Assignment 2

Josh Barber
barb4630@mylaurier.ca
215854630

Working Code:

```
import pandas as pd
from sklearn.utils import shuffle
from sklearn.impute import KNNImputer
import numpy as np
import os

df = pd.read_csv(os.path.join(os.path.abspath(''),'crx_data.csv'),
names=['a1','a2','a3','a4','a5','a6','a7','a8','a9','a10','a11','a12','a13','a14','a15','a16'])

pd.set_option("display.min_rows", df.shape[0]+1)
pd.set_option('display.max_rows', df.shape[0]+1)

df["a2"] = pd.to_numeric(df["a2"], errors='coerce')

knn_imp = KNNImputer(n_neighbors=5, weights='uniform', metric='nan_euclidean')
df['a2'] = knn_imp.fit_transform(df[['a2']])

df_shuffled = pd.DataFrame(shuffle(df))
df_shuffled['a2_bins'] = pd.cut(x=df_shuffled['a2'], bins=[0,10,20,30,40,50,60,70,80])

df_shuffled.head(3)
```

Code and Output:

```
import pandas as pd
from sklearn.utils import shuffle
from sklearn.impute import KNNImputer
import numpy as np
import os

df = pd.read_csv(os.path.join(os.path.abspath(''),'crx_data.csv'), names=['a1','a2','a3','a4','a5','a6','a7','a8',
                                                                          'a9','a10','a11','a12','a13','a14','a15','a16'])

pd.set_option("display.min_rows", df.shape[0]+1)
pd.set_option('display.max_rows', df.shape[0]+1)

df["a2"] = pd.to_numeric(df["a2"], errors='coerce')

knn_imp = KNNImputer(n_neighbors=5, weights='uniform', metric='nan_euclidean')
df['a2'] = knn_imp.fit_transform(df[['a2']])

df_shuffled = pd.DataFrame(shuffle(df))
df_shuffled['a2_bins'] = pd.cut(x=df_shuffled['a2'], bins=[0,10,20,30,40,50,60,70,80])


df_shuffled.head(3)
```

[44] ✓ 0.0s

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16	a2_bins
135	b	48.58	6.50	u	g	q	h	6.0	t	f	0	t	g	00350	0	+	(40, 50]
627	b	36.08	2.54	u	g	ff	ff	0.0	f	f	0	f	g	00000	1000	-	(30, 40]
205	a	35.42	12.00	u	g	q	h	14.0	t	t	8	f	g	00000	6590	+	(30, 40]