2.3 добав вш столбец, join merge

1 Срез по данным из вш словаря;
2 Новый столбец с вш данными:
по порядку строк(подать list) vs по совпадению индексов(Series: left.join)
3 join() и merge() - слияние столбцов.

---

case
чистим от max-min аномалий
get медиана очищенных данных(там медиана медианы)

---

1 Срез по данным из вш словаря
query in @list уже было
вместо list можно и другие типы данных
in @dict - чекает ключи
@df.index
@df.column

---

2 Добав вш столбец df2 в df1
строки заполняются по совпадению индексов

a df2 не имеет индексов df1 - такие строки None.
b df2 имеет свои уник - такие не добавятся
c df2 повторы индекса - ошибка

= **left join по индексу** (только в sql повторы индекса рождают cross join в подгруппе)

---

3 join() и merge()
merge:
suffixes=('_x',
'_y') data_x data_y

join:
можно больше двух
default - left
default on-по индексу

---

case 2 structure:

срезы
1 good_ids = too_fast<0.5
2
good_data = query id in good_ids +
time_spent [60, 1000]

групировки и джойны
id заправка
name сеть АЗС

3 good_stations_stat = good_data
gpb(**id**) - time_spent: median
медиана по всем заправкам
~~5 good_stat = good_data~~
~~gpb **name** - time_spent median~~
6 id_name = good_data
gpb(**id**) - **name**: first count

1stname это первая сеть для id(заправка) =
= сеть этой заправки

7 stations_stat_full =
= 6id_name Join 3good_stations_stat

7 = 6 J 3 можно было сделать одним pivot?
нет, left выкинул строки уникальные для good_stations
а они были? нет
получается можно было)

7 cols:
id name ~~count~~ time_median
добавили заправкам их сети

8 stations_stat_full
gpb(name) - time_median: median count
медиана медиан

4 stat = ***data***
gpb(name) - time_spent: mean
среднее по сетям АЗС

9 = 8 J 4(name time_mean)
join on name=1stname

result:
time_mean(id) vs
time_median(id)_median(1stname)
среднее заправки(плохие данные) vs медиана медиан заправок для компании(хорошие данные)