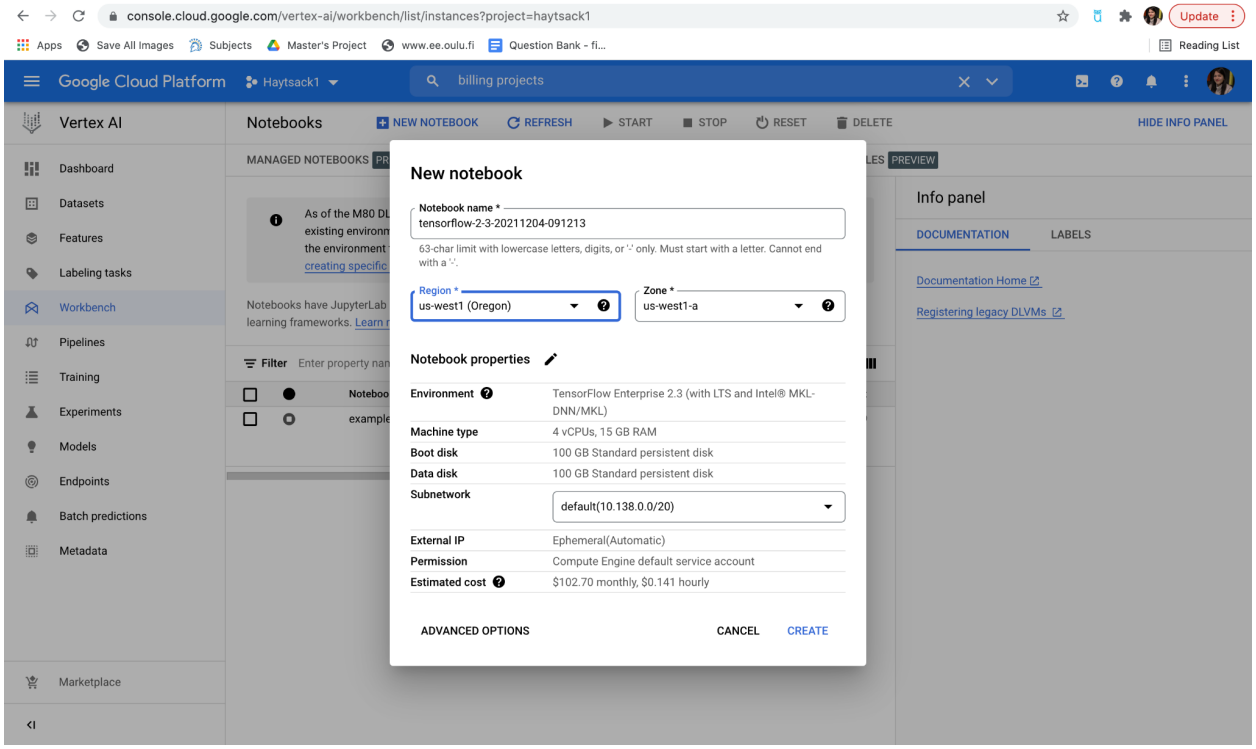
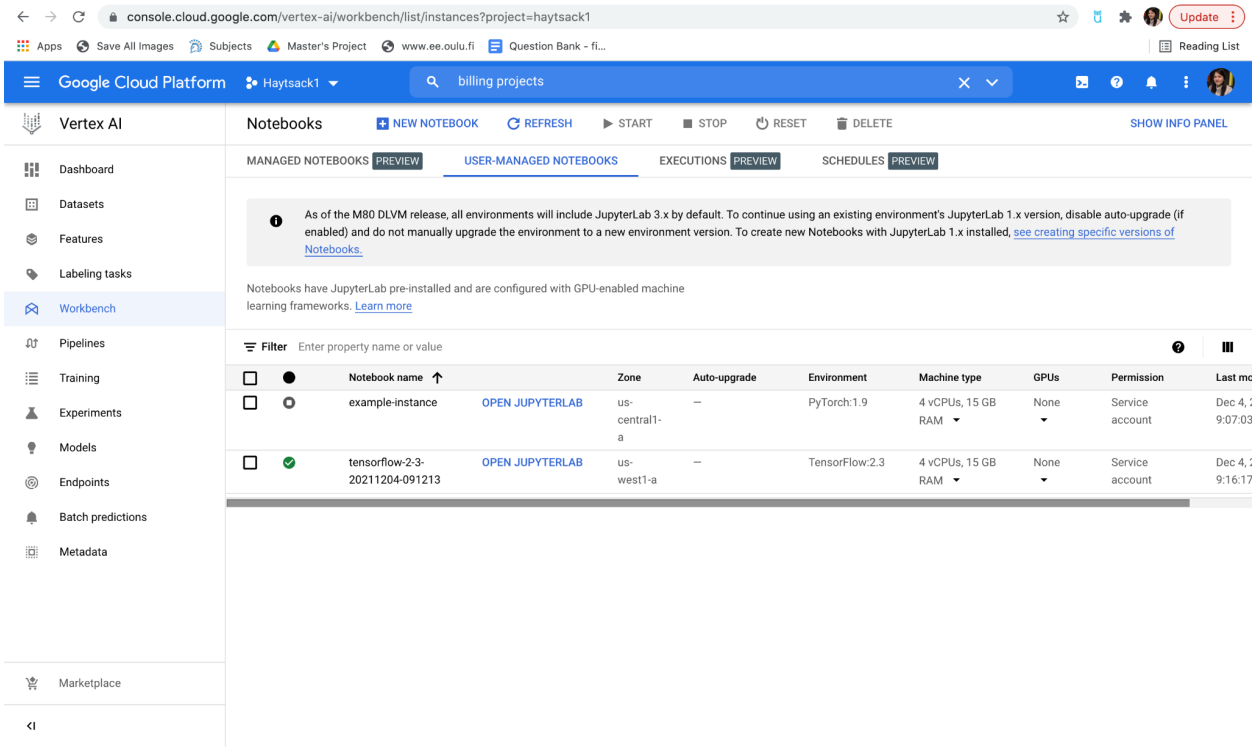
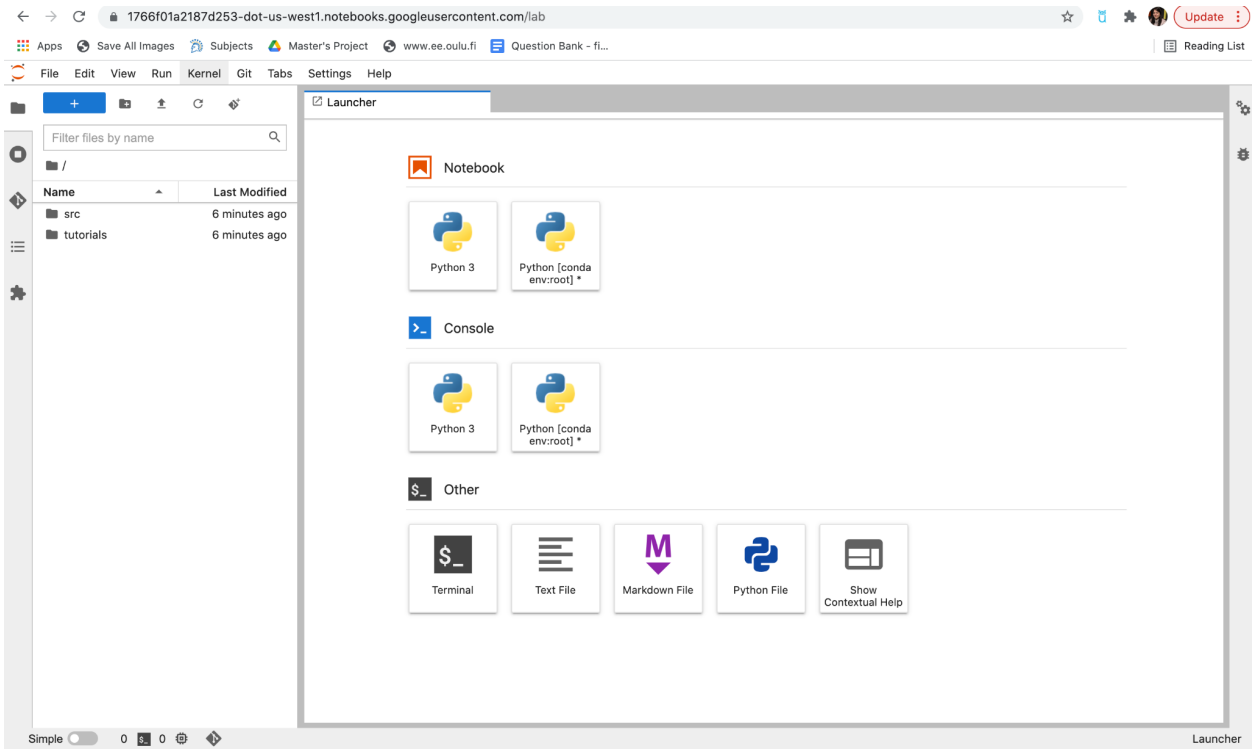
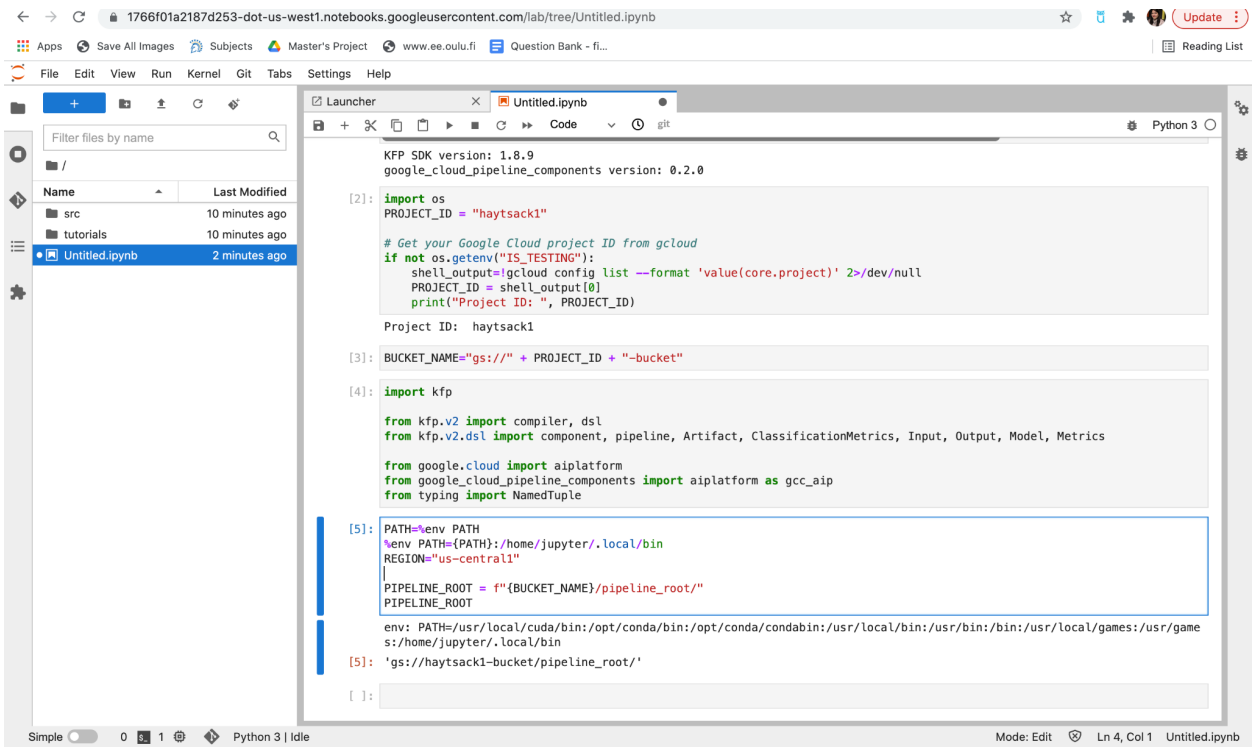
Shreya Goyal
Assignment : VERTEX AI

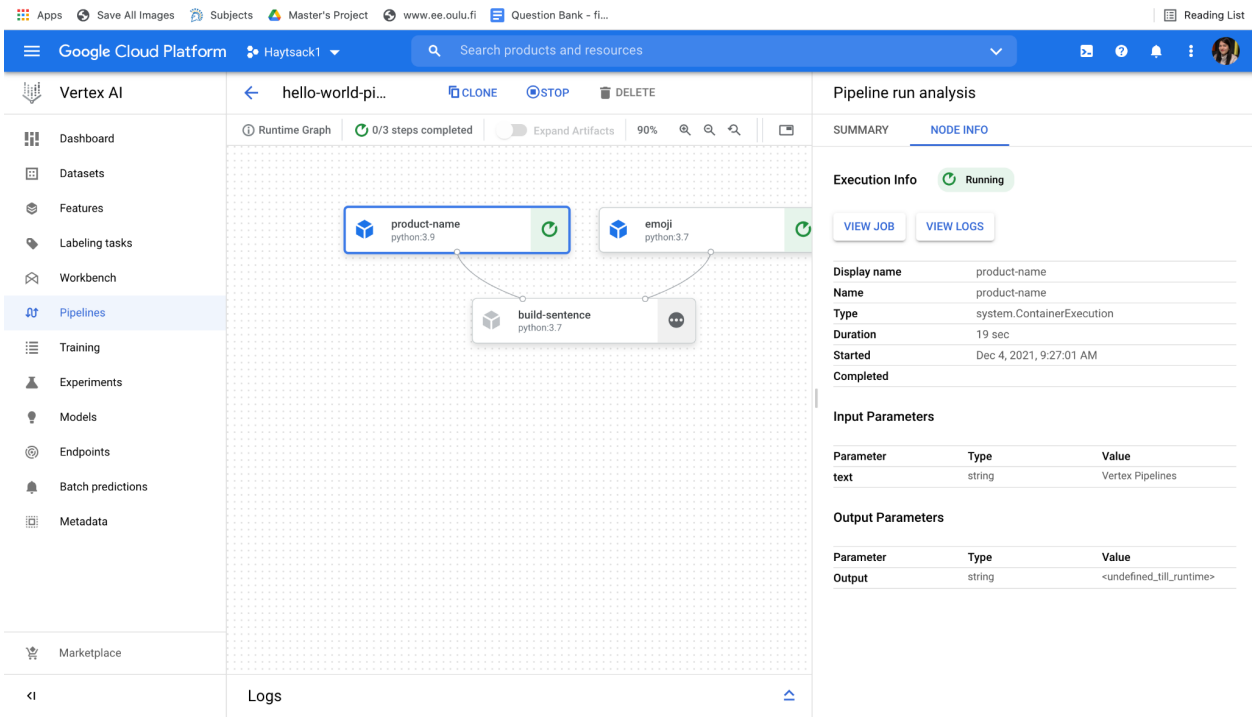
Part a) https://codelabs.developers.google.com/vertex-pipelines-intro#0



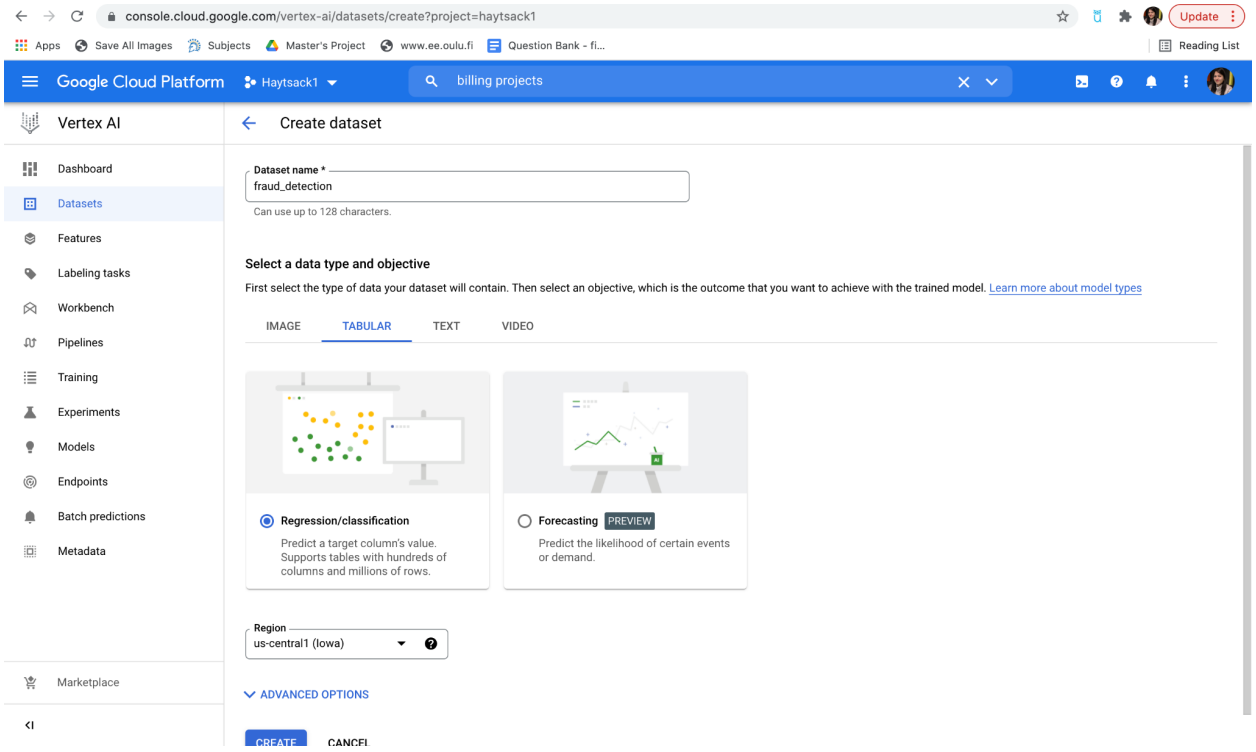

console.cloud.google.com/vertex-ai/workbench/list/instances?project=haytsack1
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
billing projects
Vertex AI
Notebooks
NEW NOTEBOOK
REFRESH
START
STOP
RESET
DELETE
SHOW INFO PANEL
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
MANAGED NOTEBOOKS PREVIEW
USER-MANAGED NOTEBOOKS
EXECUTIONS PREVIEW
SCHEDULES PREVIEW
As of the M80 DLVM release, all environments will include JupyterLab 3.x by default. To continue using an existing environment's JupyterLab 1.x version, disable auto-upgrade (if enabled) and do not manually upgrade the environment to a new environment version. To create new Notebooks with JupyterLab 1.x installed, see creating specific versions of Notebooks.
Notebooks have JupyterLab pre-installed and are configured with GPU-enabled machine learning frameworks. Learn more
Filter
Enter property name or value
Notebook name
Zone
Auto-upgrade
Environment
Machine type
GPUs
Permission
example-instance
OPEN JUPYTERLAB
us-central1-a
–
PyTorch:1.9
4 vCPUs, 15 GB RAM
None
Service account
tensorflow-2-3-20211204-091213
OPEN JUPYTERLAB
us-west1-a
–
TensorFlow:2.3
4 vCPUs, 15 GB RAM
None
Service account

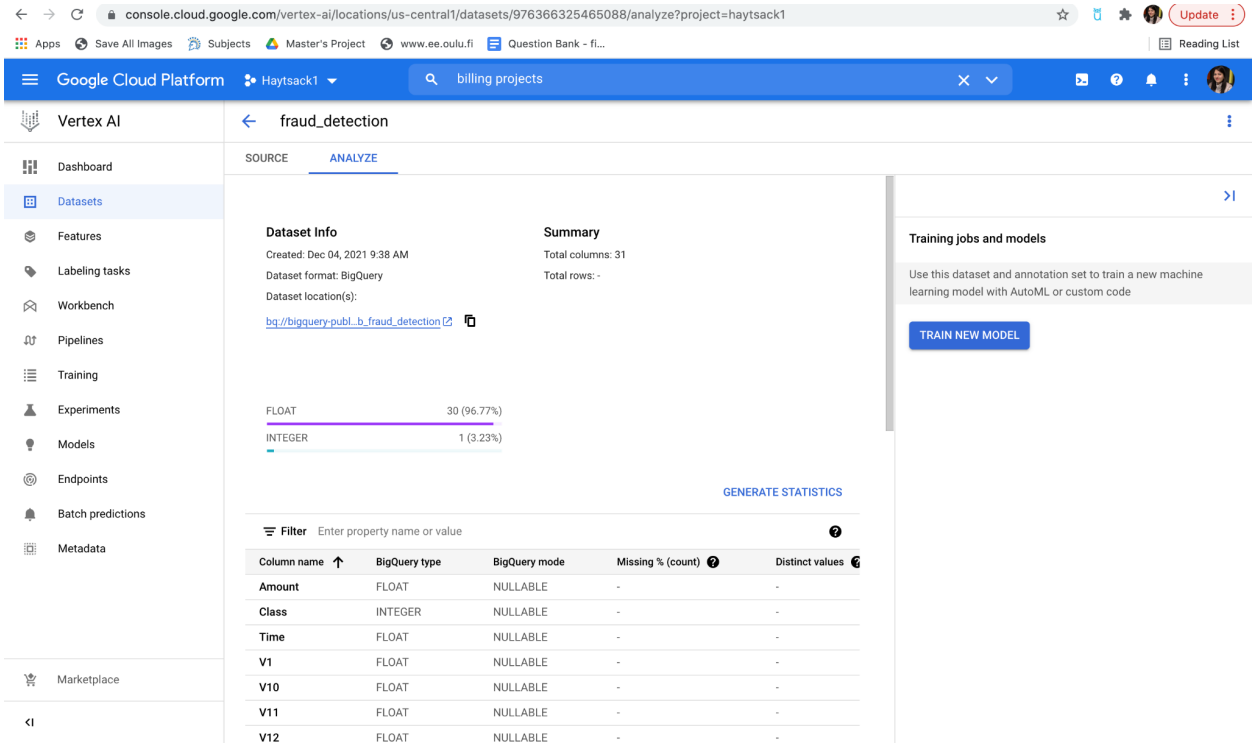

1766f01a2187d253-dot-us-west1.notebooks.googleusercontent.com/lab
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File
Edit
View
Run
Kernel
Git
Tabs
Settings
Help
Filter files by name
/
Name
Last Modified
src
6 minutes ago
tutorials
6 minutes ago
Launcher
Notebook
Python 3
Python [conda env:root] *
Console
Python 3
Python [conda env:root] *
Other
Terminal
Text File
Markdown File
Python File
Show Contextual Help
Simple
0
0
Launcher

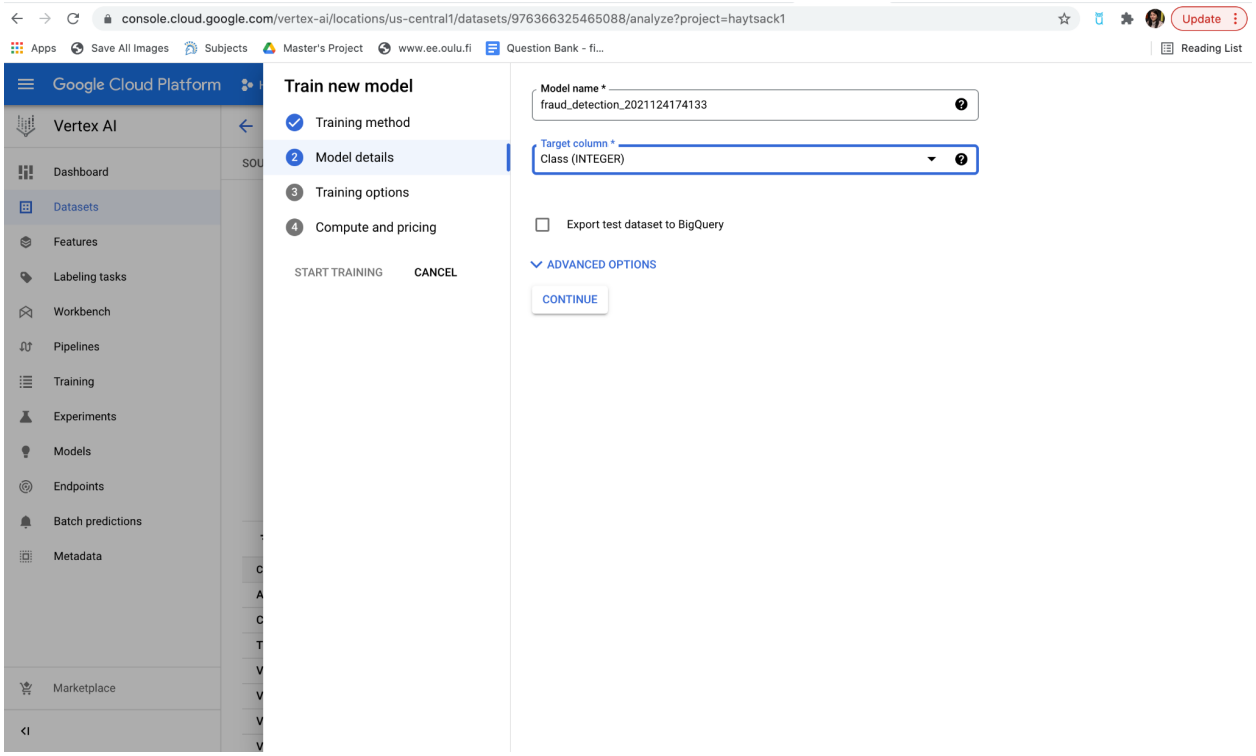


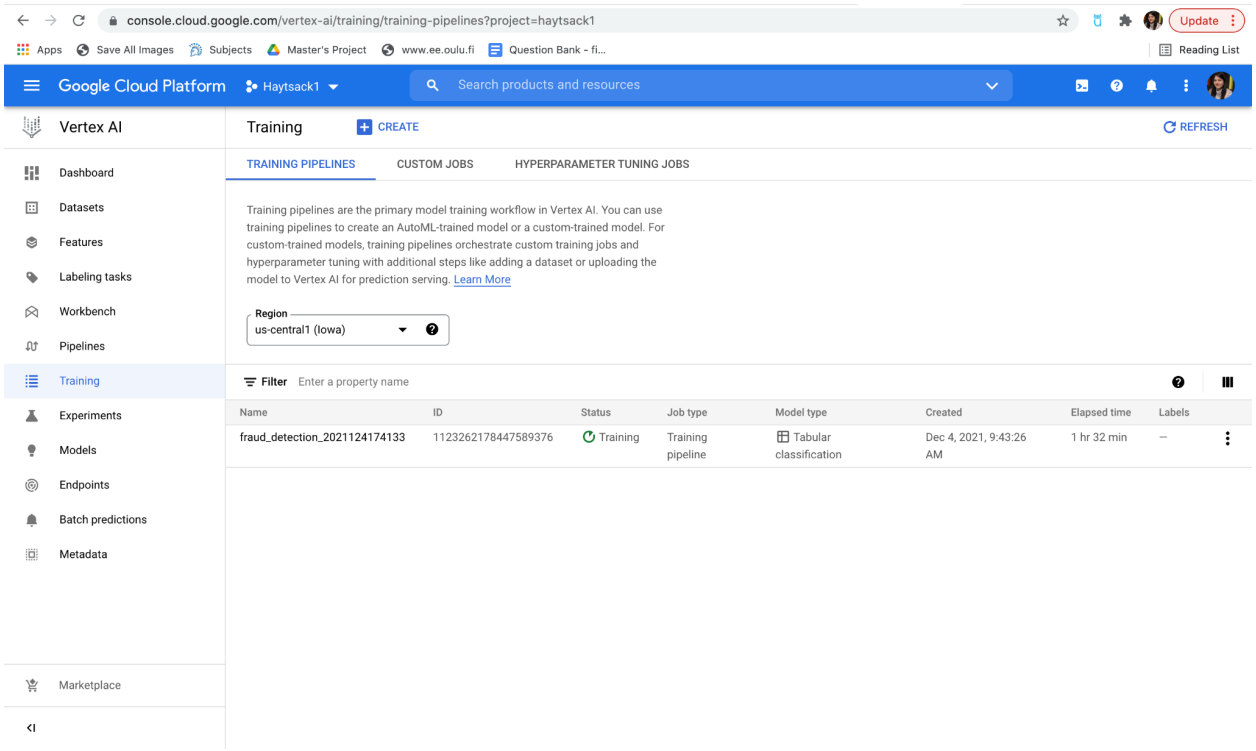


Part b) https://codelabs.developers.google.com/vertex-automl-tabular#0

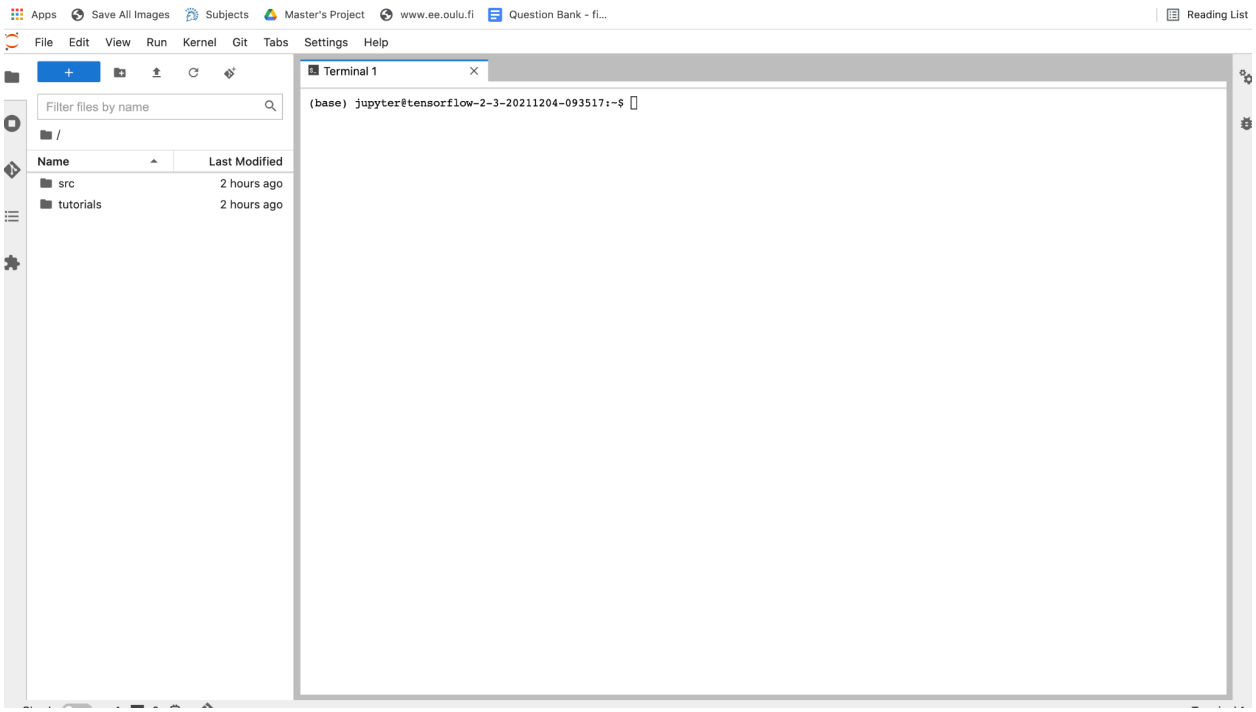

console.cloud.google.com/vertex-ai/datasets/create?project=haytsack1
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
billing projects
Vertex AI
Create dataset
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Dataset name *
fraud_detection
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
PREVIEW
Predict the likelihood of certain events or demand.
Region
us-central1 (Iowa)
ADVANCED OPTIONS
CREATE
CANCEL

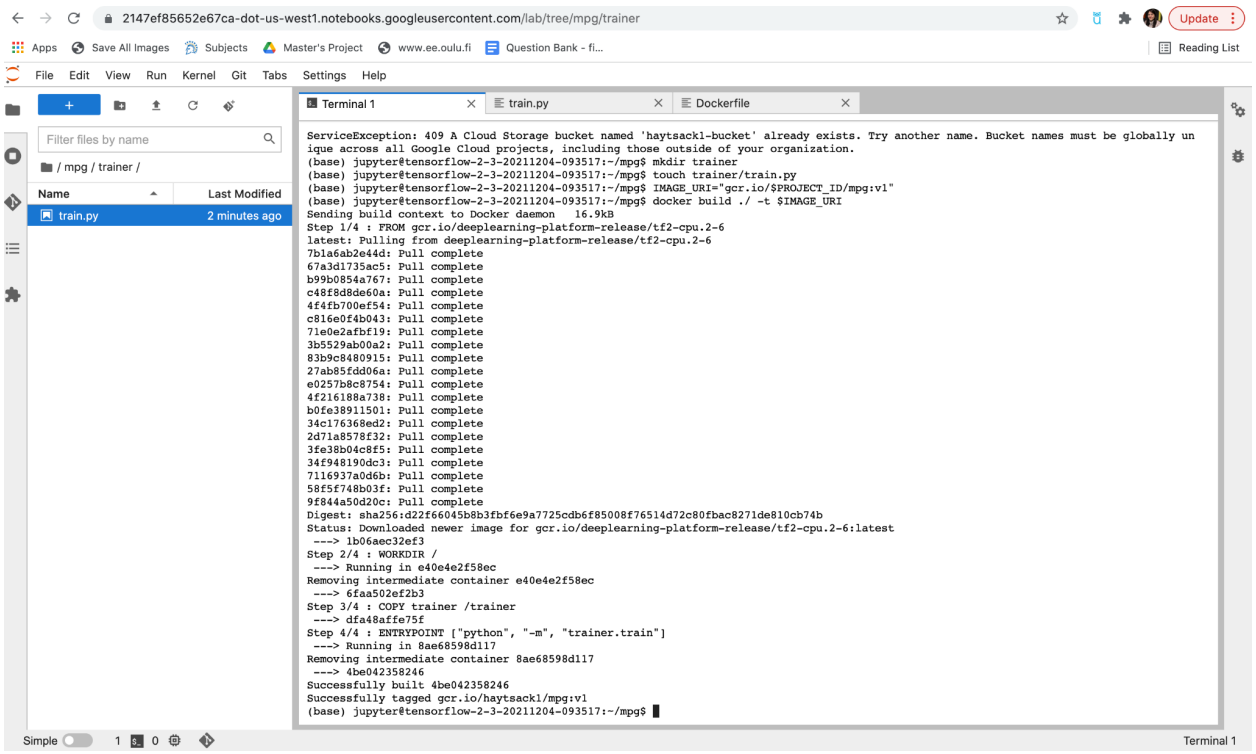

console.cloud.google.com/vertex-ai/locations/us-central1/datasets/976366325465088/analyze?project=haytsack1
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
billing projects
Vertex AI
fraud_detection
SOURCE
ANALYZE
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Dataset Info
Created: Dec 04, 2021 9:38 AM
Dataset format: BigQuery
Dataset location(s):
bq://bigquery-publ..b_fraud_detection
Summary
Total columns: 31
Total rows: -
FLOAT 30 (96.77%)
INTEGER 1 (3.23%)
GENERATE STATISTICS
Filter
Enter property name or value
Column name
BigQuery type
BigQuery mode
Missing % (count)
Distinct values
Amount FLOAT NULLABLE - -
Class INTEGER NULLABLE - -
Time FLOAT NULLABLE - -
V1 FLOAT NULLABLE - -
V10 FLOAT NULLABLE - -
V11 FLOAT NULLABLE - -
V12 FLOAT NULLABLE - -
Training jobs and models
Use this dataset and annotation set to train a new machine learning model with AutoML or custom code
TRAIN NEW MODEL

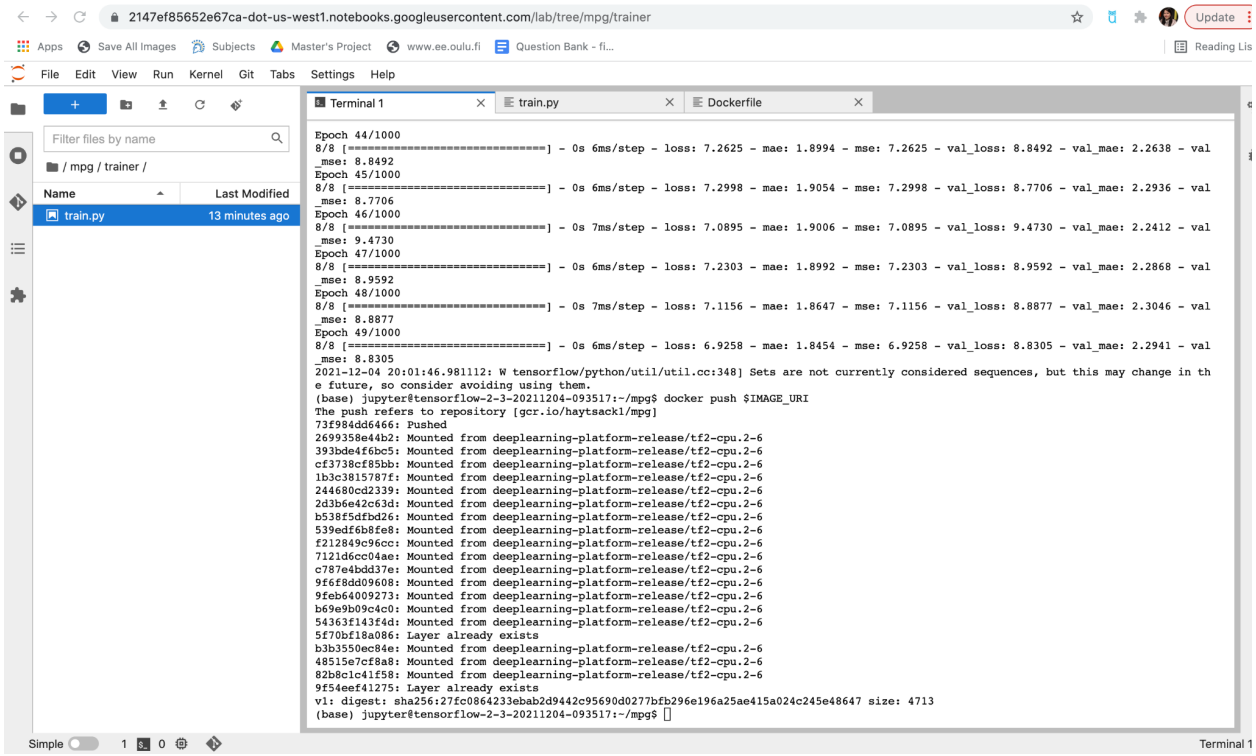




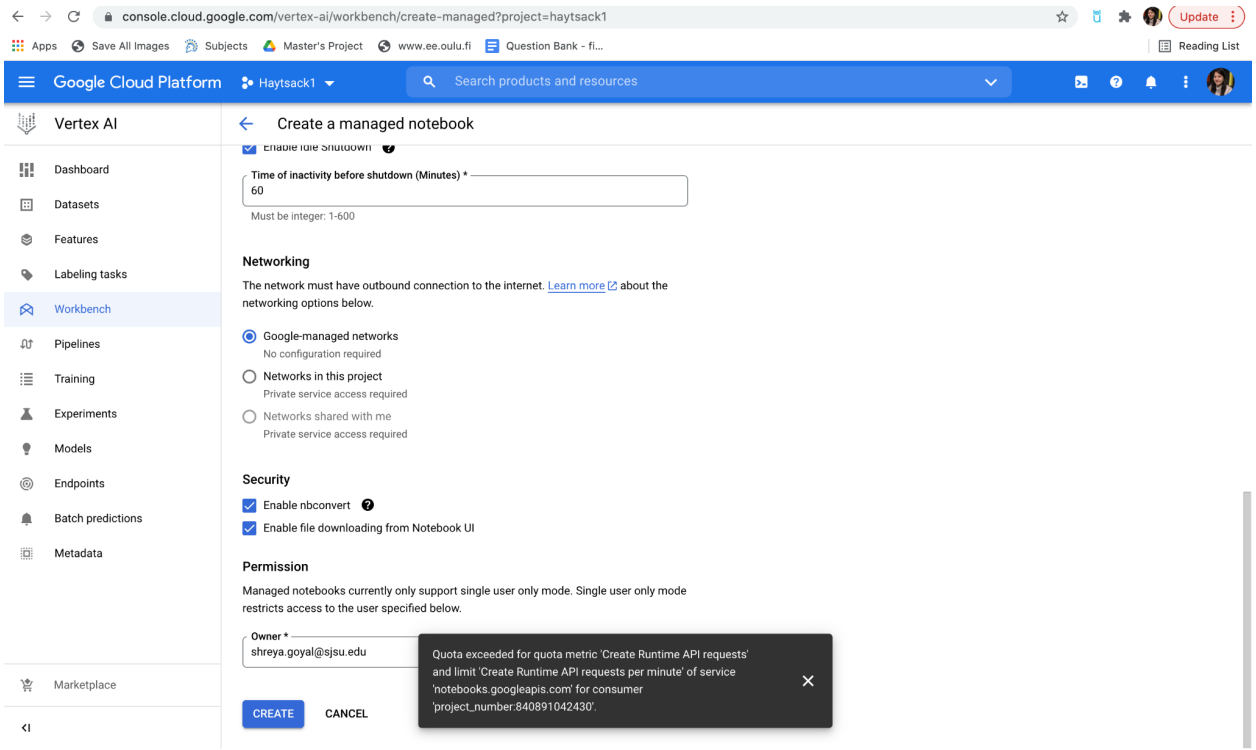
Part c) https://codelabs.developers.google.com/vertex_custom_training_prediction#0

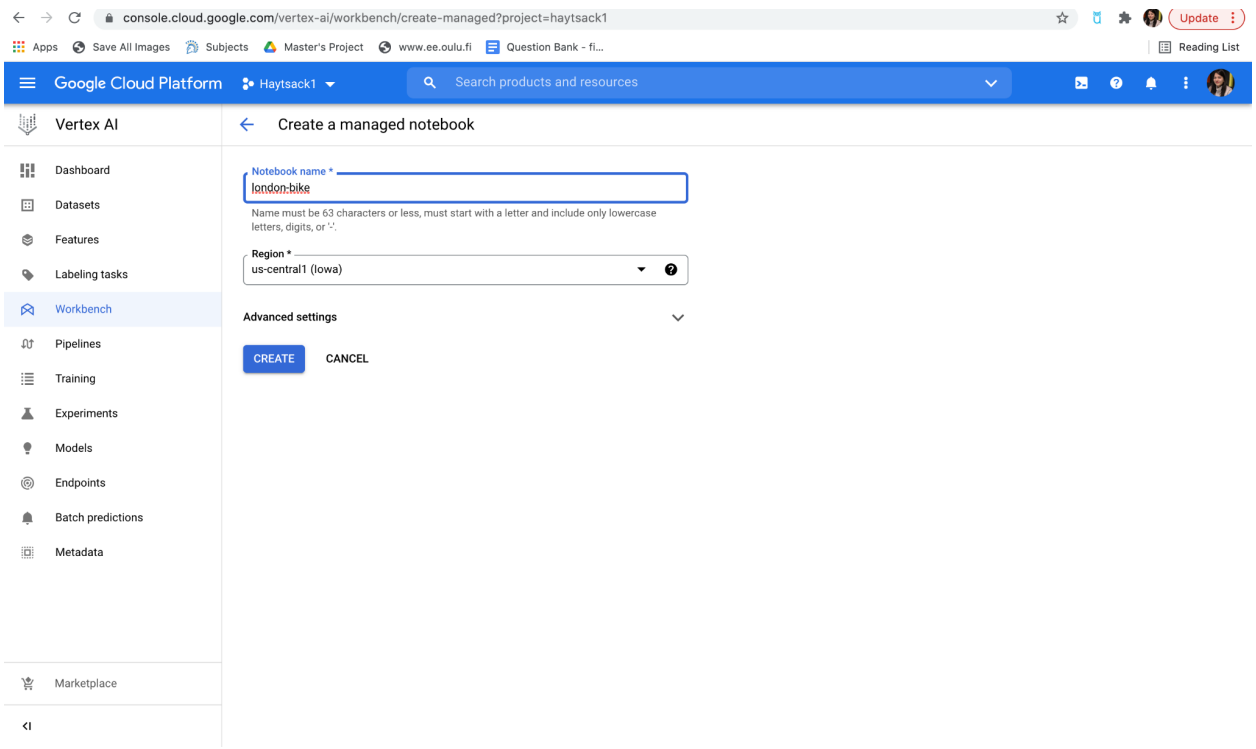

Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
(base) jupyter@tensorflow-2-3-20211204-093517:~$
Filter files by name
/
Name
Last Modified
src
2 hours ago
tutorials
2 hours ago


2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/tree/mpg/trainer
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
train.py
Dockerfile
Filter files by name
/ mpg / trainer /
Name
Last Modified
train.py
2 minutes ago
ServiceException: 409 A Cloud Storage bucket named 'haytsack1-bucket' already exists. Try another name. Bucket names must be globally un
ique across all Google Cloud projects, including those outside of your organization.
(base) jupyter@tensorflow-2-3-20211204-093517:~/mpg$ mkdir trainer
(base) jupyter@tensorflow-2-3-20211204-093517:~/mpg$ touch trainer/train.py
(base) jupyter@tensorflow-2-3-20211204-093517:~/mpg$ IMAGE_URI="gcr.io/$PROJECT_ID/mpg:v1"
(base) jupyter@tensorflow-2-3-20211204-093517:~/mpg$ docker build ./ -t $IMAGE_URI
Sending build context to Docker daemon 16.9kB
Step 1/4 : FROM gcr.io/deeplearning-platform-release/tf2-cpu.2-6
latest: Pulling from deeplearning-platform-release/tf2-cpu.2-6
7b1a6ab2e44d: Pull complete
67a3d1735ac5: Pull complete
b99b0854a767: Pull complete
c48f8d8de60a: Pull complete
4f4fb700ef54: Pull complete
c816e0f4b043: Pull complete
71e0e2afbf19: Pull complete
3b5529ab00a2: Pull complete
83b9c8480915: Pull complete
27ab85fdd06a: Pull complete
e0257b8c8754: Pull complete
4f216188a738: Pull complete
b0fe38911501: Pull complete
34c176368ed2: Pull complete
2d71a8578f32: Pull complete
3fe38b04c8f5: Pull complete
34f948190dc3: Pull complete
7116937a0d6b: Pull complete
58f5f748b03f: Pull complete
9f844a50d20c: Pull complete
Digest: sha256:d22f66045b8b3fbf6e9a7725cdb6f85008f76514d72c80fbac8271de810cb74b
Status: Downloaded newer image for gcr.io/deeplearning-platform-release/tf2-cpu.2-6:latest
---> 1b06aec32ef3
Step 2/4 : WORKDIR /
---> Running in e40e4e2f58ec
Removing intermediate container e40e4e2f58ec
---> 6faa502ef2b3
Step 3/4 : COPY trainer /trainer
---> dfa48affe75f
Step 4/4 : ENTRYPOINT ["python", "-m", "trainer.train"]
---> Running in 8ae68598d117
Removing intermediate container 8ae68598d117
---> 4be042358246
Successfully built 4be042358246
Successfully tagged gcr.io/haytsack1/mpg:v1
(base) jupyter@tensorflow-2-3-20211204-093517:~/mpg$
Simple
Terminal 1

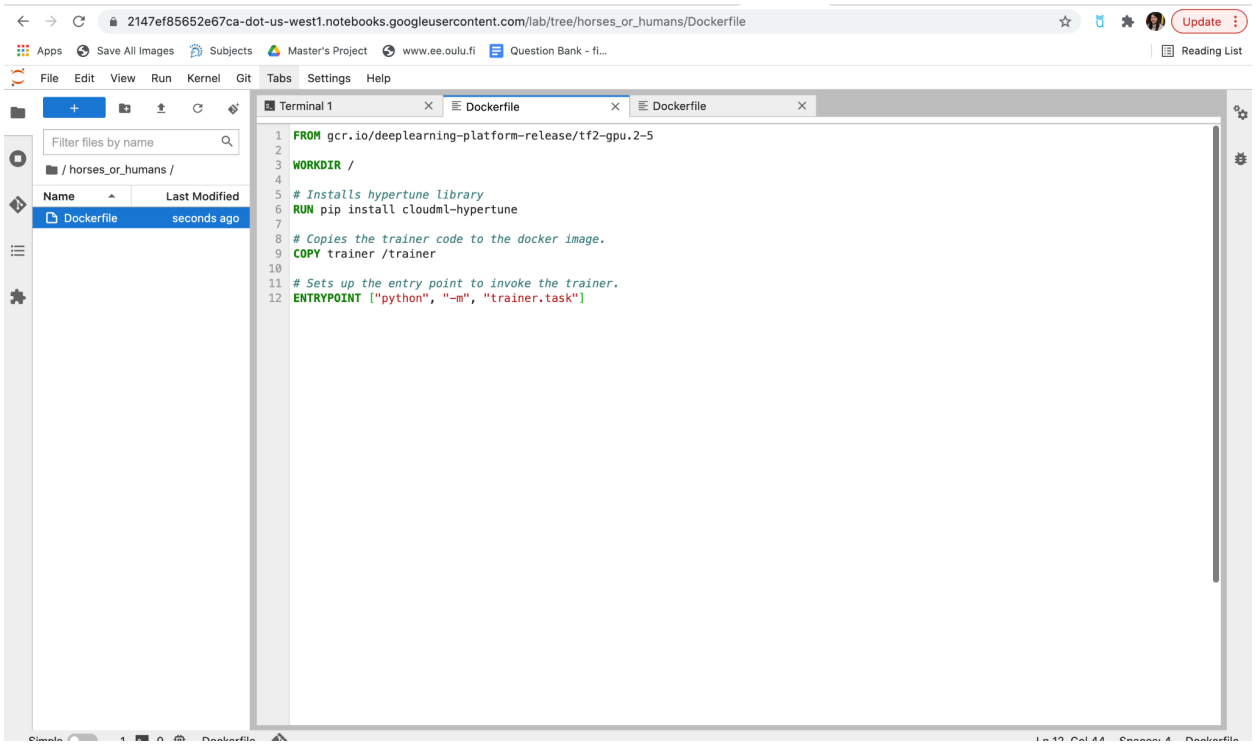


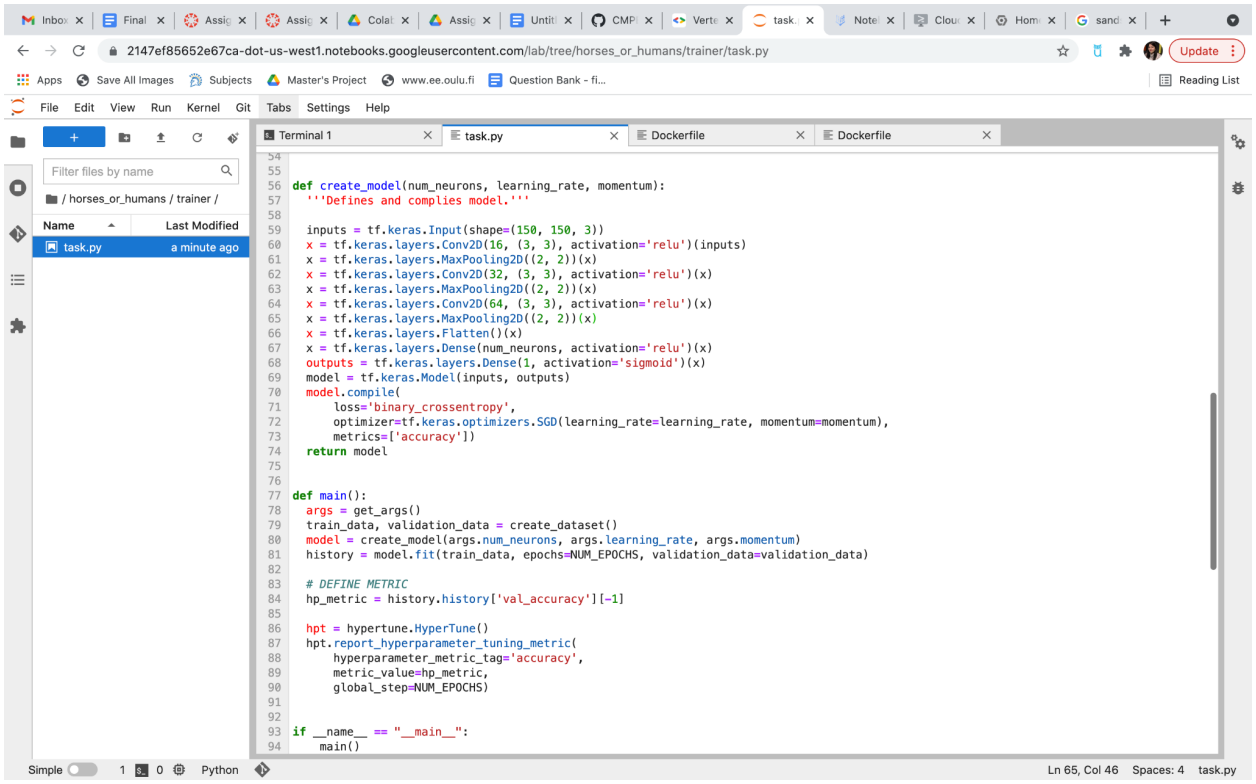
Part d) https://codelabs.developers.google.com/vertex-workbench-intro#0

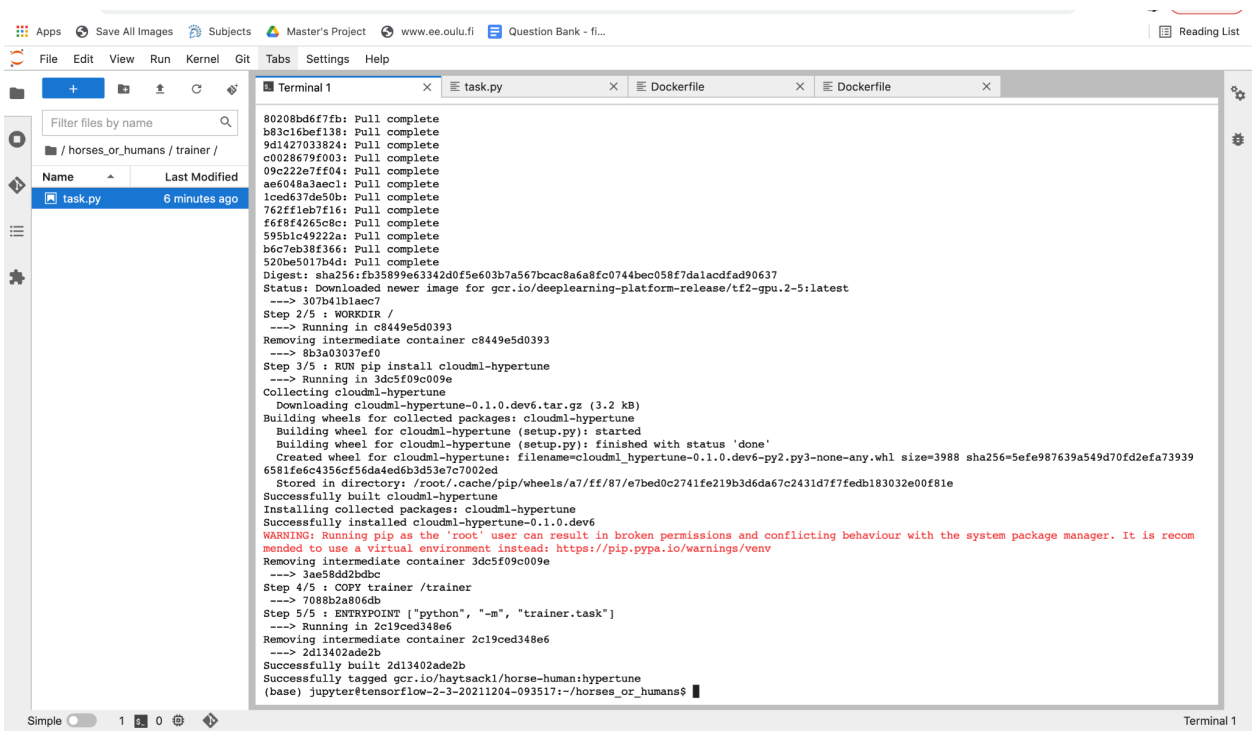




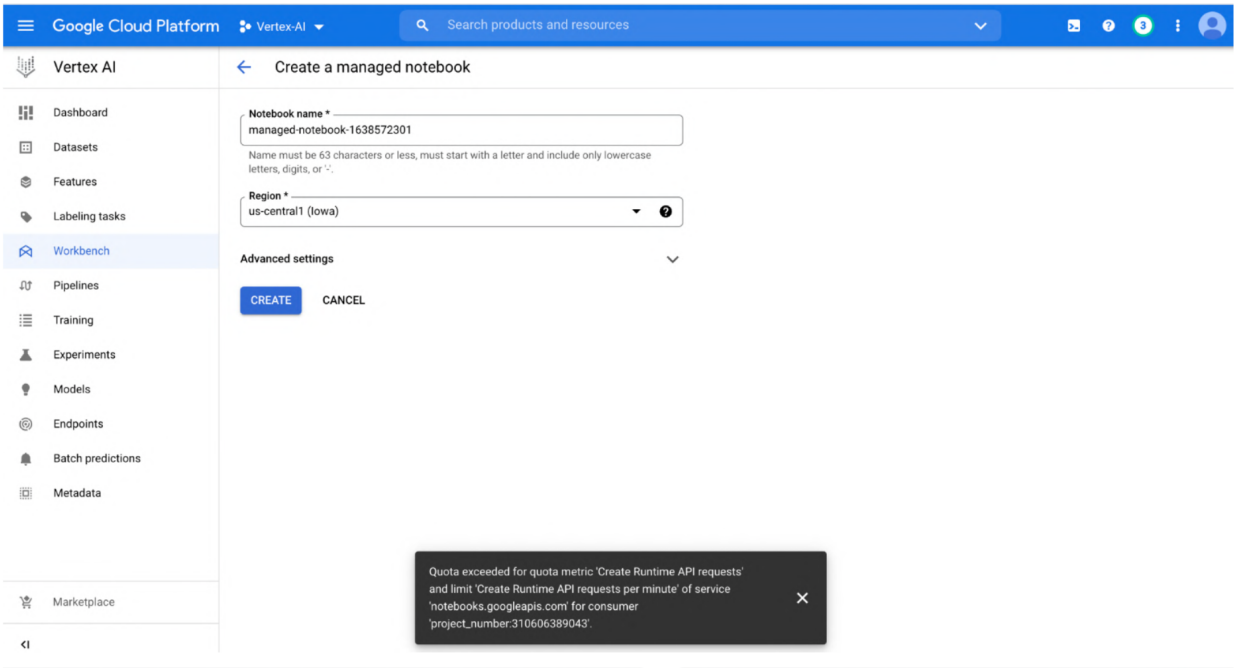
Part f)https://codelabs.developers.google.com/vertex_hyperparameter_tuning#0

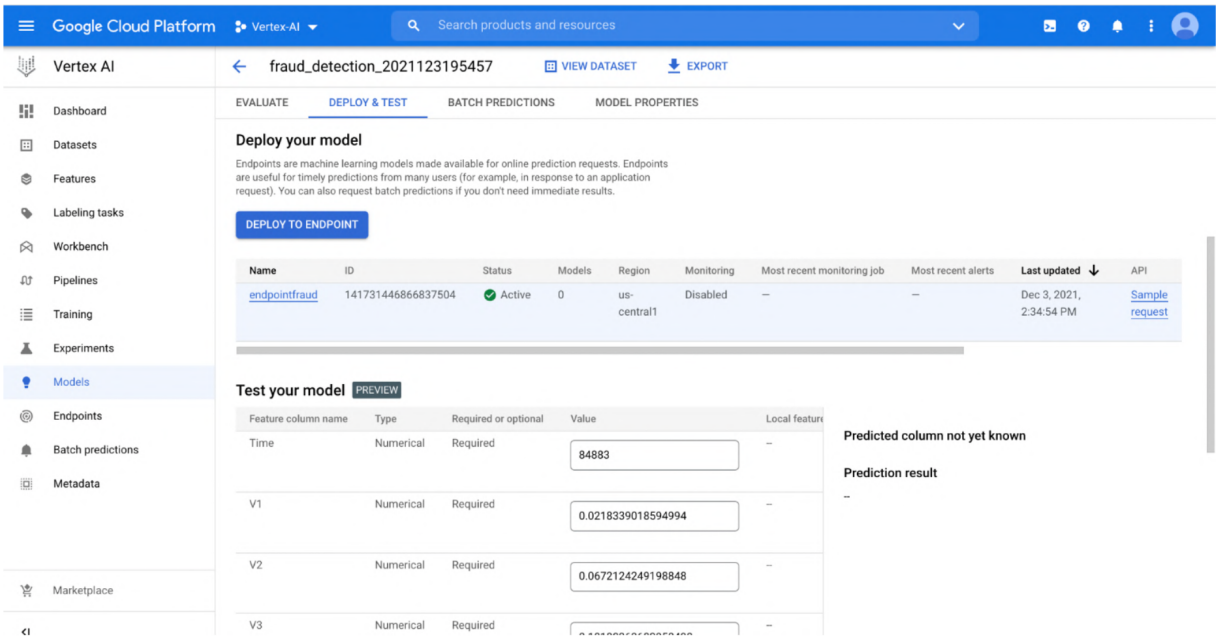




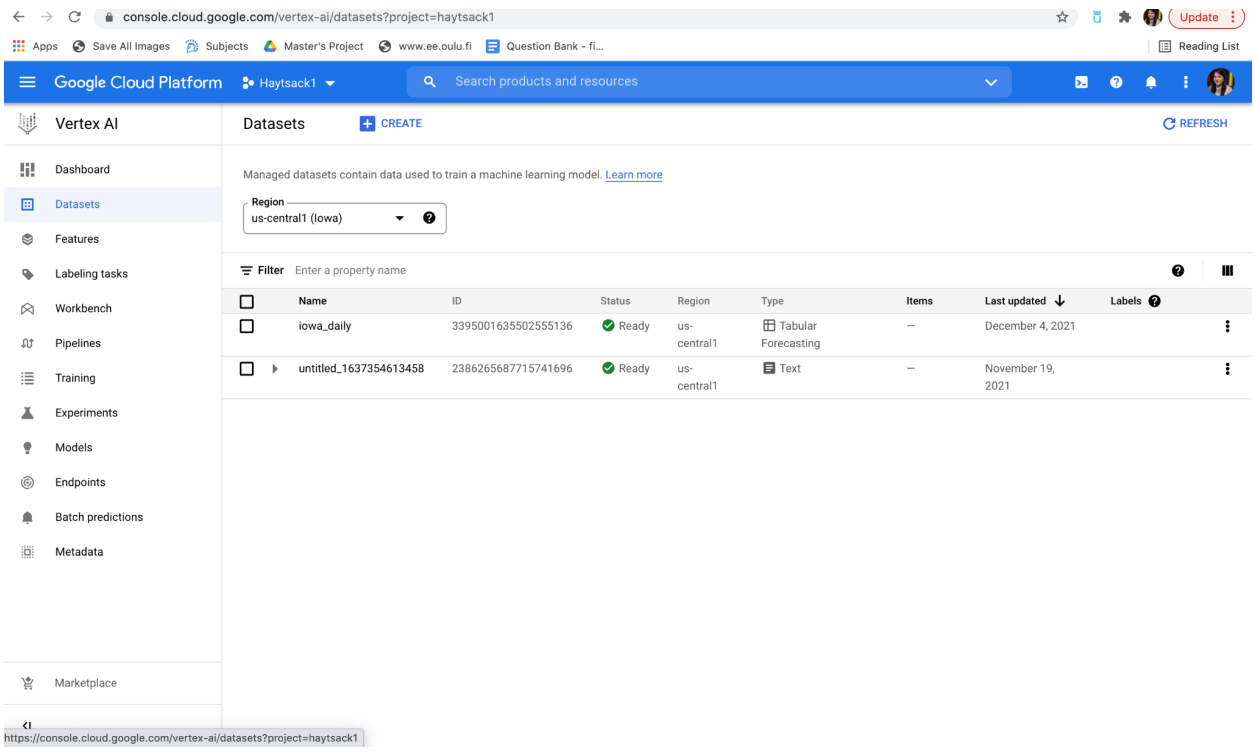


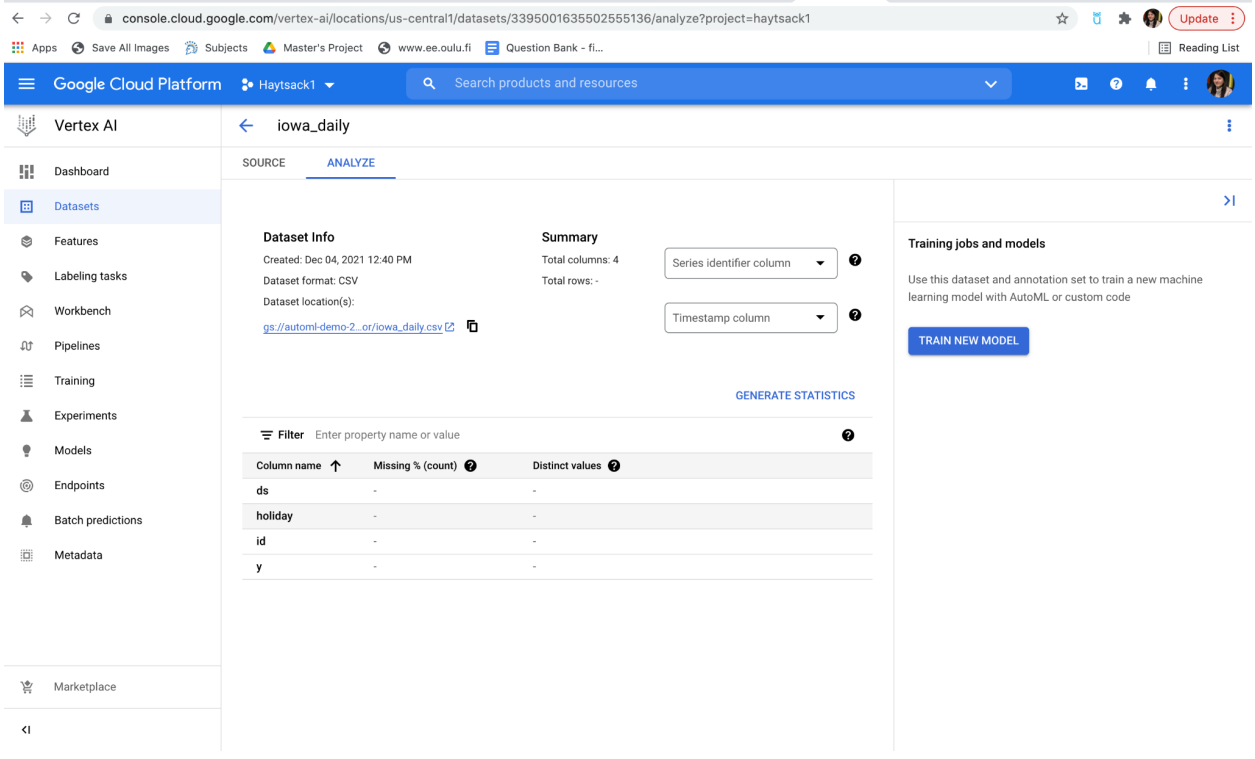
Part g) https://codelabs.developers.google.com/vertex-mlmd-pipelines#0

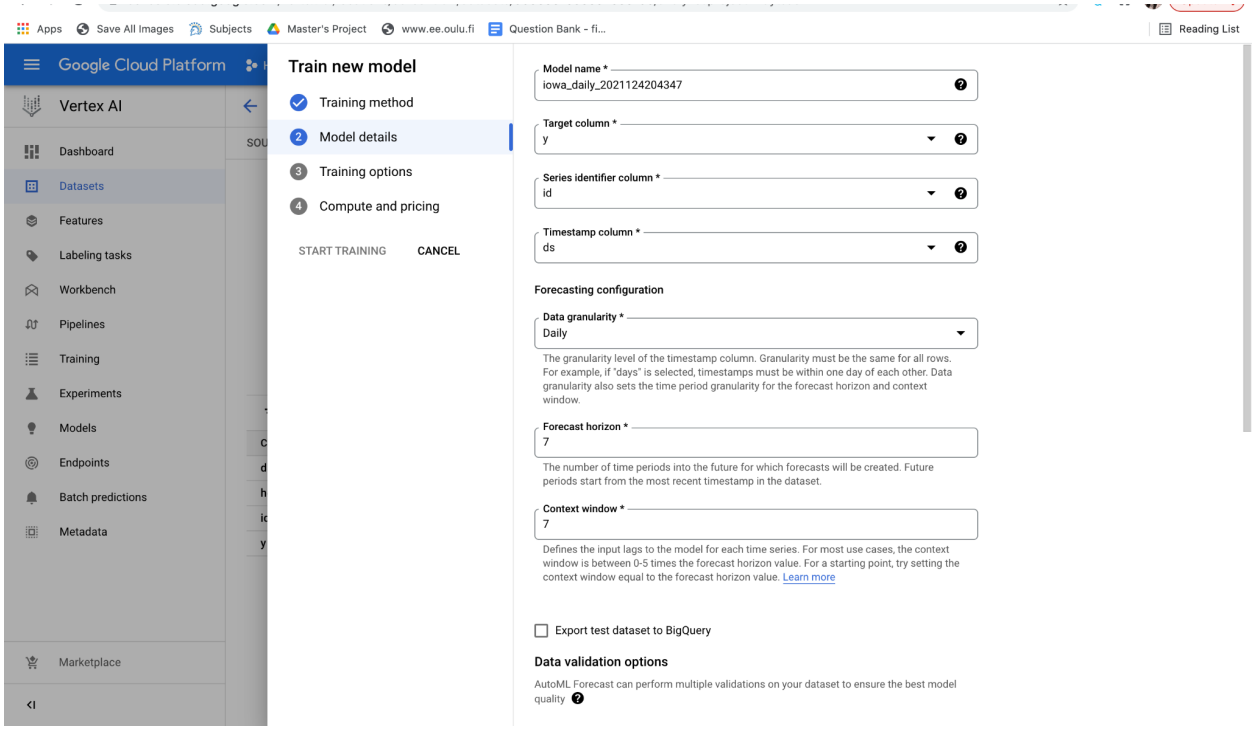


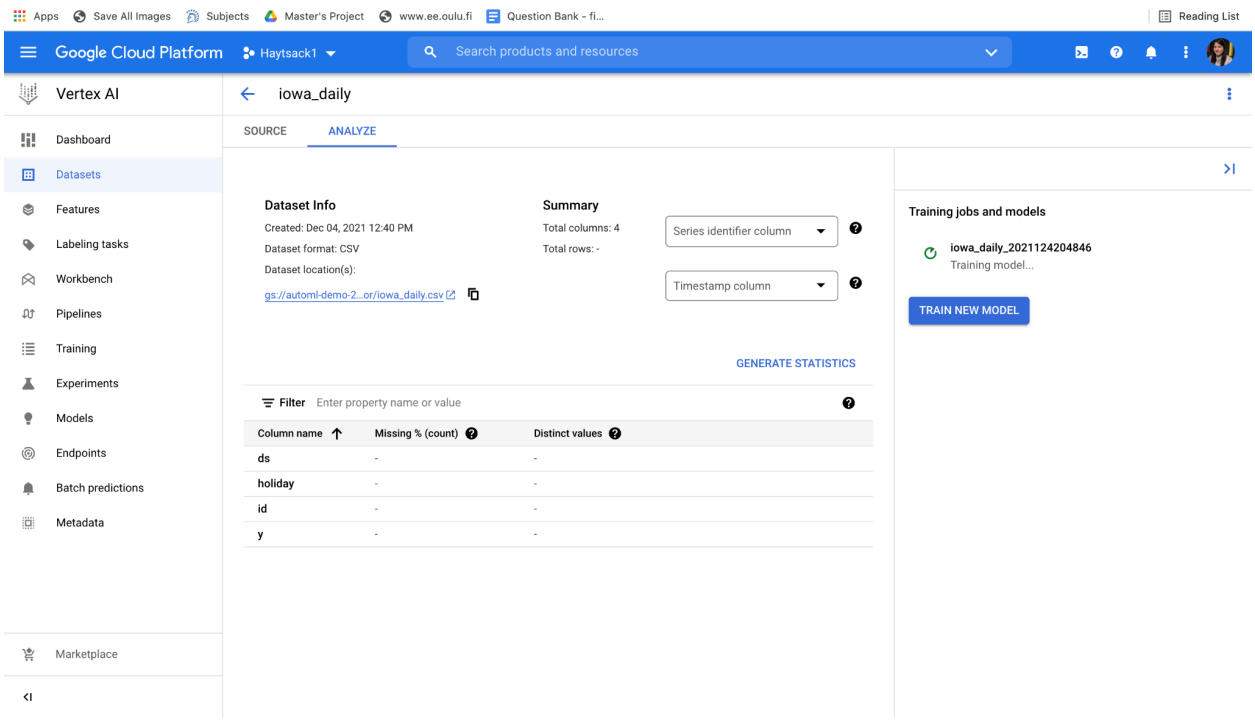


Part h)
https://codelabs.developers.google.com/codelabs/automl-forecasting-with-vertex-ai#0

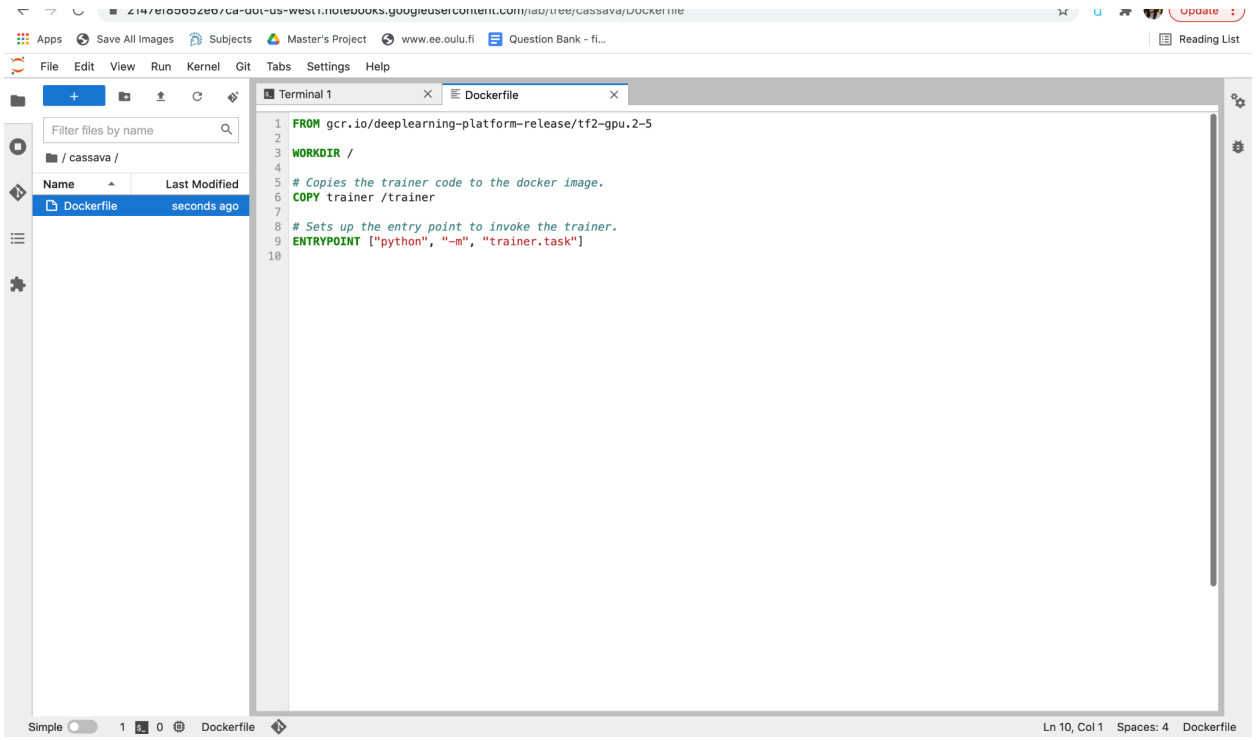

console.cloud.google.com/vertex-ai/datasets?project=haytsack1
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
Search products and resources
Vertex AI
Datasets
CREATE
REFRESH
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
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
iowa_daily
3395001635502555136
Ready
us-central1
Tabular Forecasting
–
December 4, 2021
untitled_1637354613458
2386265687715741696
Ready
us-central1
Text
–
November 19, 2021
https://console.cloud.google.com/vertex-ai/datasets?project=haytsack1

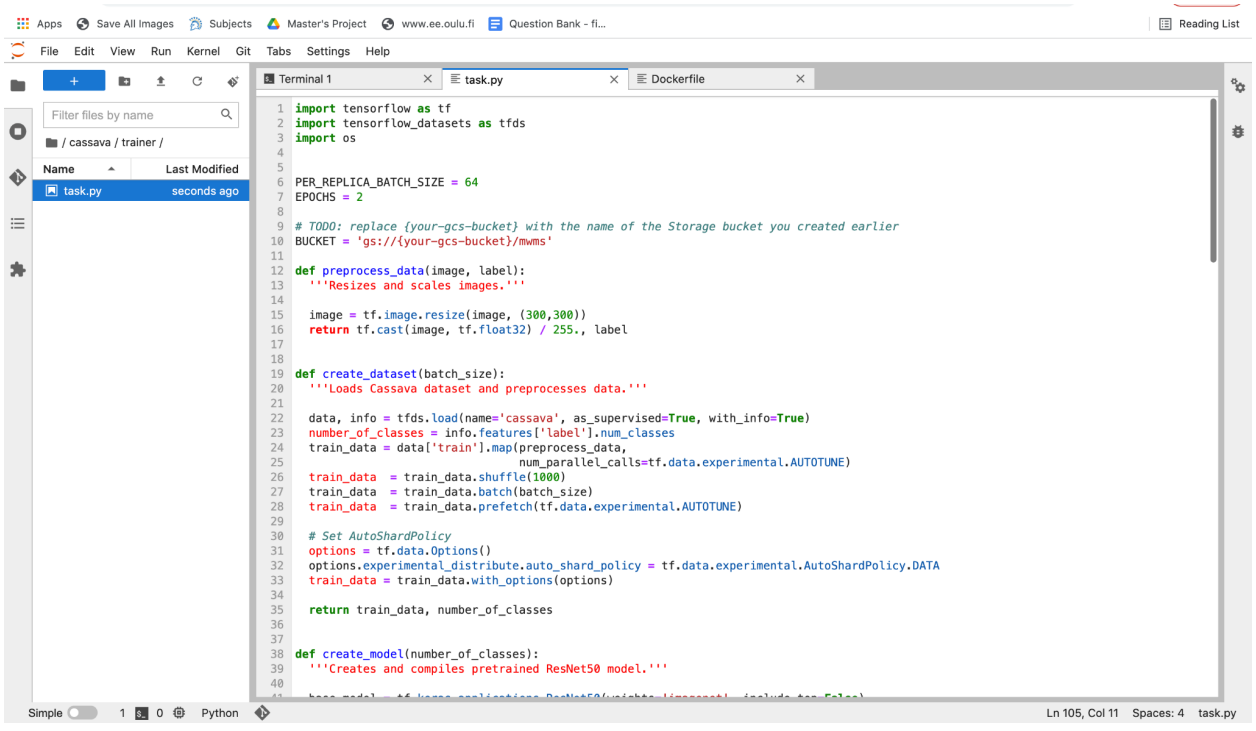

console.cloud.google.com/vertex-ai/locations/us-central1/datasets/3395001635502555136/analyze?project=haytsack1
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
Search products and resources
Vertex AI
iowa_daily
SOURCE
ANALYZE
Dashboard
Datasets
Features
Labeling tasks
Workbench
Pipelines
Training
Experiments
Models
Endpoints
Batch predictions
Metadata
Marketplace
Dataset Info
Created: Dec 04, 2021 12:40 PM
Dataset format: CSV
Dataset location(s):
gs://automl-demo-2...or/iowa_daily.csv
Summary
Total columns: 4
Total rows: -
Series identifier column
Timestamp column
GENERATE STATISTICS
Filter
Enter property name or value
Column name
Missing % (count)
Distinct values
ds
holiday
id
y
Training jobs and models
Use this dataset and annotation set to train a new machine learning model with AutoML or custom code
TRAIN NEW MODEL

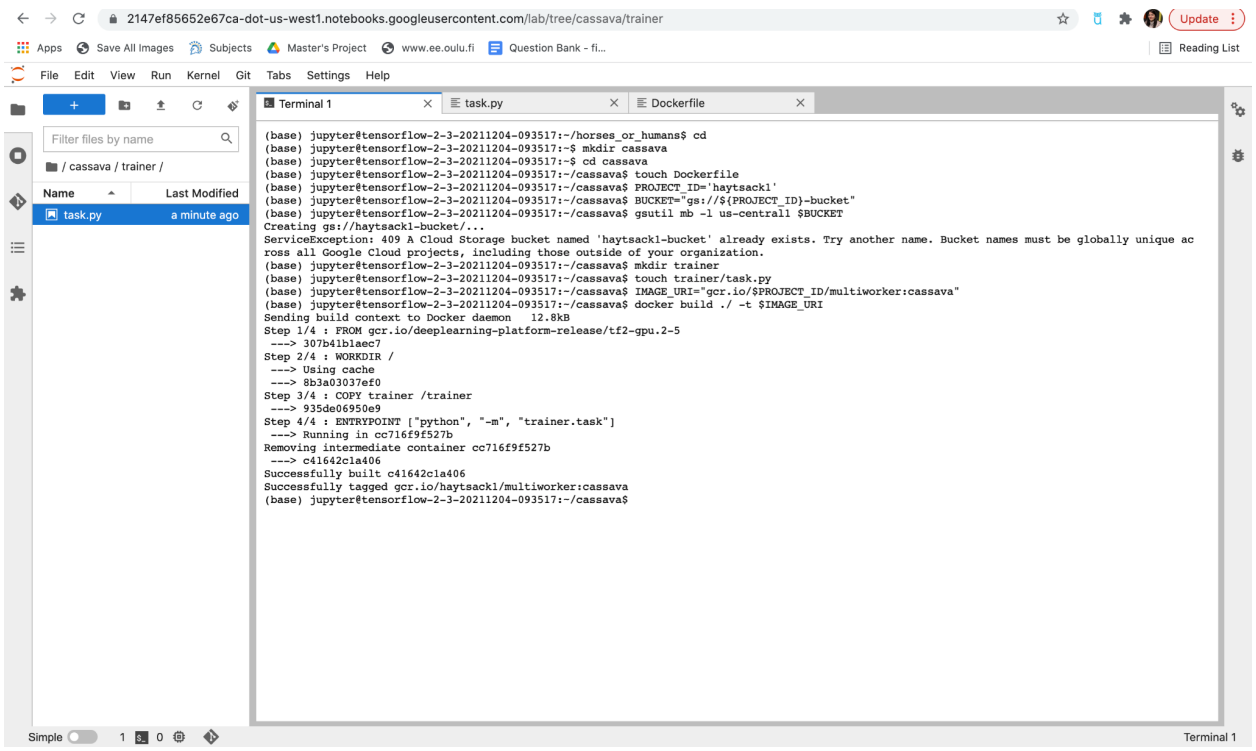


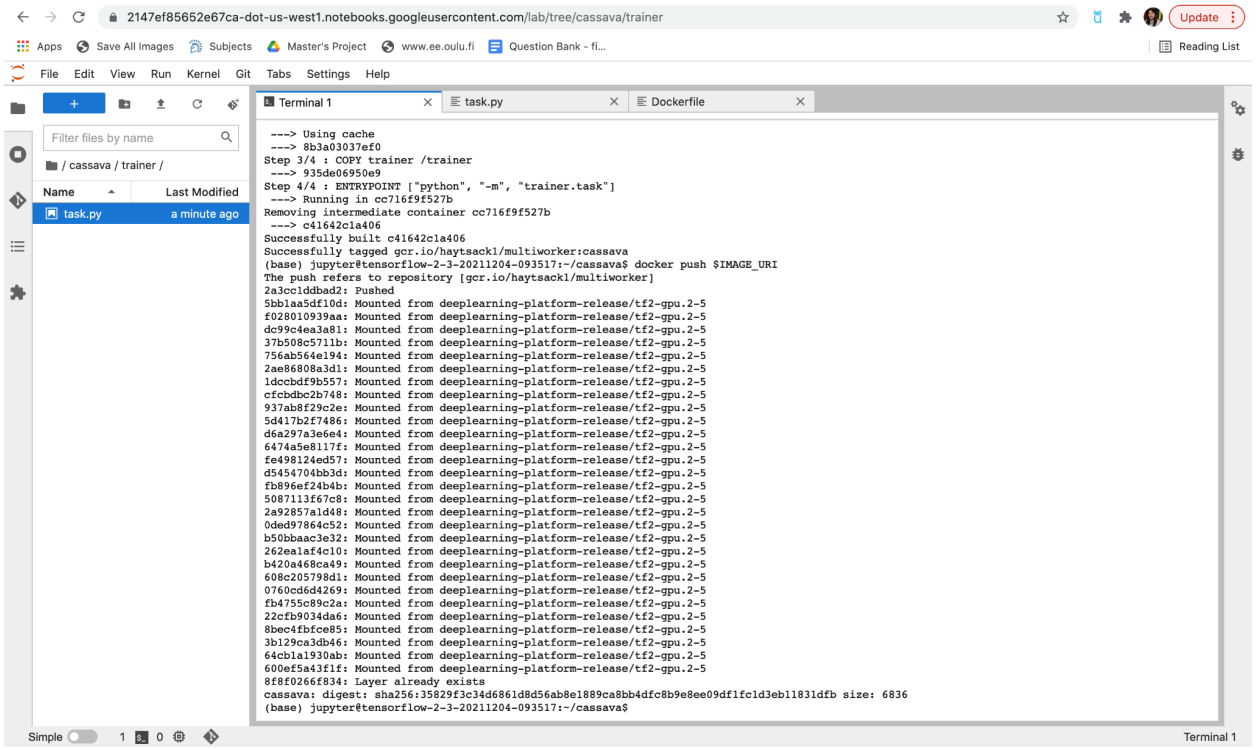


Part i) https://codelabs.developers.google.com/vertex_multiworker_training#0

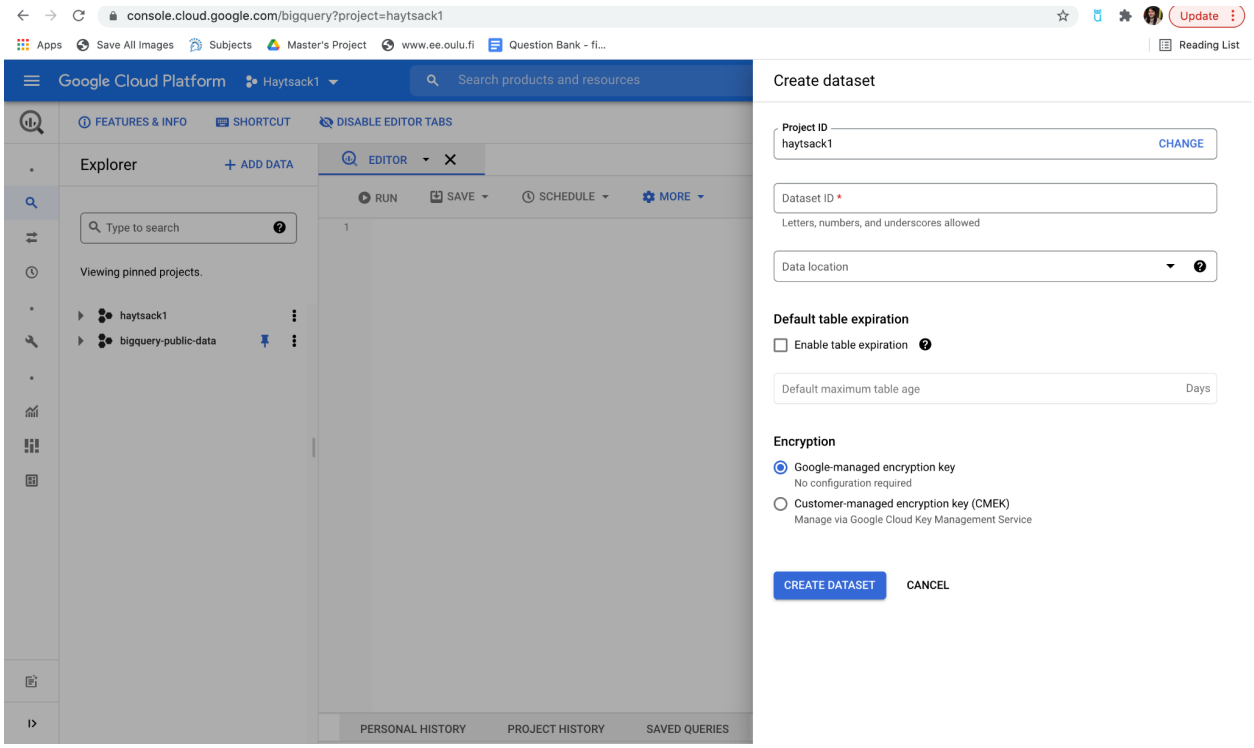

Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
Dockerfile
Filter files by name
/ cassava /
Name
Last Modified
Dockerfile
seconds ago
1 FROM gcr.io/deeplearning-platform-release/tf2-gpu.2-5
2
3 WORKDIR /
4
5 # Copies the trainer code to the docker image.
6 COPY trainer /trainer
7
8 # Sets up the entry point to invoke the trainer.
9 ENTRYPOINT ["python", "-m", "trainer.task"]
10
Simple
1
0
Dockerfile
Ln 10, Col 1
Spaces: 4
Dockerfile

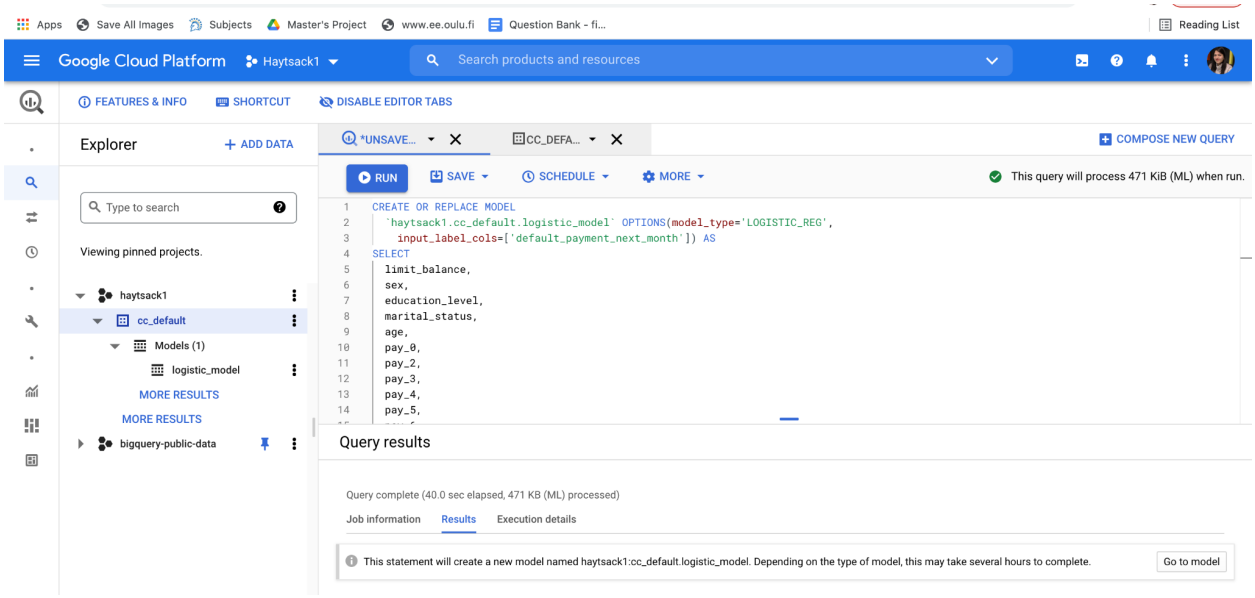

Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Terminal 1
task.py
Dockerfile
Filter files by name
/ cassava / trainer /
Name
Last Modified
task.py
seconds ago
1 import tensorflow as tf
2 import tensorflow_datasets as tfds
3 import os
4
5
6 PER_REPLICA_BATCH_SIZE = 64
7 EPOCHS = 2
8
9 # TODO: replace {your-gcs-bucket} with the name of the Storage bucket you created earlier
10 BUCKET = 'gs://{your-gcs-bucket}/mwms'
11
12 def preprocess_data(image, label):
13 '''Resizes and scales images.'''
14
15 image = tf.image.resize(image, (300,300))
16 return tf.cast(image, tf.float32) / 255., label
17
18
19 def create_dataset(batch_size):
20 '''Loads Cassava dataset and preprocesses data.'''
21
22 data, info = tfds.load(name='cassava', as_supervised=True, with_info=True)
23 number_of_classes = info.features['label'].num_classes
24 train_data = data['train'].map(preprocess_data,
25 num_parallel_calls=tf.data.experimental.AUTOTUNE)
26 train_data = train_data.shuffle(1000)
27 train_data = train_data.batch(batch_size)
28 train_data = train_data.prefetch(tf.data.experimental.AUTOTUNE)
29
30 # Set AutoShardPolicy
31 options = tf.data.Options()
32 options.experimental_distribute.auto_shard_policy = tf.data.experimental.AutoShardPolicy.DATA
33 train_data = train_data.with_options(options)
34
35 return train_data, number_of_classes
36
37
38 def create_model(number_of_classes):
39 '''Creates and compiles pretrained ResNet50 model.'''
40
Simple
1
0
Python
Ln 105, Col 11
Spaces: 4
task.py

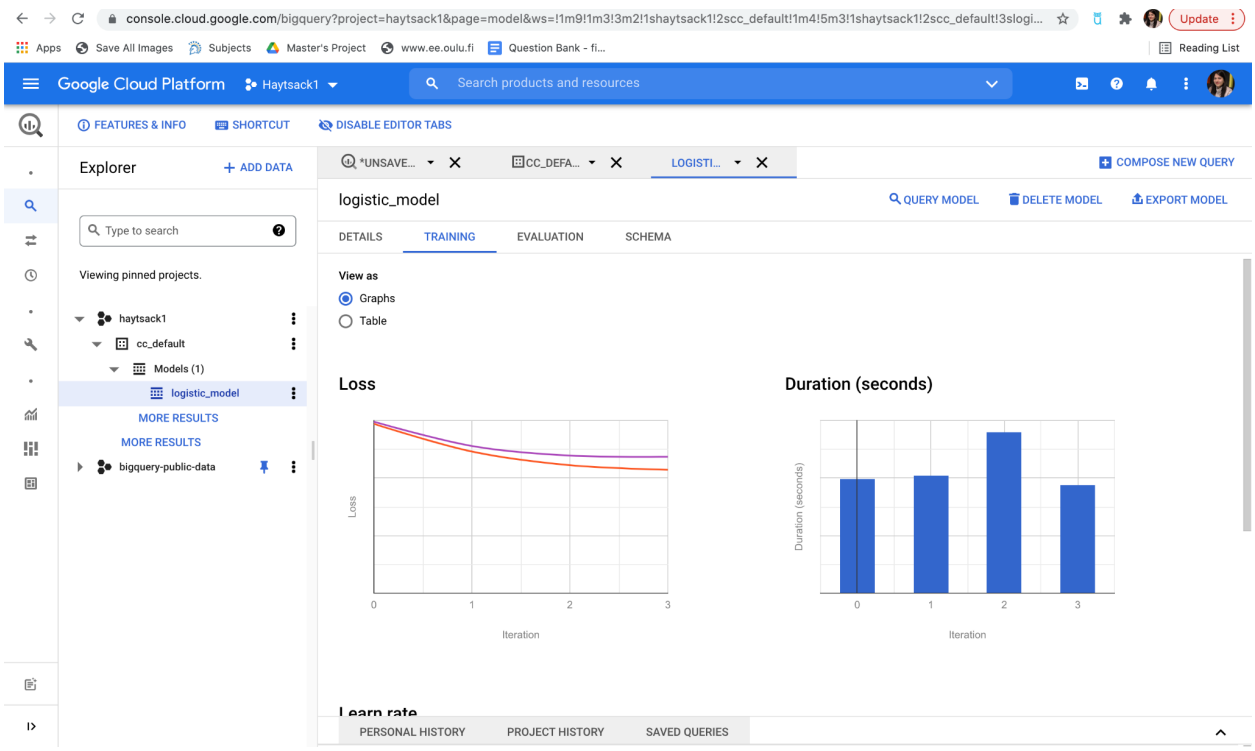




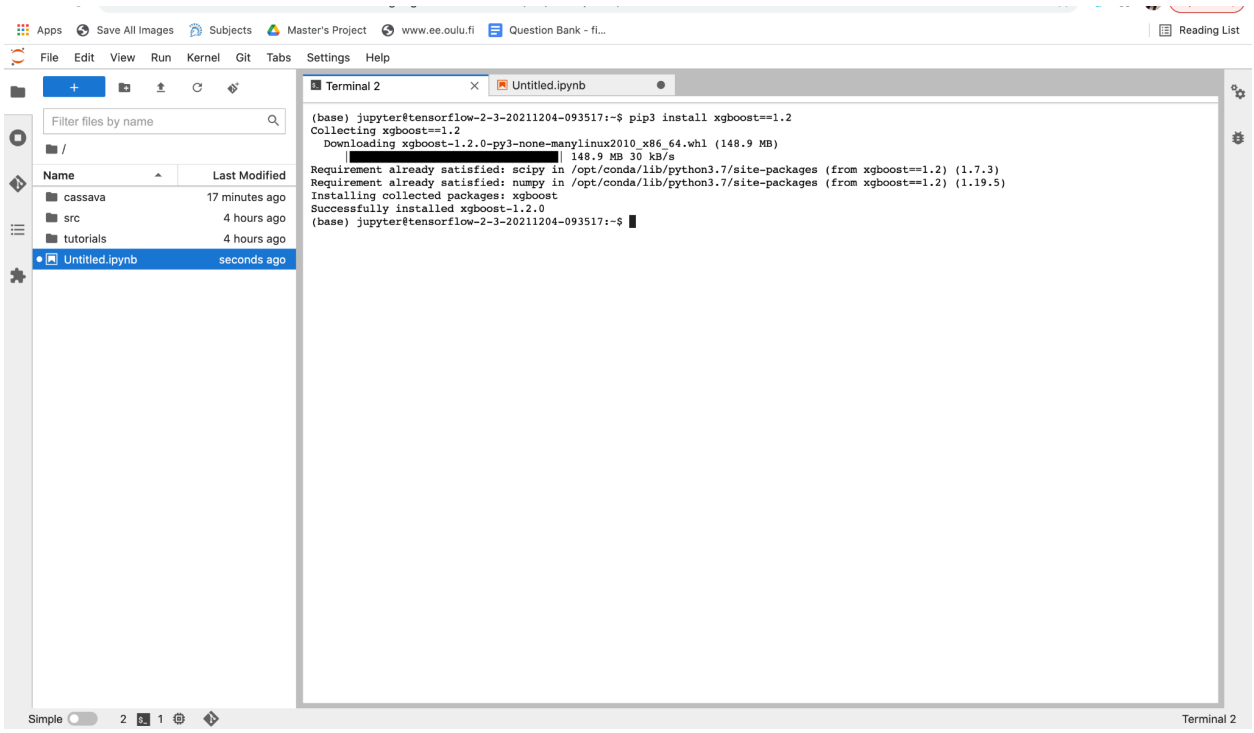
Part j) https://codelabs.developers.google.com/codelabs/bqml-vertex-prediction#0

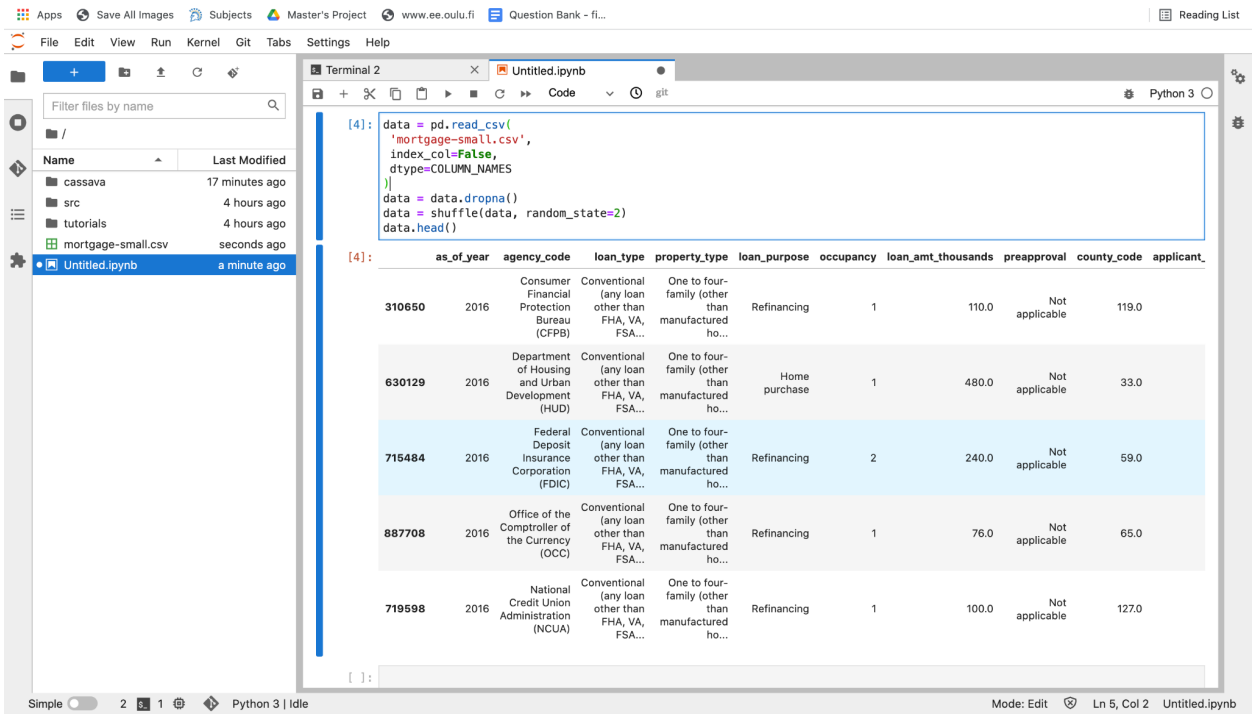

console.cloud.google.com/bigquery?project=haytsack1
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
Search products and resources
FEATURES & INFO
SHORTCUT
DISABLE EDITOR TABS
Explorer
ADD DATA
EDITOR
RUN
SAVE
SCHEDULE
MORE
Type to search
Viewing pinned projects.
haytsack1
bigquery-public-data
PERSONAL HISTORY
PROJECT HISTORY
SAVED QUERIES
Create dataset
Project ID
haytsack1
CHANGE
Dataset ID *
Letters, numbers, and underscores allowed
Data location
Default table expiration
Enable table expiration
Default maximum table age
Days
Encryption
Google-managed encryption key
No configuration required
Customer-managed encryption key (CMEK)
Manage via Google Cloud Key Management Service
CREATE DATASET
CANCEL

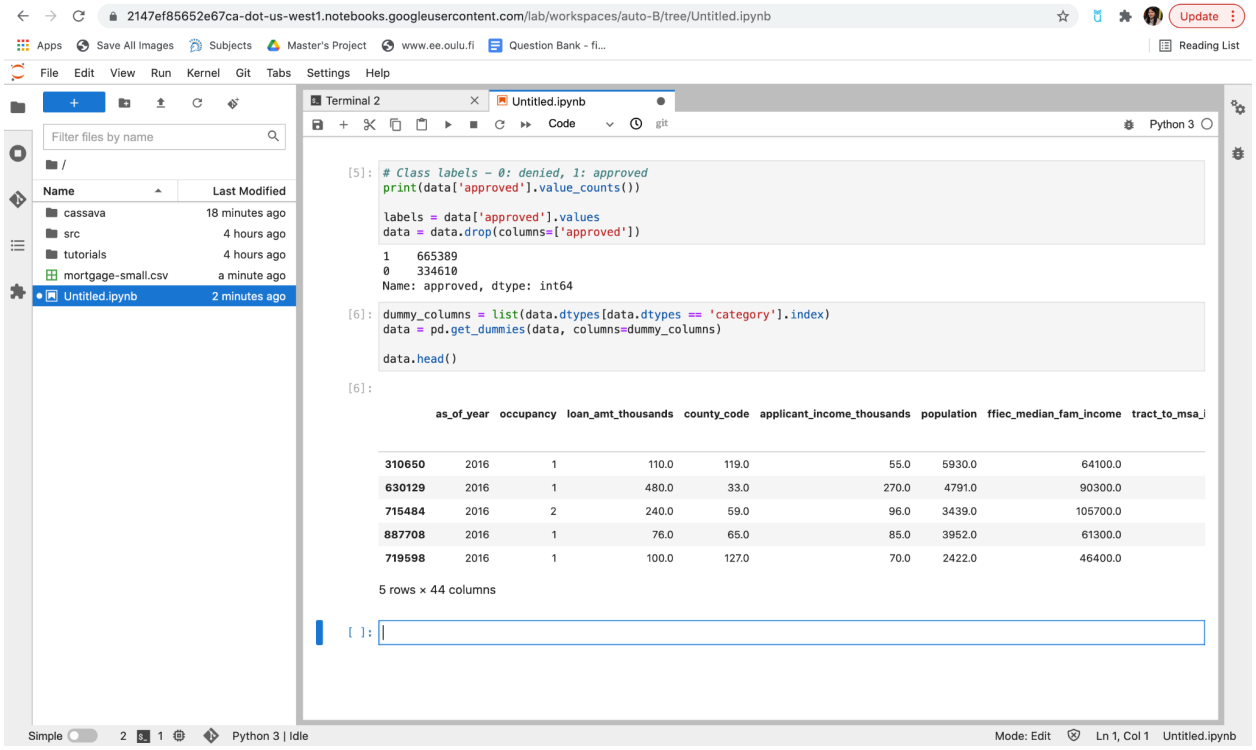

Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
Google Cloud Platform
Haytsack1
Search products and resources
FEATURES & INFO
SHORTCUT
DISABLE EDITOR TABS
Explorer
ADD DATA
*UNSAVE...
CC_DEFA...
COMPOSE NEW QUERY
RUN
SAVE
SCHEDULE
MORE
This query will process 471 KiB (ML) when run.
Type to search
Viewing pinned projects.
haytsack1
cc_default
Models (1)
logistic_model
MORE RESULTS
MORE RESULTS
bigquery-public-data
CREATE OR REPLACE MODEL
`haytsack1.cc_default.logistic_model` OPTIONS(model_type='LOGISTIC_REG',
input_label_cols=['default_payment_next_month']) AS
SELECT
limit_balance,
sex,
education_level,
marital_status,
age,
pay_0,
pay_2,
pay_3,
pay_4,
pay_5,
Query results
Query complete (40.0 sec elapsed, 471 KB (ML) processed)
Job information
Results
Execution details
This statement will create a new model named haytsack1:cc_default.logistic_model. Depending on the type of model, this may take several hours to complete.
Go to model

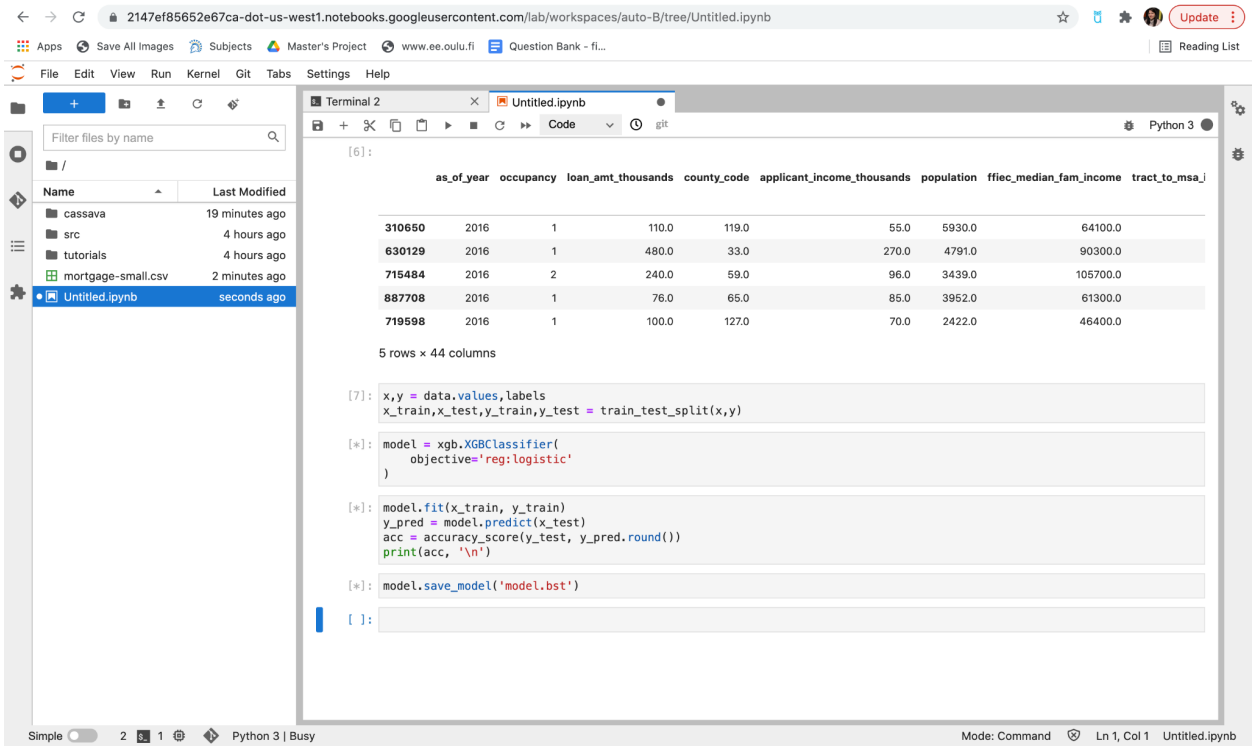


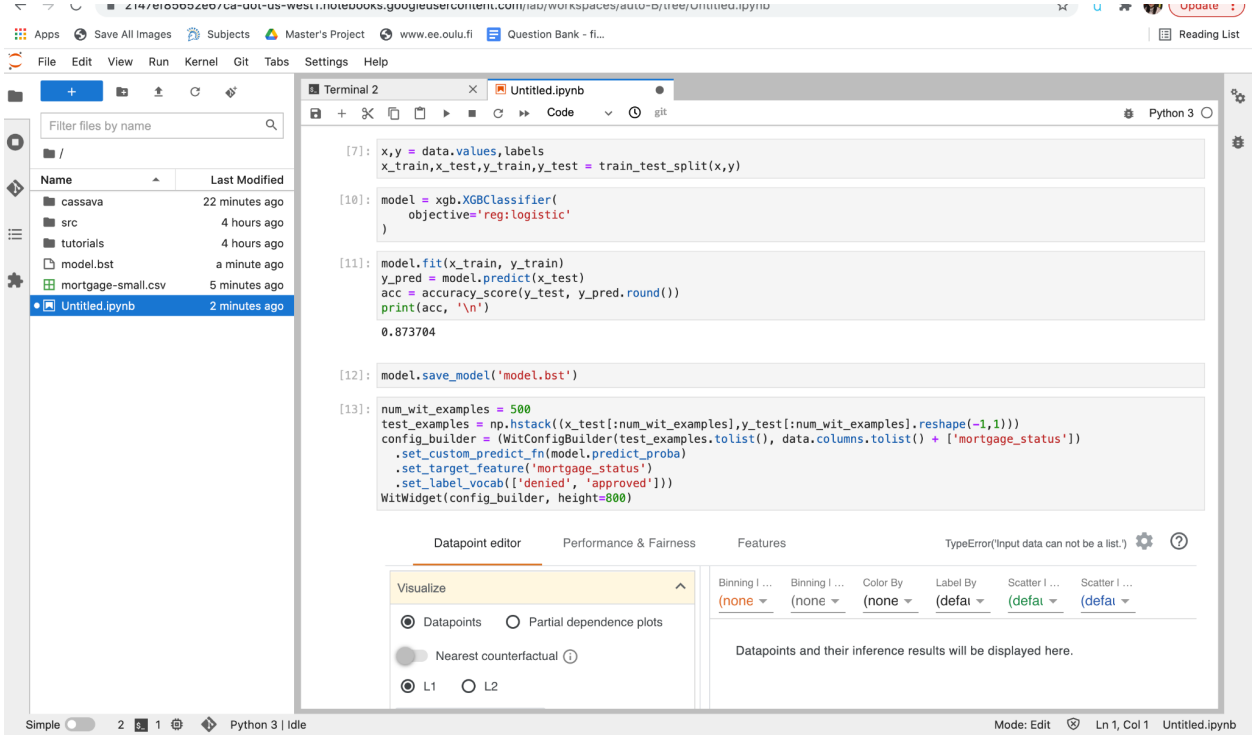
Part k) https://codelabs.developers.google.com/vertex-xgb-wit#0



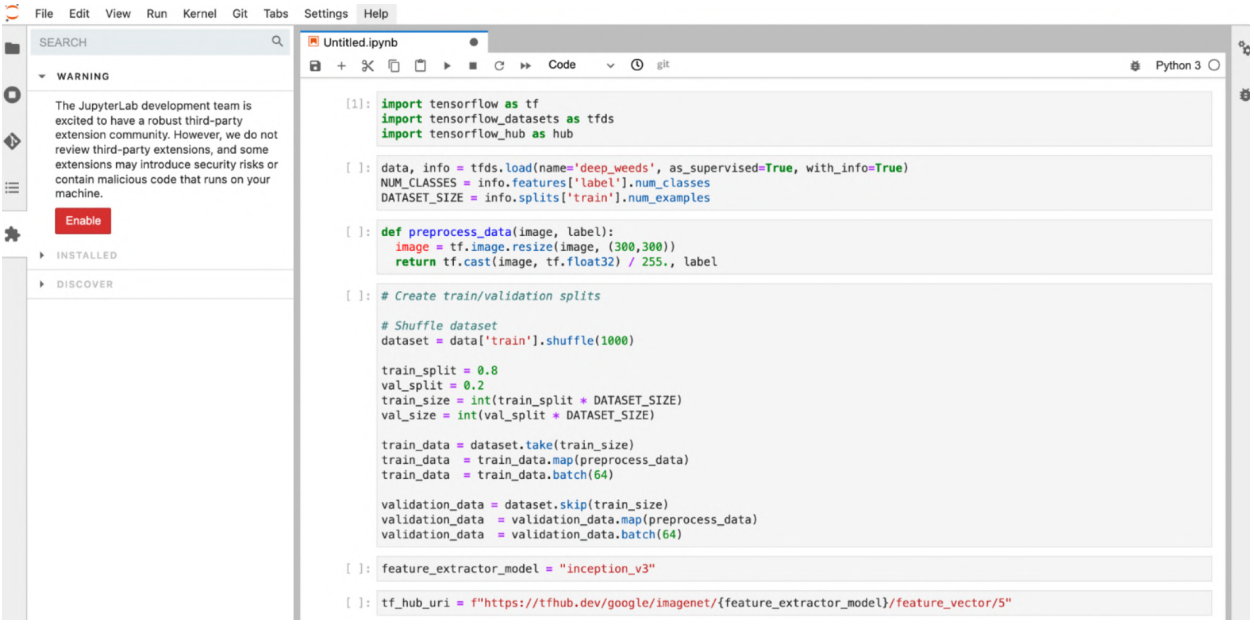

Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Terminal 2
Untitled.ipynb
Code
Python 3
Filter files by name
/
Name
Last Modified
cassava 17 minutes ago
src 4 hours ago
tutorials 4 hours ago
mortgage-small.csv seconds ago
Untitled.ipynb a minute ago
[4]:
data = pd.read_csv(
'mortgage-small.csv',
index_col=False,
dtype=COLUMN_NAMES
)
data = data.dropna()
data = shuffle(data, random_state=2)
data.head()
[4]:
as_of_year agency_code loan_type property_type loan_purpose occupancy loan_amt_thousands preapproval county_code applicant_
310650 2016 Consumer Financial Protection Bureau (CFPB) Conventional (any loan other than FHA, VA, FSA... One to four-family (other than manufactured ho... Refinancing 1 110.0 Not applicable 119.0
630129 2016 Department of Housing and Urban Development (HUD) Conventional (any loan other than FHA, VA, FSA... One to four-family (other than manufactured ho... Home purchase 1 480.0 Not applicable 33.0
715484 2016 Federal Deposit Insurance Corporation (FDIC) Conventional (any loan other than FHA, VA, FSA... One to four-family (other than manufactured ho... Refinancing 2 240.0 Not applicable 59.0
887708 2016 Office of the Comptroller of the Currency (OCC) Conventional (any loan other than FHA, VA, FSA... One to four-family (other than manufactured ho... Refinancing 1 76.0 Not applicable 65.0
719598 2016 National Credit Union Administration (NCUA) Conventional (any loan other than FHA, VA, FSA... One to four-family (other than manufactured ho... Refinancing 1 100.0 Not applicable 127.0
[ ]:
Simple 2 1 Python 3 | Idle
Mode: Edit Ln 5, Col 2 Untitled.ipynb

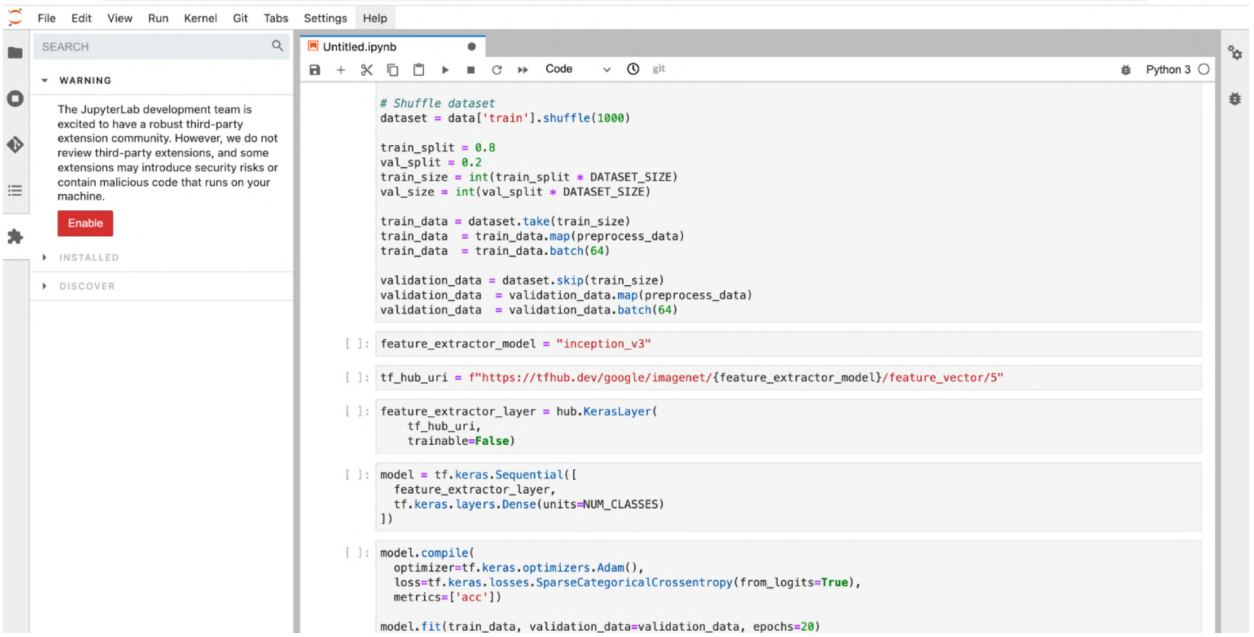

2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/workspaces/auto-B/tree/Untitled.ipynb
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Terminal 2
Untitled.ipynb
Code
Python 3
Filter files by name
/
Name
Last Modified
cassava 18 minutes ago
src 4 hours ago
tutorials 4 hours ago
mortgage-small.csv a minute ago
Untitled.ipynb 2 minutes ago
[5]:
# Class labels – 0: denied, 1: approved
print(data['approved'].value_counts())
labels = data['approved'].values
data = data.drop(columns=['approved'])
1 665389
0 334610
Name: approved, dtype: int64
[6]:
dummy_columns = list(data.dtypes[data.dtypes == 'category'].index)
data = pd.get_dummies(data, columns=dummy_columns)
data.head()
[6]:
as_of_year occupancy loan_amt_thousands county_code applicant_income_thousands population ffiec_median_fam_income tract_to_msa_i
310650 2016 1 110.0 119.0 55.0 5930.0 64100.0
630129 2016 1 480.0 33.0 270.0 4791.0 90300.0
715484 2016 2 240.0 59.0 96.0 3439.0 105700.0
887708 2016 1 76.0 65.0 85.0 3952.0 61300.0
719598 2016 1 100.0 127.0 70.0 2422.0 46400.0
5 rows × 44 columns
[ ]:
Simple 2 1 Python 3 | Idle
Mode: Edit Ln 1, Col 1 Untitled.ipynb


2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/workspaces/auto-B/tree/Untitled.ipynb
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/
Name
Last Modified
cassava 19 minutes ago
src 4 hours ago
tutorials 4 hours ago
mortgage-small.csv 2 minutes ago
Untitled.ipynb seconds ago
Terminal 2
Untitled.ipynb
Code
git
Python 3
[6]:
as_of_year occupancy loan_amt_thousands county_code applicant_income_thousands population ffiec_median_fam_income tract_to_msa_i
310650 2016 1 110.0 119.0 55.0 5930.0 64100.0
630129 2016 1 480.0 33.0 270.0 4791.0 90300.0
715484 2016 2 240.0 59.0 96.0 3439.0 105700.0
887708 2016 1 76.0 65.0 85.0 3952.0 61300.0
719598 2016 1 100.0 127.0 70.0 2422.0 46400.0
5 rows × 44 columns
[7]:
x,y = data.values,labels
x_train,x_test,y_train,y_test = train_test_split(x,y)
[*]:
model = xgb.XGBClassifier(
objective='reg:logistic'
)
[*]:
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
acc = accuracy_score(y_test, y_pred.round())
print(acc, '\n')
[*]:
model.save_model('model.bst')
[ ]:
Simple
2 1
Python 3 | Busy
Mode: Command
Ln 1, Col 1
Untitled.ipynb


Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
File Edit View Run Kernel Git Tabs Settings Help
Filter files by name
/
Name
Last Modified
cassava 22 minutes ago
src 4 hours ago
tutorials 4 hours ago
model.bst a minute ago
mortgage-small.csv 5 minutes ago
Untitled.ipynb 2 minutes ago
Terminal 2
Untitled.ipynb
Code
git
Python 3
[7]:
x,y = data.values,labels
x_train,x_test,y_train,y_test = train_test_split(x,y)
[10]:
model = xgb.XGBClassifier(
objective='reg:logistic'
)
[11]:
model.fit(x_train, y_train)
y_pred = model.predict(x_test)
acc = accuracy_score(y_test, y_pred.round())
print(acc, '\n')
0.873704
[12]:
model.save_model('model.bst')
[13]:
num_wit_examples = 500
test_examples = np.hstack((x_test[:num_wit_examples],y_test[:num_wit_examples].reshape(-1,1)))
config_builder = (WitConfigBuilder(test_examples.tolist(), data.columns.tolist() + ['mortgage_status'])
.set_custom_predict_fn(model.predict_proba)
.set_target_feature('mortgage_status')
.set_label_vocab(['denied', 'approved']))
WitWidget(config_builder, height=800)
Datapoint editor
Performance & Fairness
Features
TypeError('Input data can not be a list.')
Visualize
Datapoints
Partial dependence plots
Nearest counterfactual
L1
L2
Binning I ...
Binning I ...
Color By
Label By
Scatter I ...
Scatter I ...
(none
(none
(none
(defai
(defai
(defai
Datapoints and their inference results will be displayed here.
Simple
2 1
Python 3 | Idle
Mode: Edit
Ln 1, Col 1
Untitled.ipynb

Part I)





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

Apps Save All Images Subjects Master's Project www.ee.oulu.fi Question Bank - fi... Reading List

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

/

Name	Last Modified
src	4 hours ago
training-d...	seconds ago
tutorials	4 hours ago

Terminal 1

```
git clone https://github.com/GoogleCloudPlatform/training-data-analyst
Cloning into 'training-data-analyst'...
remote: Enumerating objects: 54506, done.
remote: Counting objects: 100% (376/376), done.
remote: Compressing objects: 100% (247/247), done.
remote: Total 54506 (delta 147), reused 266 (delta 83), pack-reused 54130
Receiving objects: 100% (54506/54506), 627.26 MiB | 31.69 MiB/s, done.
Resolving deltas: 100% (34524/34524), done.
Checking out files: 100% (12334/12334), done.
(base) jupyter@tensorflow-2-3-20211204-093517:~$
```

Simple 2 0 Terminal 1

Apps Save All Images Subjects Master's Project www.ee.oulu.fi Question Bank - fi... Reading List

File Edit View Run Kernel Git Tabs Settings Help

Terminal 1 | 01-explore.ipynb

Markdown git Python 3

Filter files by name

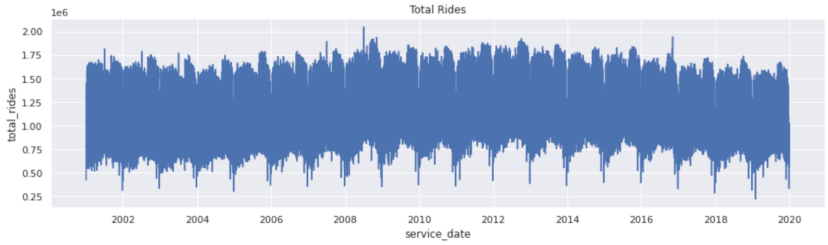
/ ... / ai-for-time-series / notebooks /

Name	Last Modified
data	4 minutes ago
01-explore...	seconds ago
02-model.i...	4 minutes ago
03-cloud-t...	4 minutes ago
cta_riders...	a minute ago

- Is the mix of bus vs rail ridership changing over time?

[9]:
```
# Initialize plotting

register_matplotlib_converters() # Addresses a warning
sns.set(rc={'figure.figsize':(16,4)})
```

[10]:
```
# Explore total rides over time

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



[11]:
```
# 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 2 1 Python 3 | Idle Mode: Command Ln 1, Col 1 01-explore.ipynb

2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb

Update

Apps | Save All Images | Subjects | Master's Project | www.ee.oulu.fi | Question Bank - fi... | Reading List

File Edit View Run Kernel Git Tabs Settings Help

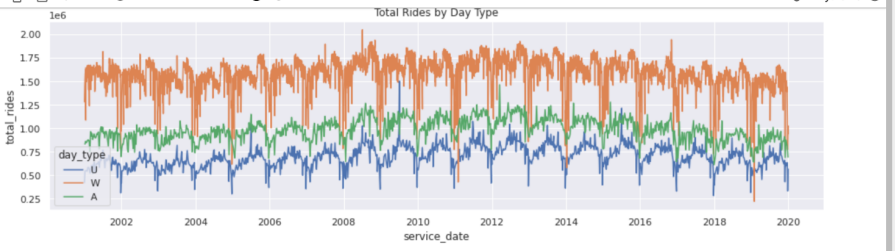
Filter files by name

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

Name	Last Modified
data	4 minutes ago
01-explore...	a minute ago
02-model.i...	4 minutes ago
03-cloud-t...	4 minutes ago
cta_riders...	a minute ago

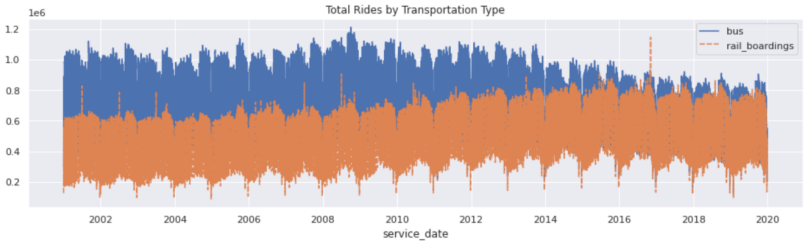
Terminal 1 | 01-explore.ipynb

Markdown | git | Python 3



[12]:

```
# Explore rides by transportation type

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



Simple | 2 | 1 | Python 3 | Idle

Mode: Command | Ln 1, Col 1 | 01-explore.ipynb

2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb

Update

Apps | Save All Images | Subjects | Master's Project | www.ee.oulu.fi | Question Bank - fi... | Reading List

File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

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

Name	Last Modified
data	5 minutes ago
01-explore...	a minute ago
02-model.i...	5 minutes ago
03-cloud-t...	5 minutes ago
cta_riders...	a minute ago

Terminal 1 | 01-explore.ipynb

Markdown | git | Python 3

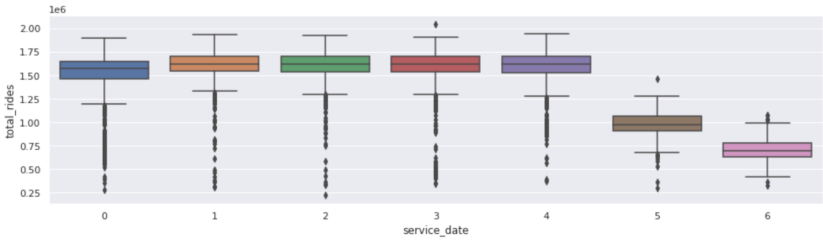
Name: total_rides, dtype: int64

TODO 3: Explore seasonality

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

[14]:

```
# 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])
```

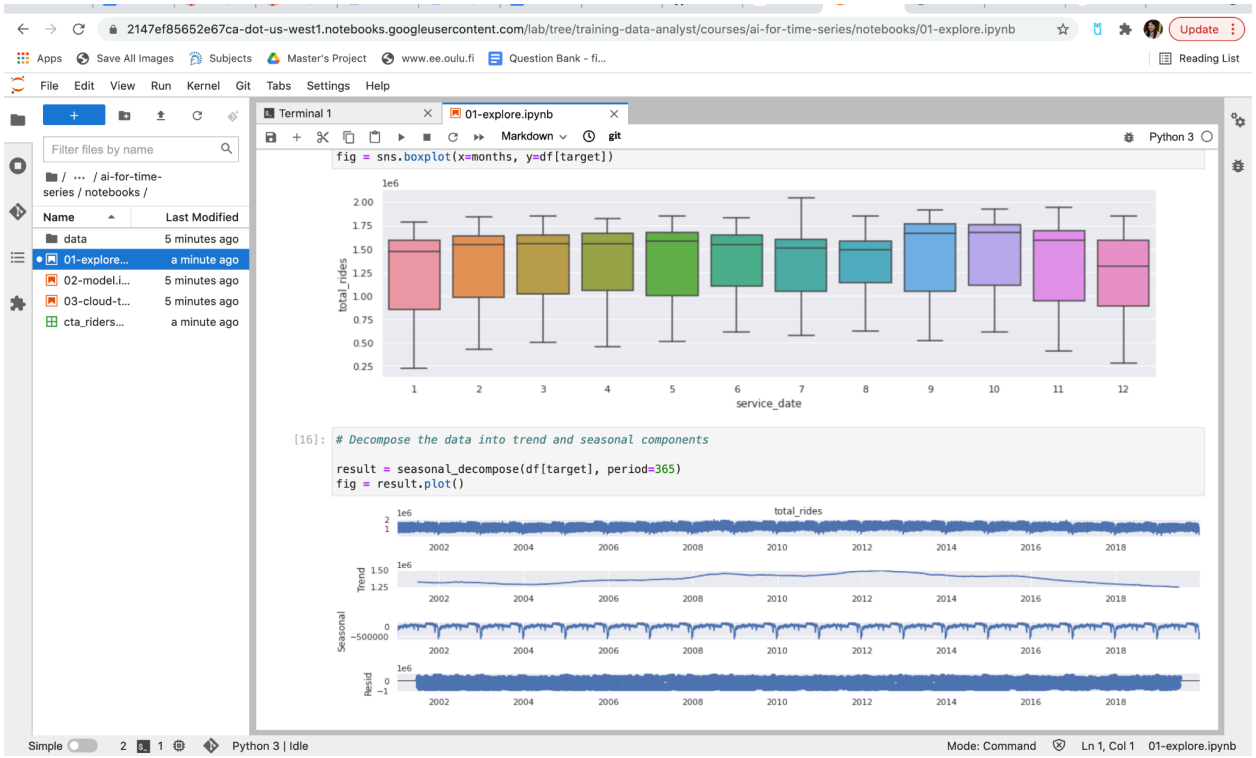


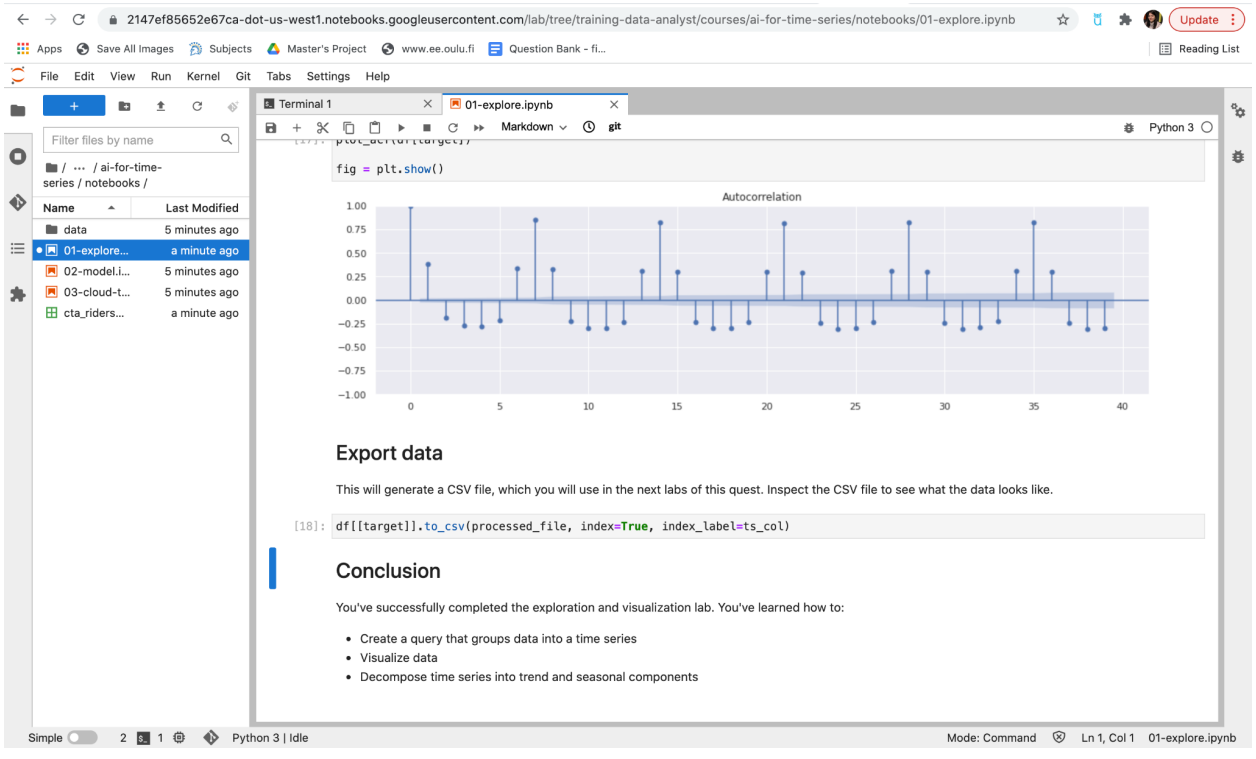
[15]:

```
# Show the distribution of values for each month in a boxplot:

months = df.index.to_series().dt.month

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

Simple | 2 | 1 | Python 3 | Idle

Mode: Command | Ln 1, Col 1 | 01-explore.ipynb


2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
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 5 minutes ago
01-explore... a minute ago
02-model.i... 5 minutes ago
03-cloud-t... 5 minutes ago
cta_riders... a minute ago
fig = sns.boxplot(x=months, y=df[target])
total_rides
service_date
[16]:
# Decompose the data into trend and seasonal components
result = seasonal_decompose(df[target], period=365)
fig = result.plot()
total_rides
Trend
Seasonal
Resid
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
01-explore.ipynb


2147ef85652e67ca-dot-us-west1.notebooks.googleusercontent.com/lab/tree/training-data-analyst/courses/ai-for-time-series/notebooks/01-explore.ipynb
Update
Apps
Save All Images
Subjects
Master's Project
www.ee.oulu.fi
Question Bank - fi...
Reading List
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 5 minutes ago
01-explore... a minute ago
02-model.i... 5 minutes ago
03-cloud-t... 5 minutes ago
cta_riders... a minute ago
fig = plt.show()
Autocorrelation
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.
[18]:
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
Simple
Python 3 | Idle
Mode: Command
Ln 1, Col 1
01-explore.ipynb