Coding Challenge

Brigham Eaquinto

Grand Question 1

Not able to make the picture

Grand Question 2

Not able to make the picture

Grand Question 3

Answer to number three rounded to 2 places is **1118.72**

Grand Question 4

Accuracy score for number four rounded to two places is **0.93**

Grand Question 5

Changes to make to the HP dataset to make it ML friendly:

1. One hot encode the house, birth_decade, and ancestry columns.
2. Fill in Professor McGonagall's ancestry as a half-blood because every decent person knows that and it's a great way to deal with that NA. Otherwise, drop it because na's wreck models.
3. Change hogwarts students true to 1 and false to 0.

Code Appendix

```
# Set Up
import pandas as pd
import altair as alt
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn import metrics

alt.data_transformers.enable('json')
```

```
url_names = 'https://github.com/byuidatascience/data4names/raw/master/data-raw/names_year/names_

dat_names = pd.read_csv(url_names)
```

```
dat_names.columns
```

```
# Question 1

(alt.Chart(dat_names)
.mark_line()
.encode(
    x = 'year:O',
    y = 'Total'
)
)
```

```
dat_names1 = dat_names.filter(['name', 'year', 'CO', 'AZ', 'Total'])

yearchart = (alt.Chart(dat_names1)
                .mark_line()
                .encode(
                        x = alt.X('year:O'),
                        y = alt.Y('Total')) )
yearchart
```

```
# Question 2

(dat_names
    .groupby('ID', 'Total')
    .sum()
```

```
)

# .groupby('name')
# .agg('sum')
# )



# Question 3

bob = pd.Series([np.nan, 18, 22, 45, 31, np.nan, 85, 38, 129, 8000, 22, 2])
standard_dev = bob.dropna().std()
bob.fillna(standard_dev).mean()


# Question 4

dwellings_ml = pd.read_csv("https://github.com/byuidatascience/data4dwellings/raw/master/data-ra

features = dwellings_ml.drop(['numbaths','parcel'], axis = 1)
target = (dwellings_ml.numbaths > 2)*1


x_train, x_test, y_train, y_test = train_test_split(features, target, test_size = .30, random_st

classifier_final = RandomForestClassifier()

classifier_final.fit(x_train, y_train)
y_predicted = classifier_final.predict(x_test)

from sklearn.metrics import accuracy_score
accuracy_score(y_test, y_predicted)


# Question 5

Changes to make to the HP dataset to make it ML friendly:

1. One hot encode the house, birth_decade, and ancestry columns.
2. Fill in Professor McGonagall's ancestry as a half-blood because every decent person knows tha
3. Change hogwarts students true to 1 and false to 0.
```