Group Members:

- Veronika Kermoshchuk
- Maria Coughlin
- Sara Conway
- Caitlin Larkin

Rule-Based Penguins in Pandas

The objective of this task is to categorize or forecast a particular type of penguins using two characteristics accessible in the dataset. Similar to the approach we demonstrated in class on Monday, you will create your own rule-oriented classifier utilizing two threshold parameters for the selected feature.

In [1]:
```
!pip install pandas
!pip install plotly
!pip install packaging
!pip install ipywidgets
```

```
WARNING: pip is being invoked by an old script wrapper. This will fail in a f
uture version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the u
nderlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running
pip directly.
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: pandas in /home/codio/.local/lib/python3.6/sit
e-packages (1.1.5)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/pytho
n3.6/dist-packages (from pandas) (2.8.0)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.6/dist
-packages (from pandas) (1.16.3)
Requirement already satisfied: pytz>=2017.2 in /home/codio/.local/lib/python
3.6/site-packages (from pandas) (2022.7.1)
Requirement already satisfied: six>=1.5 in /usr/lib/python3/dist-packages (fr
om python-dateutil>=2.7.3->pandas) (1.11.0)
WARNING: pip is being invoked by an old script wrapper. This will fail in a f
uture version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the u
nderlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running
pip directly.
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: plotly in /home/codio/.local/lib/python3.6/sit
e-packages (5.13.1)
Requirement already satisfied: tenacity>=6.2.0 in /home/codio/.local/lib/pyth
on3.6/site-packages (from plotly) (8.2.2)
WARNING: pip is being invoked by an old script wrapper. This will fail in a f
uture version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the u
nderlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running
pip directly.
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: packaging in /home/codio/.local/lib/python3.6/
site-packages (21.3)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /usr/local/lib/pyt
hon3.6/dist-packages (from packaging) (2.4.0)
WARNING: pip is being invoked by an old script wrapper. This will fail in a f
uture version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the u
nderlying issue.
To avoid this problem you can invoke Python with '-m pip' instead of running
pip directly.
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: ipywidgets in /usr/local/lib/python3.6/dist-pa
ckages (7.4.2)
Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.6/di
st-packages (from ipywidgets) (4.4.0)
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.6/d
ist-packages (from ipywidgets) (5.1.0)
Requirement already satisfied: widgetsnbextension~=3.4.0 in /usr/local/lib/py
thon3.6/dist-packages (from ipywidgets) (3.4.2)
Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.6/d
ist-packages (from ipywidgets) (4.3.2)
Requirement already satisfied: ipython>=4.0.0 in /usr/local/lib/python3.6/dis
```

```
t-packages (from ipywidgets) (6.2.1)
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.6/dis
t-packages (from ipykernel>=4.5.1->ipywidgets) (6.1.3)
Requirement already satisfied: tornado>=4.2 in /home/codio/.local/lib/python
3.6/site-packages (from ipykernel>=4.5.1->ipywidgets) (6.1)
Requirement already satisfied: decorator in /usr/local/lib/python3.6/dist-pac
kages (from ipython>=4.0.0->ipywidgets) (4.4.0)
Requirement already satisfied: pygments in /usr/local/lib/python3.6/dist-pack
ages (from ipython>=4.0.0->ipywidgets) (2.4.0)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.6/dist-p
ackages (from ipython>=4.0.0->ipywidgets) (0.7.5)
Requirement already satisfied: jedi>=0.10 in /usr/local/lib/python3.6/dist-pa
ckages (from ipython>=4.0.0->ipywidgets) (0.13.3)
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/li
b/python3.6/dist-packages (from ipython>=4.0.0->ipywidgets) (1.0.16)
Requirement already satisfied: setuptools>=18.5 in /usr/lib/python3/dist-pack
ages (from ipython>=4.0.0->ipywidgets) (39.0.1)
Requirement already satisfied: simplegeneric>0.8 in /usr/local/lib/python3.6/
dist-packages (from ipython>=4.0.0->ipywidgets) (0.8.1)
Requirement already satisfied: pexpect in /usr/local/lib/python3.6/dist-packa
ges (from ipython>=4.0.0->ipywidgets) (4.7.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/pyth
on3.6/dist-packages (from nbformat>=4.2.0->ipywidgets) (3.0.1)
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.6/dist-
packages (from nbformat>=4.2.0->ipywidgets) (4.6.3)
Requirement already satisfied: ipython-genutils in /usr/local/lib/python3.6/d
ist-packages (from nbformat>=4.2.0->ipywidgets) (0.2.0)
Requirement already satisfied: six in /usr/lib/python3/dist-packages (from tr
aitlets>=4.3.1->ipywidgets) (1.11.0)
Requirement already satisfied: notebook>=4.4.1 in /home/codio/.local/lib/pyth
on3.6/site-packages (from widgetsnbextension~=3.4.0->ipywidgets) (6.4.10)
Requirement already satisfied: parso>=0.3.0 in /usr/local/lib/python3.6/dist-
packages (from jedi>=0.10->ipython>=4.0.0->ipywidgets) (0.4.0)
Requirement already satisfied: pyrsistent>=0.14.0 in /usr/local/lib/python3.
6/dist-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets)
(0.15.1)
Requirement already satisfied: attrs>=17.4.0 in /usr/local/lib/python3.6/dist
-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets) (19.1.
0)
Requirement already satisfied: argon2-cffi in /home/codio/.local/lib/python3.
6/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets)
(21.3.0)
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.6/
dist-packages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets)
(0.6.0)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.6/dist-packag
es (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets) (2.10.1)
Requirement already satisfied: Send2Trash>=1.8.0 in /home/codio/.local/lib/py
thon3.6/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywi
dgets) (1.8.0)
Requirement already satisfied: nbconvert>=5 in /usr/local/lib/python3.6/dist-
packages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets) (5.5.
0)
Requirement already satisfied: nest-asyncio>=1.5 in /home/codio/.local/lib/py
thon3.6/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywi
dgets) (1.5.5)
Requirement already satisfied: terminado>=0.8.3 in /home/codio/.local/lib/pyt
```

```
hon3.6/site-packages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywid
gets) (0.13.0)
Requirement already satisfied: pyzmq>=17 in /usr/local/lib/python3.6/dist-pac
kages (from notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidgets) (18.0.1)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python
3.6/dist-packages (from jupyter-client->ipykernel>=4.5.1->ipywidgets) (2.8.0)
Requirement already satisfied: wcwidth in /usr/local/lib/python3.6/dist-packa
ges (from prompt-toolkit<2.0.0,>=1.0.4->ipython>=4.0.0->ipywidgets) (0.1.7)
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.6/di
st-packages (from pexpect->ipython>=4.0.0->ipywidgets) (0.6.0)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.6/dist-pa
ckages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywid
gets) (0.6.0)
Requirement already satisfied: testpath in /usr/local/lib/python3.6/dist-pack
ages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidge
ts) (0.4.2)
Requirement already satisfied: bleach in /usr/local/lib/python3.6/dist-packag
es (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywidget
s) (3.1.0)
Requirement already satisfied: mistune>=0.8.1 in /usr/local/lib/python3.6/dis
t-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.4.0->ip
ywidgets) (0.8.4)
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.
6/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.4.
0->ipywidgets) (0.3)
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python
3.6/dist-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.
4.0->ipywidgets) (1.4.2)
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.6/d
ist-packages (from jinja2->notebook>=4.4.1->widgetsnbextension~=3.4.0->ipywid
gets) (1.1.1)
Requirement already satisfied: argon2-cffi-bindings in /home/codio/.local/li
b/python3.6/site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextens
ion~=3.4.0->ipywidgets) (21.2.0)
Requirement already satisfied: typing-extensions in /home/codio/.local/lib/py
thon3.6/site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~
=3.4.0->ipywidgets) (4.1.1)
Requirement already satisfied: dataclasses in /home/codio/.local/lib/python3.
6/site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.4.0
->ipywidgets) (0.8)
Requirement already satisfied: cffi>=1.0.1 in /home/codio/.local/lib/python3.
6/site-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->wid
getsnbextension~=3.4.0->ipywidgets) (1.15.1)
Requirement already satisfied: webencodings in /usr/local/lib/python3.6/dist-
packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.
4.0->ipywidgets) (0.5.1)
Requirement already satisfied: pycparser in /home/codio/.local/lib/python3.6/
site-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>
=4.4.1->widgetsnbextension~=3.4.0->ipywidgets) (2.21)
```

In [2]:

```
import pandas as pd
import plotly as plt
import plotly.express as px
import ipywidgets as widgets
from ipywidgets import interact, interact_manual
```

In [3]:
```
plt.offline.init_notebook_mode (connected = True)
```

In [4]:
```
# install libraries if needed
# !pip3 install pandas --user codio
# !pip3 install plotly --user codio
# !pip install packaging
```

In [5]:
```
df = pd.read_csv("penguins.csv")
df.shape
```

Out[5]:
```
(344, 7)
```

In [6]:
```
df.head()
```

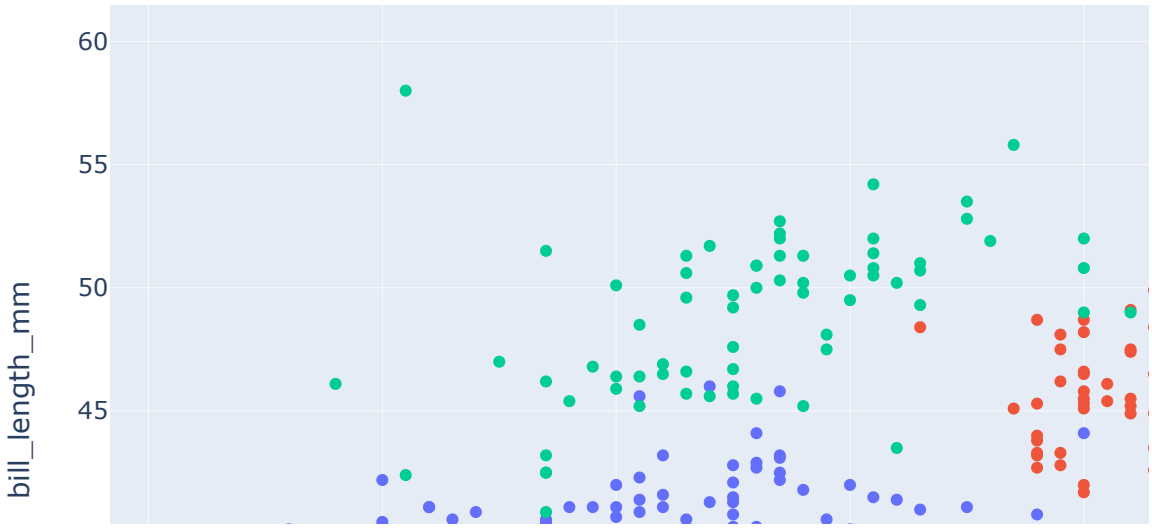
Out[6]:

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	female

Problem 1

Develop a Python function that receives three arguments, including two threshold values utilized by the rule-oriented classifier and the species type, such as Adelie penguin, you wish to forecast using the classifier. The function should return a Pandas dataframe as the confusion matrix output and precision and recall metrics for your model.

In [7]:
```
fig = px.scatter(df, x='flipper_length_mm', y='bill_length_mm', color = "species")
fig.show()
```



In [8]:
```
df1 = df.copy()
def rule_based_class(flipper_length, bill_length, species):
    predicted_species = "NA"

    if species == 'Adelie':
        if bill_length < 45 and flipper_length <= 205:
            predicted_species = 'Adelie'
        else:
            predicted_species = 'Other'

        df1['species'] = df1['species'].replace(['Chinstrap'], 'Other')
        df1['species'] = df1['species'].replace(['Gentoo'], 'Other')

    elif species == 'Chinstrap':
        if bill_length > 45 and flipper_length < 206:
            predicted_species = 'Chinstrap'
        else:
            predicted_species = 'Other'

        df1['species'] = df1['species'].replace(['Adelie'], 'Other')
        df1['species'] = df1['species'].replace(['Gentoo'], 'Other')

    elif species == 'Gentoo':
        if flipper_length >= 206:
            predicted_species = 'Gentoo'
        else:
            predicted_species = 'Other'

        df1['species'] = df1['species'].replace(['Chinstrap'], 'Other')
        df1['species'] = df1['species'].replace(['Adelie'], 'Other')

    else:
        predicted_species = 'Unknown'

    return predicted_species
```

In [9]:
```
df['predicted'] = df.apply(lambda x: rule_based_class(x.flipper_length_mm, x.b
ill_length_mm, "Chinstrap"), axis=1)
actual = df.species.value_counts().to_frame()
conf_matrix = pd.crosstab(df1.species, df["predicted"])
conf_matrix
```

Out[9]:

predicted	Chinstrap	Other
species		
Chinstrap	56	12
Other	4	272

In [10]:
```
target = conf_matrix.columns[0]
correctly_predicted = conf_matrix.loc[target, target]
all_predicted = conf_matrix.loc[:,target].sum()
all_actual = conf_matrix.loc[target,:].sum()
```

In [11]:
```
print("The precision for", target, "is", round(correctly_predicted/all_predicted,2))
```

```
The precision for Chinstrap is 0.93
```

In [12]:
```
print("The recall for", target, "is", round(correctly_predicted/all_actual,2))
```

```
The recall for Chinstrap is 0.82
```

Problem 2

Subsequently, employing the ipwidgets interactive library, enable users to construct their models by designating specific species and thresholds. Your "interact" function must generate a plot, confusion matrix, and precision/recall metrics. Ensure that the output of the model is clearly displayed by correctly color-coding the predicted and actual values and accurately labeling the results.

In [13]:
```
species = list(df.species.unique())
name1 = ['flipper_length_mm']
name2 = ['bill_length_mm']
minBill = round(df['bill_length_mm'].min(), 1)
maxBill = round(df['bill_length_mm'].max(), 1)
minFlipper = round(df['flipper_length_mm'].min(), 1)
maxFlipper = round(df['flipper_length_mm'].max(), 1)
```

In [14]:
```
@interact(type1 = name1, type2 = name2, specie_type = species)
def scatter_by_features(type1, type2, specie_type):
    df['predicted'] = df.apply(lambda x: rule_based_class(x.flipper_length_mm, x.bill_length_mm, specie_type), axis=1)
    selected_df = df[df.species == specie_type]
    fig = px.scatter(selected_df, x = type1, y = type2, color = selected_df['predicted'])
    fig.show()
```

In [ ]:
```

```