

Question 5

Question: Identify all records where the 'Billing Month' does not match the month of the 'Date In'.

Approach: Create two new columns that contain the numeric value of the months from both original columns ('Billing Month' and 'Date In'). Do this by converting each column to datetime in some form and pulling the **month** information. Finally, find the rows where the values in the two newly created columns *don't* match.

1. Nulls

```
In [35]: 1 # Check for null values
          2 charges[['Date In', 'Billing Month']].isnull().sum()
```

```
Out[35]: Date In          2
          Billing Month    27
          dtype: int64
```

```
In [36]: 1 # Replace all null cells with a string ('?') and leave all other cells as they are
          2 charges['Billing Month'] = charges['Billing Month'].map(lambda x: '?' if str(x) == 'nan' else x)
          3 charges['Date In'] = charges['Date In'].map(lambda x: '?' if str(x) == 'nan' else x)
```

2. Clean up 'Date In' columns

```
In [37]: 1 import re
```

```
In [38]: 1 # Clean up 'Date In' column
          2 for i in range(len(charges)):
          3
          4     c = charges['Date In'][i]
          5     charges['Date In'][i] = re.sub('\s+', ' ', c).strip().split(' ')
          6
          7     if i % 500 == 0:
          8         print(i)
```