

Forecasting

Predict the likelihood of certain events or demand.

PREVIEW

Region

us-central1 (Iowa)

CREATE

CANCEL

Google Cloud Platform

CMPE260

Search products and resources

Vertex AI

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Marketplace

iowa_daily

CREATE LABELING TASK

SOURCE

ANALYZE

Before you begin, read the [data guide](#) to learn how to prepare your data. Then choose a data source.

Select a data source

CSV file: Can be uploaded from your computer or on Cloud Storage. [Learn more](#)

BigQuery: Select a table or view from BigQuery. [Learn more](#)

Upload CSV files from your computer

Select CSV files from Cloud Storage

Select a table or view from BigQuery

Select CSV files from Cloud Storage

Enter the Cloud Storage path to one or more CSV files. Data from multiple files will be referenced as one dataset.

Import file path *

gs://automl-demo-240614-km/iowa_liquor/iowa_daily.csv

BROWSE

ADD ANOTHER FILE

What happens next?

The selected CSV file will be associated with your dataset. Making changes to the referenced CSV file will affect the dataset.

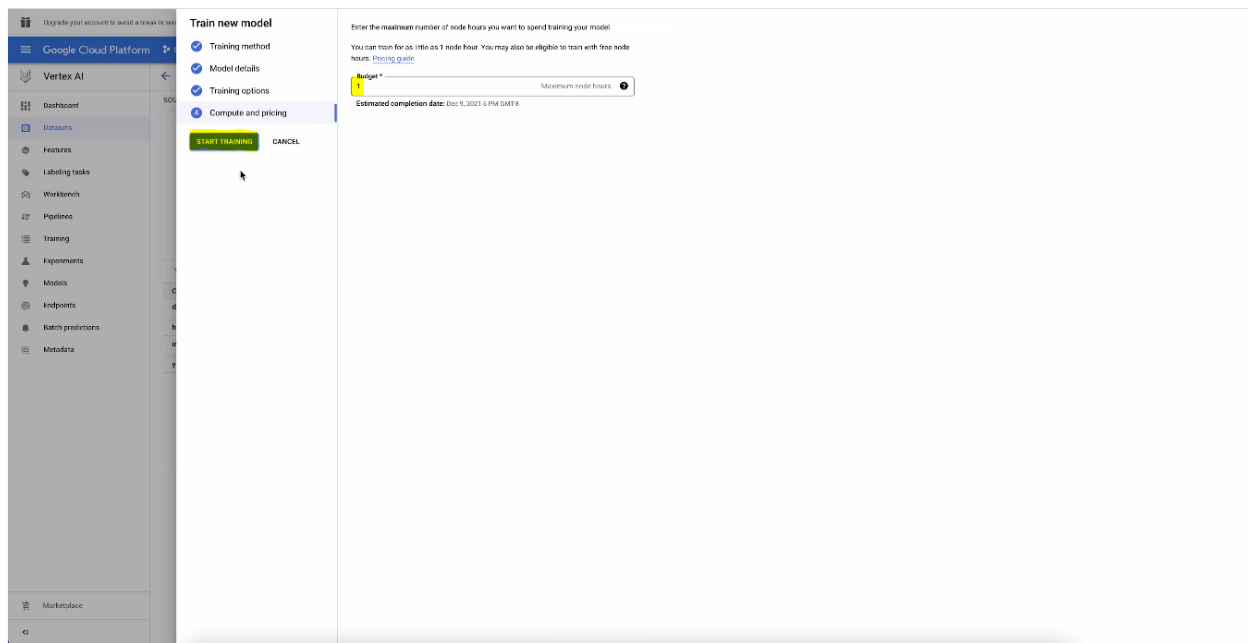
CONTINUE

Train Model

Step 1: Configure Model Features

The screenshot shows the Google Cloud Platform Vertex AI interface for a dataset named 'iowa_daily'. The left sidebar contains navigation options: Dashboard, Datasets, Features, Labeling tasks, Workbench, Pipelines, Training, Experiments, Models, Endpoints, Batch predictions, and Metadata. The main content area is divided into 'SOURCE' and 'ANALYZE' tabs, with 'ANALYZE' selected. Under 'ANALYZE', there are sections for 'Dataset Info' (Created: Dec 09, 2021 3:08 PM, Dataset format: CSV, Dataset location(s): gs://automl-demo-2_or/iowa_daily.csv), 'Summary' (Total columns: 4, Total rows: -), and 'Filter' (Enter property name or value). A table lists columns: 'ds' (Timestamp column), 'holiday', 'id' (Series Identifier), and 'y'. To the right, there are dropdowns for 'Series identifier column' (set to 'id') and 'Timestamp column' (set to 'ds'). A 'GENERATING STATISTICS...' button is visible. On the far right, the 'Training jobs and models' section includes a 'TRAIN NEW MODEL' button.

The screenshot shows the 'Train new model' configuration page in the Google Cloud Platform Vertex AI interface. The left sidebar is the same as the previous screenshot. The main content area is divided into 'Train new model' and 'Train new model' sections. The 'Train new model' section has a sidebar with options: Training method, Model details, Training options, and Compute and pricing. The 'Model details' section is selected, showing fields for 'Model name' (iowa_daily_2021121002247), 'Target column' (y), 'Series identifier column' (id), and 'Timestamp column' (ds). Below these are 'Forecasting configuration' options: 'Data granularity' (Daily), 'Forecast horizon' (7), and 'Context window' (7). There is a checkbox for 'Export test dataset to BigQuery' and a 'Data validation options' section with radio buttons for 'Cancel training', 'Ignore validation', and 'Chronological'. The 'Chronological' option is selected.



Results

Google Cloud Platform CMPE260 Search products and resources

Vertex AI iowa_daily_2021121002247 VIEW DATASET

EVALUATE BATCH PREDICTIONS MODEL PROPERTIES

Target column	MAE	MAPE	RMSE	RMSLE	r ²
y numeric	235,397.72	∞	372,578.06	-	0.735

Google Cloud PlatformCMPE260Search products and resources

FEATURES & INFOSHORTCUTDISABLE EDITOR TABS

Explorer+ ADD DATA

Q Type to search

Viewing pinned projects.

cmpe260-334300

dataset_5716044293458624512_tables_2021_12...

export_evaluated_data_items_lowa_daily_202112...

evaluated_data_items

bigquery-public-data

EDITORX

EVALUA...X

COMPOSE NEW QUERY

evaluated_data_items

QUERYSHARECOPYSNAPSHOTDELETEDEXPORT

SCHEMADetailSPREVIEW

Table schema

Filter Enter property name or value

Field name	Type	Mode	Policy Tags	Description
ds	STRING	NULLABLE		
holiday	STRING	NULLABLE		
id	STRING	NULLABLE		
predicted_on_ds	STRING	NULLABLE		
y	STRING	NULLABLE		
predicted_y	RECORD	NULLABLE		

EDIT SCHEMAVIEW ROW ACCESS POLICIES

PERSONAL HISTORYPROJECT HISTORYSAVED QUERIES

Google Cloud PlatformCMPE260Search products and resources

FEATURES & INFOSHORTCUTDISABLE EDITOR TABS

Explorer+ ADD DATA

Q Type to search

Viewing pinned projects.

cmpe260-334300

dataset_5716044293458624512_tables_2021_12...

export_evaluated_data_items_lowa_daily_202112...

evaluated_data_items

bigquery-public-data

EDITORX

EVALUA...X

COMPOSE NEW QUERY

evaluated_data_items

QUERYSHARECOPYSNAPSHOTDELETEDEXPORT

SCHEMADetailSPREVIEW

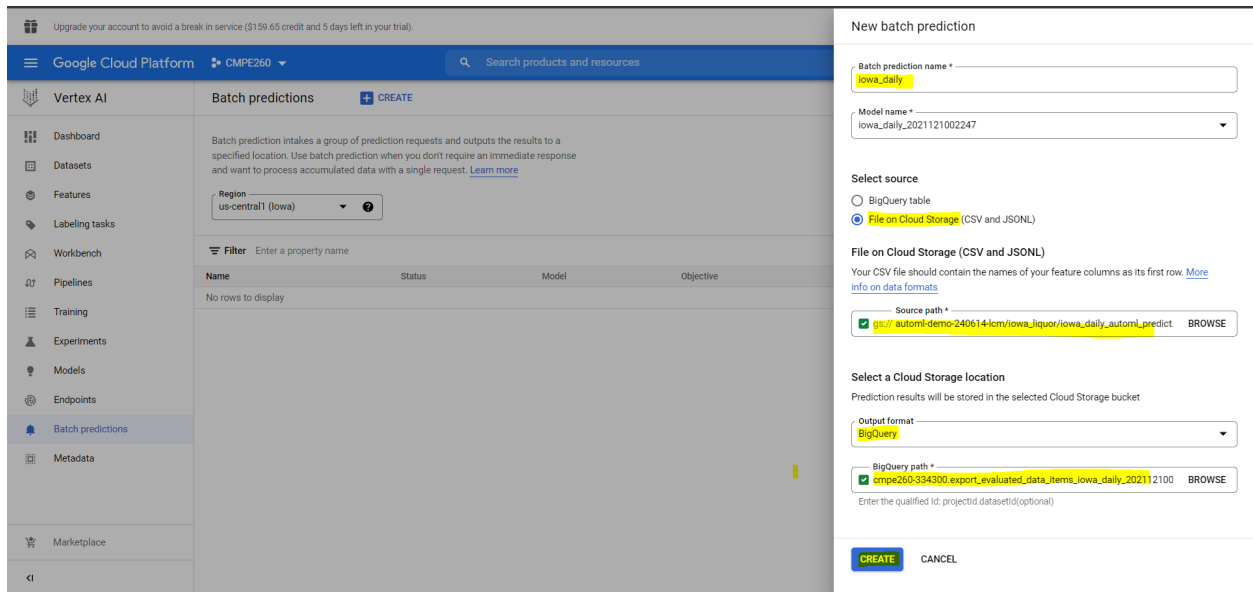
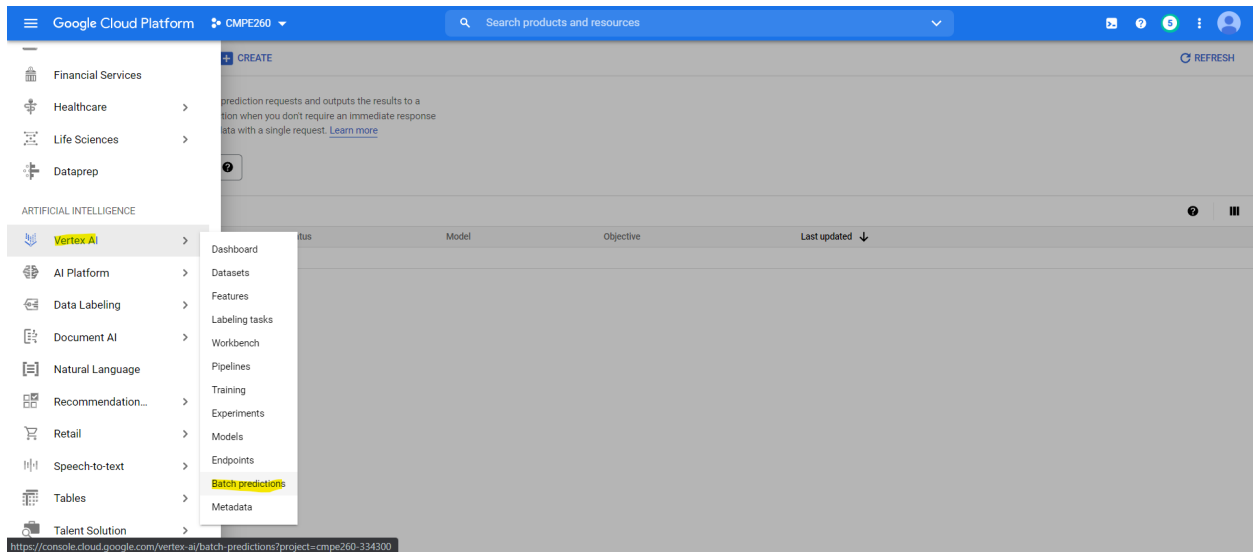
Rowdsdholidayidpredicted_on_dsypredicted_y_value

1	1/4/20	0	0	1/1/20	0	26.002988815307617
2	1/4/20	0	0	12/30/19	0	9.3418703079223633
3	1/4/20	0	0	1/2/20	0	3.0197560787200928
4	1/4/20	0	0	1/3/20	0	0.13746261596679688
5	1/4/20	0	0	1/4/20	0	18.519977569580078
6	1/4/20	0	0	12/31/19	0	5.1635699272155762
7	1/4/20	0	0	12/29/19	0	12.094734191894531
8	1/5/20	0	0	1/1/20	0	14.930747032165527
9	1/5/20	0	0	1/5/20	0	151.37579345703125
10	1/5/20	0	0	12/30/19	0	12.4290132522583
11	1/5/20	0	0	1/2/20	0	14.205635070800781
12	1/5/20	0	0	1/3/20	0	17.112651824951172
13	1/5/20	0	0	1/4/20	0	14.523508071899414
14	1/5/20	0	0	12/31/19	0	9.10609047666010

Rows per page: 1001 - 100 of 2128First page14<>>Last page

PERSONAL HISTORYPROJECT HISTORYSAVED QUERIES

Batch Predictions



Google Cloud PlatformCMPE260

Search products and resources

Vertex AI

Batch predictions

CREATE

REFRESH

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

Batch prediction intakes a group of prediction requests and outputs the results to a specified location. Use batch prediction when you don't require an immediate response and want to process accumulated data with a single request. [Learn more](#)

Region

us-central1 (Iowa)

Filter Enter a property name

Name	Status	Model	Objective	Last updated
iowa_daily	Running	iowa_daily_2021121002247	Tabular	December 9, 2021

Google Cloud PlatformCMPE260

Search products and resources

Vertex AI

← iowa_daily

Dashboard

Datasets

Features

Labeling tasks

Workbench

Pipelines

Training

Experiments

Models

Endpoints

Batch predictions

Metadata

ID5769903870545231872

Modeliowa_daily_2021121002247

ObjectiveTabular

Import locationgs://automi-demo-240614-lcm/iowa_liquor/iowa_daily_automi_predict.csv

CreatedDec 09, 2021 at 05:47PM

UpdatedDec 09, 2021 at 05:52PM

Elapsed time5 min 13 sec

StatusFinished

Export locationbq://cmpe260-334300.export_evaluated_data_items_iowa_daily_2021121002247_2021_12_09T16_29_57_962Z

Google Cloud PlatformCMPE260

Search products and resources

FEATURES & INFO

SHORTCUT

DISABLE EDITOR TABS

Explorer

+ ADD DATA

Type to search

Viewing pinned projects.

cmpe260-334300

dataset_57160442934586...

export_evaluated_data_ite...

errors_2021_12_09T17...

evaluated_data_items

predictions_2021_12_0...

bigquery-public-data

EDITOR

EXPORT...

PREDICT...

COMPOSE NEW QUERY

predictions_2021_12_09T17_47_33_624Z

QUERY

SHARE

COPY

SNAPSHOT

DELETE

EXPORT

SCHEMA

DETAILS

PREVIEW

Row	ds	holiday	id	y	predicted_y value
1	2020-05-23 00:00:00 UTC	0	0	null	4.7916741371154785
2	2020-05-24 00:00:00 UTC	0	0	null	9.0024642944335938
3	2020-05-26 00:00:00 UTC	0	0	null	1544561.75
4	2020-05-27 00:00:00 UTC	0	0	null	1652441.875
5	2020-05-28 00:00:00 UTC	0	0	null	1459899.25
6	2020-05-29 00:00:00 UTC	0	0	null	1481845.625
7	2020-05-25 00:00:00 UTC	1	0	null	142.09532165527344

PERSONAL HISTORY

PROJECT HISTORY

SAVED QUERIES