Problem 1:

```
import pandas as pd
from sklearn.feature_extraction import DictVectorizer
from sklearn import preprocessing
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from apyori import apriori

bankData = pd.read_csv("bank_data.csv")

X = bankData.drop("pep", axis=1, inplace = False)
y = bankData.pep

X.drop("id", axis = 1, inplace=True)

X.replace(to_replace = ['YES', 'NO'], value = [1.0, 0.0], inplace=True)

def model_matrix(df , columns):
    dummified_cols = pd.get_dummies(df[columns])
    df = df.drop(columns, axis = 1, inplace=False)
    df_new = df.join(dummified_cols)
    return df_new

X = model_matrix(X, ['gender', 'region'])

Scaler = MinMaxScaler()
X[['age', 'income', 'children']] = Scaler.fit_transform(X[['age', 'income', 'children']])

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,random_state = 17, stratify = y)

association_rules = apriori(X_train, min_support=0.0045, min_confidence=0.2, min_lift=3,
min_length=2)
association_results = list(association_rules)

print (association_results)
```

        results is unable to convert data set into a list format, I tried so many different type of data separation and processing , but I always ended up with unable to convert data set into a list format, which is needed to run the apriori algorithm for association mining rule.

Problem 2:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.cluster import KMeans
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
```

```
inFile = pd.read_csv('Disputed_Essay_data.csv')

dataSetPandaFrame = pd.DataFrame(inFile)

dataSetPandaFrame = dataSetPandaFrame.drop(dataSetPandaFrame.index[62:72])

dataSetPandaFrame = dataSetPandaFrame.drop('filename', axis= 1)


disputeVariableDataSet = dataSetPandaFrame[:11]

disputeVariableDataSet = disputeVariableDataSet.drop('author', axis=1)


dataSet = dataSetPandaFrame[11:]


targetVariable = dataSet.author

dataSet = dataSet.drop('author',axis = 1)

x_train, x_test, y_train, y_test = train_test_split(dataSet,targetVariable,test_size=0.5,random_state=1)

kmeans_model = KMeans(n_clusters=3, random_state=1).fit(x_train)

labels = kmeans_model.labels_

print (metrics.calinski_harabasz_score(x_train, y_test) )
```

results:

my K means model performance score results is in the screen capture below.

```
  File "/home/chris/anaconda3/lib/python3.7/site-packages/pandas/core/arrays/num
py_.py", line 171, in __array__
    return np.asarray(self._ndarray, dtype=dtype)
  File "/home/chris/anaconda3/lib/python3.7/site-packages/numpy/core/numeric.py"
, line 538, in asarray
    return array(a, dtype, copy=False, order=order)
ValueError: could not convert string to float: 'Hamilton'
(base) chris@ChrisUbuntu:~/Downloads$ python3 homework1\(1\).py
Traceback (most recent call last):
  File "homework1(1).py", line 39, in <module>
    print (metrics.calinski_harabasz_score(x_train, x_test) )
  File "/home/chris/anaconda3/lib/python3.7/site-packages/sklearn/metrics/cluste
r/unsupervised.py", line 269, in calinski_harabasz_score
    X, labels = check_X_y(X, labels)
  File "/home/chris/anaconda3/lib/python3.7/site-packages/sklearn/utils/validati
on.py", line 724, in check_X_y
    y = column_or_1d(y, warn=True)
  File "/home/chris/anaconda3/lib/python3.7/site-packages/sklearn/utils/validati
on.py", line 760, in column_or_1d
    raise ValueError("bad input shape {0}".format(shape))
ValueError: bad input shape (32, 70)
(base) chris@ChrisUbuntu:~/Downloads$ python3 homework1\(1\).py
1.134652902929946
(base) chris@ChrisUbuntu:~/Downloads$
```