

Introduction to

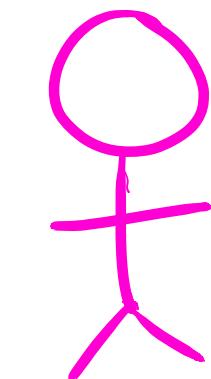
Data,
Analytics &
AI

with Python

What does the "old world" of Analytics look like?

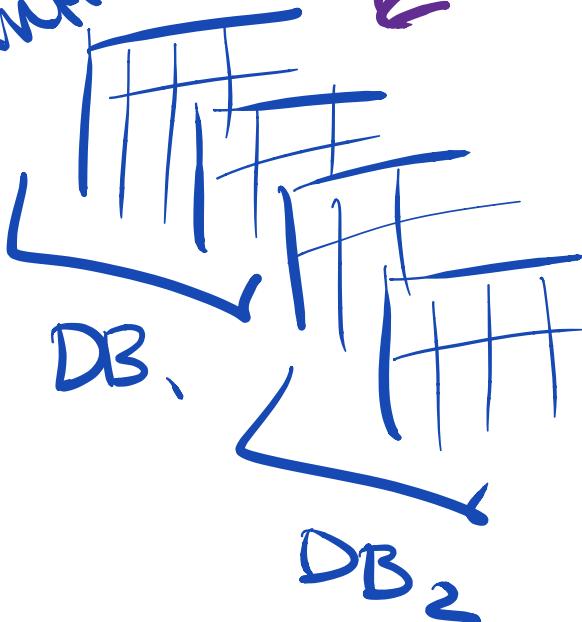
- * Highly Specific (Reporting)
- * SQL
 - Relational DBs
- * Domain Experts (e.g. clinicians)
- * Managers & Politicians, ...
- * Audience: Executives

... & from a unstructured pov?

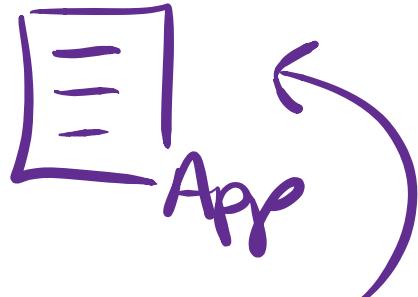


Analyst

Static -
inflexible



no cross ref.



S. Eng

What does the New world look like?

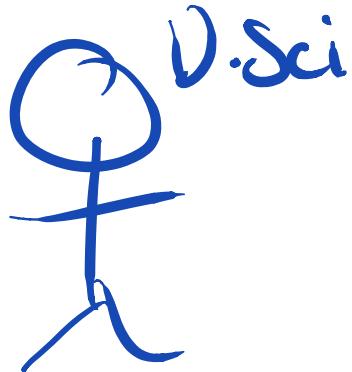
- Aims:
- ④ Technical Integration → consistent
 - ④ Data in useful structures
 - ④ Data available → common
 - ④ Faster Route to Market, Vocal
Delivery --
 - ④ Repeatable → Automatable
 - ④ Sharing of techniques [& code]

... new world infrastructure.

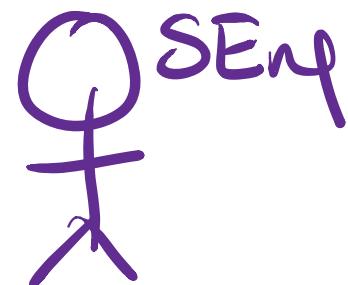
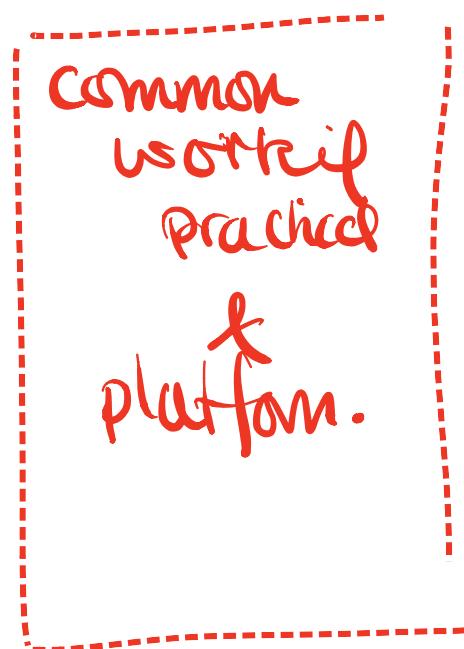
AIM: Democratize data & solutions



D.Analyst



D.Sci



SEng



D.Eng

What tool enables
all these roles

to Share

& work in common :

Python

What are these Roles?

<u>Title</u>	<u>YearsToLearn</u>	<u>Summary</u>
D. Analysts	1+	Querying + Reporting
S. Engineers	3-5 yrs	<u>apps</u> + automation
D. Engineers	3+ yrs	DB Design + Ingestion
D. Science	7+	ALL ABOVE Exp. automated stats & General Problem Solv.

How do these Rules use Python?

Software Engineer : Python

```
name = input()           } user interaction  
age = input()
```

```
if len(name) <= 2 :    } business rules  
    raise Error()      & validation
```

```
else:  
    package = [name, float(age)]  } Processing  
    SubmitToDb(package, "user")  
    DB Use
```

Data Engineer: Python

data1 = query()

} Data
injection

data2 = execProcess()

If valid(data1, data2) :
 VALIDATION ↴

 reshape(data1, data2)

else:

 Raise Error

 Data Process ↴

 Error Cases.

Data Scientist : Python

```
data1 = query1            # data sources  
data2 = query2            # preparation  
combined = process(data1, data2)  
model = fit(LinearReg, combined)    # stats  
if model.predict() == "DEAD":  
    print("WARNING")  
"Software" → interact with user
```

Data Analyst : Python

df = pd.read_csv() ← Acquiring Data

r = df[col].groupby(df[col[2]])

if missing(r): ← Validation
Validation

sns.scatterplot(r) ← Visualization

else:

sns.lineplot(r) ← Exploration

Why not use SQL?

Analysts just want
CSVs (tables) & SQL ...

Why use python?

→ Why?

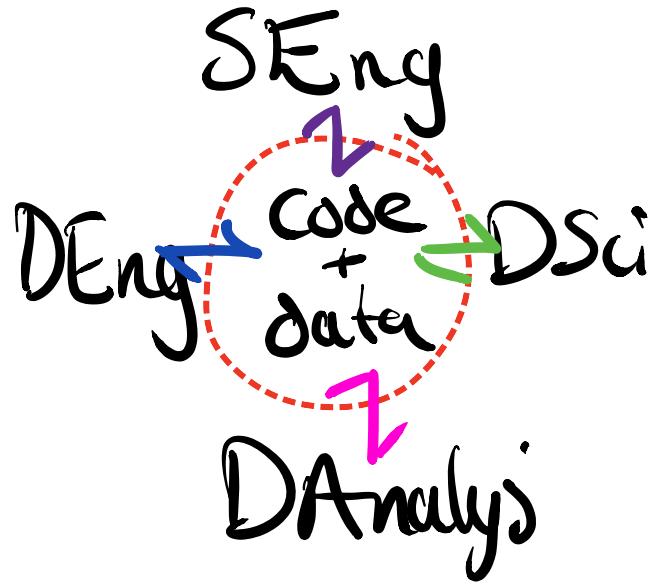
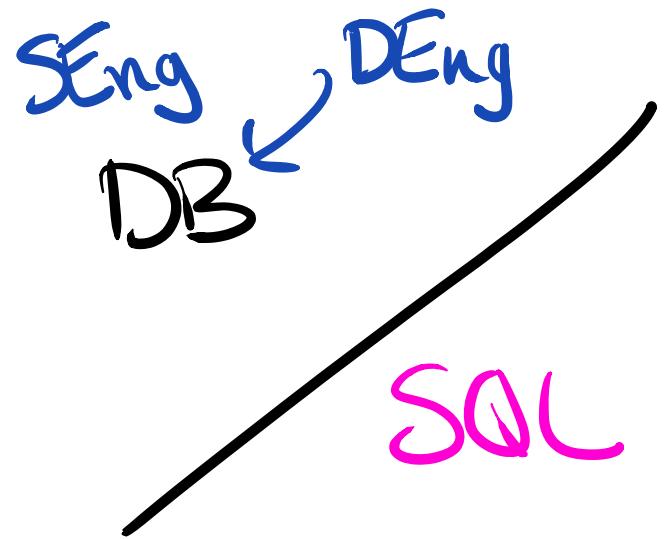
Not to advantage
Analysts

↓ the job is HARDER
with python

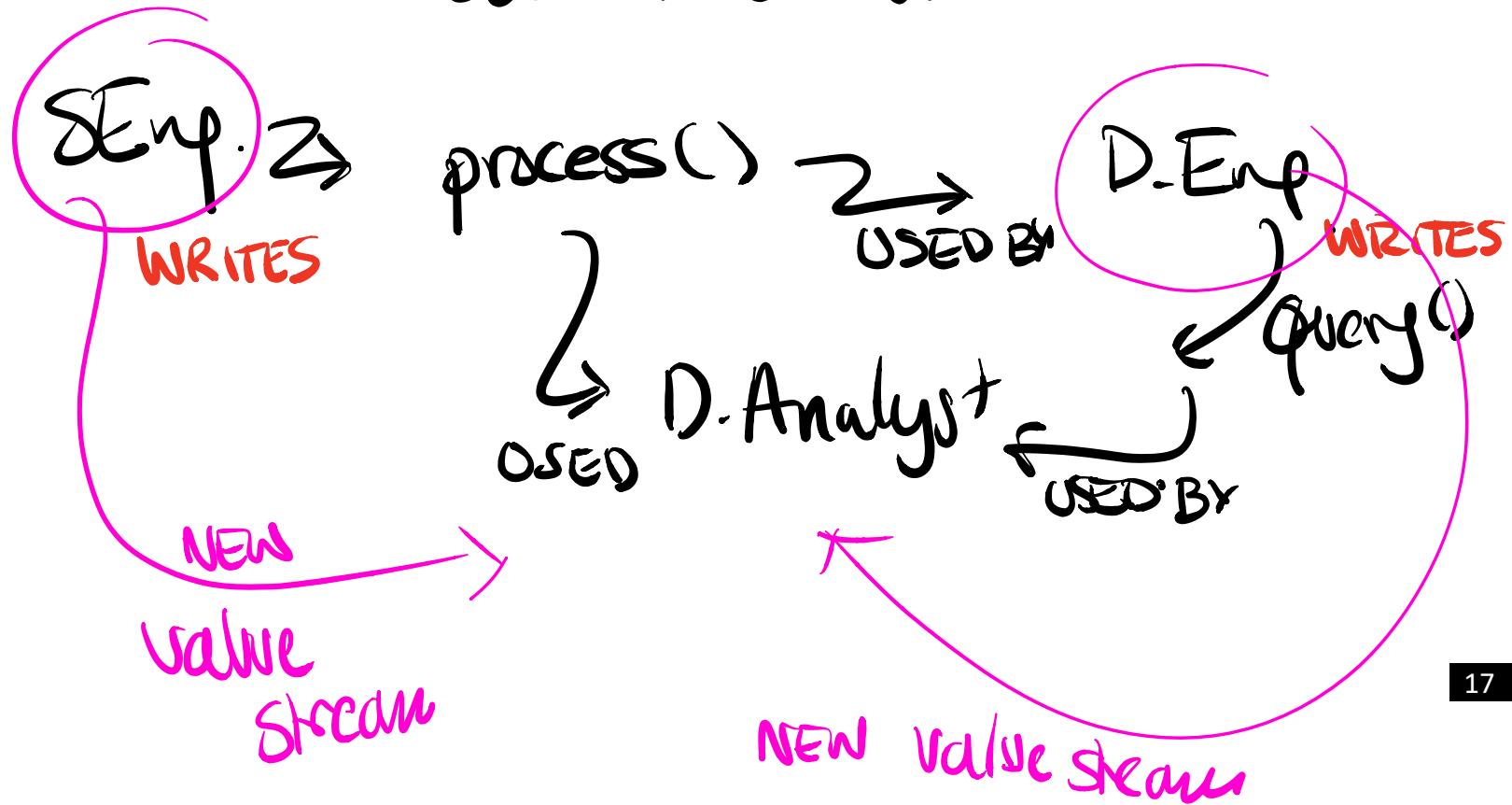
2> But...

You can do more ...

& all nodes can talk
to each other. . .



Eg. This Enables sharing & automation



Problem: Analyst Suffers..

Python Software Eng. hangs.
→ General Purpose.

↓ imposes stronger knowledge required on

- Analyst - -

Learning Python for Analysis

- ① Basic SEng. : if, def, for ...
- ② Basic DEng. : .connect(), ...
- ③ Data Analysis
PROVIDE CODE
- ④ Adv. SEng. : class, ...

SEng.
Provide
libraries

