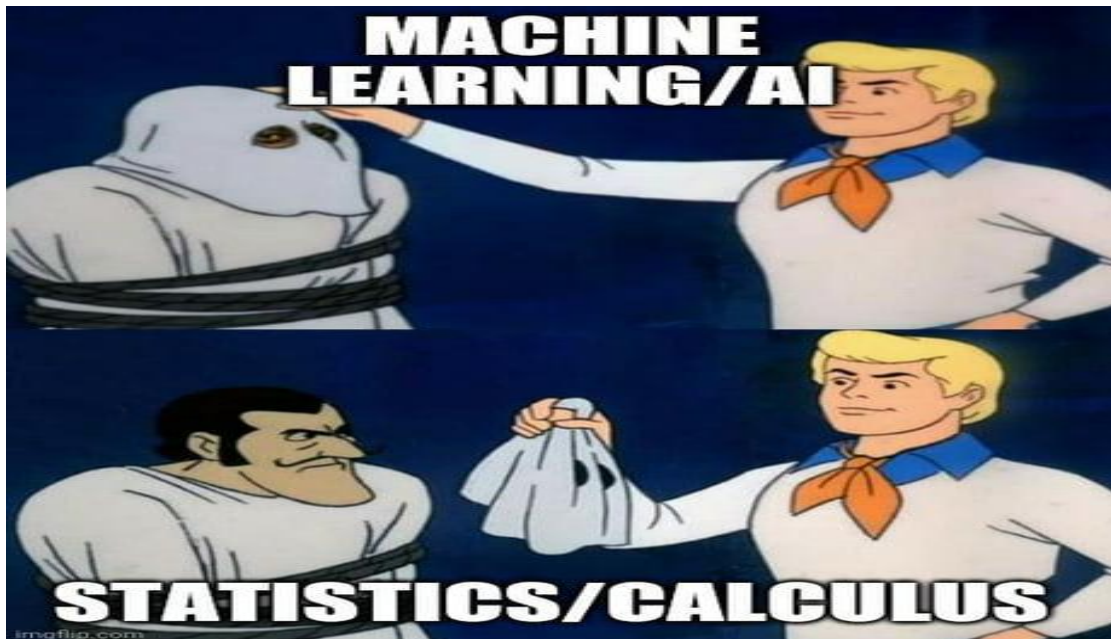


Numpy And Pandas And Data Visualization



:Numpy<=

- ايه الفرق ما بين ال Library بتاعت ال math وال library بتاعت Numpy ؟
- كمية ال functions والحاجات الي ممكن اعمالها ب numpy اكتر
- مش بتدي مساحة للداتا dynamic يعني احنا عارفين إنه اي variable في بايثون بياخذ مساحه كبيره وخلاص لإن بايثون لغة مش typed
- فا Numpy قالتلك انا مينفعش معايا الكلام ده انا عاوزه احدد ال data types الأول عشان كل حاجه تبقى واضحه
- سموها numpy array لإن انت فيها مينفعش تعمل كذا variable في array من different types لا لازم كلهم بيقوا نفس ل type
- وكمان عملوا datatype casting في numpy يعني كل int مثلا فيها , int 8 int 16 , int 32 وهكذا
- طب إزاي بنقدر نحجز مساحات قليله في بايثون ب numpy ؟
 - وده لإن numpy معموله ب cpp
 - و numpy بتخزن الداتا وره بعض في الميموري الى حدا ما
- ال shuffle بيعمل ترتيب عشوائي للأرقام
- في Numpy على الأقل الحاجه الواحده تقدر تعملها ب 3 طرق

- **flatten** ال بتحول اي عدد من ال dimension ل 1D
- **data set** اسمها كده ؟
- عشان احنا علوزين ن indicate إنه مفيش حاجه في الداتا دي متكرره

=> Pandas :

- عيوب ال Tools انها limited مبتقدرش تعمل بيها كل حاجه
- الناس بره بداءت تنقل من **pandas** ل **dask** ليه؟
 - لإن **pandas** عادية , يعني ايه عاديه ؟
 - لو انت معاك **data** مساحتها 30G وانت حجم الرامات بتاعتك 16G هل هتعرف تقراء الداتا دي ؟ لا.
 - **Dask** بتحل المشكله دي عن طريق **parallel processing**.
 - بتستخدم ال **virtualization** مع الميموري الي معاك عشان تعرف تقراء الداتا.
 - وهما بره بيستخدموه عشان يقللوا استخدامهم لل **cloud** .
- **Py spark** ميزتها إنها بتعمل حاجات الداتا ساينس + **ml algorithms** في نفس ال **library**.
- **csv** اختصار ل **comma separated values** يعني **values** مفصوله عن بعض ب **comma** "سبحان الله" .
- **NAN** دي معناها **missing value** , ومش معنى ان ال **value** بتاعت **row** ب **nan** انه الداتا **missing** لا عادي ممكن تكون ولا حاجه من الداتا الي موجوده زي مثلا الداتا بتاعتي ال **values** الي فيها يأما في مطرة او عاصفة او يوم مشمس ممكن ال **value** بتاعتي تكون ولا حاجه من دول ويطلع الجو مغيم "اه زي حياتك بالظبط".

=> Data Visualization :

- هي اني ب **visualize** ال **data** يعني انا بعرض الداتا بشكل بسيط وسهل يخلي الانسان العادي يفهمها
- هل ال **visualization** ينفع أطلع منه **analysis** ؟

- لا المفروض إنني اتجنب ال visualization في ال analysis يعني موصلش لمعلومه عن طريق graph
- وده لإن الجراف في الحقيقة خادع
- الأفضل والأصح إنه بعد ما اخلص استخدم ال visualization في إن انا الحاجات الي انا عارفها أمثلها بشكل أبسط

● عشان اعمل visualization في بايثون عندنا 2 library

1. Matplotlib

2. seaborn

