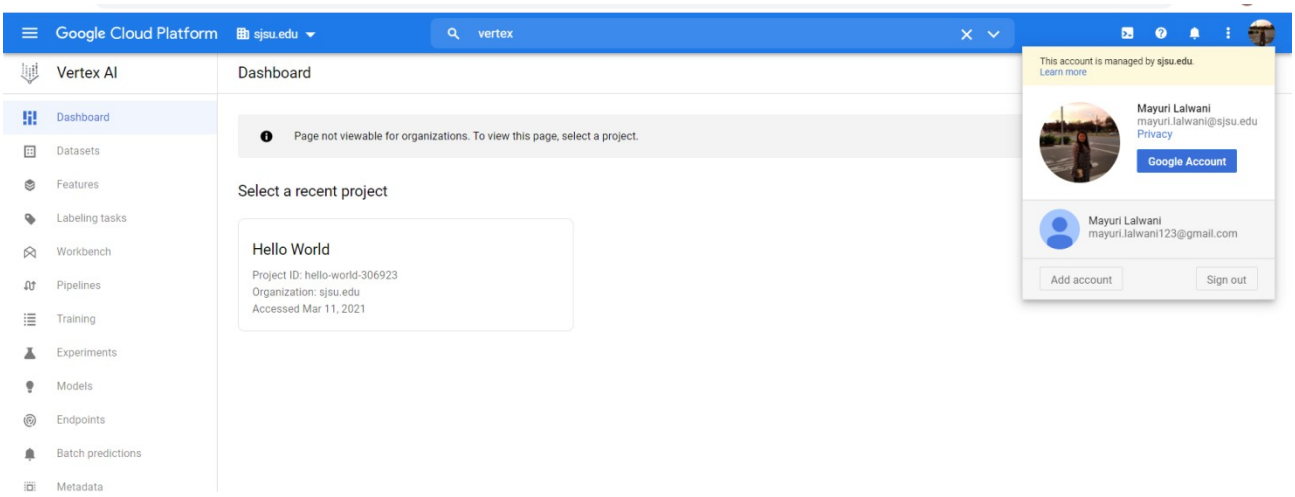
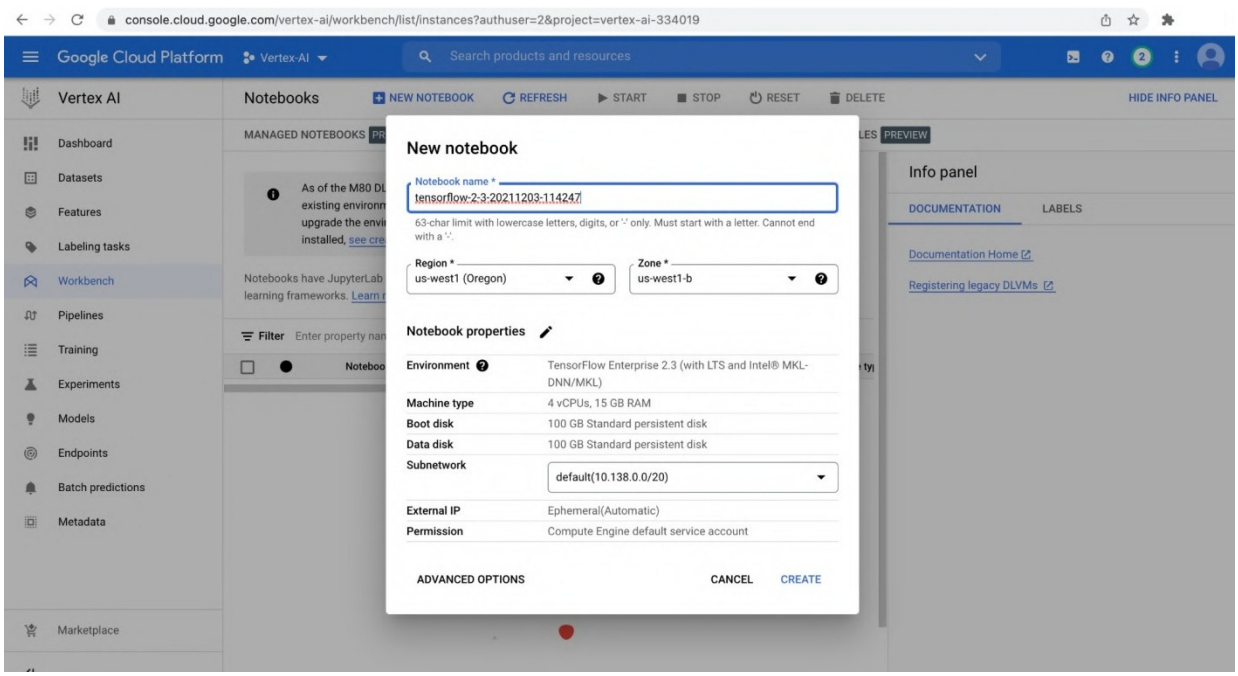
Name - Mayuri Lalwani

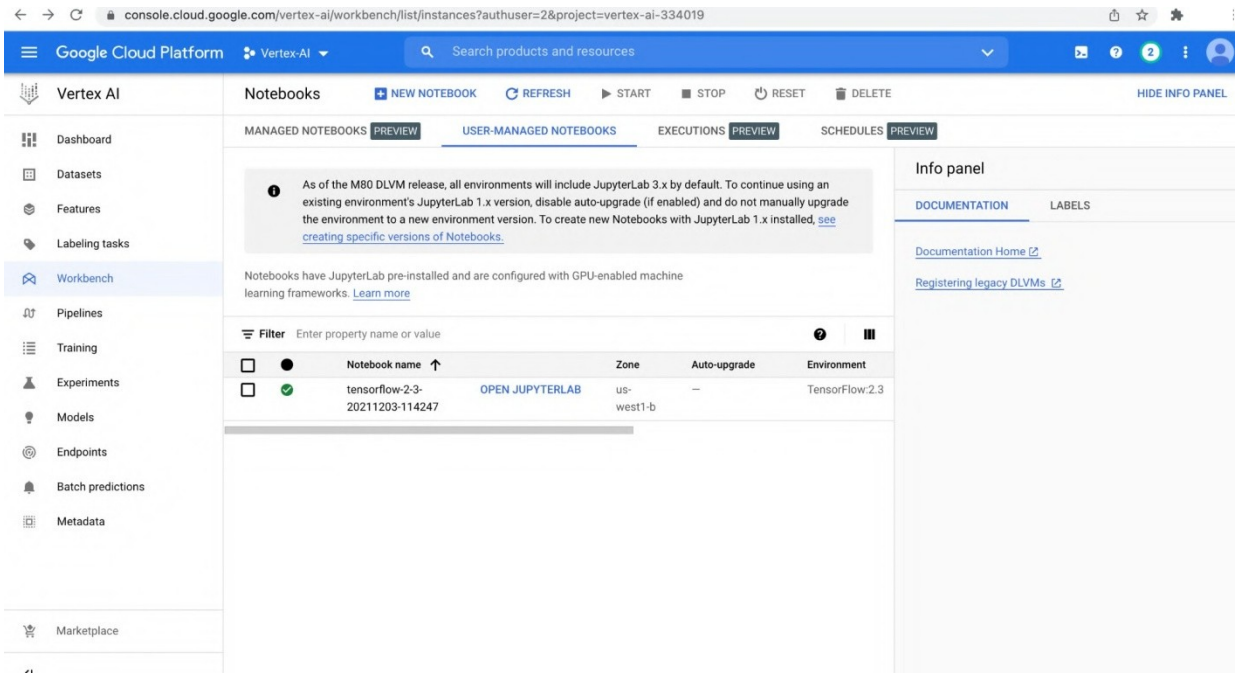
Assignment 7 - VERTEX AI

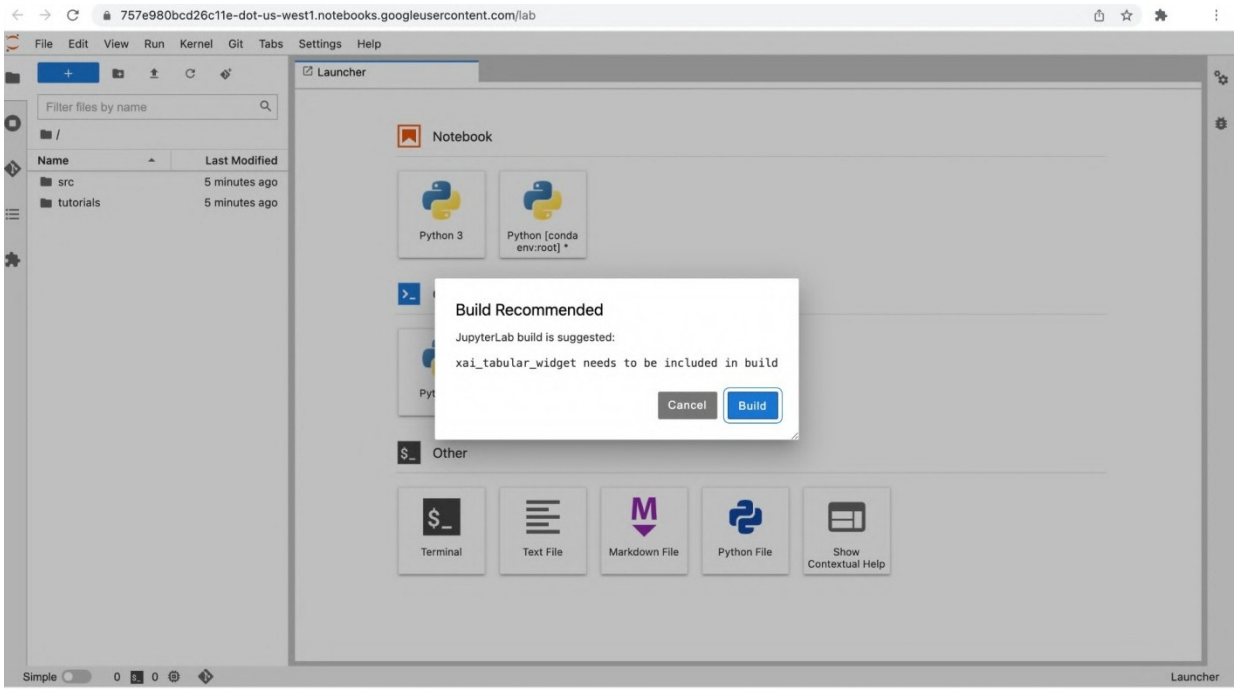
a)

https://codelabs.developers.google.com/vertex-pipelines-intro#0



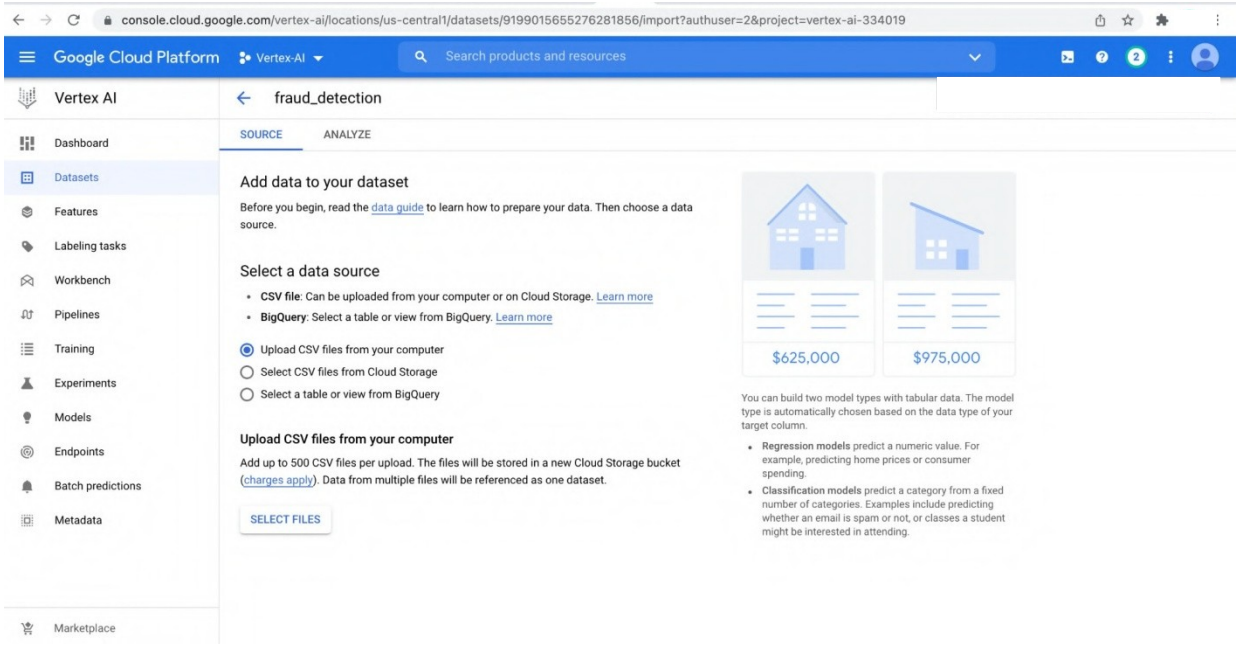


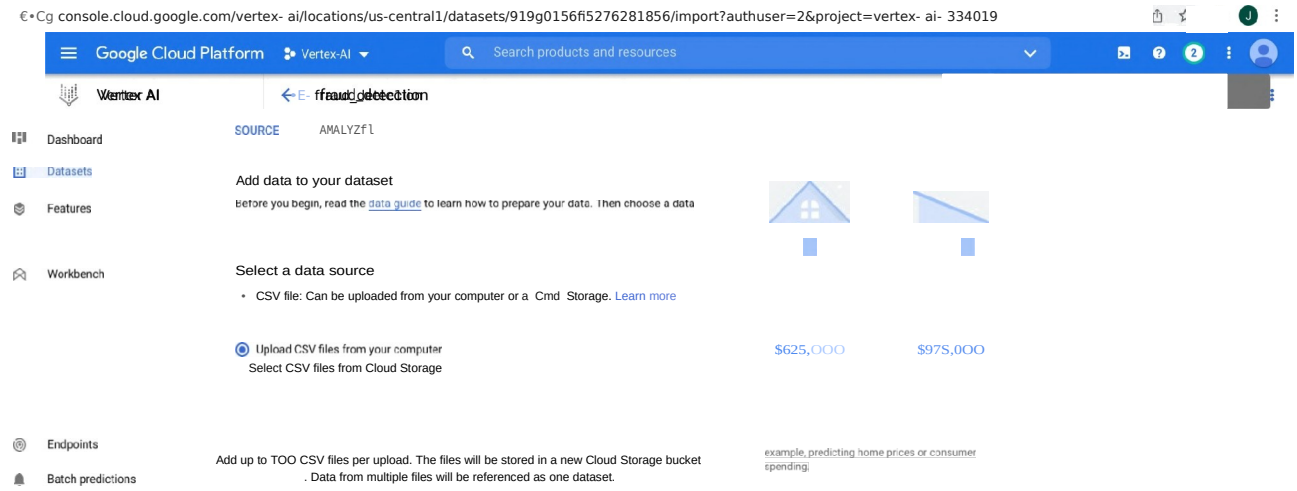


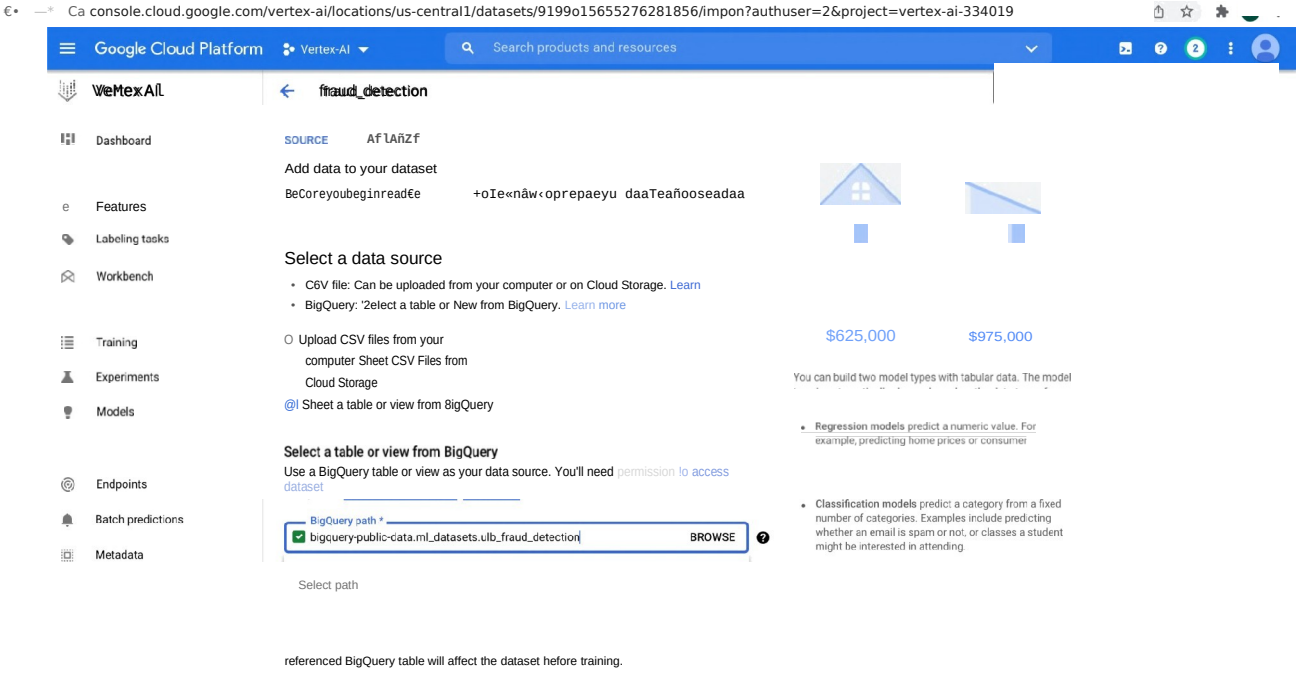


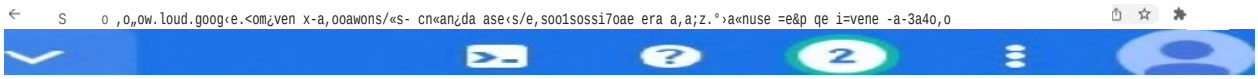
b)

https://codelabs.developers.google.com/vertex-automl-tabular#0


console.cloud.google.com/vertex-ai/locations/us-central1/datasets/9199015655276281856/import?authuser=2&project=vertex-ai-334019
Google Cloud Platform
Vertex-AI
Search products and resources
Vertex AI
fraud_detection
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
SOURCE
ANALYZE
Add data to your dataset
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
Upload CSV files from your computer
Add up to 500 CSV files per upload. The files will be stored in a new Cloud Storage bucket (charges apply). Data from multiple files will be referenced as one dataset.
SELECT FILES
$625,000
$975,000
You can build two model types with tabular data. The model type is automatically chosen based on the data type of your target column.
Regression models predict a numeric value. For example, predicting home prices or consumer spending.
Classification models predict a category from a fixed number of categories. Examples include predicting whether an email is spam or not, or classes a student might be interested in attending.


€•Cg console.cloud.google.com/vertex- ai/locations/us-central1/datasets/919g0156fi5276281856/import?authuser=2&project=vertex- ai- 334019
Google Cloud Platform
Vertex-AI
Search products and resources
Vertex AI
fraud_detection
Dashboard
Datasets
Features
Workbench
Endpoints
Batch predictions
SOURCE
AMALYZfl
Add data to your dataset
Before you begin, read the data guide to learn how to prepare your data. Then choose a data
Select a data source
• CSV file: Can be uploaded from your computer or a Cmd Storage. Learn more
Upload CSV files from your computer
Select CSV files from Cloud Storage
$625,OOO
$97S,0OO
Add up to TOO CSV files per upload. The files will be stored in a new Cloud Storage bucket
. Data from multiple files will be referenced as one dataset.
example, predicting home prices or consumer spending.


€• —* Ca console.cloud.google.com/vertex-ai/locations/us-central1/datasets/9199o15655276281856/impon?authuser=2&project=vertex-ai-334019
Google Cloud Platform
Vertex-AI
Search products and resources
Vertex AI
fraud_detection
Dashboard
Features
Labeling tasks
Workbench
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
SOURCE
AflAñZf
Add data to your dataset
BeCoreyoubeginread€e
+oIe«nãw‹oprepaeyu daaTeañooseadaa
Select a data source
• C6V file: Can be uploaded from your computer or on Cloud Storage. Learn
• BigQuery: '2elect a table or New from BigQuery. Learn more
O Upload CSV files from your computer Sheet CSV Files from Cloud Storage
@I Sheet a table or view from 8igQuery
Select a table or view from BigQuery
Use a BigQuery table or view as your data source. You'll need permission lo access dataset
BigQuery path *
bigquery-public-cata.ml_datasets.ulb_fraud_detection
BROWSE
Select path
referenced BigQuery table will affect the dataset hefore training.
$625,000
$975,000
You can build two model types with tabular data. The model
Regression models predict a numeric value. For example, predicting home prices or consumer
Classification models predict a category from a fixed number of categories. Examples include predicting whether an email is spam or not, or classes a student might be interested in attending.



Vertex AI fraud_detection

SOURCE A NALYZE

Features

Dataset Info

Created Oec 03, 20?1 11 II All

Dataset format: BigQuery

Summary

Training jobs and models

TRAIN NEW MODEL

GENERATE STATISTICS

Batch predictions

Column name	BigQuery type	BigQuery mode	Missing a f<<t> e	Disqmm values @
Amount	FLOAT	NULLAB.E		
Class	irJTEñER	hT ULLAd.E		
Time	FUA I	NULLAd .L		
VI	F LOA I	hT		
ULLAB.L VI 0	FLOAT	hlULLAB.E		
VII	FLOAT	NULLAB.E		

O C B console.cloud.google.com}vertex-a +/l ocations/ us-central 4 /datasets}919904 SO 55 27G 284 8 56t ana ! yze*a u th user=2&project=vertez- at -8 34019

Train new model

fraud_detection_202112319 s457

4 Compute and pricing

Export test dataset to BigQuery

START TRAINING

W ADVANCED OPTIONS

CONTINUE

g console.cloud.google.com/vertex-ai/locations/us-central1/datasets/9199015655276 2B1856/analyze"authuser= 2&project=vertex-ai- 334019

Train new model

@ Training method

@ Model details

@ Training options

@ Compute and pricing

0AflMb

Enter the maximum number of node hours you wan1 to spend training your model.

You can train for as little as node hour. You may also be eligible to train with free node

Budget *

1 Maximum node hours

Estimated completion date: Dec 3, 2021 1 PM GMT-8

Enable early stopping

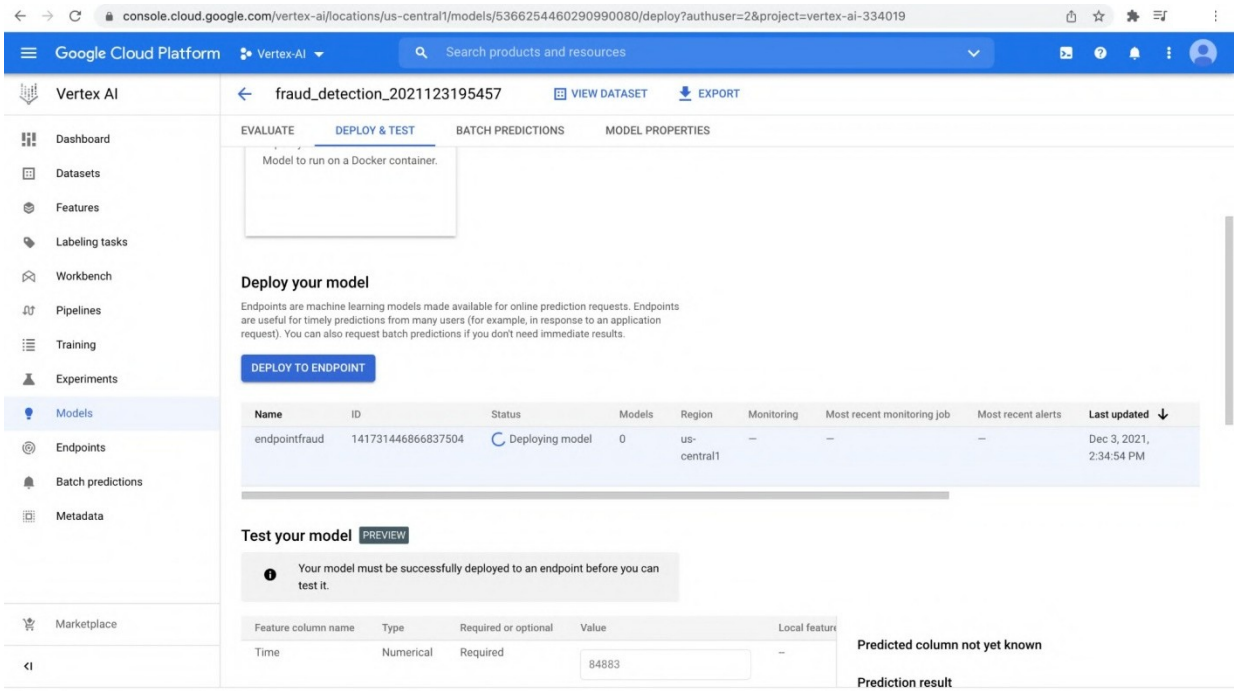
Ends model training when no more improvements can be made and refunds leftover training budget. If early stopping is disabled, training continues until the budget is

Workbench

Pipelines

Training

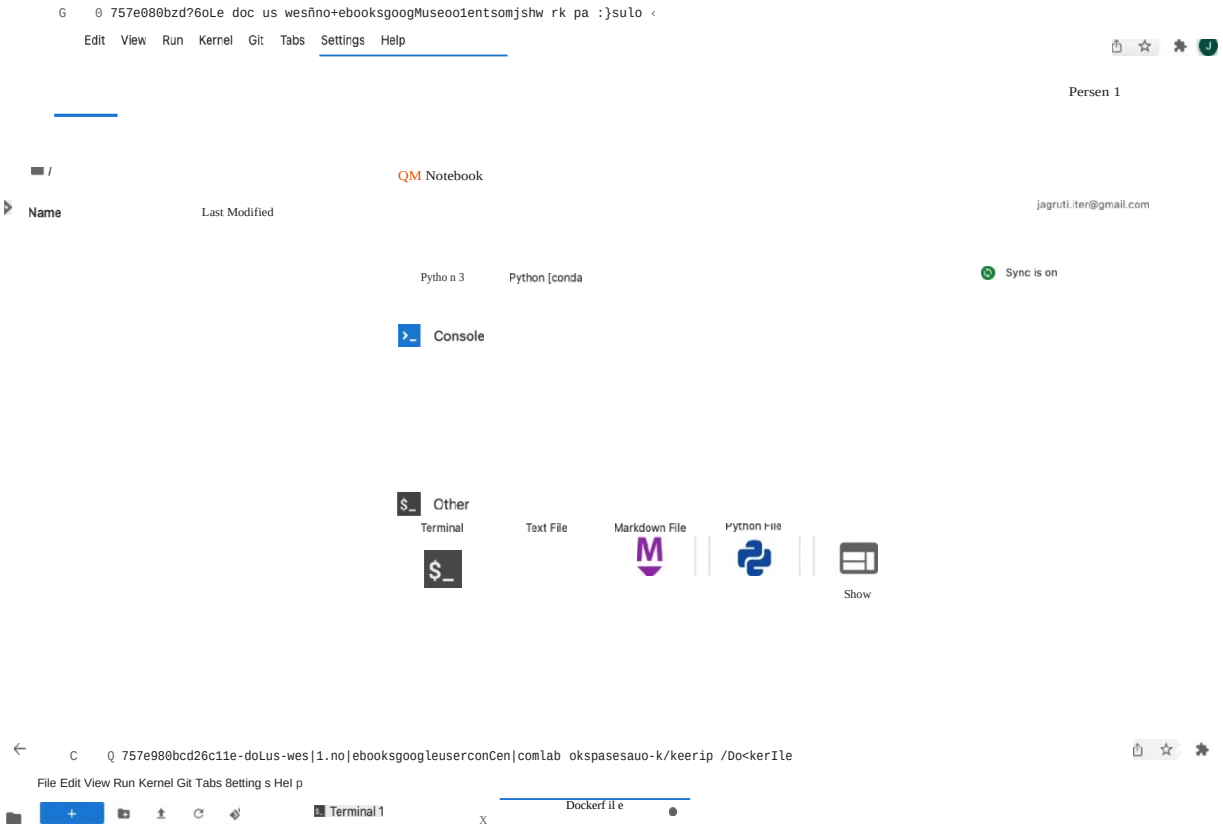
Experiments

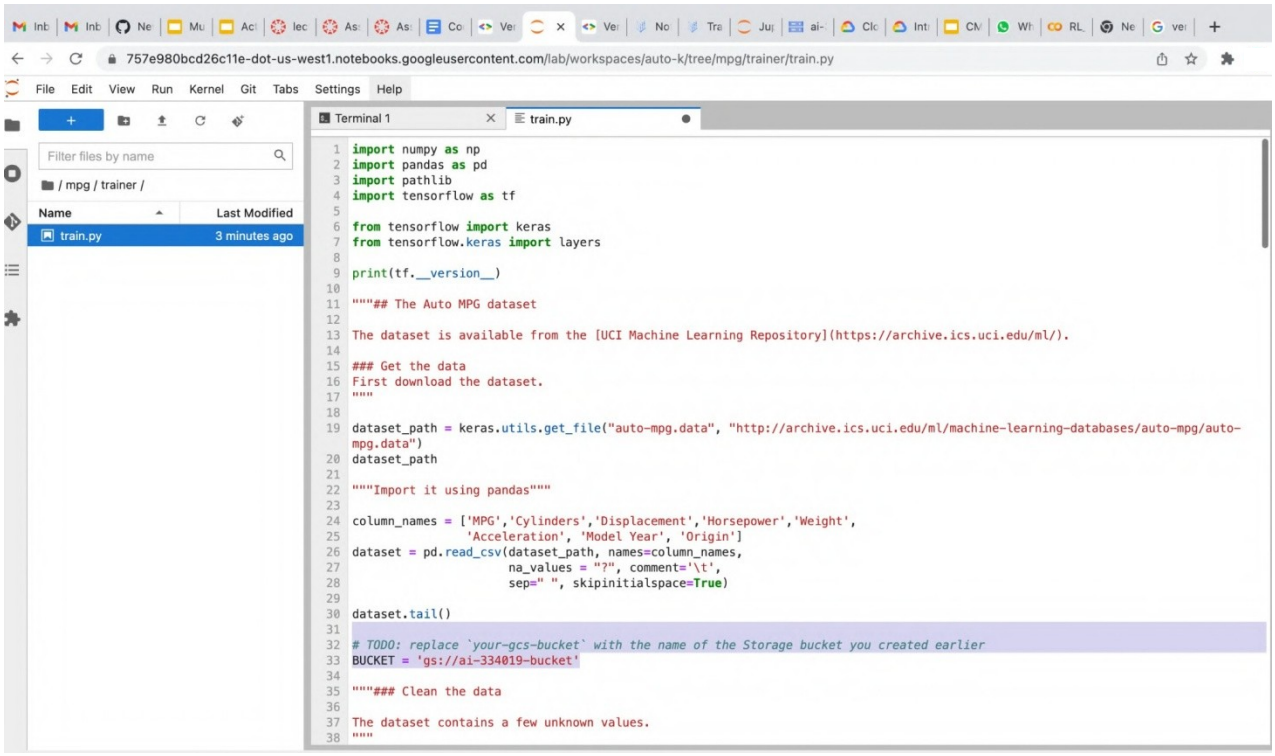


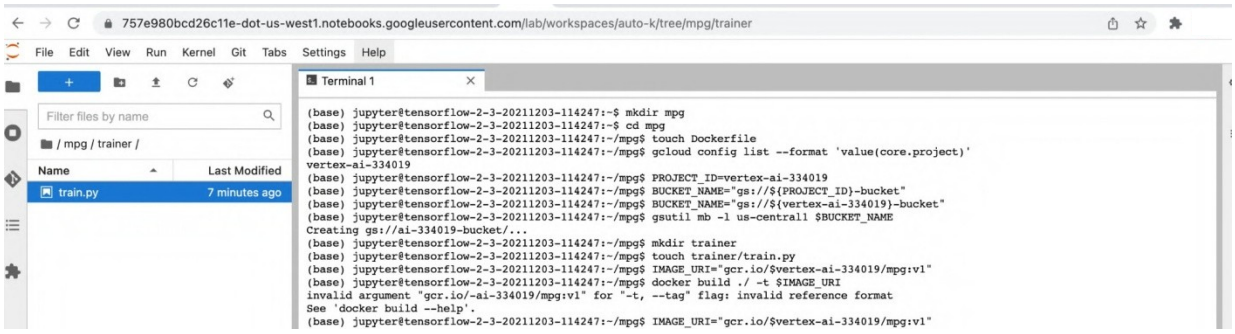
c)

https://codelabs.developers.google.com/vertex_custom_training_prediction#0

(Links to an external site.)

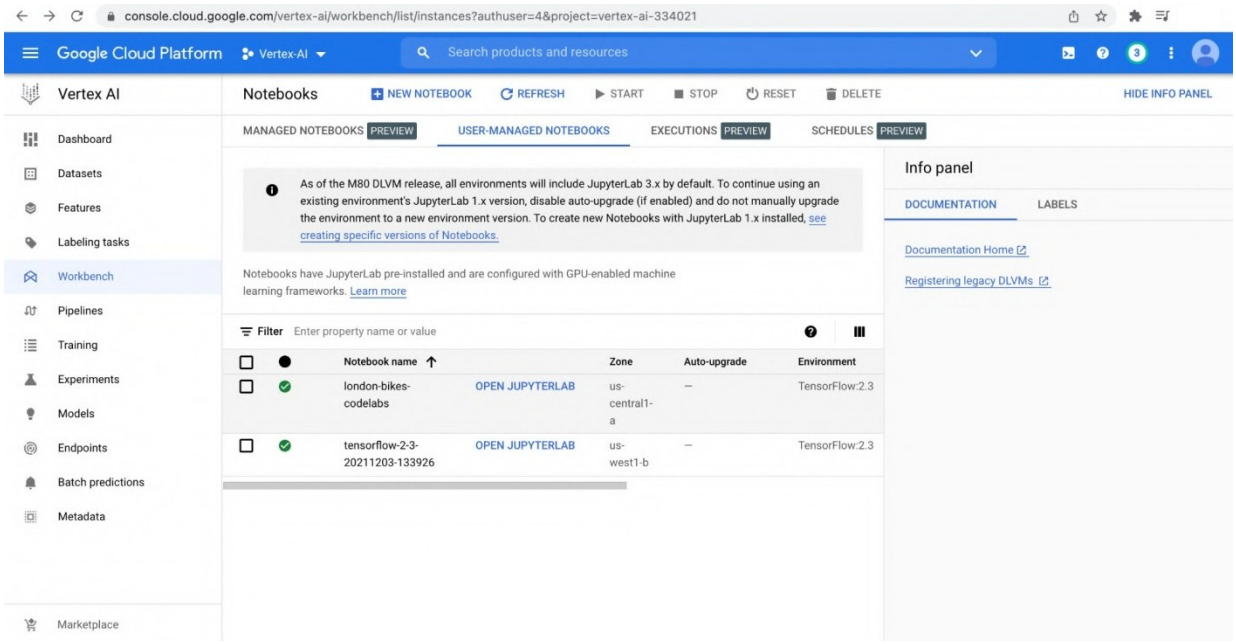


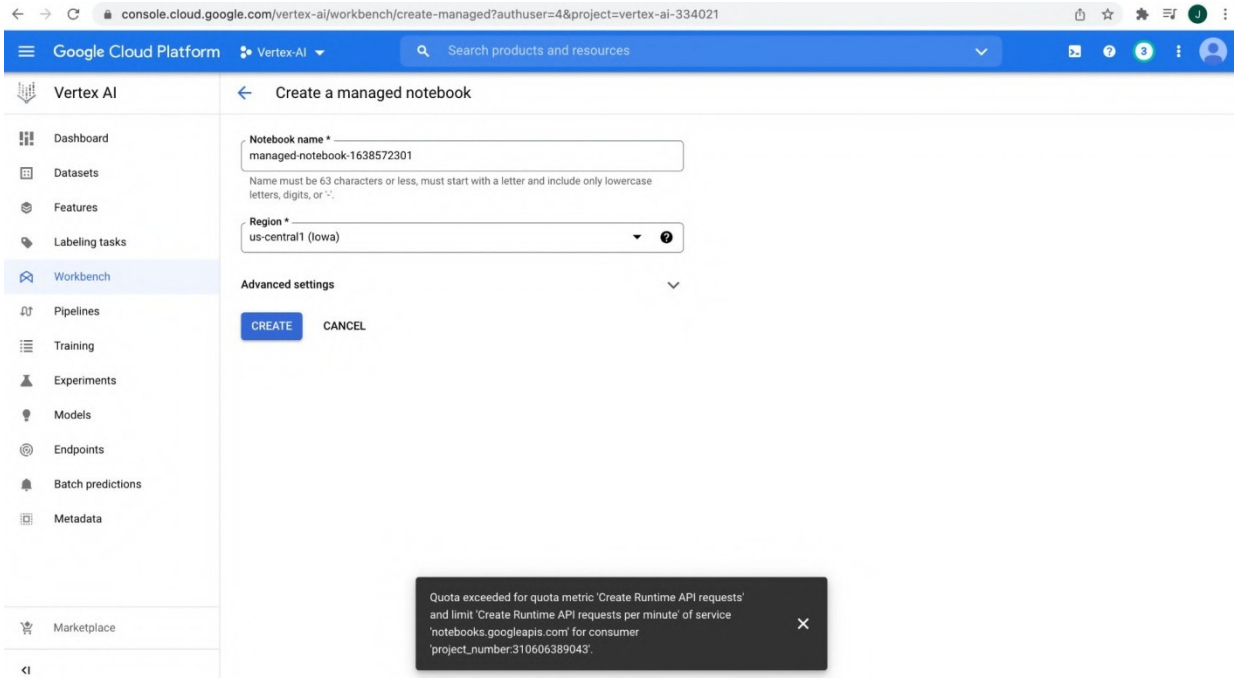




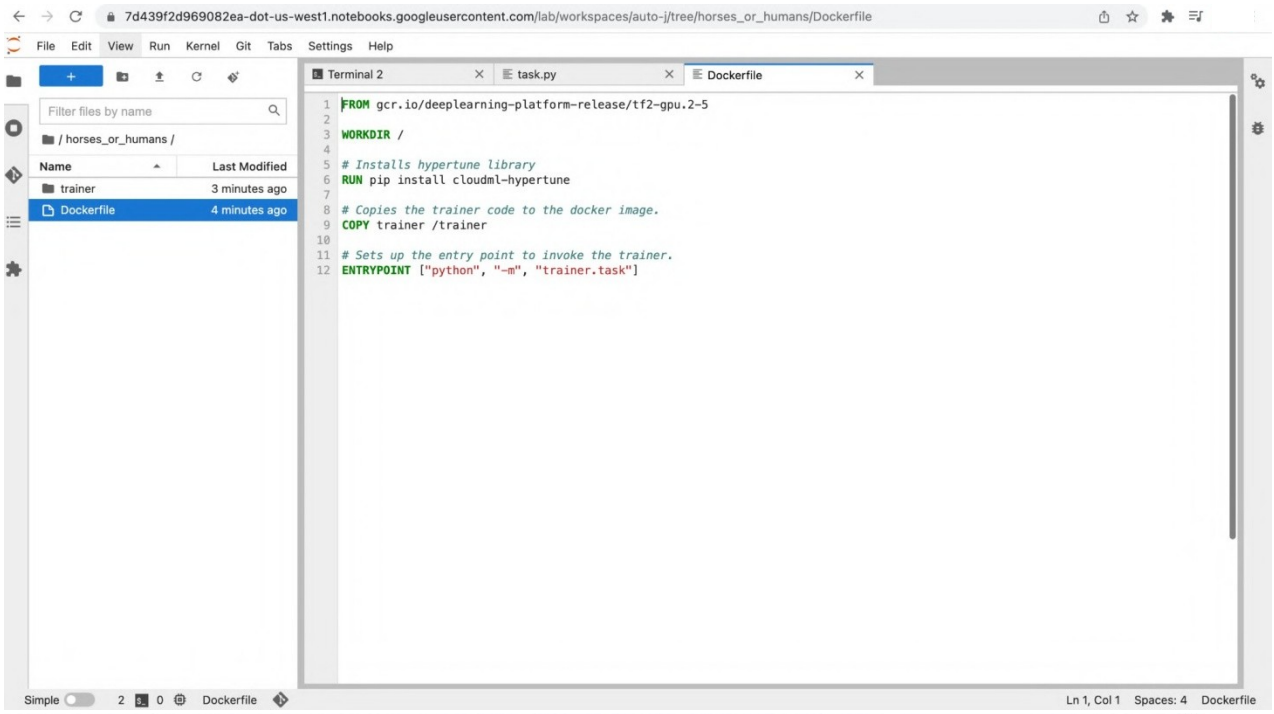
d) Big query :

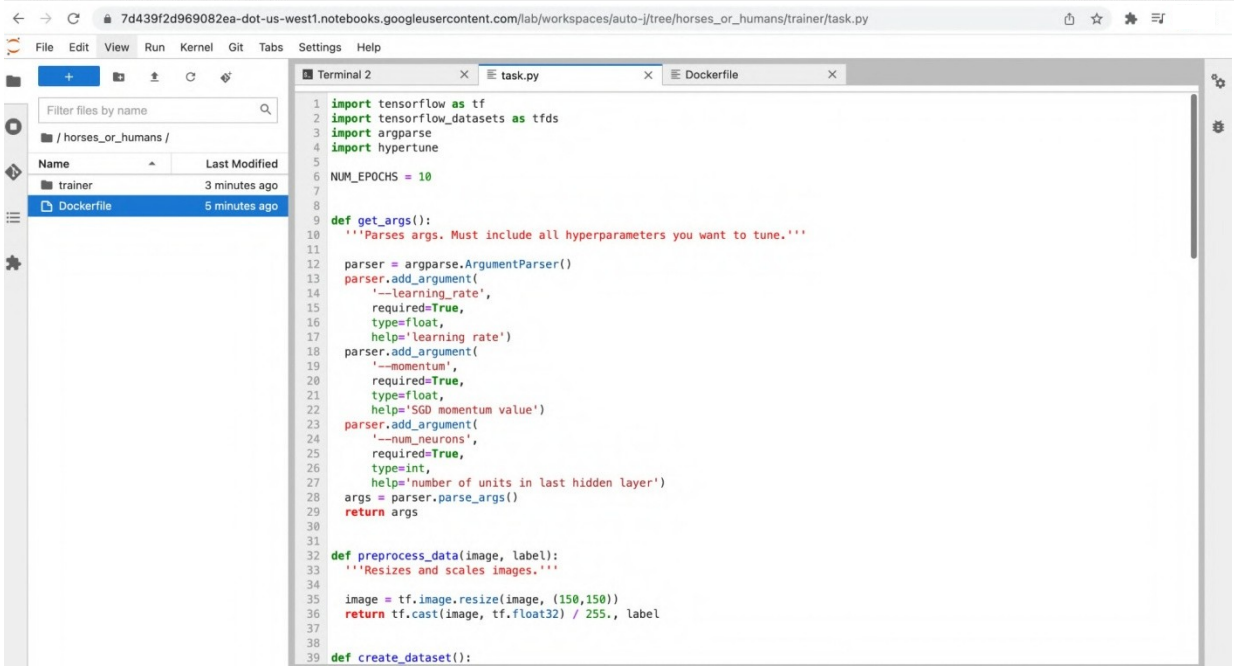
https://codelabs.developers.google.com/vertex-workbench-intro#0

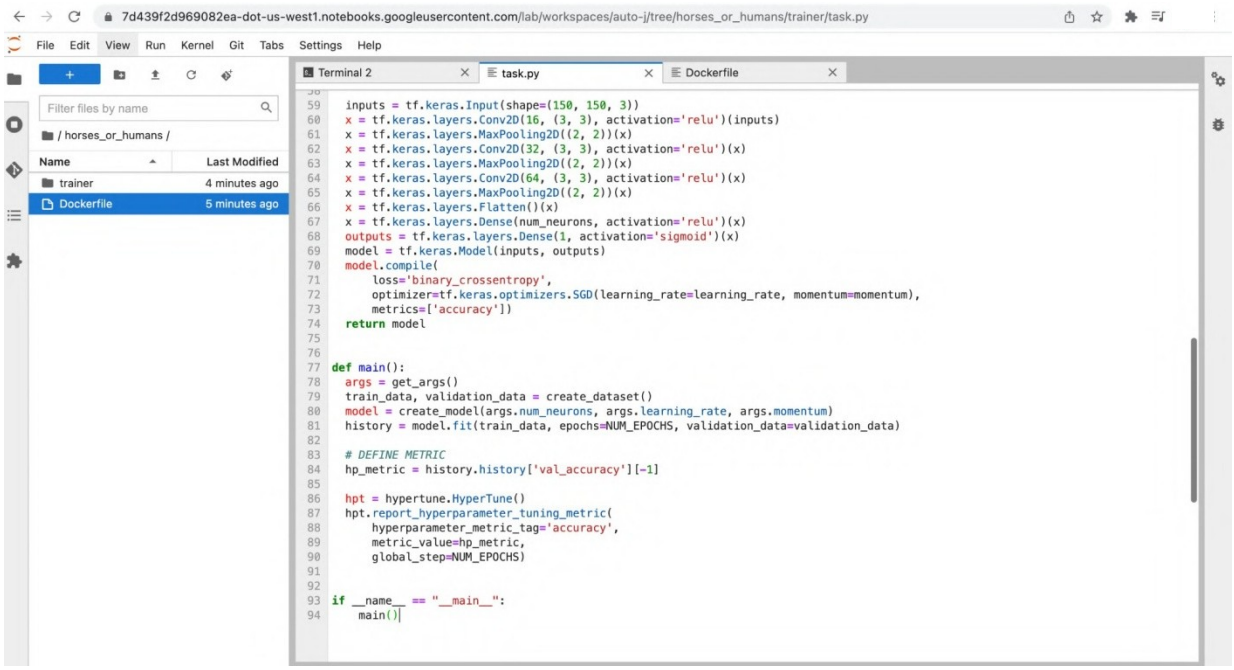


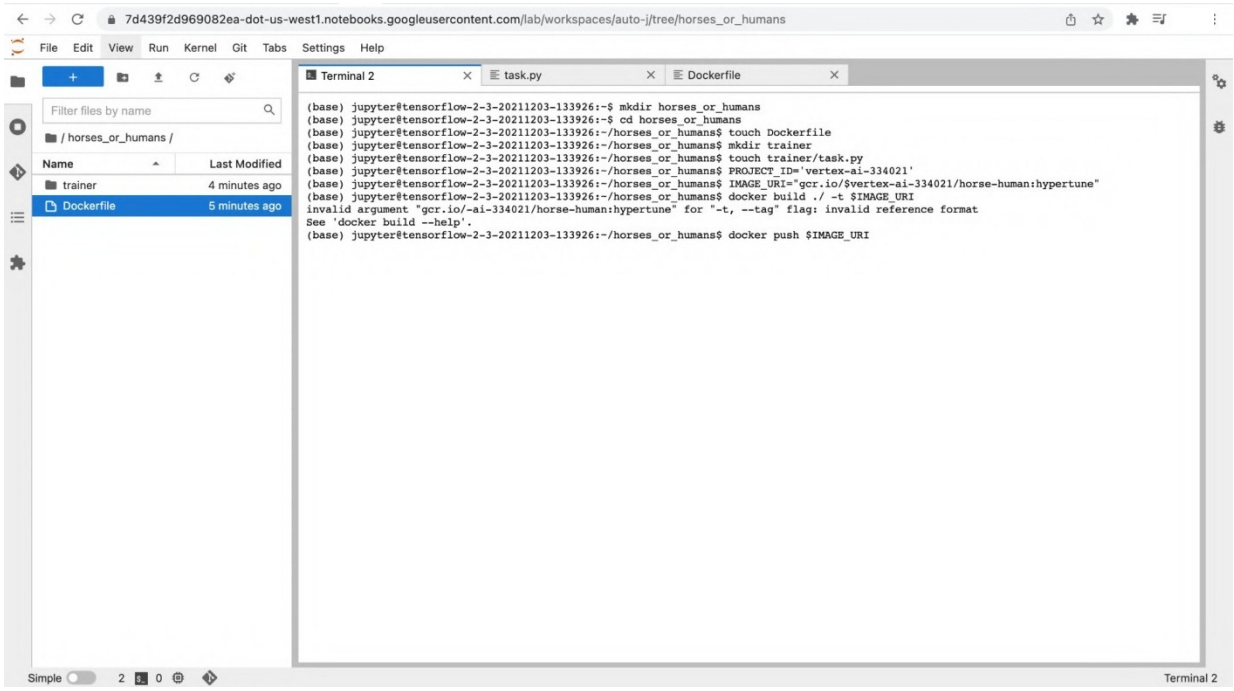


f) https://codelabs.developers.google.com/vertex_hyperparameter_tuning#0

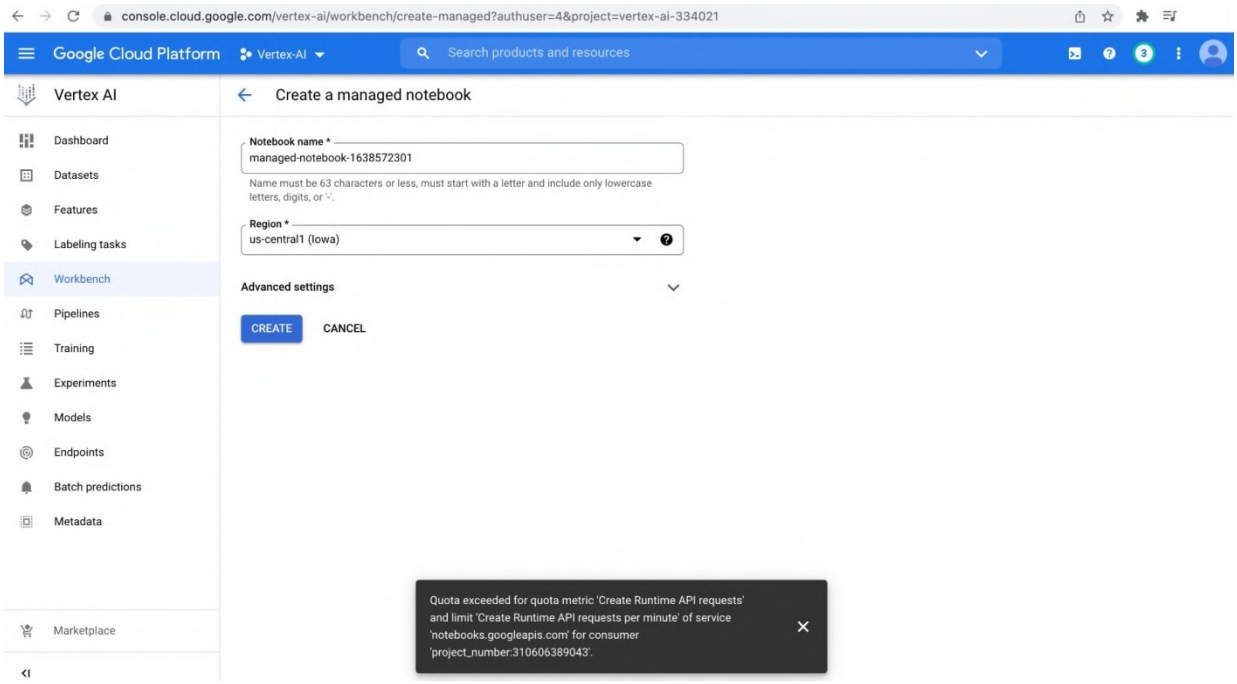

7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/workspaces/auto-j/tree/horses_or_humans/Dockerfile
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ horses_or_humans /
Name
Last Modified
trainer 3 minutes ago
Dockerfile 4 minutes ago
Terminal 2
task.py
Dockerfile
1 FROM gcr.io/deeplearning-platform-release/tf2-gpu.2-5
2
3 WORKDIR /
4
5 # Installs hypertune library
6 RUN pip install cloudml-hypertune
7
8 # Copies the trainer code to the docker image.
9 COPY trainer /trainer
10
11 # Sets up the entry point to invoke the trainer.
12 ENTRYPOINT ["python", "-m", "trainer.task"]
Simple 2 0 Dockerfile
Ln 1, Col 1 Spaces: 4 Dockerfile

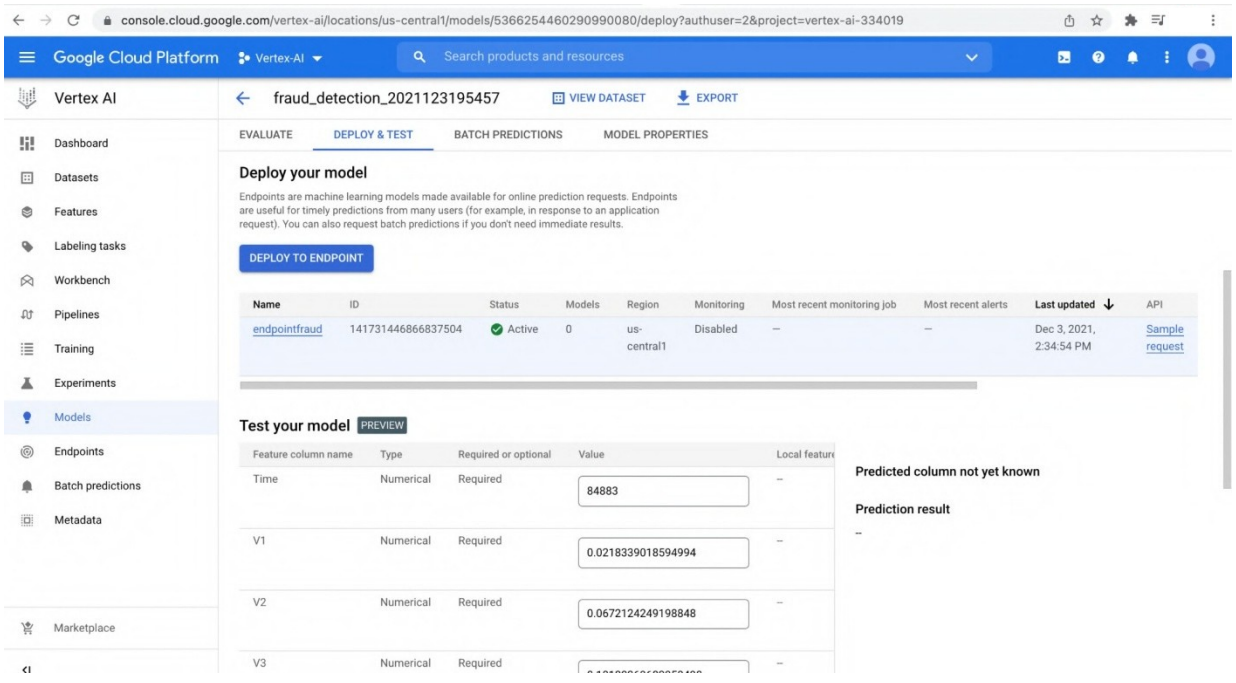

7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/workspaces/auto-j/tree/horses_or_humans/trainer/task.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ horses_or_humans /
Name
Last Modified
trainer 3 minutes ago
Dockerfile 5 minutes ago
Terminal 2
task.py
Dockerfile
1 import tensorflow as tf
2 import tensorflow_datasets as tfds
3 import argparse
4 import hypertune
5
6 NUM_EPOCHS = 10
7
8
9 def get_args():
10 '''Parses args. Must include all hyperparameters you want to tune.'''
11
12 parser = argparse.ArgumentParser()
13 parser.add_argument(
14 '--learning_rate',
15 required=True,
16 type=float,
17 help='learning rate')
18 parser.add_argument(
19 '--momentum',
20 required=True,
21 type=float,
22 help='SGD momentum value')
23 parser.add_argument(
24 '--num_neurons',
25 required=True,
26 type=int,
27 help='number of units in last hidden layer')
28 args = parser.parse_args()
29 return args
30
31
32 def preprocess_data(image, label):
33 '''Resizes and scales images.'''
34
35 image = tf.image.resize(image, (150,150))
36 return tf.cast(image, tf.float32) / 255., label
37
38
39 def create_dataset():





g) https://codelabs.developers.google.com/vertex-mlmd-pipelines#0





h) https://codelabs.developers.google.com/codelabs/automl-forecasting-with-vertex-ai#0

7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb

File Edit View Run Kernel Git Tabs Settings Help

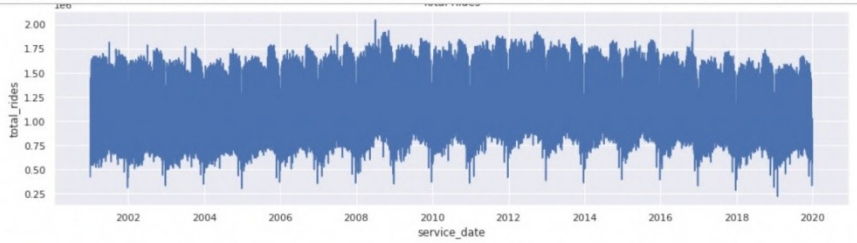
Filter files by name

/ ··· / ai-for-time-series / notebooks /

Name	Last Modified
data	11 minutes ago
01-explore.ipynb	3 minutes ago
02-model.ipynb	11 minutes ago
03-cloud-training.i...	11 minutes ago
cta_ridership.csv	4 minutes ago

Terminal 1 | 01-explore.ipynb

Markdown | git | Python 3



[15]:
```
# Explore rides by day type: Weekday (W), Saturday (A), Sunday/Holiday (U)

sns.lineplot(data=df, x=df.index, y=df[target], hue=df['day_type']).set_title('Total Rides by Day Type')
fig = plt.show()
```

Simple | 1 | 1 | Python 3 | Idle | Mode: Command | Ln 1, Col 1 | 01-explore.ipynb

7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/ ··· / ai-for-time-series / notebooks /

Name	Last Modified
data	11 minutes ago
01-explore.ipynb	3 minutes ago
02-model.ipynb	11 minutes ago
03-cloud-training.i...	11 minutes ago
cta_ridership.csv	4 minutes ago

Terminal 1 | 01-explore.ipynb

Markdown | git | Python 3

[16]:
```
# Explore rides by transportation type

sns.lineplot(data=df[['bus','rail_boardings']]).set_title('Total Rides by Transportation Type')
fig = plt.show()
```

TODO 2: Review summary statistics

- How many records are in the dataset?
- What is the average # of riders per day?

[17]:
```
df[target].describe().apply(lambda x: round(x))
```

[17]:
```
count      6939
mean    1368761
std      391443
min      222071
25%     1005394
50%     1548343
```

G é 7d4fl9I?d96908?ea-donus-west.no+ebooksgoogIeuserconientcom/abjr e,'waiing-dna-anas ouwu \idor-liu«ri 'uo+ book\O4-Score.p ab

fe Edit View Ru Kerel Gis labs Ss4ings Help

RTermia4 X ñJbe,Qoe.ipab X

Markdown Python 3

Name - Last Modified

02- mod .ºL +p ynb 11 minutes ago

cta_ridership.csv 5 minutes ago

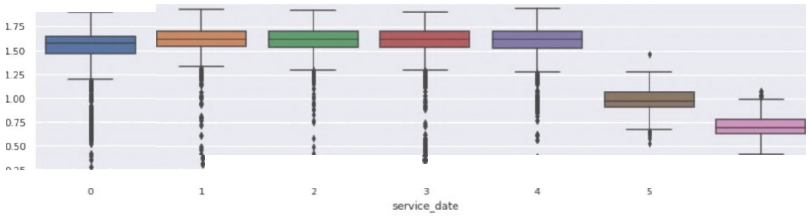
min 222071

75% 1660947

TODO 3: Explore seasonality

- Is there much difference between months*
- can you extract the trend and seasonal pattern from the datas

d ay s o free k = d I . ndex . to se r ze s ( ) . d I . da y ofree k



ntent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb

File Edit View Run Kernel G it Tabs Se ttings Help

Terminal 1

01-explore.ipynb

'e

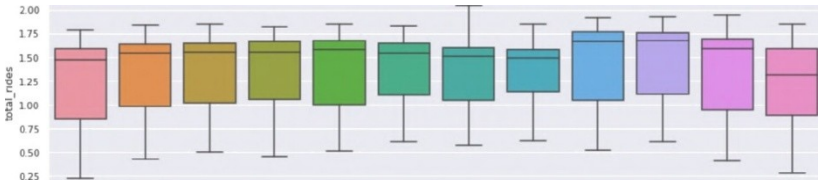
«meseries}xébooñ,'

Name Last Modified

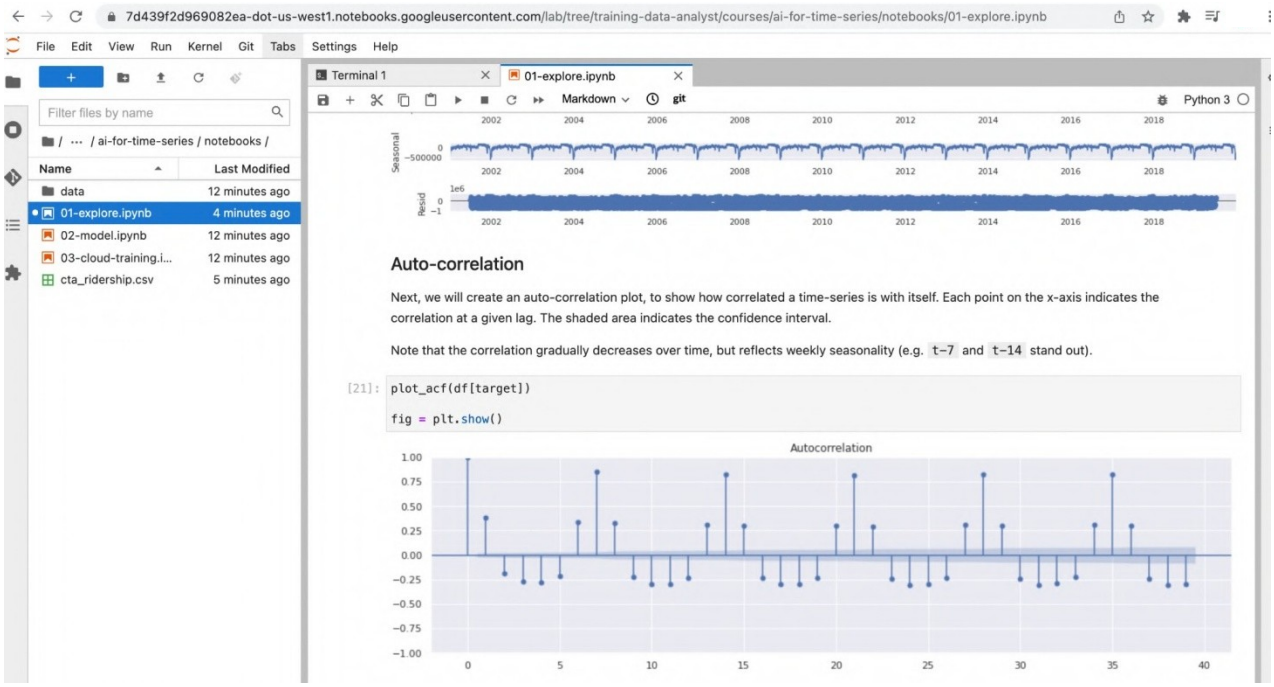
02-model.i ab 12 minutes ago

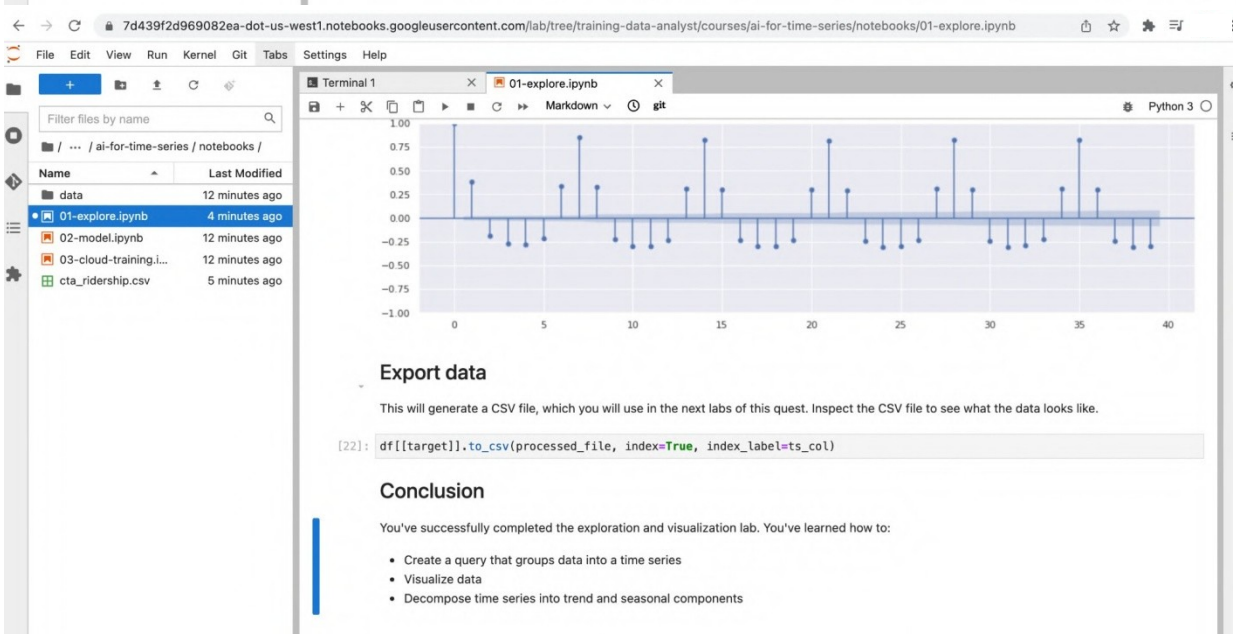
FB cta ridership.csv 5 minutes ago

```
 fo fe %s:ridnios of vales far eat svrh » s bo piñ:
moths = d .nde to_series).d:.nook
```

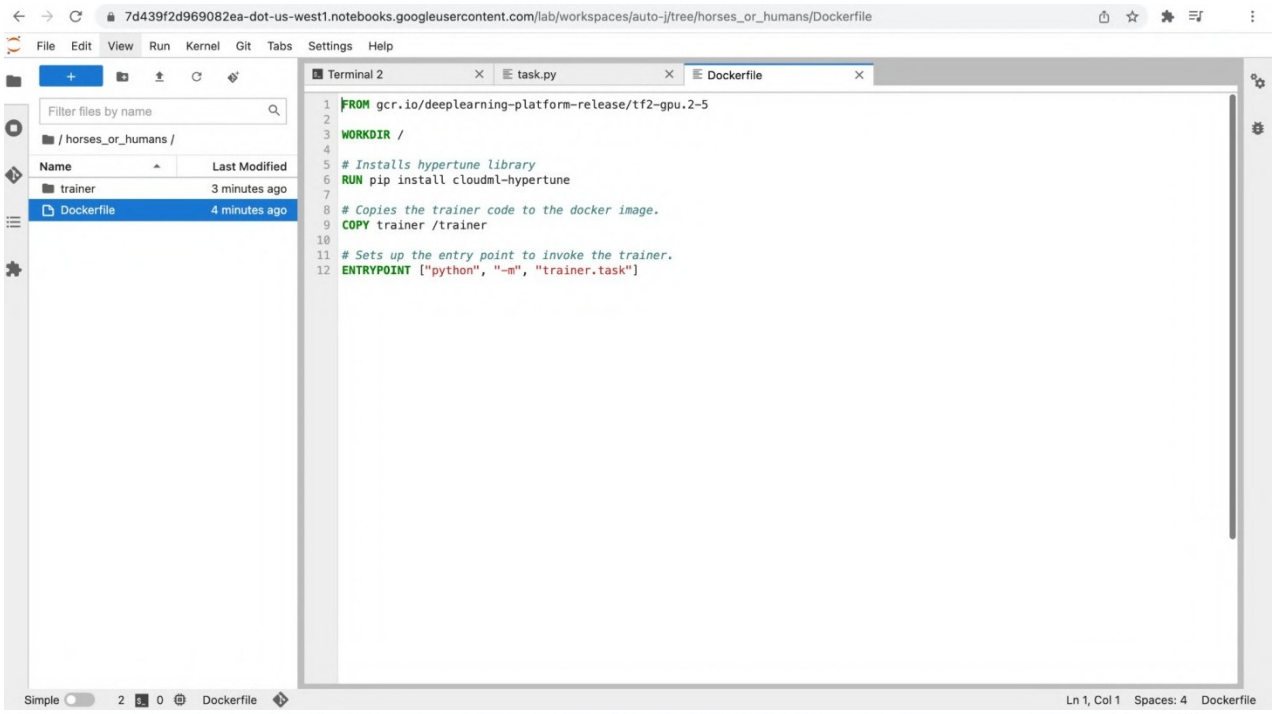


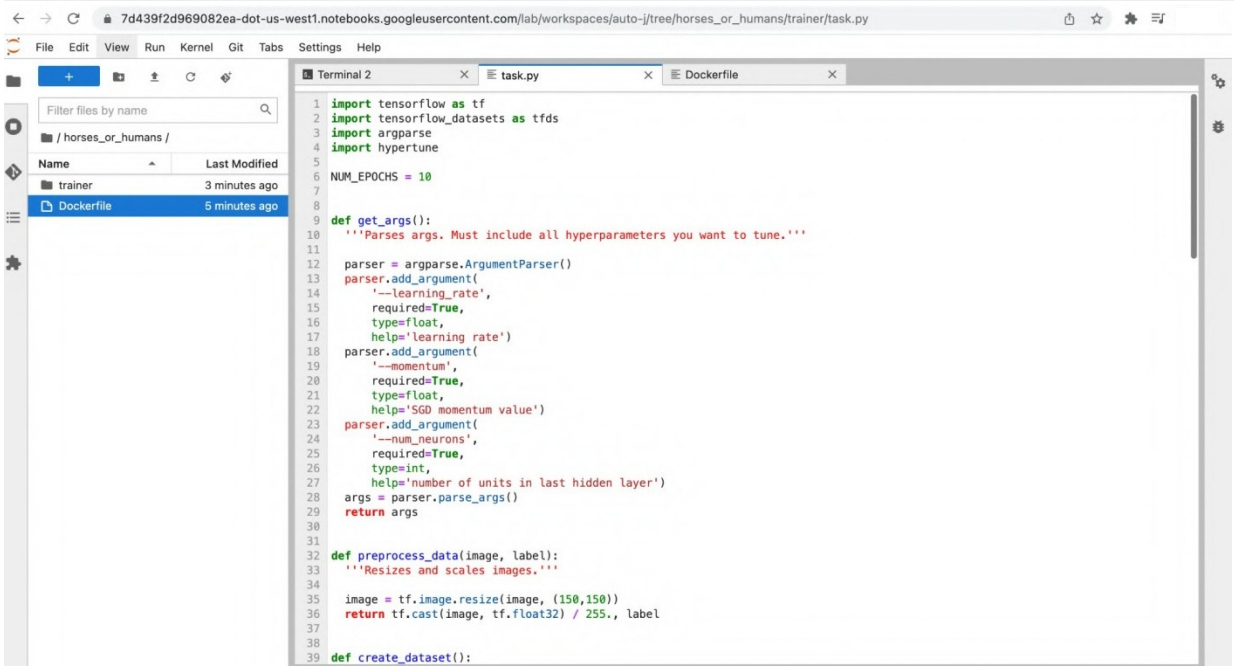

```
reññ = seasonal de omyse{d[arge), perisd=3f
```

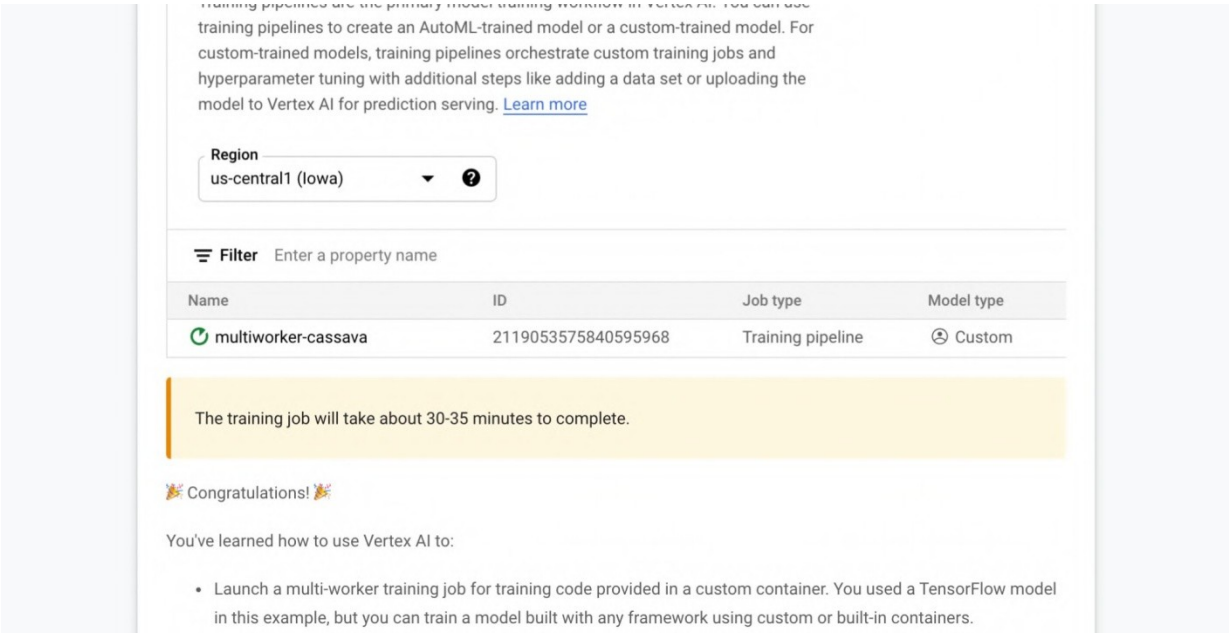




i) https://codelabs.developers.google.com/vertex_multiworker_training#0


7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/workspaces/auto-j/tree/horses_or_humans/Dockerfile
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ horses_or_humans /
Name
Last Modified
trainer
3 minutes ago
Dockerfile
4 minutes ago
Terminal 2
task.py
Dockerfile
1 FROM gcr.io/deeplearning-platform-release/tf2-gpu.2-5
2
3 WORKDIR /
4
5 # Installs hypertune library
6 RUN pip install cloudml-hypertune
7
8 # Copies the trainer code to the docker image.
9 COPY trainer /trainer
10
11 # Sets up the entry point to invoke the trainer.
12 ENTRYPOINT ["python", "-m", "trainer.task"]
Simple
2
0
Dockerfile
Ln 1, Col 1
Spaces: 4
Dockerfile


7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/workspaces/auto-j/tree/horses_or_humans/trainer/task.py
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/ horses_or_humans /
Name
Last Modified
trainer
3 minutes ago
Dockerfile
5 minutes ago
Terminal 2
task.py
Dockerfile
1 import tensorflow as tf
2 import tensorflow_datasets as tfds
3 import argparse
4 import hypertune
5
6 NUM_EPOCHS = 10
7
8
9 def get_args():
10 '''Parses args. Must include all hyperparameters you want to tune.'''
11
12 parser = argparse.ArgumentParser()
13 parser.add_argument(
14 '--learning_rate',
15 required=True,
16 type=float,
17 help='learning rate')
18 parser.add_argument(
19 '--momentum',
20 required=True,
21 type=float,
22 help='SGD momentum value')
23 parser.add_argument(
24 '--num_neurons',
25 required=True,
26 type=int,
27 help='number of units in last hidden layer')
28 args = parser.parse_args()
29 return args
30
31
32 def preprocess_data(image, label):
33 '''Resizes and scales images.'''
34
35 image = tf.image.resize(image, (150,150))
36 return tf.cast(image, tf.float32) / 255., label
37
38
39 def create_dataset():



j) https://codelabs.developers.google.com/codelabs/bqml-vertex-prediction#0

training pipelines to create an AutoML-trained model or a custom-trained model. For custom-trained models, training pipelines orchestrate custom training jobs and hyperparameter tuning with additional steps like adding a data set or uploading the model to Vertex AI for prediction serving. Learn more

Region
us-central1 (Iowa)

Filter Enter a property name

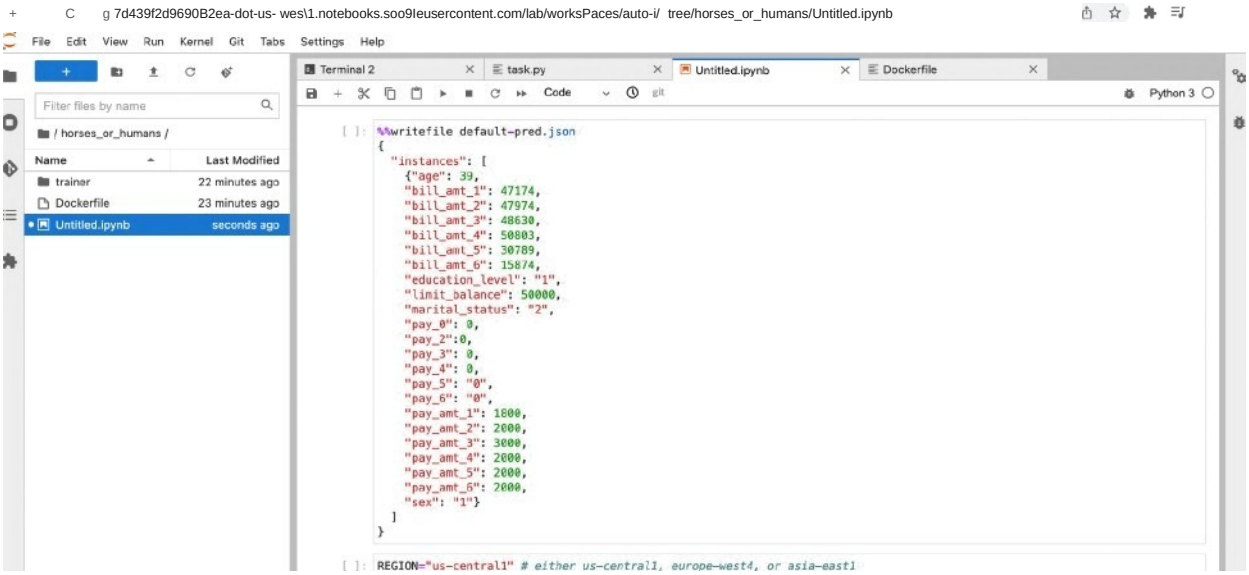
Name	ID	Job type	Model type
j multiworker-cassava	2119053575840595968	Training pipeline	@ CUStOft\

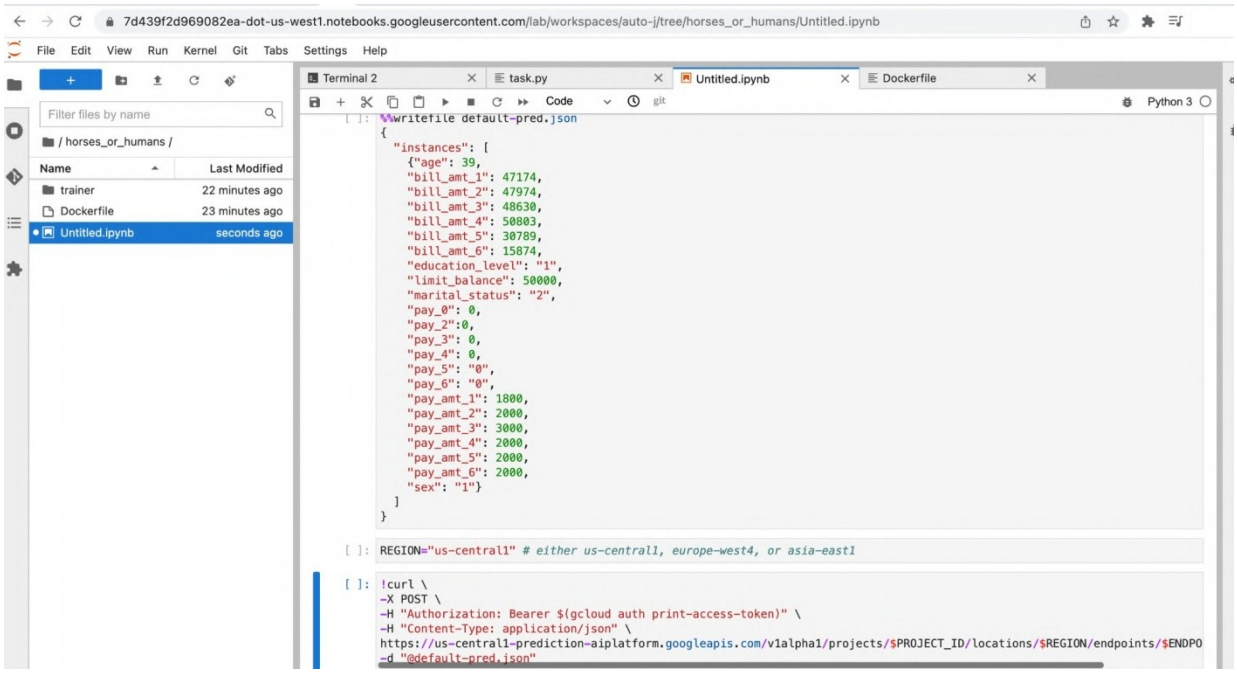
The training job will take about 30-35 minutes to complete.

@ Congratulations! 6

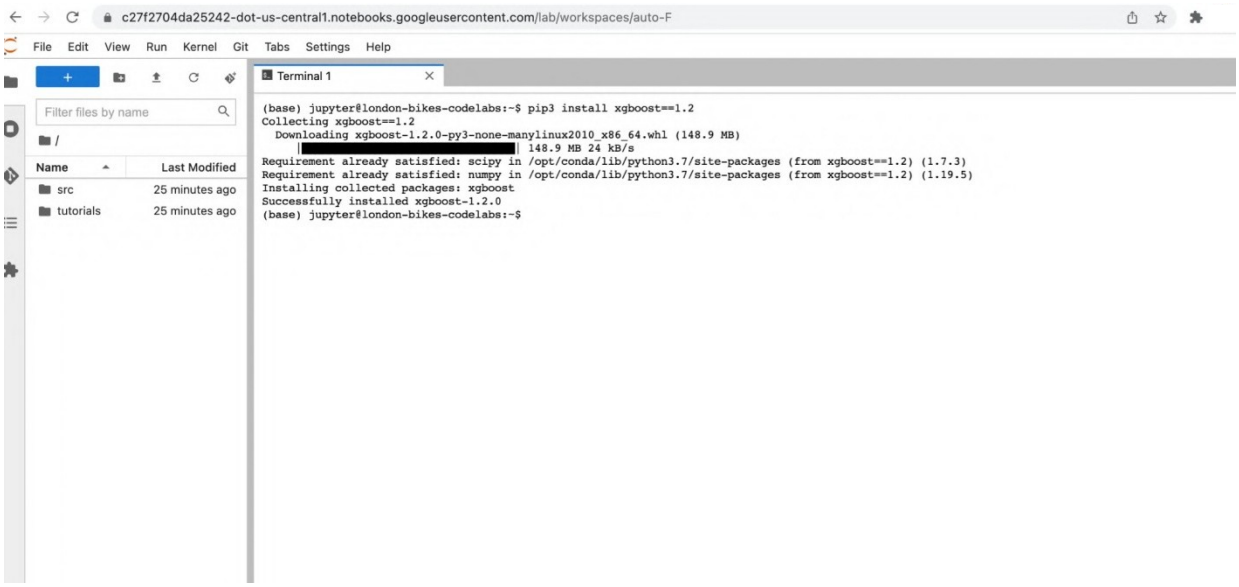
You've learned how to use Vertex AI to:

- Launch a multi-worker training job for training code provided in a custom container. You used a TensorFlow model in this example, but you can train a model built with any framework using custom or built-in containers.





k) https://codelabs.developers.google.com/vertexx-xgb-wit#0



g c27f2704da25242-dot- us- central1.notebooks.googleusercontent.com /lab/Tree/e ntitled.ipynb

Code

Python 3

extension community. However, we do n
review third-party extensions, and some
extensions may introduce security risks or
contain malicious code that runs on your

```
import collections
import witwidget
```

```
f rom with/tdg et . no tebo ok . v zsua "-I z a l No n ?mpo rt WTtW id get , III Co n f zgBu i Ide r
```

```
cocuru mAmzs = cone<‹io»s.ordeceJoi<‹(‹
```

```
'purchaser_type': 'category',
'hoepa_status': 'category',
```

G a cz7fz7o4dazsz4:z-dot-us-central1.notebooks.googieusercontent.com,'lau/i‹ee/ui titieu ipyou

rile Edit view Ran kernel cii cabs sewings i-Jelp

review ! hird-par/ y*exiensions, and"»me"'
extensions may introduce security' risks cr

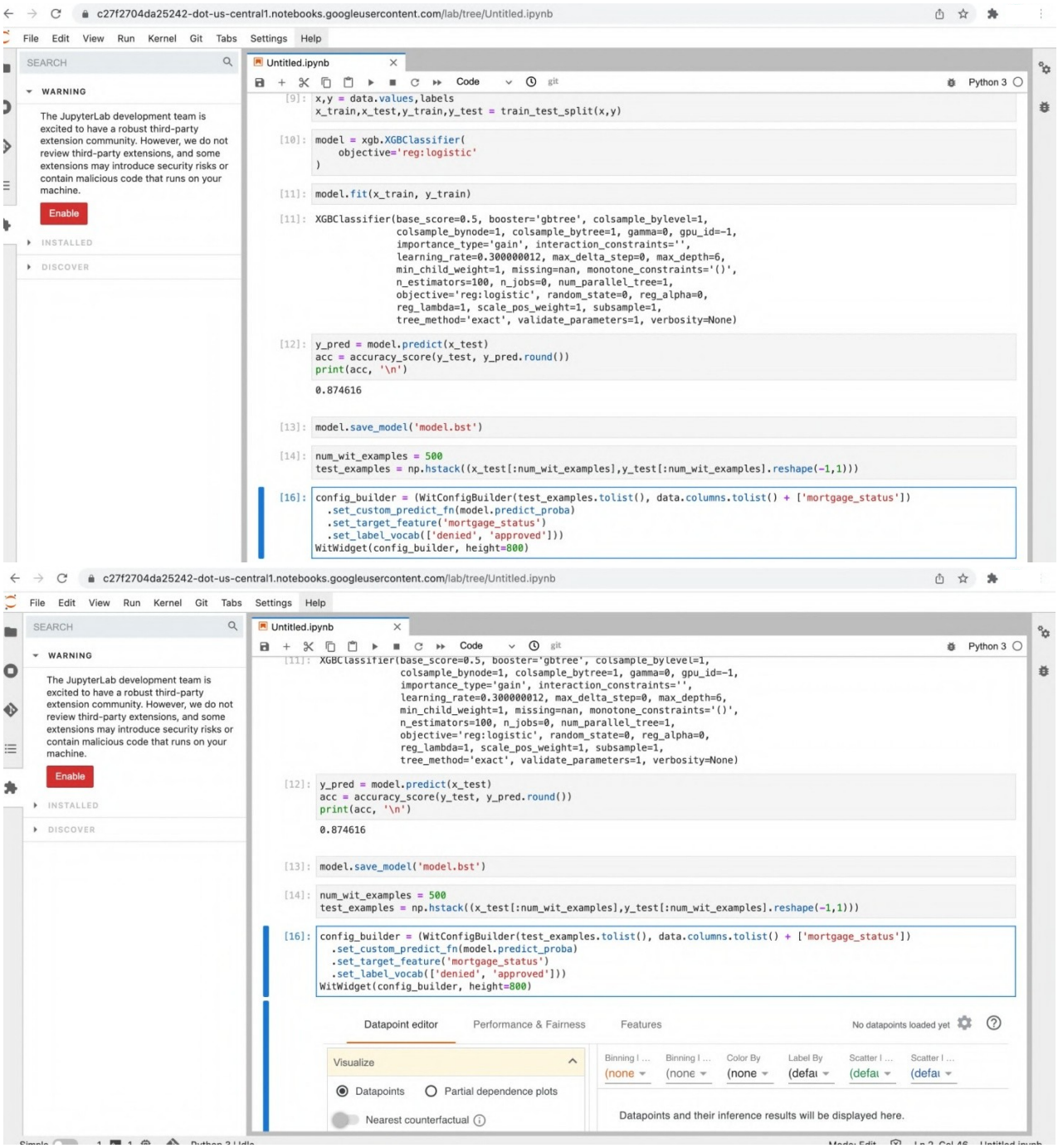
Enable

```
'preapproval': 'category',
'county_code': np.float64,
```

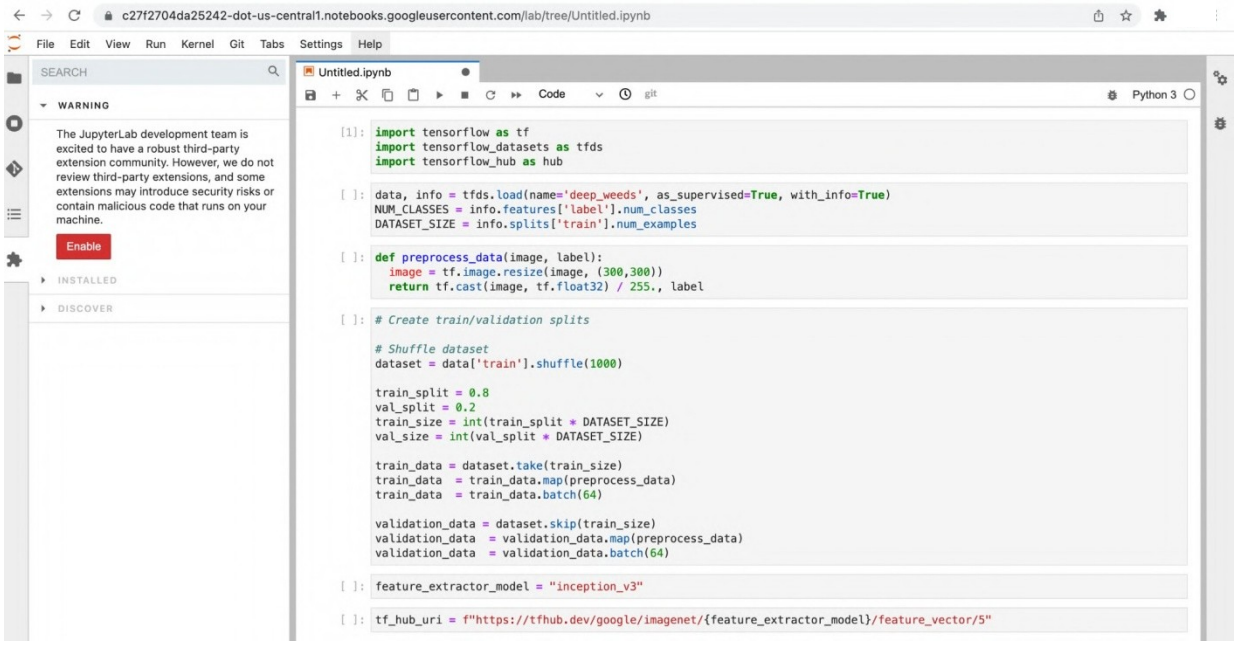
```
'lien_status': 'category',
```

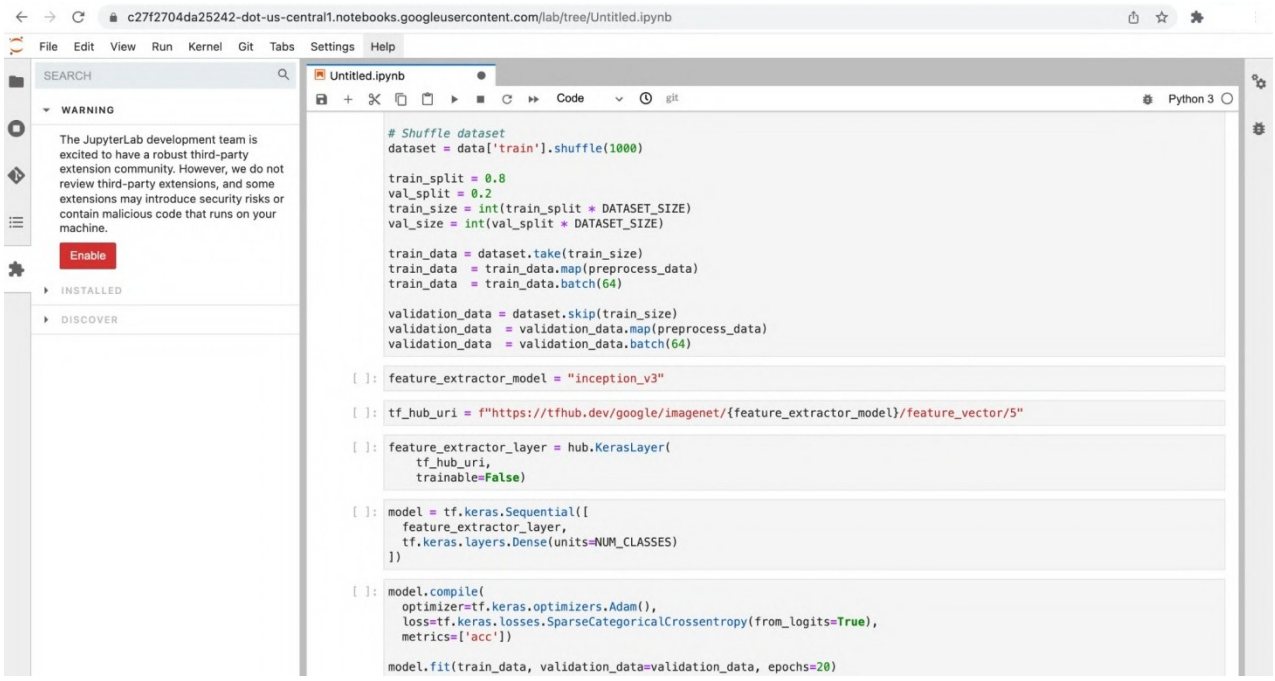
```
data = pd . rea d c s v(
```


W g c27f27 04da 25242-dot-u s—central1. notebooks.googleusercо n tenI.c om,'ia b/I ree/U ntitle d.ipynb
ri|e Ewe yiez Fun Kerne| sir Rao› stings neip
Untitled.ipyr b
Code
Administration (NCUA)
FHA, VA, FSA...
manufactured ho...
applicable
WARNING
extension community. Howe\'er we do not
extensions may introduce security risks or
Enable
DISCOVER
[7]:
dummy_columns = list(data.dtypes[data.dtypes == 'category'].index)
data = pd.get_dummies(data, columns=dummy_columns)
as=yar o u#aa loanamq=«sa=s =u=¿‹»e app isan_ nome_i ousa s g Pm aion me me°n amp ome iras¿n «
719598 2016 1 10 0 0 127 0 70.0 2422 0 46400.0

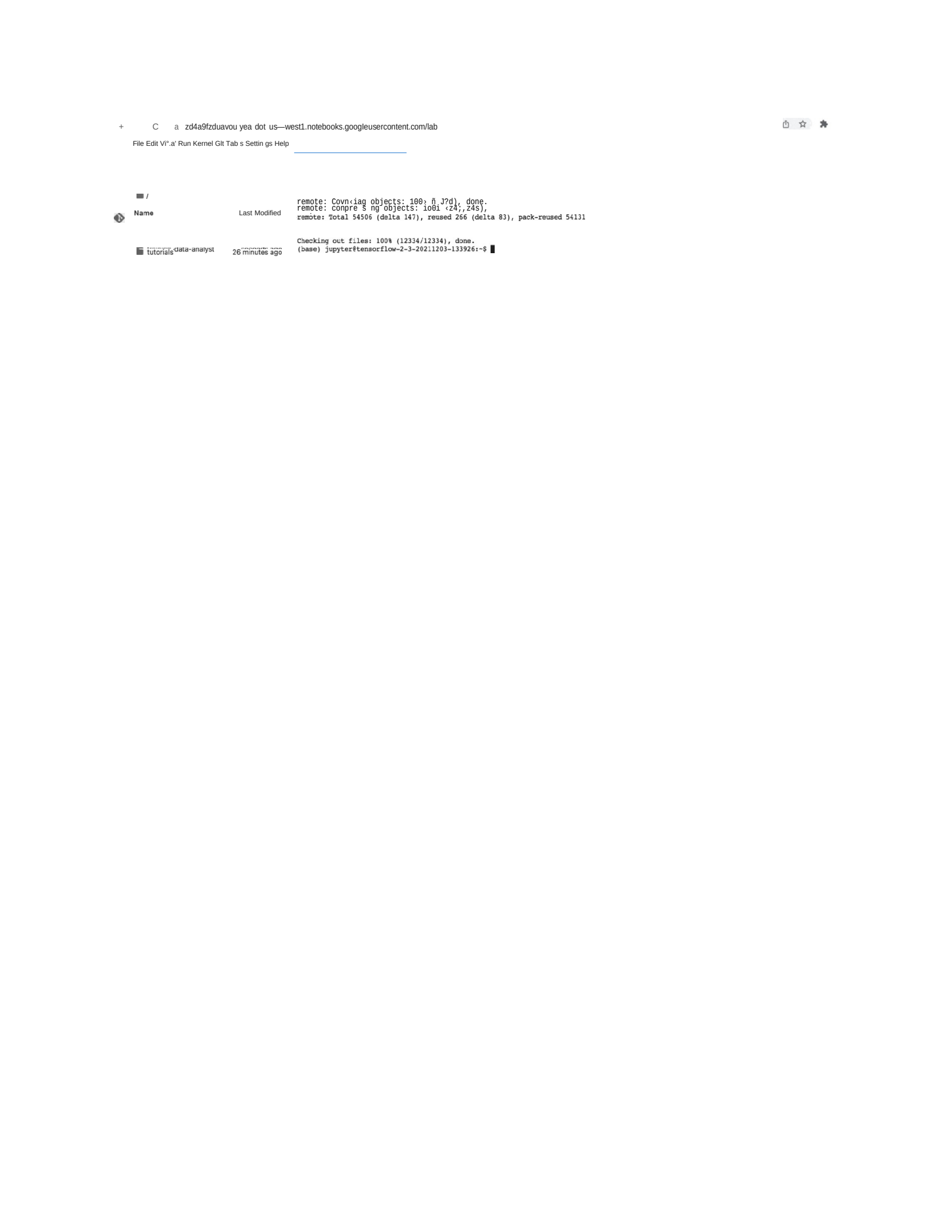


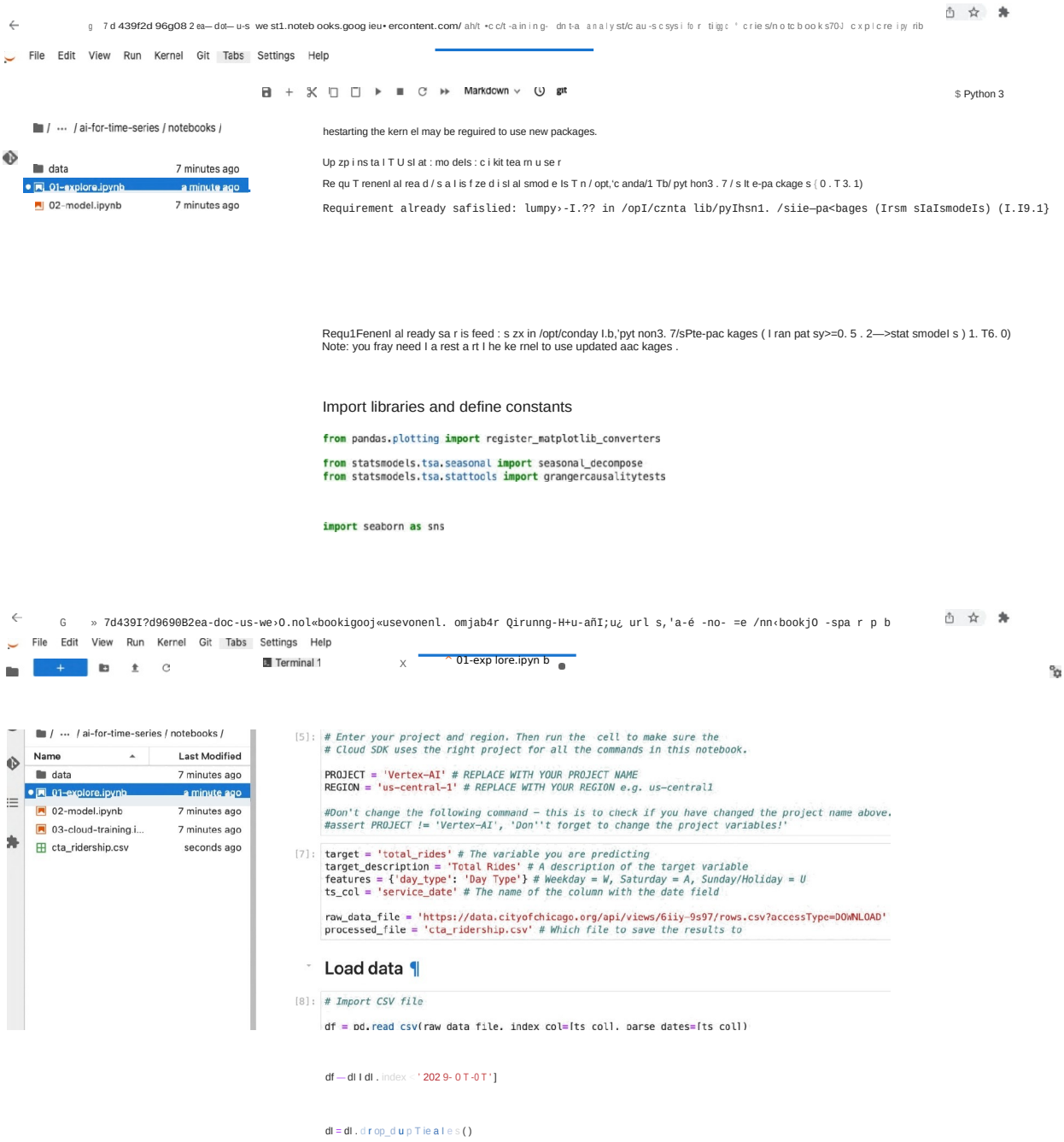
l) h ttps://codelabs.developers.google.com/vertex_notebook_executor#0



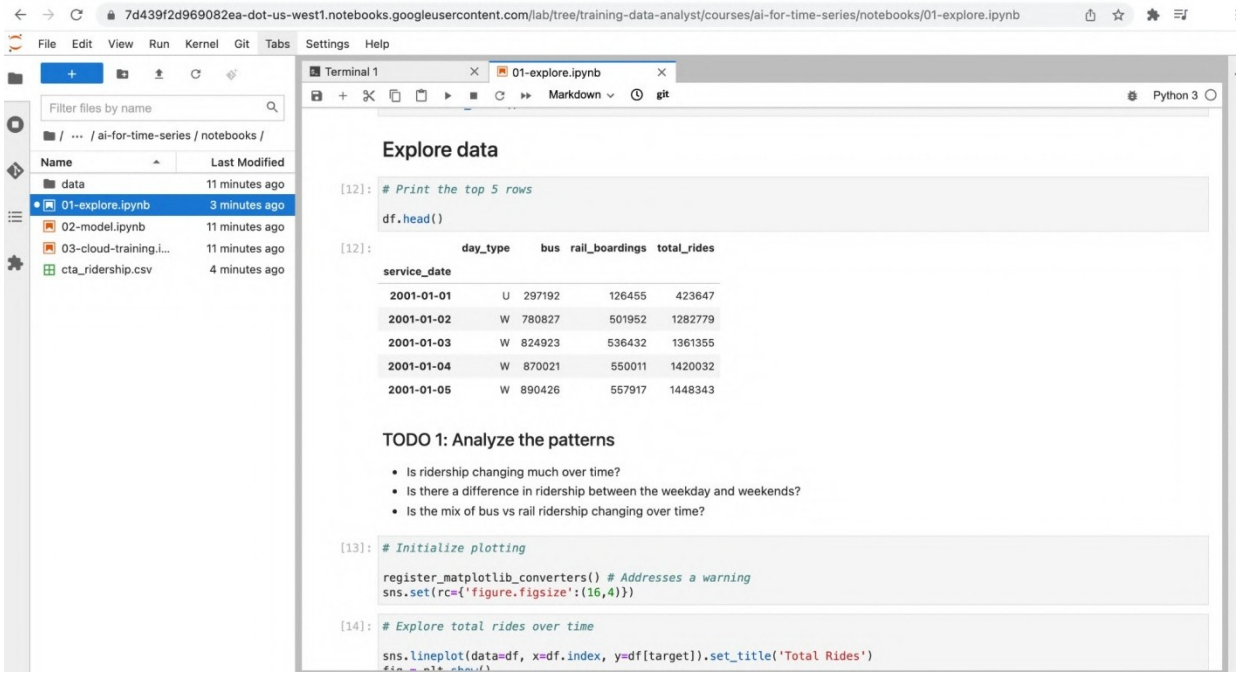


m) https://codelabs.developers.google.com/codelabs/time-series-forecasting-with-cloud-ai-platform#0


+ C a zd4a9fzduavou yea dot us—west1.notebooks.googleusercontent.com/lab
File Edit Vi°.a' Run Kernel Glt Tab s Settin gs Help
/
Name
Last Modified
tutorials
data-analyst
26 minutes ago
remote: Covn‹iag objects: 100› ñ J?d), done.
remote: conpre S ng objects: io0i ‹z4',,z4s),
remote: Total 54506 (delta 147), reused 266 (delta 83), pack-reused 54131
Checking out files: 100% (12334/12334), done.
(base) jupyter@tensorflow-2-3-20211203-133926:~$


File Edit View Run Kernel Git Tabs Settings Help
Markdown
$ Python 3
/ ... / ai-for-time-series / notebooks /
data 7 minutes ago
01-explore.ipynb a minute ago
02-model.ipynb 7 minutes ago
hestarting the kern el may be reguired to use new packages.
Up zp i ns ta l T U sl at : mo dels : c i kit tea m u se r
Re qu T renenl al rea d / s a l is f ze d i sl al smod e ls T n / opt,'c anda/1 Tb/ pyt hon3 . 7 / s lt e-pa ckage s (0 . T 3. 1)
Requirement already safislied: lumpy›-I.?? in /opI/cznta lib/pyIhsn1. /siie—pa<bages (Irsm sIaIsmodeIs) (I.I9.1}
Requ1Fenenl al ready sa r is feed : s zx in /opt/conday l.b,'pyt non3. 7/sPte-pac kages (I ran pat sy>=0. 5 . 2—>stat smodel s) 1. T6. 0)
Note: you fray need I a rest a rt I he ke rnel to use updated aac kages .
Import libraries and define constants
from pandas.plotting import register_matplotlib_converters
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import grangercausalitytests
import seaborn as sns

File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
01-exp lore.ipyn b
/ ... / ai-for-time-series / notebooks /
Name
Last Modified
data 7 minutes ago
01-explore.ipynb a minute ago
02-model.ipynb 7 minutes ago
03-cloud-training i... 7 minutes ago
cta_ridership.csv seconds ago
[5]:
# Enter your project and region. Then run the cell to make sure the
# Cloud SDK uses the right project for all the commands in this notebook.
PROJECT = 'Vertex-AI' # REPLACE WITH YOUR PROJECT NAME
REGION = 'us-central-1' # REPLACE WITH YOUR REGION e.g. us-central1
#Don't change the following command - this is to check if you have changed the project name above.
#assert PROJECT != 'Vertex-AI', 'Don''t forget to change the project variables!'
[7]:
target = 'total_rides' # The variable you are predicting
target_description = 'Total Rides' # A description of the target variable
features = {'day_type': 'Day Type'} # Weekday = W, Saturday = A, Sunday/Holiday = U
ts_col = 'service_date' # The name of the column with the date field
raw_data_file = 'https://data.cityofchicago.org/api/views/6iiy-9s97/rows.csv?accessType=DOWNLOAD'
processed_file = 'cta_ridership.csv' # Which file to save the results to
Load data ¶
[8]:
# Import CSV file
df = pd.read_csv(raw_data_file, index_col=[ts_col], parse_dates=[ts_col])
df — dI I dI . index ' 202 9- 0 T -0 T ']
dI = dI . d r op_d u p T ie a I e s ( )


7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
01-explore.ipynb
Markdown
git
Python 3
Filter files by name
/ ··· / ai-for-time-series / notebooks /
Name
Last Modified
data 11 minutes ago
01-explore.ipynb 3 minutes ago
02-model.ipynb 11 minutes ago
03-cloud-training.i... 11 minutes ago
cta_ridership.csv 4 minutes ago
Explore data
[12]:
# Print the top 5 rows
df.head()
[12]:
day_type bus rail_boardings total_rides
service_date
2001-01-01 U 297192 126455 423647
2001-01-02 W 780827 501952 1282779
2001-01-03 W 824923 536432 1361355
2001-01-04 W 870021 550011 1420032
2001-01-05 W 890426 557917 1448343
TODO 1: Analyze the patterns
Is ridership changing much over time?
Is there a difference in ridership between the weekday and weekends?
Is the mix of bus vs rail ridership changing over time?
[13]:
Initialize plotting
register_matplotlib_converters() # Addresses a warning
sns.set(rc={'figure.figsize':(16,4)})
[14]:
# Explore total rides over time
sns.lineplot(data=df, x=df.index, y=df[target]).set_title('Total Rides')


7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
01-explore.ipynb
Markdown
git
Python 3
Filter files by name
/ ··· / ai-for-time-series / notebooks /
Name
Last Modified
data 11 minutes ago
01-explore.ipynb 3 minutes ago
02-model.ipynb 11 minutes ago
03-cloud-training.i... 11 minutes ago
cta_ridership.csv 4 minutes ago
1e6
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
total_rides
2002 2004 2006 2008 2010 2012 2014 2016 2018 2020
service_date
[15]:
# Explore rides by day type: Weekday (W), Saturday (A), Sunday/Holiday (U)
sns.lineplot(data=df, x=df.index, y=df[target], hue=df['day_type']).set_title('Total Rides by Day Type')
fig = plt.show()
Total Rides by Day Type
1e6
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
total_rides
day_type
U
W
A
2002 2004 2006 2008 2010 2012 2014 2016 2018 2020
service_date
Simple
1 1
Python 3 | Idle
Mode: Command
Ln 1, Col 1
01-explore.ipynb

7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/ ... / ai-for-time-series / notebooks /

Name	Last Modified
data	11 minutes ago
01-explore.ipynb	3 minutes ago
02-model.ipynb	11 minutes ago
03-cloud-training.i...	11 minutes ago
cta_ridership.csv	4 minutes ago

Terminal 1 | 01-explore.ipynb

Markdown | git | Python 3

[16]:
```
# Explore rides by transportation type

sns.lineplot(data=df[['bus','rail_boardings']]).set_title('Total Rides by Transportation Type')
fig = plt.show()
```

TODO 2: Review summary statistics

- How many records are in the dataset?
- What is the average # of riders per day?

[17]:
```
df[target].describe().apply(lambda x: round(x))
```

[17]:
```
count       6939
mean     1368761
std       391443
min       222071
25%      1005394
50%      1548343
75%      1660947
max      2049519
Name: total_rides, dtype: int64
```

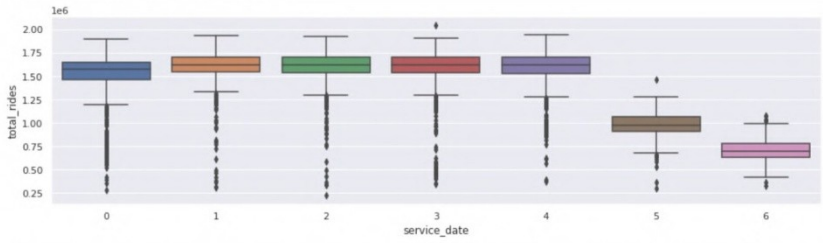
TODO 3: Explore seasonality

- Is there much difference between months?
- Can you extract the trend and seasonal pattern from the data?

[18]:
```
# Show the distribution of values for each day of the week in a boxplot:
# Min, 25th percentile, median, 75th percentile, max

daysofweek = df.index.to_series().dt.dayofweek

fig = sns.boxplot(x=daysofweek, y=df[target])
```



File Edit View Run Kernel Git Tabs Se ttings Help

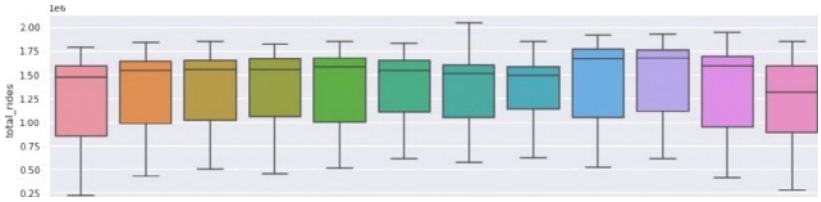
Terminal 1 | 01-explore.ipynb

'e

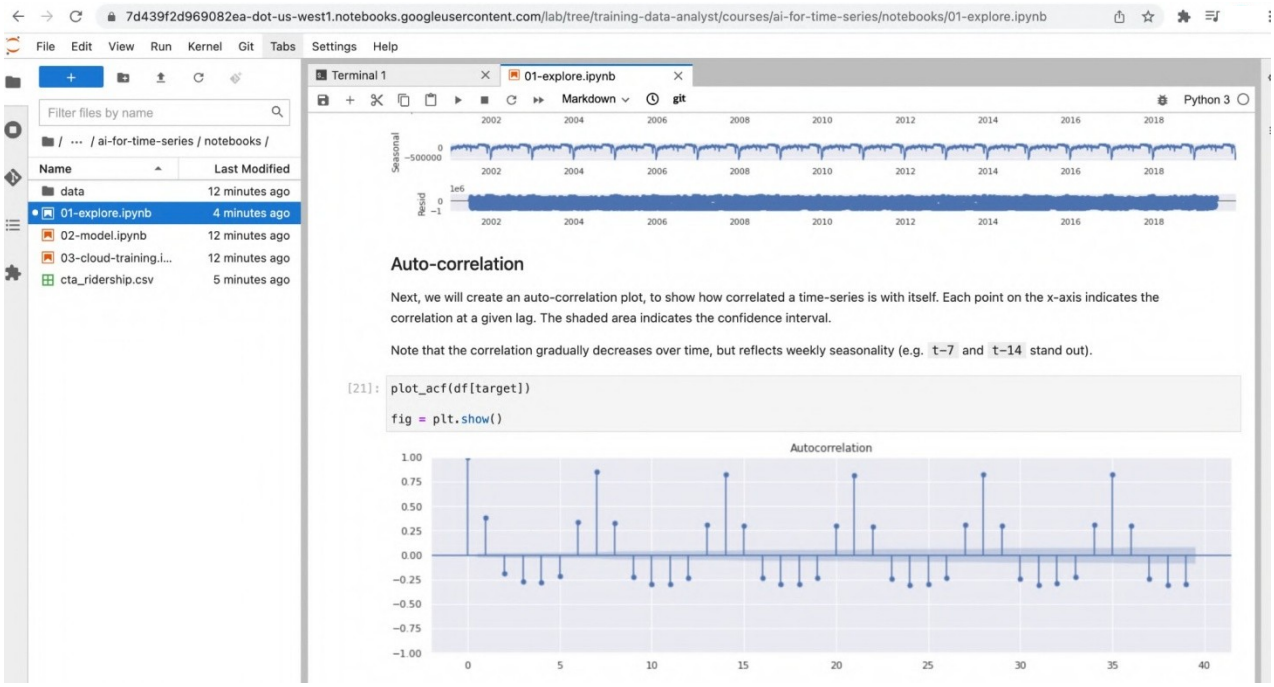
«meseries}xébooñ,'

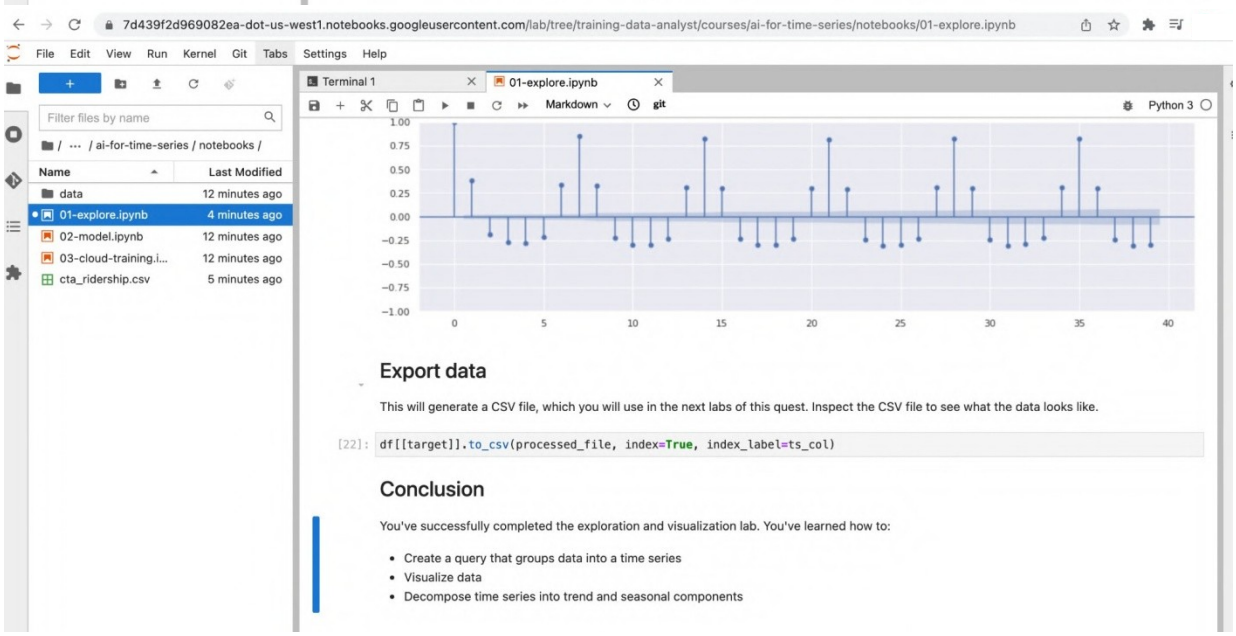
Name	Last Modified
02-model.i ab	12 minutes ago
FB cta ridership.csv	5 minutes ago

```
 fo  fe %s:ridnios of vales far eat svrh » s bo piñ:
moths = d .nde to_series).d:.nook
```




```
reññ = seasonal de omyse{d[arge), perisd=3f
```


7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
01-explore.ipynb
Markdown
git
Python 3
Filter files by name
/ ··· / ai-for-time-series / notebooks /
Name
Last Modified
data 12 minutes ago
01-explore.ipynb 4 minutes ago
02-model.ipynb 12 minutes ago
03-cloud-training.i... 12 minutes ago
cta_ridership.csv 5 minutes ago
Seasonal
Resid
Auto-correlation
Next, we will create an auto-correlation plot, to show how correlated a time-series is with itself. Each point on the x-axis indicates the correlation at a given lag. The shaded area indicates the confidence interval.
Note that the correlation gradually decreases over time, but reflects weekly seasonality (e.g. t-7 and t-14 stand out).
[21]:
plot_acf(df[target])
fig = plt.show()
Autocorrelation


7d439f2d969082ea-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
01-explore.ipynb
Markdown
git
Python 3
Filter files by name
/ ··· / ai-for-time-series / notebooks /
Name
Last Modified
data 12 minutes ago
01-explore.ipynb 4 minutes ago
02-model.ipynb 12 minutes ago
03-cloud-training.i... 12 minutes ago
cta_ridership.csv 5 minutes ago
Export data
This will generate a CSV file, which you will use in the next labs of this quest. Inspect the CSV file to see what the data looks like.
[22]:
df[[target]].to_csv(processed_file, index=True, index_label=ts_col)
Conclusion
You've successfully completed the exploration and visualization lab. You've learned how to:
Create a query that groups data into a time series
Visualize data
Decompose time series into trend and seasonal components