

■図2-13：欠損値チェック結果(補完金額の検証)

```
In [14]: for trg in list(uriage_data["item_name"].sort_values().unique()):
         print(trg + "の最大額:" + str(uriage_data.loc[uriage_data["item_name"]==trg]["item_price"].max)
         + "の最小額:" + str(uriage_data.loc[uriage_data["item_name"]==trg]["item_price"].min)

商品Aの最大額: 100.0の最小額: 100.0
商品Bの最大額: 200.0の最小額: 200.0
商品Cの最大額: 300.0の最小額: 300.0
商品Dの最大額: 400.0の最小額: 400.0
商品Eの最大額: 500.0の最小額: 500.0
商品Fの最大額: 600.0の最小額: 600.0
商品Gの最大額: 700.0の最小額: 700.0
商品Hの最大額: 800.0の最小額: 800.0
商品Iの最大額: 900.0の最小額: 900.0
商品Jの最大額: 1000.0の最小額: 1000.0
商品Kの最大額: 1100.0の最小額: 1100.0
商品Lの最大額: 1200.0の最小額: 1200.0
商品Mの最大額: 1300.0の最小額: 1300.0
商品Nの最大額: 1400.0の最小額: 1400.0
商品Oの最大額: 1500.0の最小額: 1500.0
商品Pの最大額: 1600.0の最小額: 1600.0
商品Qの最大額: 1700.0の最小額: 1700.0
商品Rの最大額: 1800.0の最小額: 1800.0
商品Sの最大額: 1900.0の最小額: 1900.0
商品Tの最大額: 2000.0の最小額: 2000.0
商品Uの最大額: 2100.0の最小額: 2100.0
商品Vの最大額: 2200.0の最小額: 2200.0
商品Wの最大額: 2300.0の最小額: 2300.0
商品Xの最大額: 2400.0の最小額: 2400.0
商品Yの最大額: 2500.0の最小額: 2500.0
商品Zの最大額: 2600.0の最小額: 2600.0
```

1行目ですべての商品に対してループ処理を実施しています。

ループの中で商品に設定されている金額の最大額と最小額を出力する処理を実行し、結果が画面上に表示されます。

結果を見る限り、すべての商品の最大と最小が一致している事が確認できましたので、金額の補完は成功している事が分かります。

処理中の`.min(skipna=False)`における「`skipna`」はNaNデータを無視するかを設定できます。今回はFalseを明示指定しているので、NaNが存在する場合、最小値はNaNと表示されます。

応用で、欠損値補完処理をする前に、この処理を実行してみましょう。最小値にNaNと表示されることが確認できるはずです。

⑩ ノック16： 顧客名の揺れを補正しよう

顧客台帳の顧客名の揺れを補正していきましょう。
まずデータの確認から行います。

```
kokyaku_data["顧客名"].head()
```

■図2-14：顧客台帳の顧客名

```
In [15]: kokyaku_data["顧客名"].head()

Out[15]: 0    須賀ひとみ
         1    岡田 敏也
         2    芳賀 希
         3    荻野 愛
         4    栗田 憲一
         Name: 顧客名, dtype: object
```

```
uriage_data["customer_name"].head()
```

■図2-15：売上履歴の顧客名

```
In [16]: uriage_data["customer_name"].head()

Out[16]: 0    深井菜々美
         1    浅田賢二
         2    南部慶二
         3    麻生莉緒
         4    平田鉄二
         Name: customer_name, dtype: object
```

顧客台帳と売上履歴の顧客名を比較してみると、顧客台帳の顧客名には姓名の間にスペースが含まれていますが、売上履歴の顧客名にはスペースが含まれず、姓名が一つになっています。

さらに詳しく見ると、顧客台帳の顧客名のスペースは全角・半角が混じっている、スペースがない等、書式が混在してしまっています。

このまま、売上履歴と顧客台帳を結合しても正しく結合できません。

顧客名以外に2つのデータをつなげるキーが存在しないため、顧客名の補正は必須となります。