

مراجعة
DA 2

بسم الله توكلت على الله ولا حول ولا قوة الا بالله

* بسم الله وليس بسم الله (بسم الله مكتوب على ابي لينة لظلال فقط)

Intro to Pandas

Relap de oop → سوال شغل git ازاي ss

Statistical

- ① mean
- ② median
- ③ mode

variance

- ④ Variance
- ⑤ Std
- ⑥ Co-Variance
- ⑦ Correlation

① → mean → المتوسط (المجموع ÷ العدد)
1, 2, 3, 4, 5 → Mean = $\frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$

outlier ← # ينتمونه لو الـ Normal distribution ومفترض

← لما بيبي مفترض outlier بيبي في الـ Normal

② → median → المتوسط (الترتيب)

← ياخذ العنصر الأوسط بعد ترتيب الـ data

← # من بيبي تر بالـ outlier

ex: 1, 2, 3, 4, 5 → الحل
الترتيب مرتبة 1
عنصر الأوسط 2

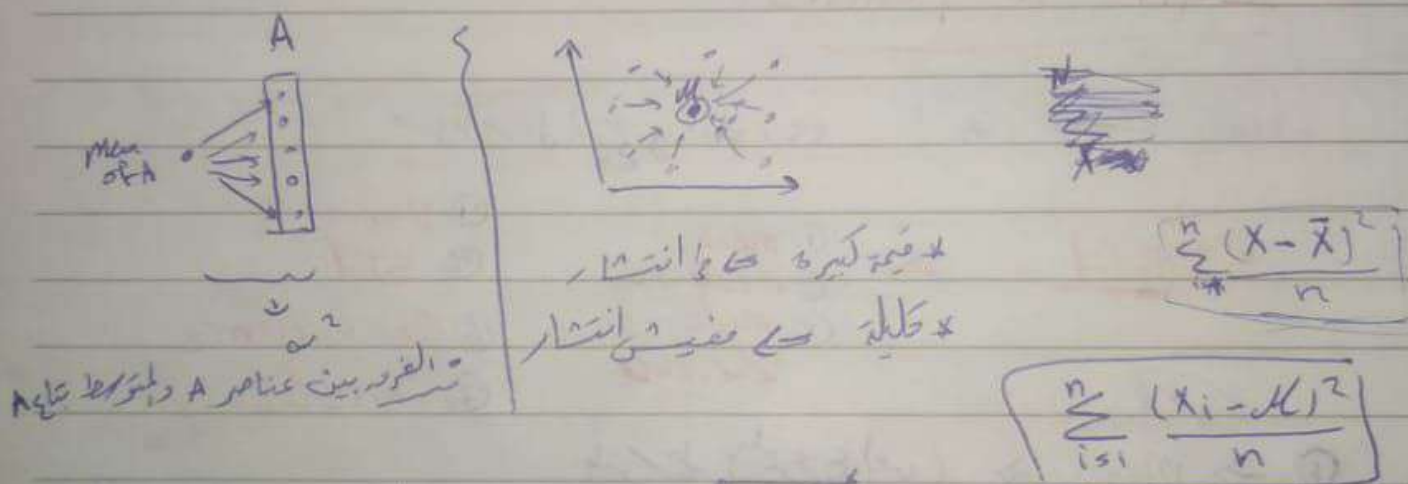
الحل: Median = 3

③ mode : القيمة الأكثر تكراراً في البيانات

يفضل استخدامها لو يتعامل مع بيانات Categorical .
• يختار القيمة الأكثر تكراراً في البيانات .

④ Variance σ^2 : يمكن نقول انه انتشار البيانات حول μ (mean)

التي يبعد عن المنتشر . (# علاقة البيانات بالمتوسط بتأثيرها)



⑤ Std (Standard deviation) : σ

$$\sqrt{\text{Variance}} = \sqrt{\sigma^2} = \sigma$$

Std : أكثر استمرارية مع الأرقام الصغيرة

Variance : أكثر استمرارية مع الأرقام الكبيرة (أي مجالها أكبر)

لإرشين شقيين ونحسب بينهم

distribution : توزيع البيانات وازاي يتطابق

⑥ Co Variance: مشتق العلاقات بين الجداول

مشتق عليها PCA

؟ # الذي يستخرجها feature engineer أو feature selection لا نستخرج
نحكم في العلاقات بين المتغيرات وكذا .

① Strong ⊕ ② Strong ⊖

③ weak ⊖ weak negative

مشتق العلاقات بين الجداول
negative ← R.s. strong
positive ← R.s. weak
R.s. Relationship

حقيقة - عمل زي كذا لما احسب للأعمدة V_1 و V_2 وبعدين
أقوم حنار به في بعين V_1 و V_2 .

لو لقيت طالع كبر ← مترابطين اولاً وبتعدد نوعه اى الظبط

⑦ Correlation:

map



العلاقة مع نفسها Main diagonal

← مشتق ، لباني بعين هو العلاقات بين الجداول و مشتق

Normalized R.s. بين الجداول مصحورة بين $[-1, 1]$

مشتق المتغيرات

Confusion Matrix
Correlation Matrix
Heat map

① + رابط قوي
② - رابط قوي
③ = مشتق رابط