In [104…
```
# Câu 1 (10 điểm)
# Sử dụng thư viện pandas, đọc toàn bộ dữ liệu này vào 1 dataframe tên là churn_df
import pandas as pd
churn_df = pd.read_csv('customer_churn.csv')
```

In [105…
```
# Câu 2 (5 điểm)
# In ra 10 dòng dữ liệu đầu tiên của churn_df
churn_df.head(10)
```

Out[105…

	state	account_length	area_code	international_plan	voice_mail_plan	number_vmail_messages
0	OH	107	area_code_415	no	yes	26
1	NJ	137	area_code_415	no	no	0
2	OH	84	area_code_408	yes	no	0
3	OK	75	area_code_415	yes	no	0
4	MA	121	area_code_510	no	yes	24
5	MO	147	area_code_415	yes	no	0
6	LA	117	area_code_408	no	no	0
7	WV	141	area_code_415	yes	yes	37
8	IN	65	area_code_415	no	no	0
9	RI	74	area_code_415	no	no	0

In [106…
```
# Câu 3 (10 điểm)
# Dùng các câu lệnh hợp lý của pandas để in ra thông tin các cột của bộ dữ liệu
# Cho biết những cột thuộc tính nào KHÔNG PHẢI dạng số (numeric)
churn_df.dtypes
```

Out[106…
```
state                          object
account_length                  int64
area_code                      object
international_plan             object
voice_mail_plan                object
number_vmail_messages           int64
total_day_minutes             float64
total_day_calls                 int64
total_day_charge              float64
total_eve_minutes             float64
total_eve_calls                 int64
total_eve_charge              float64
total_night_minutes           float64
total_night_calls               int64
total_night_charge            float64
total_intl_minutes            float64
total_intl_calls                int64
total_intl_charge             float64
number_customer_service_calls   int64
churn                          object
dtype: object
```

In [107…
```
print('Các cột không phải dạng số là: state, area_code, international_plan, voice_ma
```

```
Các cột không phải dạng số là: state, area_code, international_plan, voice_mail_pla
n, churn
```

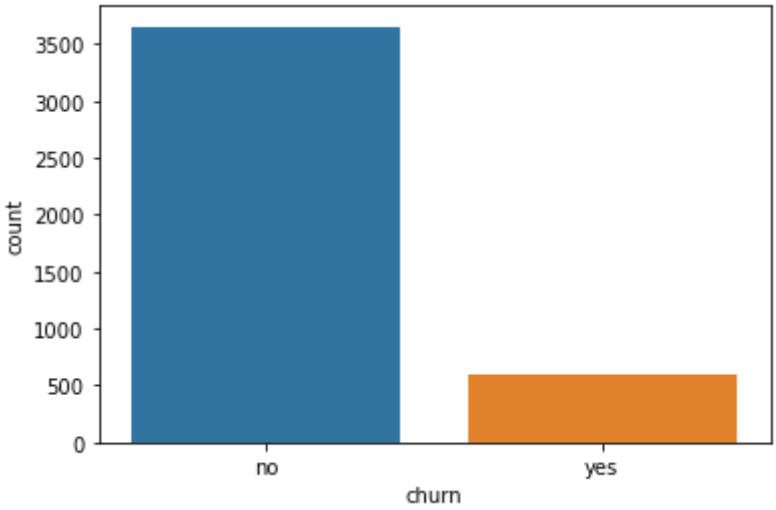
In [108…

```
# Câu 4 (15 điểm)
# Vẽ biểu đồ cột thể hiện tương quan số khách hàng rời bỏ / không rời bỏ dịch vụ
# Gợi ý: Có thể dùng matplotlib hoặc seaborn
import seaborn as sns
sns.countplot(churn_df['churn'])
```

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning:
Pass the following variable as a keyword arg: x. From version 0.12, the only valid p
ositional argument will be `data`, and passing other arguments without an explicit k
eyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[108…

```
<AxesSubplot:xlabel='churn', ylabel='count'>
```



In [109…

```
# Câu 5 (10 điểm)
# Chọn cột cuối làm y, các cột trước đó, trừ cột [account_length, area_code, interna
features=['state','number_vmail_messages','total_day_minutes','total_day_calls','tot
          'total_eve_minutes','total_eve_calls','total_eve_charge','total_night_minu
          'total_night_charge','total_intl_minutes','total_intl_calls','total_intl_c
target=['churn']
X=churn_df[features]
y=churn_df[target]
# Dùng LabelEncoder để chuyển dữ liệu cột state từ dạng categorical sang dạng numeri
from sklearn.preprocessing import LabelEncoder
X['state'] = le.fit_transform(X['state'])
```

```
C:\Users\Admin\AppData\Local\Temp/ipykernel_11564/3563749754.py:11: SettingWithCopyW
arning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/u
ser_guide/indexing.html#returning-a-view-versus-a-copy
  X['state'] = le.fit_transform(X['state'])
```

In [110…

```
# Câu 6 (10 điểm)
# Chia dữ liệu thành 2 phần (X_train, y_train), (X_test, y_test) với tỷ lệ 80/20
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test=train_test_split(X, y, test_size=0.2)
print(X_train)
print(y_train)
print(X_test)
```

```
print(y_test)
# In ra X_train, y_train để kiểm tra
print(X_train)
print(y_train)
```

```
      state  number_vmail_messages  total_day_minutes  total_day_calls  \
4117     41                      0              165.8               94
628      43                      0              109.6               88
1100     45                      0              222.2              127
540      34                      0              235.5               81
2551      1                      0              242.8               90
...     ...                    ...                ...              ...
938       9                      0              199.2              122
2012      0                      0              262.3              114
1840     45                     35              205.5               86
2409     39                     20              211.9              110
651      28                      0              219.1              100

      total_day_charge  total_eve_minutes  total_eve_calls  total_eve_charge  \
4117             28.19              192.8              135             16.39
628              18.63              137.6              108             11.70
1100             37.77              153.1              125             13.01
540              40.04              257.2              130             21.86
2551             41.28              234.1               80             19.90
...                ...                ...              ...               ...
938              33.86              214.7              114             18.25
2012             44.59              198.9               96             16.91
1840             34.94              298.5              119             25.37
2409             36.02              215.1              120             18.28
651              37.25              242.9               90             20.65

      total_night_minutes  total_night_calls  total_night_charge  \
4117                247.8                 64               11.15
628                 159.7                121                7.19
1100                227.4                 80               10.23
540                 103.1                111                4.64
2551                211.5                104                9.52
...                   ...                ...                 ...
938                 150.9                105                6.79
2012                165.9                 90                7.47
1840                214.2                104                9.64
2409                238.5                107               10.73
651                 168.9                101                7.60

      total_intl_minutes  total_intl_calls  total_intl_charge  \
4117                12.4                 7               3.35
628                 11.0                 5               2.97
1100                12.9                 4               3.48
540                 11.5                 4               3.11
2551                 6.0                 3               1.62
...                  ...               ...                ...
938                 11.8                 7               3.19
2012                 6.6                 5               1.78
1840                 6.9                 4               1.86
2409                 9.4                 2               2.54
651                 10.1                 4               2.73

      number_customer_service_calls
4117                              3
628                               2
1100                              1
540                               2
2551                              5
...                             ...
```

```
938                             1
2012                            3
1840                            1
2409                            0
651                             2

[3400 rows x 15 columns]
     churn
4117    no
628     no
1100    no
540     no
2551    no
...    ...
938     no
2012    no
1840    no
2409    no
651     no

[3400 rows x 1 columns]
      state  number_vmail_messages  total_day_minutes  total_day_calls  \
3481     40                      0               92.1              109
4174     23                      0               86.0              105
2865     41                      0              307.2               65
2685     43                      0              225.9              110
822      43                      0              188.9               94
...     ...                    ...                ...              ...
2619     41                      0              226.3               95
3280     11                      0              140.3              144
3173     44                      0              157.8               96
2513      6                      0              234.9              136
3535     39                      0              241.5              114

      total_day_charge  total_eve_minutes  total_eve_calls  total_eve_charge  \
3481             15.66              163.0               83             13.86
4174             14.62              215.5              102             18.32
2865             52.22              138.6               97             11.78
2685             38.40              299.1               86             25.42
822              32.11              203.9              104             17.33
...                ...                ...              ...               ...
2619             38.47              274.3              109             23.32
3280             23.85              294.8               89             25.06
3173             26.83              160.0              120             13.60
2513             39.93              270.8              134             23.02
3535             41.06              195.2               94             16.59

      total_night_minutes  total_night_calls  total_night_charge  \
3481                133.3                102                6.00
4174                185.7                 83                8.36
2865                381.6                 99               17.17
2685                251.3                 81               11.31
822                 151.8                124                6.83
...                   ...                ...                 ...
2619                242.7                119               10.92
3280                153.5                126                6.91
3173                198.8                112                8.95
2513                219.3                101                9.87
3535                201.6                 93                9.07

      total_intl_minutes  total_intl_calls  total_intl_charge  \
3481                 9.7                11               2.62
4174                 8.2                 3               2.21
2865                10.2                 4               2.75
```

```
2685               11.2                4               3.02
822                11.6                8               3.13
...                 ...              ...                ...
2619                8.2                3               2.21
3280               11.7                4               3.16
3173               13.7                6               3.70
2513               13.9                2               3.75
3535               14.1                3               3.81

      number_customer_service_calls
3481                              2
4174                              3
2865                              2
2685                              1
822                               3
...                             ...
2619                              2
3280                              2
3173                              3
2513                              1
3535                              3

[850 rows x 15 columns]
     churn
3481    no
4174    no
2865   yes
2685   yes
822     no
...    ...
2619   yes
3280    no
3173    no
2513   yes
3535    no

[850 rows x 1 columns]
      state  number_vmail_messages  total_day_minutes  total_day_calls  \
4117     41                      0              165.8               94
628      43                      0              109.6               88
1100     45                      0              222.2              127
540      34                      0              235.5               81
2551      1                      0              242.8               90
...     ...                    ...                ...              ...
938       9                      0              199.2              122
2012      0                      0              262.3              114
1840     45                     35              205.5               86
2409     39                     20              211.9              110
651      28                      0              219.1              100

      total_day_charge  total_eve_minutes  total_eve_calls  total_eve_charge  \
4117             28.19              192.8              135             16.39
628              18.63              137.6              108             11.70
1100             37.77              153.1              125             13.01
540              40.04              257.2              130             21.86
2551             41.28              234.1               80             19.90
...                ...                ...              ...               ...
938              33.86              214.7              114             18.25
2012             44.59              198.9               96             16.91
1840             34.94              298.5              119             25.37
2409             36.02              215.1              120             18.28
651              37.25              242.9               90             20.65

      total_night_minutes  total_night_calls  total_night_charge  \
```

```
4117                247.8                 64                11.15
628                 159.7                121                 7.19
1100                227.4                 80                10.23
540                 103.1                111                 4.64
2551                211.5                104                 9.52
...                   ...                ...                  ...
938                 150.9                105                 6.79
2012                165.9                 90                 7.47
1840                214.2                104                 9.64
2409                238.5                107                10.73
651                 168.9                101                 7.60

      total_intl_minutes  total_intl_calls  total_intl_charge  \
4117                12.4                 7               3.35
628                 11.0                 5               2.97
1100                12.9                 4               3.48
540                 11.5                 4               3.11
2551                 6.0                 3               1.62
...                  ...               ...                ...
938                 11.8                 7               3.19
2012                 6.6                 5               1.78
1840                 6.9                 4               1.86
2409                 9.4                 2               2.54
651                 10.1                 4               2.73

      number_customer_service_calls
4117                              3
628                               2
1100                              1
540                               2
2551                              5
...                             ...
938                               1
2012                              3
1840                              1
2409                              0
651                               2

[3400 rows x 15 columns]
     churn
4117    no
628     no
1100    no
540     no
2551    no
...    ...
938     no
2012    no
1840    no
2409    no
651     no

[3400 rows x 1 columns]
```

In [111…

```python
# Câu 7 (10 điểm)
# Sử dụng Logistic Regression, xây dựng mô hình dự đoán khách hàng rời bỏ dịch vụ vớ
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

model = LogisticRegression(solver='newton-cg', max_iter=150)
model.fit(X_train, y_train)
```

```
C:\Users\Admin\anaconda3\lib\site-packages\sklearn\utils\validation.py:63: DataConve
```

```
rsionWarning: A column-vector y was passed when a 1d array was expected. Please chan
ge the shape of y to (n_samples, ), for example using ravel().
  return f(*args, **kwargs)
```

Out[111…
```
LogisticRegression(max_iter=150, solver='newton-cg')
```

In [112…
```
# Câu 8 (5 điểm)
# In ra độ chính xác của mô hình vừa huấn luyện trên tập (X_test, y_test)
from sklearn.metrics import accuracy_score
pred2 = model.predict(X_test)
accuracy2 = accuracy_score(y_test, pred2)
print('Độ chính xác của mô hình là:',accuracy2)
```

```
Độ chính xác của mô hình là: 0.8564705882352941
```

In [113…
```
# Câu 9 (5 điểm)
# Cho khách hàng có thông tin như sau:
# state: NJ=31, number_vmail_message: 24, các cột sau lần lượt: 208,    88,     35,
# Dùng model vừa huấn luyện để dự đoán khả năng rời bỏ của khách hàng này
x=[[31,24,208,88,35,312,108,33,212.6,118,9.57,8.5,7,2.4,3]]
prediction = model.predict(x)
print(prediction)
```

```
['no']
```

In [ ]: