

## 2.4 корр-ии

---

- 1 scatter
  - 2 corr coef
  - 3 pairplot
- 

case

- 1 матрица диаграмм рассеяния по исходным данным vs по отфильтрованным(аном макс мин)
  - 2 снизили влияние подозрительных данных на итоговый результат.
- 

plot(, kind='scatter')

**fix наложение точек scatter**

- 1 **alpha** - прозрачность
  - 2 **hexbin** - плотность точек = цвет шестиугольника
- kind='hexbin'
- ячеечная диаграмма
- 

df.corr()

- до 0,2 Очень слабая
- до 0,5 Слабая
- до 0,7 Средняя
- до 0,9 Высокая
- свыше 0,9 Очень высокая корреляция

Коэффициент корреляции не рассчитывается:

- 1 не лин
  - 2 в данных имеется больше 1-го наблюдения по каждому случаю;
  - 3 имеются выбросы,
  - 4 данные содержат ярко выраженные подгруппы**
- 

pd.plotting.scatter\_matrix(data)

sns.pairplot

---

**попарный коэфф**

тоже .corr()