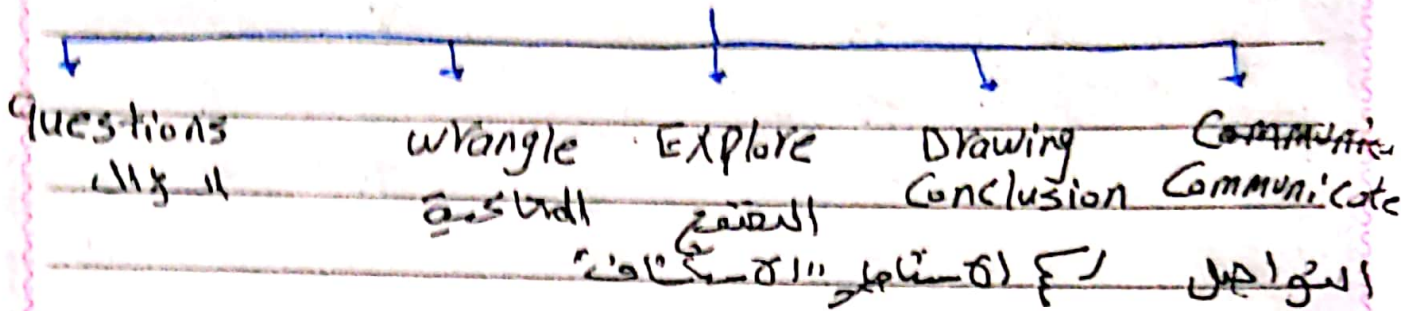
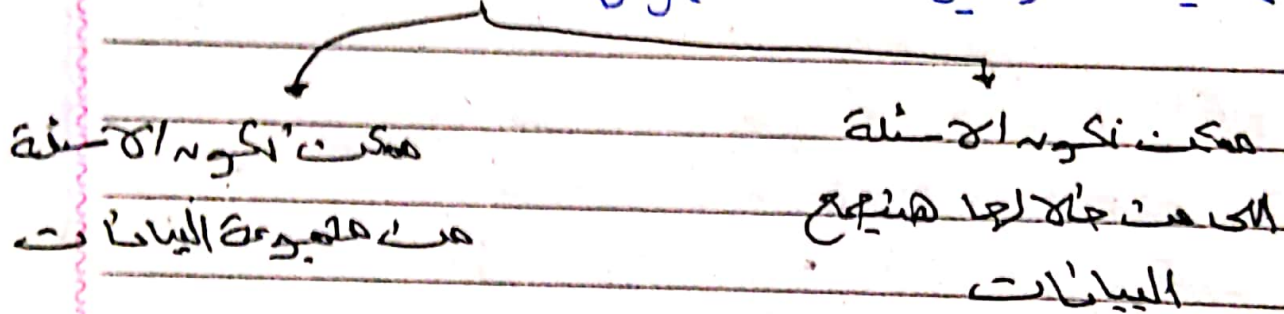


* Data analysis Process *

← نقف الى خمس خطوات



← تبدأ عملية تحليل البيانات، بمرحلة الأسئلة



← بعد جمع البيانات و طرح الأسئلة، نحتاج الى معالجة البيانات، لتساعدهم في الاستجابة عليها، المقصود هنا اننا نتأكد اننا قمنا بحصولنا على جميع البيانات التي نحتاجها لجودة عالية

وتوجد ثلاث اجزاء لهذه الخطوة

1 Gather → جمع البيانات

2 Assess → تقييم البيانات لتحديد جودتها

3 Clean → تنظيف البيانات / ازالة

او ازالة او نقل البيانات التي لا نحتاجها من مجموعة البيانات ذات جودة عالية

في هذا المستند البيانات، أو تحليل البيانات الاستكشافي
هو من حيث المستند EDA وتفرغ بياناتك لتكرار ما بينك
في ملاحظة وفهمها، وفهمها وفهمها وفهمها

← يمثل الاستنتاج على إيجاد المعاد في بياناتك وبناء حسب
عوامله

← بعد استكشاف البيانات، تفسر نتائج التحليل العامر، حيث
وانشاء مميزات جديدة وأكثر دقة من البيانات الحالية
وهي تعرف أيضًا بـ "مهندسة المميزات" Feature engineering

← كخطوة عملية تحليل البيانات، لا ينبغي أن ننسى

← في التحليل، تحتاج إلى فهم الاستنتاجات، أو وضع الشواهد

في تحقيق أهداف الخطوات عند فهمها

→ machine learning, inferential statistics

Descriptive statistics

← في التحليل، تحتاج إلى تحويل نتائج التحليل إلى

لقد تم تحويل النتائج إلى لغة P

← المقايير والمقاييس والخصائص والسمات والسمات =

موضوع الدرس: التاريخ / /

ED A
wrangling
⇒ Gather
⇒ Assess
⇒ Clean
⇒ Explore
⇒ Augment

⇒ CSV "Comma separated values"

* Assessing Data with python *

df.shape[0] عدد الصفوف
df.shape[1] عدد الأعمدة
df.dtypes columns data type
df.isnull().count() عدد القيم غير موجودة
df.nameof.unique() أسماء الأعمدة الفريدة
df.nameof.max() الحد الأقصى
df.nameof.min() الحد الأدنى
df.nameof.std() الانحراف المعياري
df.nameof.count() عدد القيم

df.nameof.describe()['%540']

summary