**特征工程**

1. country 计算取消比率中位数 country\_rate，缺失值以均值补充
2. agent 计算取消比率均值/中位数 agent\_rate，缺失值以均值补充
3. company 计算取消比率均值/中位数 company\_rate，缺失值以均值补充

**需要注意的变量分布**

1. babies is highly skewed (γ1 = 24.64654483)
2. previous\_cancellations is highly skewed (γ1 = 24.45804872)
3. previous\_bookings\_not\_canceled is highly skewed (γ1 = 23.53979995)
4. lead\_time is highly skewed; has 5.23% zeors

**需要注意的变量**

1. reservation\_status\_date only contains datetime values, but is categorical. Consider applying pd.to\_datetime()

**疑问**

1. reserved\_room\_type, assigned\_room\_type 定类还是定序