



## Notebooks got your modern data back

Elizabeth Dlha | Deepnote Data Council Austin 2022



## Agenda

#### **01** Data science notebooks are *purple*

## Agenda

**01** Data science notebooks are *purple* 

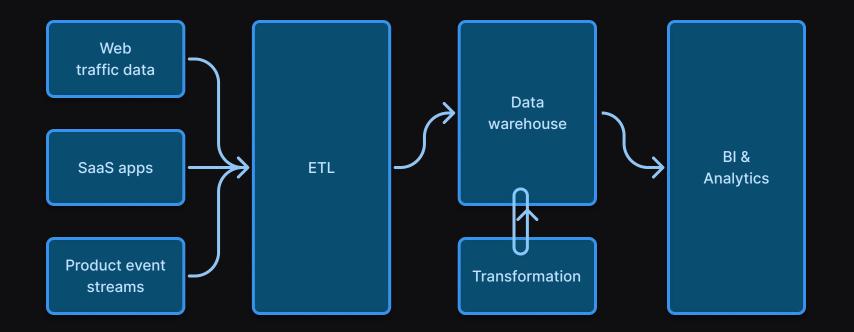
**02** Best practices for working in notebooks

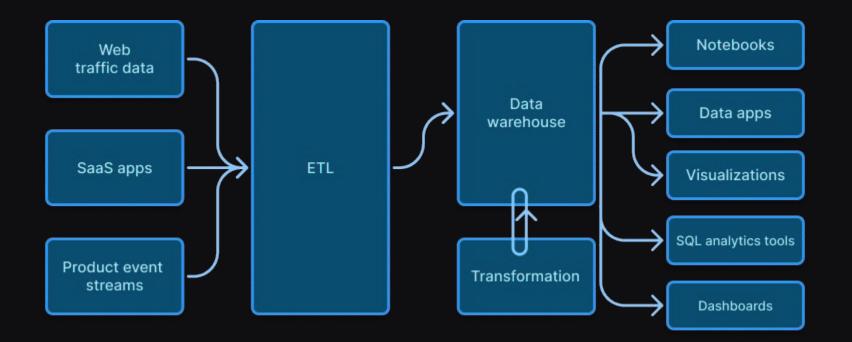
Agenda

Data science notebooks are *purple* 

Best practices for working in notebooks

**o** under the hood of Deepnote

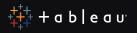




#### MACHINE LEARNING, ARTIFICIAL INTELLIGENCE, AND DATA (MAD) LANDSCAPE 2021 INFRACTRUCTURE - ANALYTICS -MACHINE LEARNING & ARTIFICIAL INTELLIGENCE - APPLICATIONS - ENTERPRISE DATA SCIENCE -DATA SCIENCE ------- MI PLATICIPALS ------CALES ------- MARYTING -> MARYTING -ROT -------COSTOMER EXPERIENCE/SERVICE PLATFORM 100--- 10-0 sdatabricks loöker 🤝 — 🖬 💷 🎲 + + + h le e o 🖓 Frances distance brance articled cather Character Calante Stranssites della Sure sdatabricks DataRobot 🙆 teres () Binder colab OT\_\_\_\_ DRACH B ...... a straight perdets Oursens Oberheinin Sugaran everystered @Australia **10** 12 Cianta On-0-0 mm (2 mm) 10.21 T. Theughtiget Sint. Otre Common H.O.ai Anna Supervision Greene Senar sounds Cases a financia -Barren OPerti Anne peoses 0 - GE 0-0-E-113mot ------ 2-unter 8-----mya e est Constitute Owned T 6 dremio ----- 0----- 0---- 0---Protect Color Concerns (inConcil ATSCALE COR. OF THE MOUNT 1933 THEO F casat tactual 🧕 Anosity & Omenan A Descarte Jethre Construction ASAPP Quele afert metter comme Queuet ----Total Interest Etunicae in historic Ginnen Mannie Stolegt Pinnen O Observable + abuier 201 Williacons source? Judge Gausse Later Livers Hyperts CRESTA diment and Haribarral Domar () course 0 Pachyderm 😳 Octore, biguazio stradigis 10P1 an and a state Protei Manuria enigena wiAmus Manny LARY County -----(PERSON THE AVER THE AVER THE AVER THE AVER THE AVER Minuthatage Atentic Scillo Spread Warners III Quanta 58.mm B HITTART Atlant Aladrah Excitives dation Sumi Scone observes LEGAL REGECH & PRUNCE AUTOMATION & RPA SECURITY -DATA ANALYST PLATFORMS DATA GENERATION -----Annalis ann Uper cours gran Channel Channel Canada Sanata BEALTIME - GRANGERS MISMENTED ANALYTICS Annual Control Comment Oling Page 0 ---- 0 Parts 10 man 540 ------Non CODUL Anima Net analasenstatututut denive 190100 8-0000 DUCTION A OPCCOV Microsoft @periodes alteryx ThoughtSpot. **O** SingleStore pice nation manage while manager and the descent F.3 ----IBM Same Termine 127 ---O---- Pivetol O inaly Datameer 🚽 📰 👩 anod?t -outliker colla laswork sopper Page dego 🌑 🙆 Google Claut Arthur 10000..... TILTHE ACCOUNT ADDRESS dotobricka PWITNEISHIPS - OneTrust ALCOPENIN Viere -----ROOKSET CRACLE -biense Xerentia Gens upermost ubeles Officier - ----------------- 900 -----8 1 1000 B 100 0 ...... CHARLE Marklight @Coathase on tearrol apice ATTIVIO al reserves bulleter Open stress a person de la constante ..... ENDOR S ..... YOUTH Californe 9: 4 introduce -cortex -----Gamma ALE Sabolo SFRANT ALE ----truera (Trail Seman gumus @ebee efocet machage S. SELDON BERTEN SCYLLA. 8 Mail sector ----incorto. Charles Standards Second Standard ∧ orize O alterna weighthound obviewsty ai 🎝 Aintoble diff.or -----ALACCONTRA A CRATE.IS WHYLARS 4.98 4 £ ...... MPP DEs REVERSE IT. --- DATA INTEGRATION ------SYNTHETIC MEDIA ADVERTISING \_\_\_\_\_ REALESTATE \_ GOVY A \_\_\_\_\_\_ COMMERCE \_\_\_\_\_ FINANCE -LENDING \_\_\_\_\_ INSURANCE -\_\_\_\_\_ splunk> 🥙 Elebrand Anne Canada Ca O DEDREAM AND THE Census 227 ----- 2 ----- @ -----exondr MedaMath REDFIN affirm A Hoveto Mabe talend atteyo (a stants (Distants) M metaphor Official Operation SHEIN FALLER G Contras 8--- 0 MM 8 mm 8 mm # G----- D-ID. criteo lastres. Distan OVTE Care ZEST® Annak Bern VERTICA hightouch Torren. Hunter Francisk scalage IEM atlan Trees Bond Anthenion -----Trace C. Sunthesis D dans I town and the second +KNEWTON Orchard PV8 -----Conners de terre ----men detti sonu Abline Formanippiers ----Renative PRIMER etropio 👧, sauranda Hoono AUpstert SOOLHER Sintice 9990H D----(D-money Amoune Chil Channe Barris & Superprain O despondie Phone man Summan M presspiero Geclara FINANCE - INVESTING Ballin Arthony & Anderson The sendbard be Oction steama America and Star Share Organization Streamy ZALENI Impart.in DATES The Sheptie Coopina 4 olbert pumpum Sector Parablete Avenue AYASDI KENSHC Omercie Ballio V HISTORY e' transform + SELECT STAR de genantit met an LD maner Koonina Seale porte DEPT supertone @ -----Communic Plaquetures atlan Arounden Informatio securitor @ futfile Esereda di Castor IQL Option of Parliances and Children Children Children Evasal The second strategy (ALLER AL MINACYA -HORIZONTAL AL ----OPUIDES & CLOUD --- ALHARDWARE-LIFE SCIENCES TRANSPORTATION MINT/MONDENG ----HEALTHCARE \_\_\_\_ AGASCIETURE - INDUCTION Netron Status Sacrage Tearing & USER TEALS M AVTUA SIEMENS COM Pottolin ------Ownert: MAN TRM kinetica Geogle TPU arm A HERE GOODE SECUR With Watson da OpenAT O Devider -----this cases hims (Bennin utreas Ratte 0 steen the ground ground ground 0- -----O Paperapace CONTRACT CHAPHCON Statute 2044 ----- with anima diameter dameter anima algolia coveo subidua States Casto Base Stoweds States sections Habitat Dallas Company 0-ma 📑 1.0.000 044 ----- 0 mill 0-- 9000 mill 0---- .0. A 1634 5 Granular State - institute 20 The I ----100mg A 1111 Lucidworks Oswiftype ATTIV/O A.Marrenta 🔛 natinger Petrone toreans Arm man Area Official GT (R. REALENSE Siles Elbert DMA GUALITY >Cribt @Magain distance ahans Share Come HAILD (One phones little surgest Con Barrers Y FORMA BENERA Opennie Schwarzen Unter Ferneren Unter Schwarzen Unter Ferneren Unter Schwarzen Unter S skoflow (C) ratend some Grove generating 5 zongs nuclo 0--GERTER COLORED IN Affective -----packet services Piston duetto Characteristics of College addressings Secondarian Openicy CT Privary Departure Pueter P II NI254 Christoff Sealers Distance Change State and Annual State Pressent O States One de ---- 2 byte | August Anna anne anne ----WARADA MAANA & QUICKWIT CHAOSSEARCH B comine C Kabrts ---COID Hanta OPEN SOURCE FRAMEWORKS -FORMET - OUFPY | DATE ROW - DATE -MILLINGES CONTRACT CO References and States, second and second to a second to be the second se Antonia 1907 Antonia 19990 Esgan Oriente Year State Sat where So and the state Brist SSN TE25 Oktowers and Tanta Alanta Salaria Di cara Anaran (1) g----- 4---- 10200# • Yum Down William Beener bokeh @ O second A regts. Denne \varTheta ellie 5- 0 0- W. 1 op Land Land Contra C a 🖻 🛴 4 mener \$----- HELDK - DATA SOURCES & APIK ----- DATA RESOURCES -IATA MARKETPLACES -FINANCIAL & ECONOMIC DATA AR/SPACE/SEA LOCATION INTELLIGENCE OTHER -DATA SERVICES INCUBATORS & SCHOOLS -RESEARCH PEOPLE / ENTITIES -Boorberg Ci rescourses D | ore mers Quark Differen Z merite action thereas rounsquare Omaphox america actional O menter Addedtes Aster . MADELLEY SHOW MADELLET AND C COMPANY BACK BACK | Marridge @ OpenA1 Gouge Research Sacebook research C PLURAL DENT O SENERAL ASSEMBLY C DataCarre 10 - 0 - 2 Jacks main Singleton Standard Piccold Gesti Mana Ander Angelary MILL VALUE AND DEL MO CRUX Souther Gamers xgnite Emer enricht preise au mei Trans Quercast Charty Startister Incole Electrifi Instales 701 A Datacolo galvaniza (P) wmv "mattert AIZ ANTHROPIC Salk I X-martine martines. S--- ()(102) marches Contactions Output Sources Dennet melissa Sause CURDIQ Reportsonate Visitione entracore DataRinit intoncos The Desireday FIRSTMARK © Matt Turck (@mattturck), John Wu (@john\_d\_wu) & FirstMark (@firstmarkcap) Version 3.0 - November 2021 mattturck.com/data2021

SOURCE: MAD LANDSCAPE BY MATT TURCK







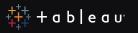
Power BI

Ő











Power BI









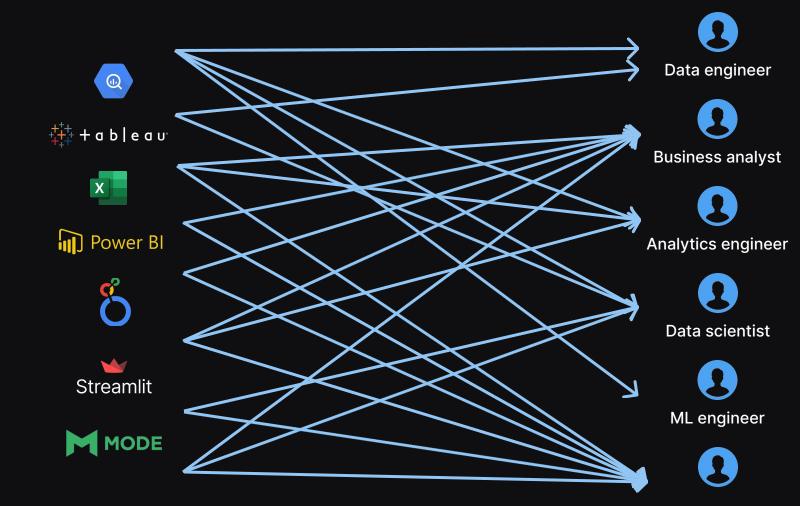
Business analyst

Analytics engineer

Data scientist

**M**L engineer





6



# 1. How can we *retain & share* organizational knowledge?



SOURCE: WE THE PURPLE PEOPLE BY ANNA FILIPPOVA



## Tech people (data producers)

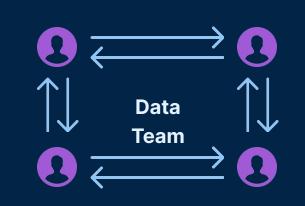
## 🤎 = Business people (data consumers)



**\*** + **\*** = **\*** 

#### Engineering

Data Engineer



#### Marketing

2

Ω

Marketing analyst

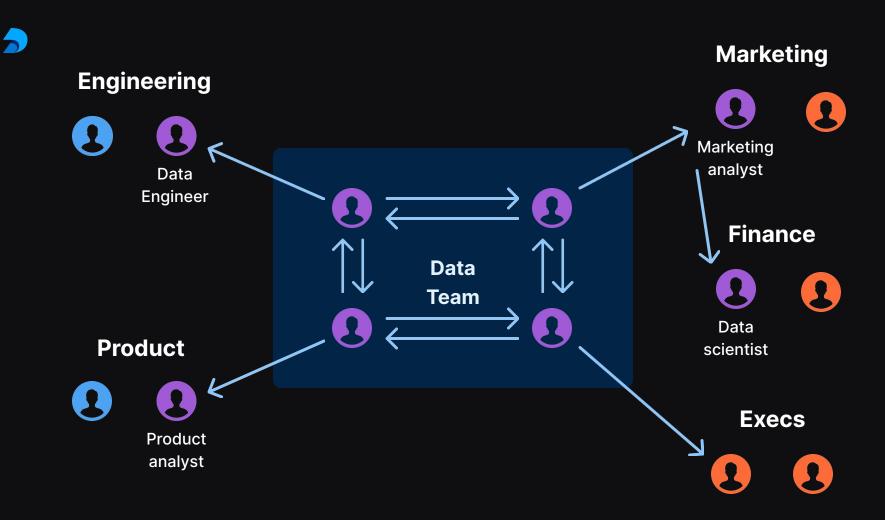
#### Finance

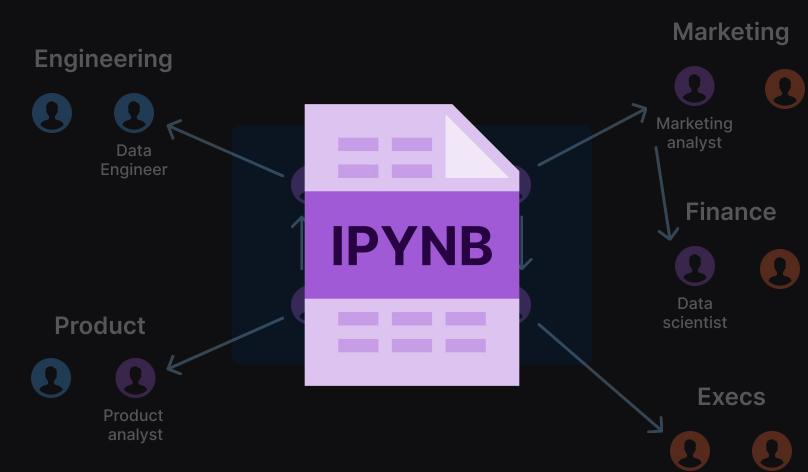
Data scientist

Product

Product analyst

Execs







# Notebooks can be the *purple interface* for the *purple people*



#### Notebooks 1.0



## Explainable, $\checkmark$ transparent Shared context all in one place

 $\checkmark$ 



## 2. How can we use notebooks *collaboratively?*

• Follow style guides, write documentation, write tests

• Follow style guides, write documentation, write tests

#### **02** When training models, log all experiments automatically

• Follow style guides, write documentation, write tests

#### **02** When training models, log all experiments automatically

#### **03** Split development and production environments

• Sanitize user data in development environment

• Follow style guides, write documentation, write tests

#### **02** When training models, log all experiments automatically

#### **03** Split development and production environments

• Sanitize user data in development environment



• Follow style guides, write documentation, write tests

#### **02** When training models, log all experiments automatically

#### **03** Split development and production environments

- Sanitize user data in development environment
- **04** Parametrize your notebooks





## 3. How do we make notebooks more *purple*?







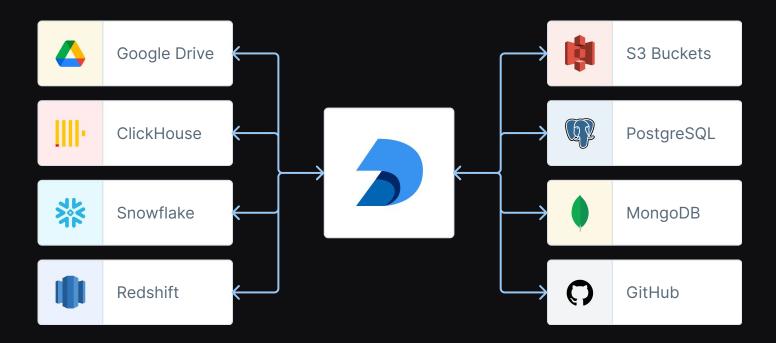
## Start with the *blue*

### Language interoperability

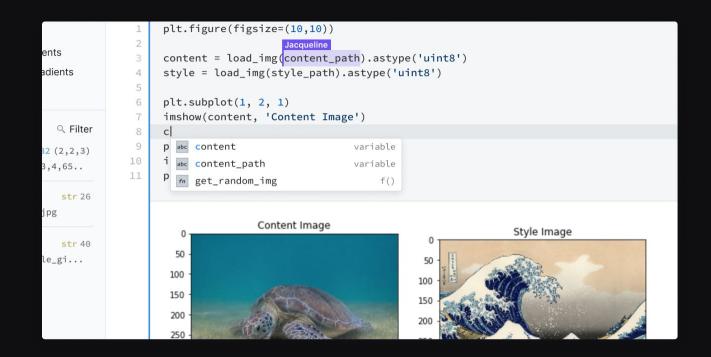
#### COMBINE PYTHON AND SQL

sta	tion_name $\frac{\exists p}{d-1}$							
В	UFFALO NIAGARA	INTERNATIONAL 🗸	0					
** :	snowflake_demo	Saved to variable df_1 •						<mark>ال</mark> ال
FRO	OM demo_db.pub	ne, temp as mean_temp, wds blic.ny_weather			ain			
	ERE date > '20 DER BY date DE	015-12-31' and country = '	US' and name = {{	station_name}}				
	MIT 50	100						
0	<u> </u>					Preview	•	🗠 Visualize
			<pre>mean_temp float64</pre>	<pre>wind_speed object</pre>	<b>rain</b> obj	ect		
	2020-08 2% 2020-08 2%	BUFFALO NIAGARA IN 100%	64.2 - 82.0	5.4	0 1	00%		
	48 others 96%			4.3				
0	2020-08-18	BUFFALO NIAGARA INTERNATIONAL	64.6	4.3	0			
1	2020-08-17	BUFFALO NIAGARA INTERNATIONAL	71.3	5.4	0			

#### Integrations



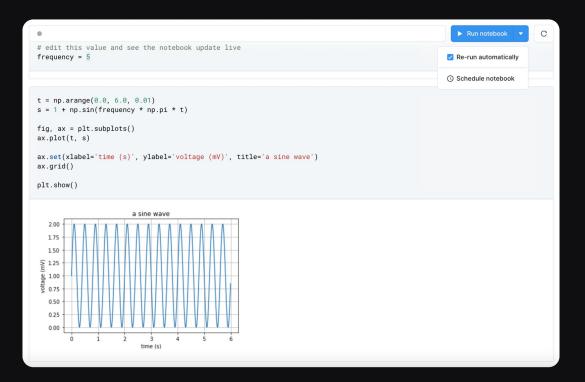
#### **Code intelligence**





## Mix in the *red*

#### Reactivity



#### Seamless editing (WYSIWYG)

#### The new editing experience

Gastropub chillwave umami lyft. Poke austin direct trade, marfa raclette letterpress actually. Chartreuse sriracha pinterest twee lo-fi try-hard. Meditation banh mi kitsch, prism organic hot chicken literally heirloom occupy af semiotics food truck.

Aesthetic asymmetrical gluten-free, health goth shaman meh bespoke kinfolk helvetica vaporware fashion axe freegan. Pour-over hammock succulents disrupt chartreuse raw denim. Brunch aesthetic

	-				•				• •					ccie plaid pork
В	i	<u>U</u>	÷	$\odot$	<u>-</u> •	Ì≡∙	≔∙	<	<u>لمة</u>	Θ	D	$\sim$	$\overline{a}$	o. VHS blog

pickled scenester venmo hashtag lo-fi.

#### Data apps

company\_name 🗄

#### **SEC Annual Reports**

Look for annual revenue for a company in a given fiscal year.

Companies pick different names for the measure to file the number, so we look for all measures containing the word "revenue".

NVIDI	A CORP		Apply					
🔕 Data	Warehouse • Sa	wed to variable filte	r_annual_report ▼					
fi me FROM bi WHERE co	mpany_name, .scal_year, .asure_tag, .gquery-public-c mpany_name = {{ ID LOWER(measure)		r_financials.quick_summary nue%'					
	company_name o	2012 - 2012	measure_tag ob NumberOfCu20% Revenues	i⊻ Visualize				
0	NVIDIA CORP	2012	NumberOfCustome rsWithSignific					
1	NVIDIA CORP	2012	DeferredRevenue Noncurrent					
2	NVIDIA CORP	2012	Revenues					
3	NVIDIA CORP	2012	Revenues					
4	NVIDIA CORP	2012	CostOfRevenue					

#### **SEC Annual Reports**

Look for annual revenue for a company in a given fiscal year.

Companies pick different names for the measure to file the number, so we look for all measures containing the word "revenue".

company\_name

NVIDIA CORP

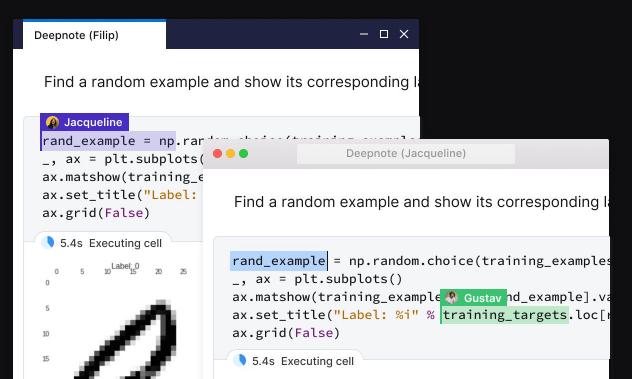
		<i>a</i>			
	company_name o	fiscal_year int 2012 - 2012	measure_tag obj   NumberOfCu20%   Revenues20%   6 others66%		
0	NVIDIA CORP	2012	NumberOfCustome rsWithSignific		
1	NVIDIA CORP	2012	DeferredRevenue Noncurrent		
2	NVIDIA CORP	2012	Revenues		
3	NVIDIA CORP	2012	Revenues		
4	NVIDIA CORP	2012	CostOfRevenue		
5	NVIDIA CORP	2012	NumberOfCustome rsWithSignific		
6	NVIDIA CORP	2012	AllocatedShareB asedCompensati		
7	NVIDIA CORP	2012	NetWarrantyChar geAgainstCost0		
8	NVIDIA CORP	2012	DeferredRevenue Current		
9	NVIDIA CORP	2012	RevenueFromSign ificantCustome		



## Dive into *purple*

## Collaboration

#### **REAL-TIME COLLABORATION**



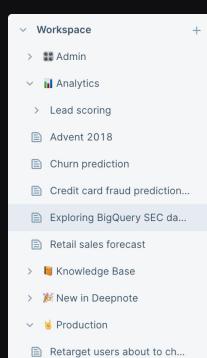
## Collaboration

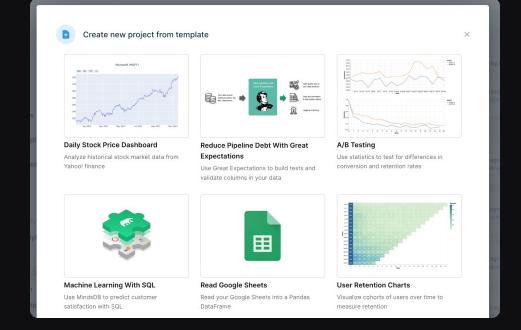
#### g = E 2 stocks=pd.read\_csv('stocks.csv') stocks Show 2 resolved threads Allan Campopiano (You) 5 months ago 30 date object AAPL float64 **AMZN** float64 GOOG flc 2000-01-01 .... 0.8% 7.07 - 223.02 5.97 - 135.91 102.37 -Can you transform this data into long form? I'll start 2000-02-01 .... 0.8% writing the code for a chart. 121 others 98.4% Sam Salvatore 5 months ago 0 2000-01-01 25.94 64.56 nan Absolutely I can do that! 28.66 68.87 1 2000-02-01 nan 2 67 2000-03-01 33.95 nan + Add comment $\Re + \sqrt{C} + C$ ✓ Resolve 3 55.19 99.95 28.37 2000-04-01 31.01 nan

COMMENTS

### **Knowledge organization**

#### **FOLDERS & SHARING**





**TEMPLATES** 



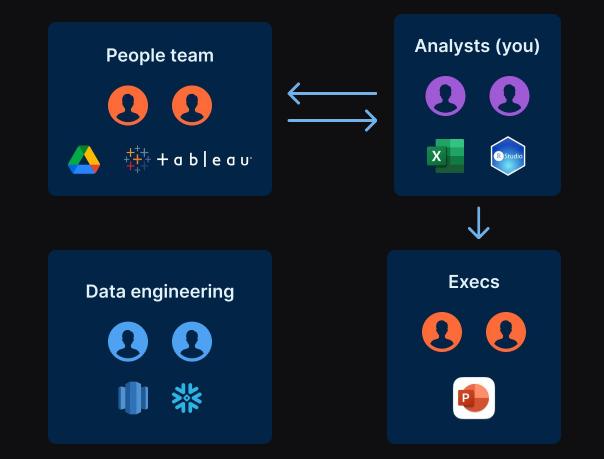




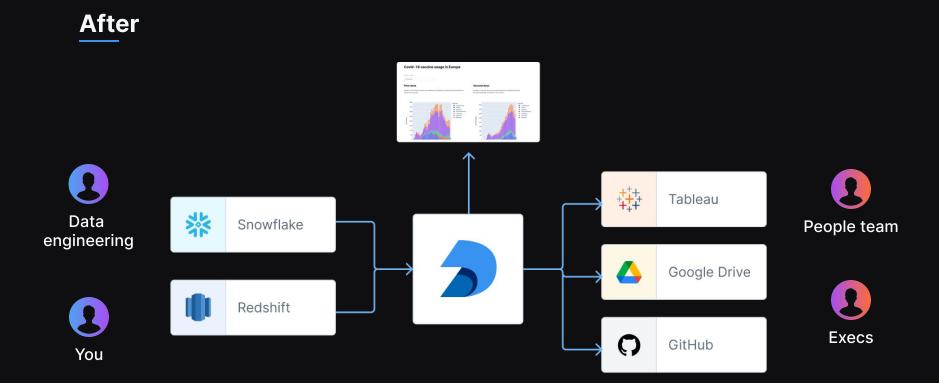
# 4. What does this *purple* world look like?



Before









#### Outcomes



Streamlined workflows



Reduced time-to-insight by breaking down silos



Hues of purple everywhere



## **Questions?**

Elizabeth Dlha elizabeth@deepnote.com







#### Case study: How Gusto scales their data capabilities

You're an analyst on the People team, helping your company make data-driven decisions about workforce.

Questions such as...

- 1. What are the key drivers of attrition?
- 2. How does remote impact employee well-being?



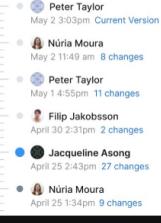
### Versioning

Let's create methods that will allow us to load and preprocessing process as are expected according to trained on image with each channel normalized by me channels BGR.

def load\_and\_process\_img(path\_to\_img): img = load\_img(path\_to\_img) img = tf.keras.applications.vgg19.preprocess\_ir return image

In order to view to look at the outputs of our optimiza step. Furthermore, since our optimized image may tal must clip to maintain our values from within the 0-25!

def deprocess\_img(processed\_img):



### **Productionize instantly**

#### SCHEDULING

