

Question 3

Question: Identify all patients with an invalid Date of Birth.

```
In [15]: 1 # Change from string to datetime format
        2 charges['DOB'] = pd.to_datetime(charges['DOB'])
```

```
In [16]: 1 charges['DOB'].sort_values().tail(45)
```

```
Out[16]: 2181    1998-05-11
        330    1999-08-02
        224    2016-03-20
        354    2016-12-14
        1631   2018-06-25
        585    2019-01-21
        1074   2019-03-15
        2082   2055-05-11
        1908   2055-05-20
        2087   2058-06-17
        613    2060-06-05
        2021   2061-07-25
        614    2061-08-06
        1595   2063-04-13
        1492   2063-04-18
        704    2063-04-27
        514    2063-12-25
        1541   2064-05-03
        1483   2064-08-31
        1034   2064-11-05
```

Answer: When changing the datatype from string to datetime, any row with a DOB *before the year of 1969* get's incorrectly converted to future dates that aren't currently valid birthdays.

For example, `11/25/68` incorrectly becomes `2068-11-25` , whereas `11/20/69` correctly becomes `1969-11-20` .