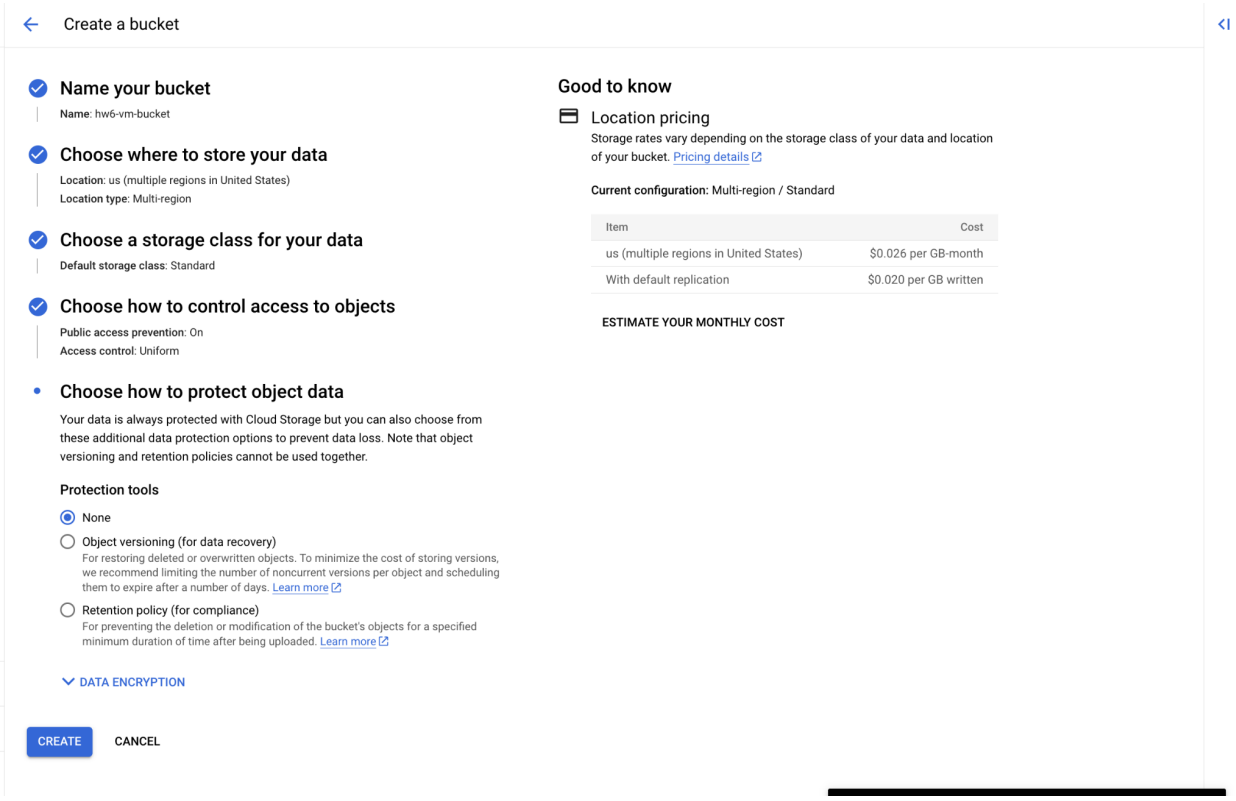
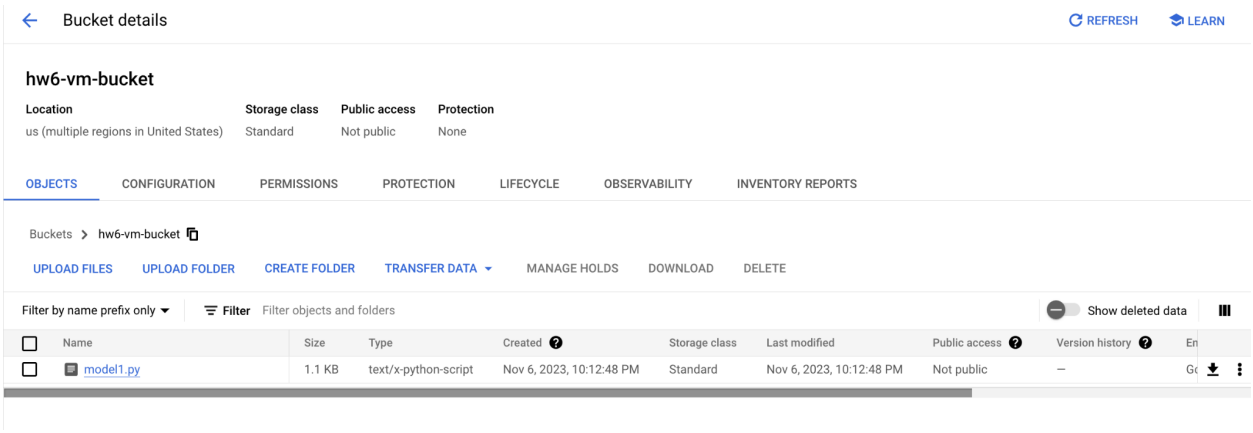
Collaborated with James Wong
Bryce Hiraoka

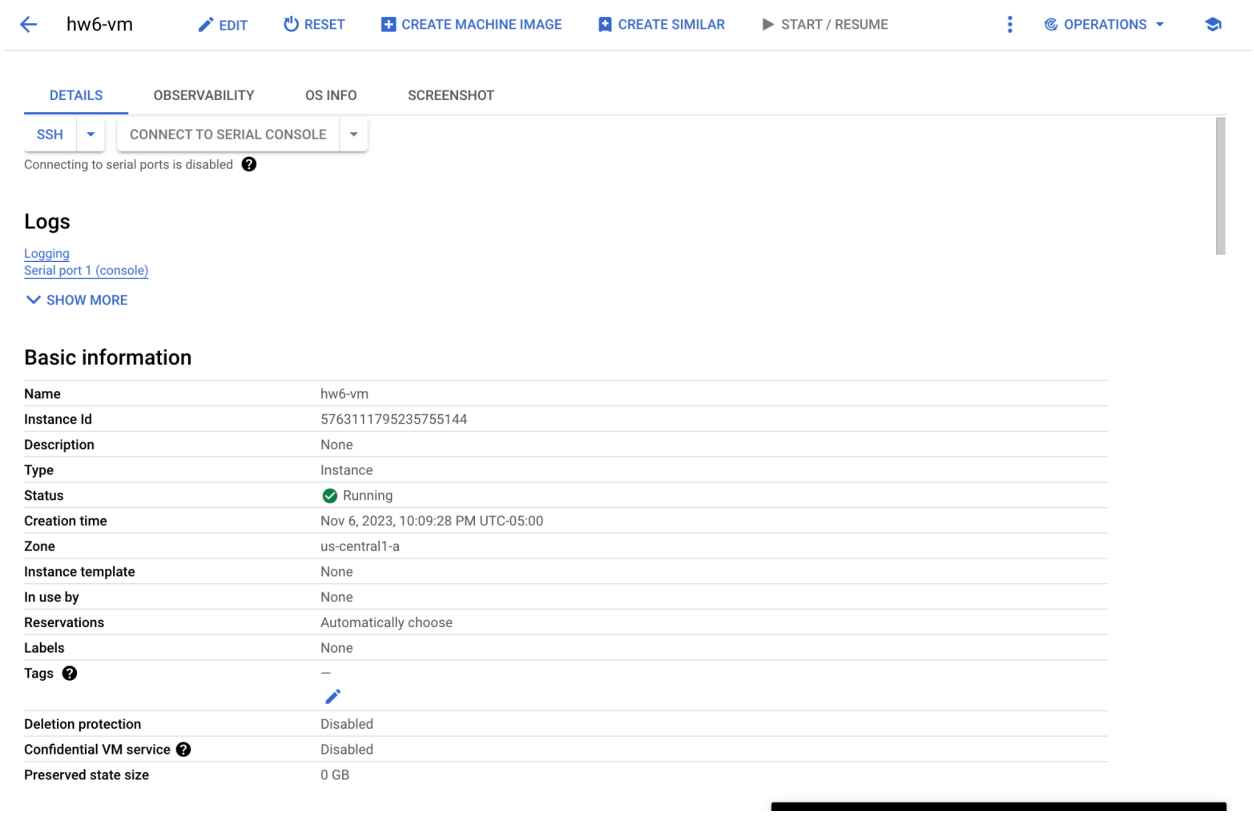
HW 6

- First I added my python file to a bucket so that it could be run on a VM





- Next I creates a VM to run the python file



- I then SSH into the vm and run

```
1.if [ -d "/home/bhiraoka/hw6-vm-bucket/model1.py" ]; then
      echo "Directory /home/bhiraoka/hw5-vm-bucket/Web-server exists"
  else
      gsutil -m cp -r gs://hw6-vm-bucket/ /home/bhiraoka/
  fi
2. cd /home/bhiraoka/hw6-vm-bucket
3. sudo apt install python3-pip -y
4. pip install cloud-sql-python-connector==0.9.3
5. pip install PyMySQL
6. pip install pandas
7. pip install -U scikit-learn
8. python3 model1.py
9. Python3 model2.py
```

- This is the accuracy of my first model

```
Installing collected packages: PyMySQL
Successfully installed PyMySQL-1.1.0
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 % python3 model1.py
Your instance connection name is: ds-561-398918:us-east1:hw5-database
/Users/brycehiraoka/Desktop/DS_561_2/6-HW-DS561/model1.py:24: UserWarning: panda
s only supports SQLAlchemy connectable (engine/connection) or database string UR
I or sqlite3 DBAPI2 connection. Other DBAPI2 objects are not tested. Please cons
ider using SQLAlchemy.
  df = pd.read_sql(query, con=conn)
Model accuracy: 0.9964363774588996
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 %
```

- For my second model I was not able to get the accuracy to be the 80% I could only get my code to get up to 20%. I tried every combination of inputs to outputs bu none of them worked.
- I analyzed the http client that creates the income and I found that the income is chosen randomly using the random.randrange(). I know from previous classes that random is not truly random and it is based off of system time. This is why my first thought was to use the timestamp and correlate it to the income.

1. This is the result when I used TimeStamp compared to Income

```
/Users/brycehiraoka/Desktop/DS_561_2/6-HW-DS561/model2.py:24: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2
 objects are not tested. Please consider using SQLAlchemy.
  data = pd.read_sql(query, con=conn)
Model accuracy: 0.1306186448731350 4
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 %
```

- I then realized that both Gender and Age use the same random.randrange() so I tried to use all or them correlated with the income.

2. This is the result when I compared TimeStamp, Age, Gender to Income

```
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 % python3 model2.py
Your instance connection name is: ds-561-398918:us-east1:hw5-database
/Users/brycehiraoka/Desktop/DS_561_2/6-HW-DS561/model2.py:24: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2
 objects are not tested. Please consider using SQLAlchemy.
  data = pd.read_sql(query, con=conn)
Model accuracy: 0.12534448351230637
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 %
```

- At this point nothing was working so I just compared random columns to income

3. This is the result when I use Country, TimeStamp, Gender, Age compared to Income

```
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 % python3 model2.py
Your instance connection name is: ds-561-398918:us-east1:hw5-database
/Users/brycehiraoka/Desktop/DS_561_2/6-HW-DS561/model2.py:24: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2
 objects are not tested. Please consider using SQLAlchemy.
  data = pd.read_sql(query, con=conn)
Model accuracy: 0.12724508220089328
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 %
```

4. This is the result when I use Country compared to Income

```
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 % python3 model2.py
Your instance connection name is: ds-561-398918:us-east1:hw5-database
/Users/brycehiraoka/Desktop/DS_561_2/6-HW-DS561/model2.py:24: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2
 objects are not tested. Please consider using SQLAlchemy.
  data = pd.read_sql(query, con=conn)
Model accuracy: 0.12762520193861066
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 %
```

5. This is the result when I use Country, Age, Gender to Income

```
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 % python3 model2.py
Your instance connection name is: ds-561-398918:us-east1:hw5-database
/Users/brycehiraoka/Desktop/DS_561_2/6-HW-DS561/model2.py:24: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection) or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2
 objects are not tested. Please consider using SQLAlchemy.
  data = pd.read_sql(query, con=conn)
Model accuracy: 0.12867053121733346
(base) brycehiraoka@crc-dot1x-nat-10-239-9-185 6-HW-DS561 %
```

- At this point I could not get it to be above 13% so I am guessing that my model idea does not work however I could not get any other model to even return an input. I will try more models and turn in late if I figure it out.

I started at $167 and ended at $165 so I spent $2 while working on this project

Git repository https://github.com/Bryce-Hiraoka/DS_561/tree/main/6-HW-DS561