Python y ciencia de datos en la ingeniería, un enfoque practico

Luis Enrique Navarro Morales

Universidad Nacional Autónoma de México
SPE Student Chapter

Curso realizado con el apoyo de :

ORKAHUB ENERGY

Agenda
27 – 29 Enero 2021 -> 17 – 18:30hrs

- Dia 1
  - Introducción
  - Entorno de programación
  - Datos a disposición
  - Live Coding
- Dia 2
  - Live Coding
  - Ciencia de datos
  - Representaciones graficas e interpretaciones

- Dia 3
  - Relevancia
  - Conclusiones

Repositorio del curso

- Github

LuisNavarro93/SPE-UNAM-Python-y-Ciencia-de-Datos: Curso de capitulo estudiantil SPE-UNAM (github.com)

Introducción

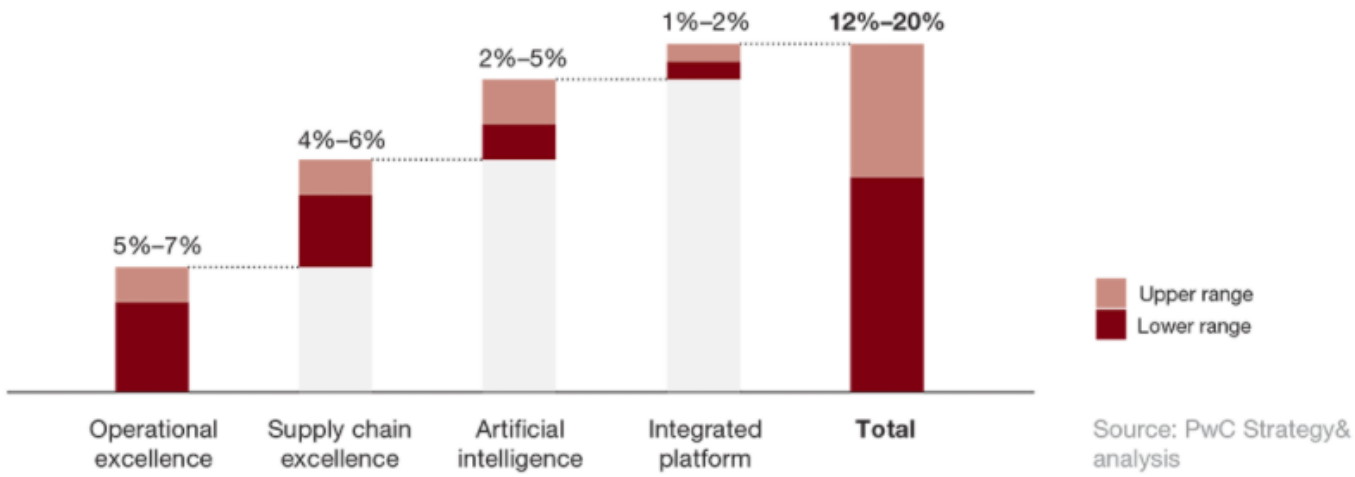
strategy&
Part of the PwC network

Industries Functions Solutions Insights About us Careers

Home › Insights › Drilling for data: Digitizing upstream oil and gas

Drilling for data: Digitizing upstream oil and gas

Efficiency increase from digitization (% saving on total operating expenditure)



[Drilling for data: Digitizing upstream oil and gas | Strategy& (pwc.com)]

Introducción

IHS Markit

SOLUTIONS

Energy & Natural Resources Research & Analysis >

Data: Is the oil and gas industry's most valuable resource being overlooked?

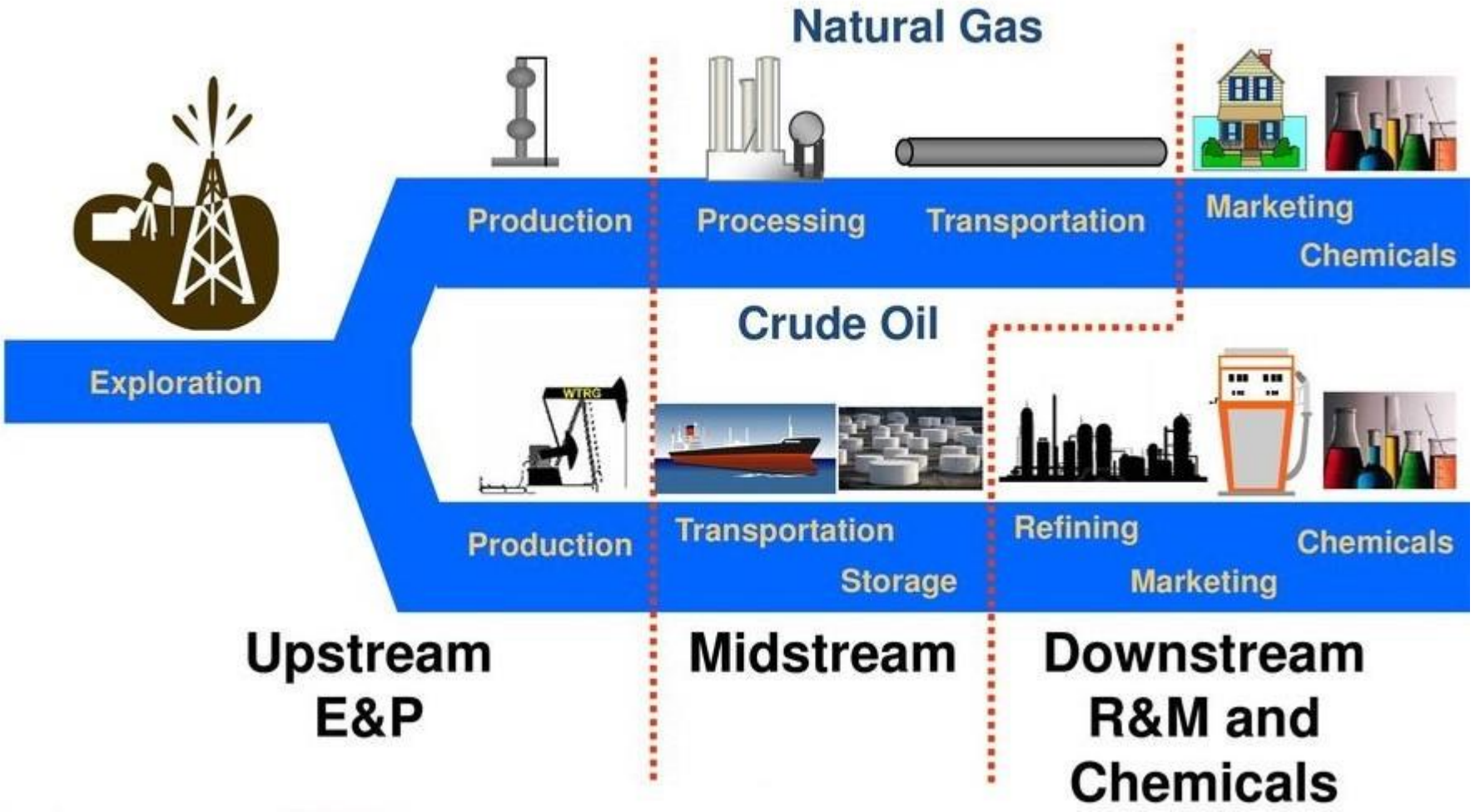
21 February 2018 | Nathan Amery

The typical technology architecture at oil and gas firms consists of four separate systems: ETL (extract, translate and load), MDM (Master Data Management), a data warehouse (or store) and some kind of analytics tool. As a result, there are multiple handoffs between disparate databases and systems with data often being manually moved between applications, and, in many cases, through an intermediate step, such as Excel.

Not only is this multi-application environment costly and cumbersome to maintain, but it also increases the potential for error in the data and introduces complexity in identifying, tracking and fixing data. Business users may be able to generate visually appealing reports in their business intelligence tools, but how accurate are they? Can users trace the source of data, the changes made to that data, by whom and when?

Data: Is the oil and gas industry's most valuable resource being overlooked? (ihsmarkit.com)

Introducción



Introducción

Society of Petroleum Engineers

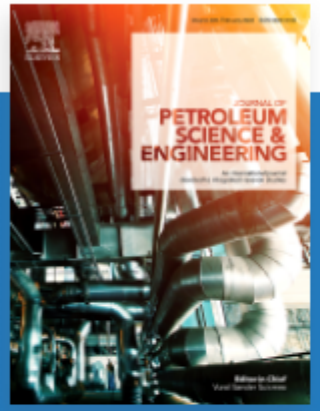
SPE-193776-MS

The Role of Big Data Analytics in Exploration and Production: A Review of Benefits and Applications

Christine I. Noshi, Ahmed I. Assem, and Jerome J. Schubert, Texas A&M University

It is undeniable that Big Data performs a progressively vital role in the E&P value chain. With the continuous acquisition of real-time data, huge volumes of data are being assembled from drilling, logging operations, seismic surveys, and production. Accordingly, valuable information is currently accessible from high frequency surface and downhole sensors. Web-based monitoring pipelines and platforms for instantaneous surveillance of producing assets are being deployed. These assembled workflows support field personnel and engineers with their day to day monitoring and surveillance operations for reservoirs, production systems, wells, and various fields. Not to mention, helps faster team collaborations for faster decision making. The entirety of collecting, processing, and analyzing metadata is known as analytics (Bravo et al. 2014). Data Analytics allows the extraction of maximal value from data through trend

Introducción



Journal of Petroleum Science and Engineering

Supports *open access*

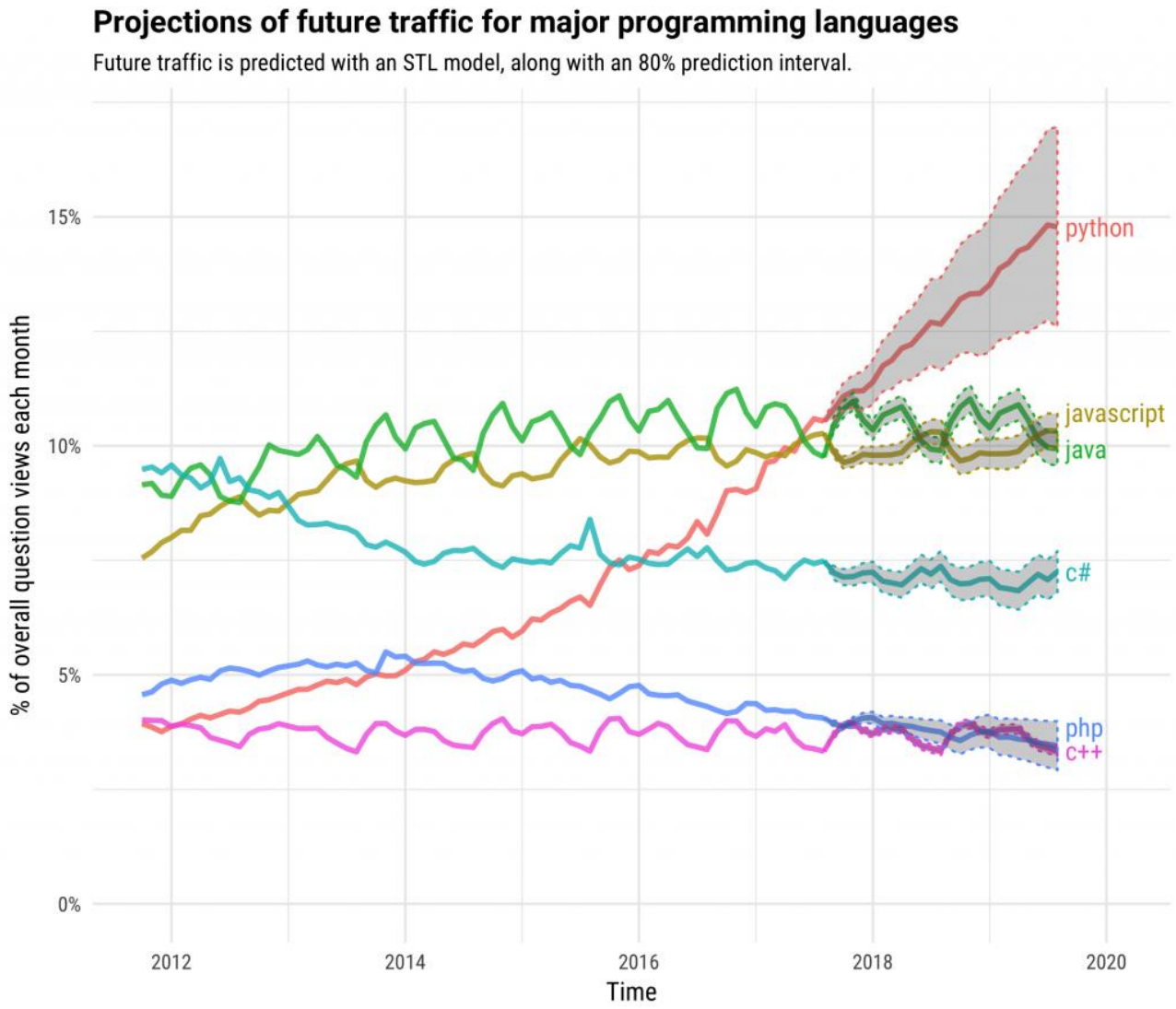
Articles & Issues | About | Publish | Order journal | Search in this journa

Petroleum Data Science and Machine Learning

The Special Issue focus will be on machine learning (ML) experimentation and operationalization ML experimentation refers to the efforts centered on data preparation, algorithm selection and model validation and verification. The issue will accept submissions on the following topics, as applied to petroleum engineering problems: data preparation, descriptive analytics, predictive analytics, prescriptive analytics, hyper-parameter tuning, and automated ML.

Entorno de desarrollo

- Python



IDE “Interface Development Enviroment”

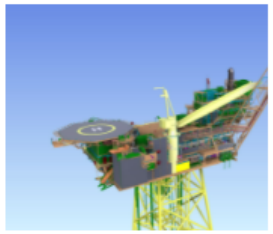
- Live Coding
  - PyCharm
  - Spyder
  - GoogleColab
  - Jupyter Notebook
    - Fundamentos

Datos disponibles

equinor

Forum FAQ Luis Enrique Navarro Morales

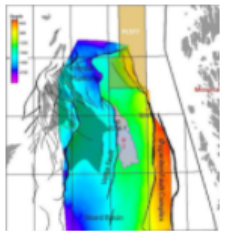
Find open data



Plant 3D models

01/11/21

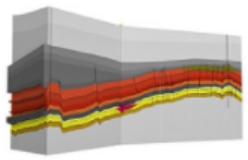
Equinor has a large portfolio of 3D models representing all our topside assets. To enable innovations within new ways of working and also use of 3D models within learning institutions, we will now share the 3D model of our Huldra asset.



Smeaheia dataset

11/03/21

The Smeaheia dataset is a reference dataset containing subsurface data, reports and geomodels related to assessment of proposed CO2 storage sites in the Smeaheia region, Hordaland Platform offshore Norway.



Northern Lights

28/09/20

The Northern Lights partnership releases all relevant well data and reports from 31/5-7 CCS verification well in exploitation license EL001.

Read more Northern Lights Homepage

Sleipner CO2 reference dataset

03/02/20

Equinor and its partners will disclose datasets from the Sleipner field; the world's first offshore CCS plant, in a push to advance innovation and development on the field of CO2 storage.

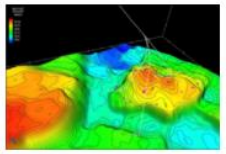
Read more

Hywind Scotland operational data

28/11/19

Equinor and ORE Catapult collaborating to share Hywind Scotland operational data.

Read more

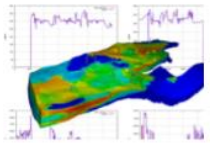


Volve Data Village

14/06/18

For the first time all subsurface and production data from a field on the Norwegian continental shelf (NCS) will be disclosed.

Read more



Norne benchmark case

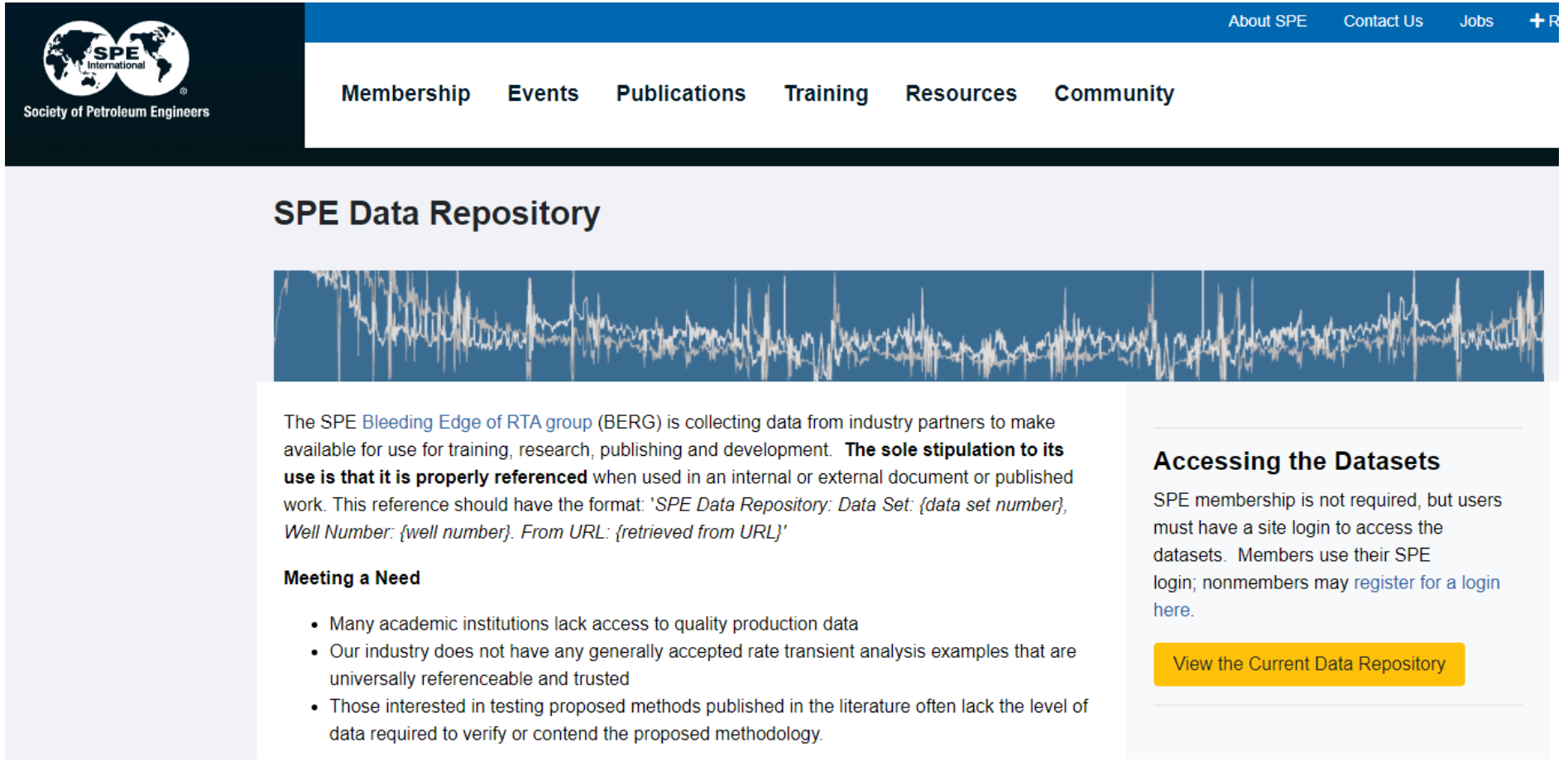
23/04/18

Equinor and the the Norne license partners has released a subsurface dataset for educational purposes.

Read more

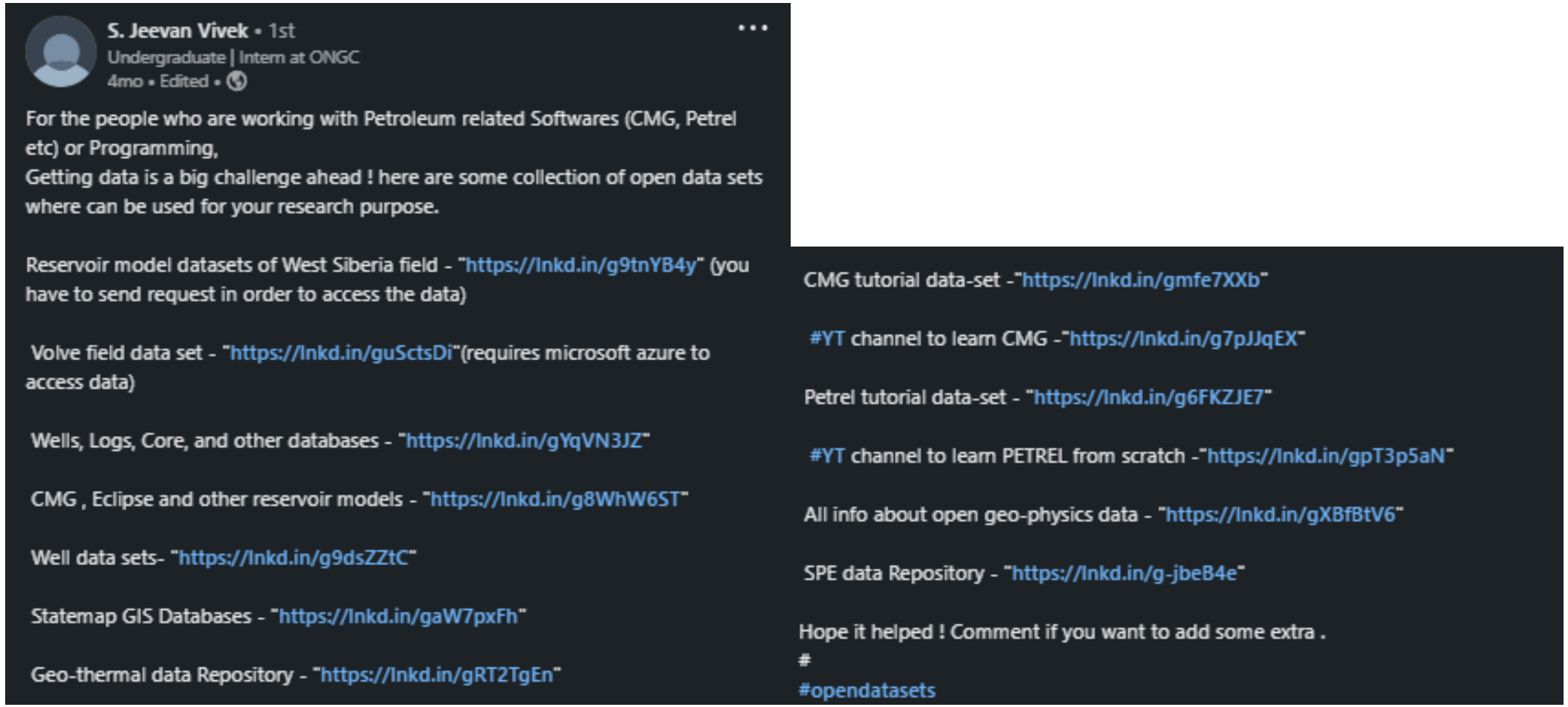
[Equinor Open Data]

Datos disponibles



SPE Data Repository

Datos disponibles



S. Jeevan Vivek • 1st
Undergraduate | Intern at ONGC
4mo • Edited •

For the people who are working with Petroleum related Softwares (CMG, Petrel etc) or Programming,
Getting data is a big challenge ahead ! here are some collection of open data sets where can be used for your research purpose.

Reservoir model datasets of West Siberia field - "https://lnkd.in/g9tnYB4y" (you have to send request in order to access the data)

Volve field data set - "https://lnkd.in/guSctsDi"(requires microsoft azure to access data)

Wells, Logs, Core, and other databases - "https://lnkd.in/gYqVN3JZ"

CMG , Eclipse and other reservoir models - "https://lnkd.in/g8WhW6ST"

Well data sets- "https://lnkd.in/g9dsZZtC"

Statemap GIS Databases - "https://lnkd.in/gaW7pxFh"

Geo-thermal data Repository - "https://lnkd.in/gRT2TgEn"

CMG tutorial data-set -"https://lnkd.in/gmfe7XXb"

#YT channel to learn CMG -"https://lnkd.in/g7pJJqEX"

Petrel tutorial data-set - "https://lnkd.in/g6FKZJE7"

#YT channel to learn PETREL from scratch -"https://lnkd.in/gpT3p5aN"

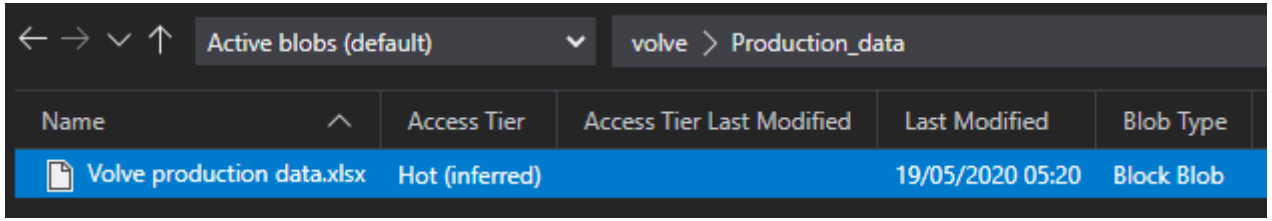
All info about open geo-physics data - "https://lnkd.in/gXBfBtV6"

SPE data Repository - "https://lnkd.in/g-jbeB4e"

Hope it helped ! Comment if you want to add some extra .
#
#opendatasets

https://www.linkedin.com/posts/jeevan-vivek-58805880_yt-opendatasets-activity-6841770285658730496-tCe8

Excel *(.xlsx)*



Librería : Pandas

pd.read_excel(*args)

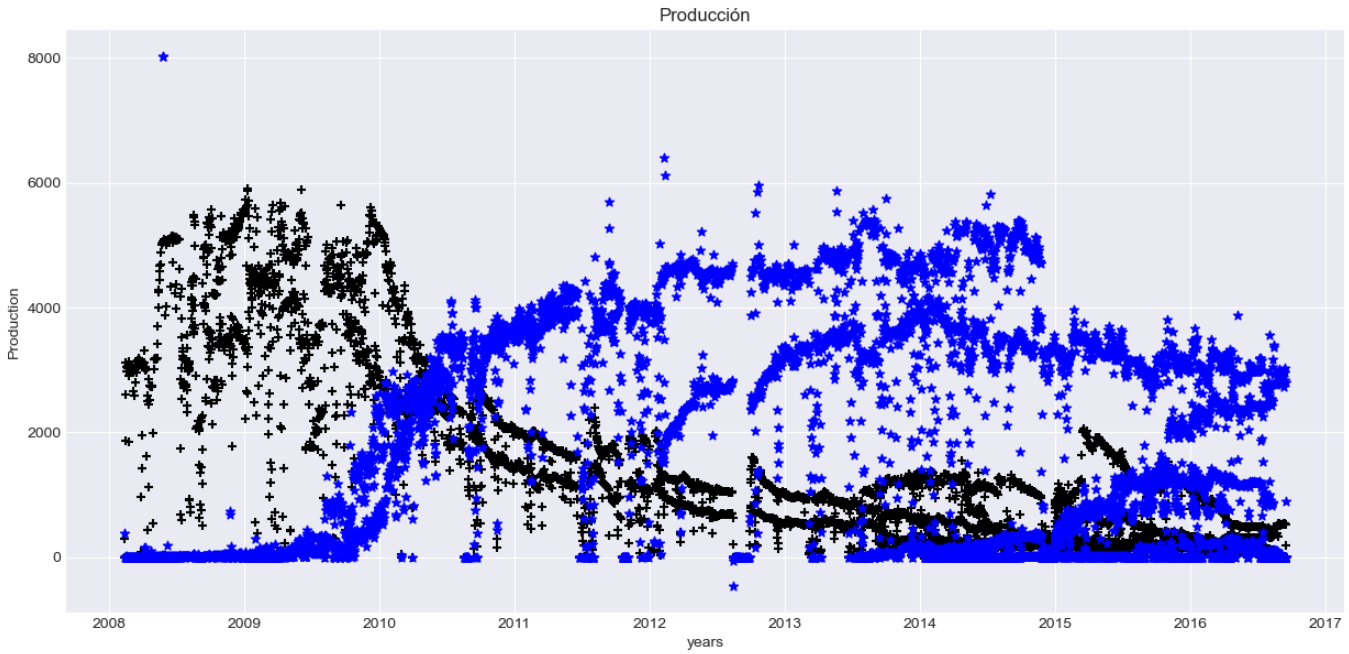
```
Tamaño DP : (15634, 24)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15634 entries, 0 to 15633
Data columns (total 24 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   DATEPRD                    15634 non-null  datetime64[ns]
 1   WELL_BORE_CODE             15634 non-null  object
 2   NPD_WELL_BORE_CODE         15634 non-null  int64
 3   NPD_WELL_BORE_NAME         15634 non-null  object
 4   NPD_FIELD_CODE             15634 non-null  int64
 5   NPD_FIELD_NAME             15634 non-null  object
 6   NPD_FACILITY_CODE          15634 non-null  int64
 7   NPD_FACILITY_NAME          15634 non-null  object
 8   ON_STREAM_HRS              15349 non-null  float64
 9   AVG_DOWNHOLE_PRESSURE      8980 non-null   float64
 10  AVG_DOWNHOLE_TEMPERATURE   8980 non-null   float64
 11  AVG_DP_TUBING              8980 non-null   float64
 12  AVG_ANNULUS_PRESS          7890 non-null   float64
 13  AVG_CHOKE_SIZE_P           8919 non-null   float64
 14  AVG_CHOKE_UOM              9161 non-null   object
 15  AVG_WHP_P                  9155 non-null   float64
 16  AVG_WHT_P                  9146 non-null   float64
 17  DP_CHOKE_SIZE              15340 non-null  float64
 18  BORE_OIL_VOL               9161 non-null   float64
 19  BORE_GAS_VOL               9161 non-null   float64
 20  BORE_WAT_VOL               9161 non-null   float64
 21  BORE_WI_VOL                5706 non-null   float64
 22  FLOW_KIND                  15634 non-null  object
 23  WELL_TYPE                  15634 non-null  object
dtypes: datetime64[ns](1), float64(13), int64(3), object(7)
memory usage: 2.9+ MB
```

Excel

	DATEPRD	WELL_BORE_CODE	NPD_WELL_BORE_CODE	NPD_WELL_BORE_NAME	NPD_FIELD_CODE	NPD_FIELD_NAME	NPD_FACILITY_CODE	NPD_F
0	2014-04-07	NO 15/9-F-1 C	7405	15/9-F-1 C	3420717	VOLVE	369304	MA
1	2014-04-08	NO 15/9-F-1 C	7405	15/9-F-1 C	3420717	VOLVE	369304	MA
2	2014-04-09	NO 15/9-F-1 C	7405	15/9-F-1 C	3420717	VOLVE	369304	MA
3	2014-04-10	NO 15/9-F-1 C	7405	15/9-F-1 C	3420717	VOLVE	369304	MA
4	2014-04-11	NO 15/9-F-1 C	7405	15/9-F-1 C	3420717	VOLVE	369304	MA
...	...	...	...	...	...	...	...	
15629	2016-09-14	NO 15/9-F-5 AH	5769	15/9-F-5	3420717	VOLVE	369304	MA
15630	2016-09-15	NO 15/9-F-5 AH	5769	15/9-F-5	3420717	VOLVE	369304	MA
15631	2016-09-16	NO 15/9-F-5 AH	5769	15/9-F-5	3420717	VOLVE	369304	MA
15632	2016-09-17	NO 15/9-F-5 AH	5769	15/9-F-5	3420717	VOLVE	369304	MA
15633	2016-09-18	NO 15/9-F-5 AH	5769	15/9-F-5	3420717	VOLVE	369304	MA

15634 rows × 24 columns

Librería : matplotlib.pyplot
plt.scatter(*args)



Registros *(.LAS)*

WL_RAW_BHPR-GR-MECH_MWD_1 - Notepad

File Edit Format View Help

```
~Version Information
VERS.              2.0 : CWLS log ASCII standard - Version 2.0
WRAP.              NO  : One line per depth step
#
~Well Information Block
#MNEM.UNIT                  Data Type:Information
#_____________________________________________________________
 STRT .M            910.13280  :START INDEX
  STOP.M           1083.1068 : STOP INDEX
  STEP.M              0.1524 : STEP
  NULL.              -999.25 : NULL VALUE
  COMP.           StatoilHydro : COMPANY
  WELL.            15/9-F-9  : WELL
   FLD.                Volve : FIELD
   LOC.           Norwegian North Sea : LOCATION
  PROV.                     : PROVINCE
  CNTY.                Norway : COUNTY
  STAT.           Maersk Inspirer : RIG
  CTRY.                     : COUNTRY
  SRVC.           Schlumberger D&M : SERVICE COMPANY
  DATE.            30-Aug-08 : LOG DATE
   UWI.                     : UNIQUE WELL ID
   API.                     : API NUMBER
#
~Curve Information Block
#MNEM.UNIT                  : Curve Description
#---------                  --------------------
DEPT        .M              : Bit Depth 2hz
ROP5        .M/HR           : 5ft ROP 2hz
ARC_GR_RT   .GAPI           : ARC Gamma Ray, Real-Time
Stick_RT    .RPM            : MWD PKtoPK RPM 2hz
SWOB        .KKGF           : Surface Weight on Bit 2hz
SHKRSK_RT   .----           : MWD SHKRSK 2hz
RPM         .RPM            : SRF SRPM 2hz
TRPM_RT     .RPM            : MWD TUR_RPM 2hz
TFLO        .LPM            : HSPM Total Pump Flow 2hz
ATMP        .DEGC           : ARC Annular Temperature
TQA         .KMN            : SRF STOR 2hz
ECD_ARC     .G/C3           : Equivalent Circulating Density
SPPA        .BAR            : SRF PUMPRS 2hz
~
```

Librería : lasio

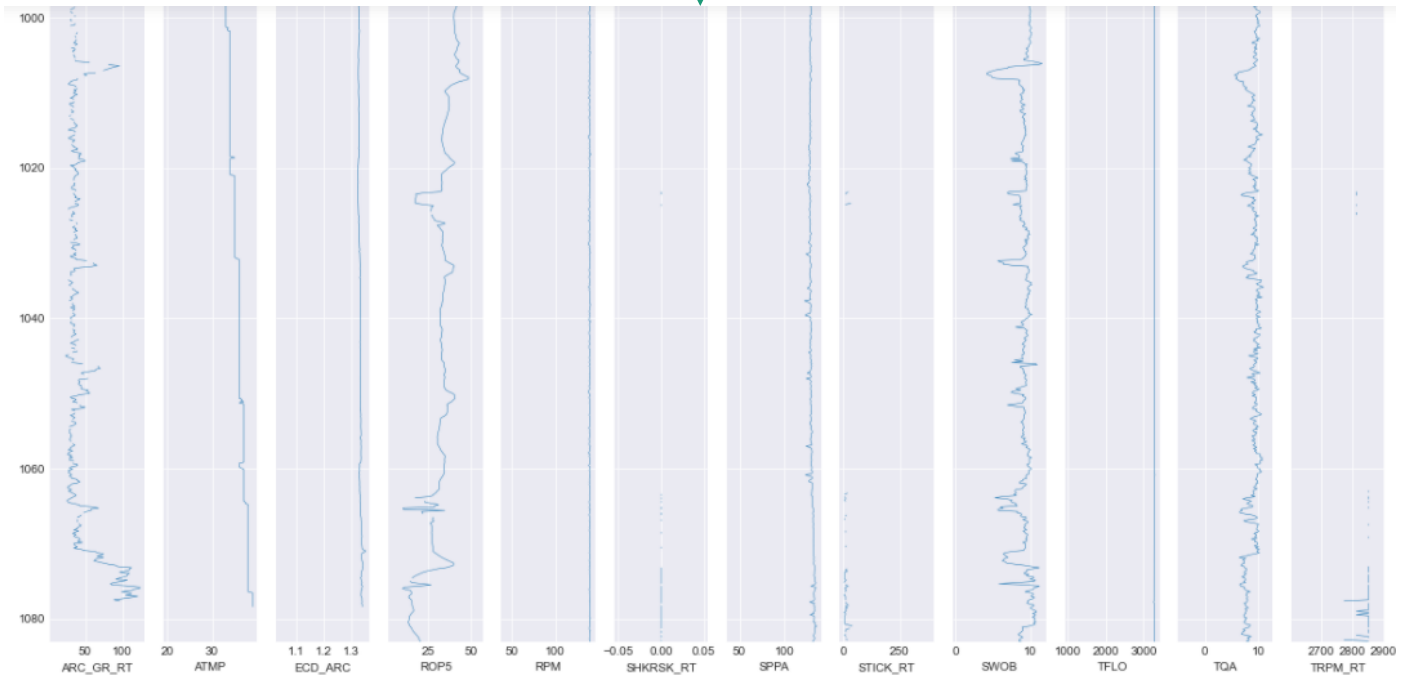
lasio.read(*args)

DEPT	ROP5	ARC_GR_RT	STICK_RT	SWOB	SHKRSK_RT	RPM	TRPM_RT	TFLO	ATMP	TQA	ECD_ARC	SPPA
910.1328	4.2626	NaN	105.0	2.4109	0.0	64.7053	2695.3125	3101.6666	20.0	6.0463	1.0502	95.6870
910.2852	4.6394	NaN	45.0	0.4642	0.0	64.2314	2695.3125	3101.6666	20.0	6.1035	1.0544	96.4459
910.4376	6.4096	NaN	108.0	0.7414	0.0	64.2660	2695.3125	3101.6666	20.0	5.5775	1.0550	96.6515
910.5900	6.3423	NaN	84.0	1.0050	0.0	64.4379	2695.3125	3101.6666	20.0	5.8363	1.0548	96.6629
910.7424	6.5074	NaN	138.0	1.1867	0.0	64.6786	2695.3125	3101.6666	20.0	5.3887	1.0547	96.2187
...	...	...	...	...	...	...	...	...	...	...	...	...
1082.4972	19.3254	NaN	NaN	8.9563	0.0	139.4464	2851.5625	3278.9046	NaN	8.0129	NaN	133.8932
1082.6496	19.5981	NaN	12.0	8.6458	NaN	139.0000	NaN	3278.9046	NaN	8.3624	NaN	133.5826
1082.8020	19.8582	NaN	18.0	8.4776	0.0	139.8823	2773.4375	3278.9046	NaN	8.2228	NaN	133.1638
1082.9544	20.2412	NaN	NaN	8.4618	0.0	139.4694	2851.5625	3278.9046	NaN	7.8519	NaN	131.7679
1083.1068	20.4969	NaN	39.0	8.9094	NaN	139.4902	NaN	3278.9046	NaN	8.4750	NaN	132.6377

1136 rows × 12 columns

Librería : matplotlib.pyplot

plt.subplots(*args)



Registros *(.DLIS)*

```
Logical file ID          : CPI_OUTPUT.logdata
File set name and number : PETROLOG-TO-DLIS / 305463381
File number and type     : 1 / PETROLOG-FILE

Field                    : VOLVE
Well (id/name)           :  / 15/9-F-15
Produced by (code/name)  : 126 / CROCKER DATA PROCESSING
Produced for             : STATOILHYDRO ASA
Created                  : 2009-09-07 10:56:21.625000

Created by               : PETROLOG, (version: Version 10.5)
Other programs/services  : PETROLOG-TO-DLIS
```

Librería : dlisio

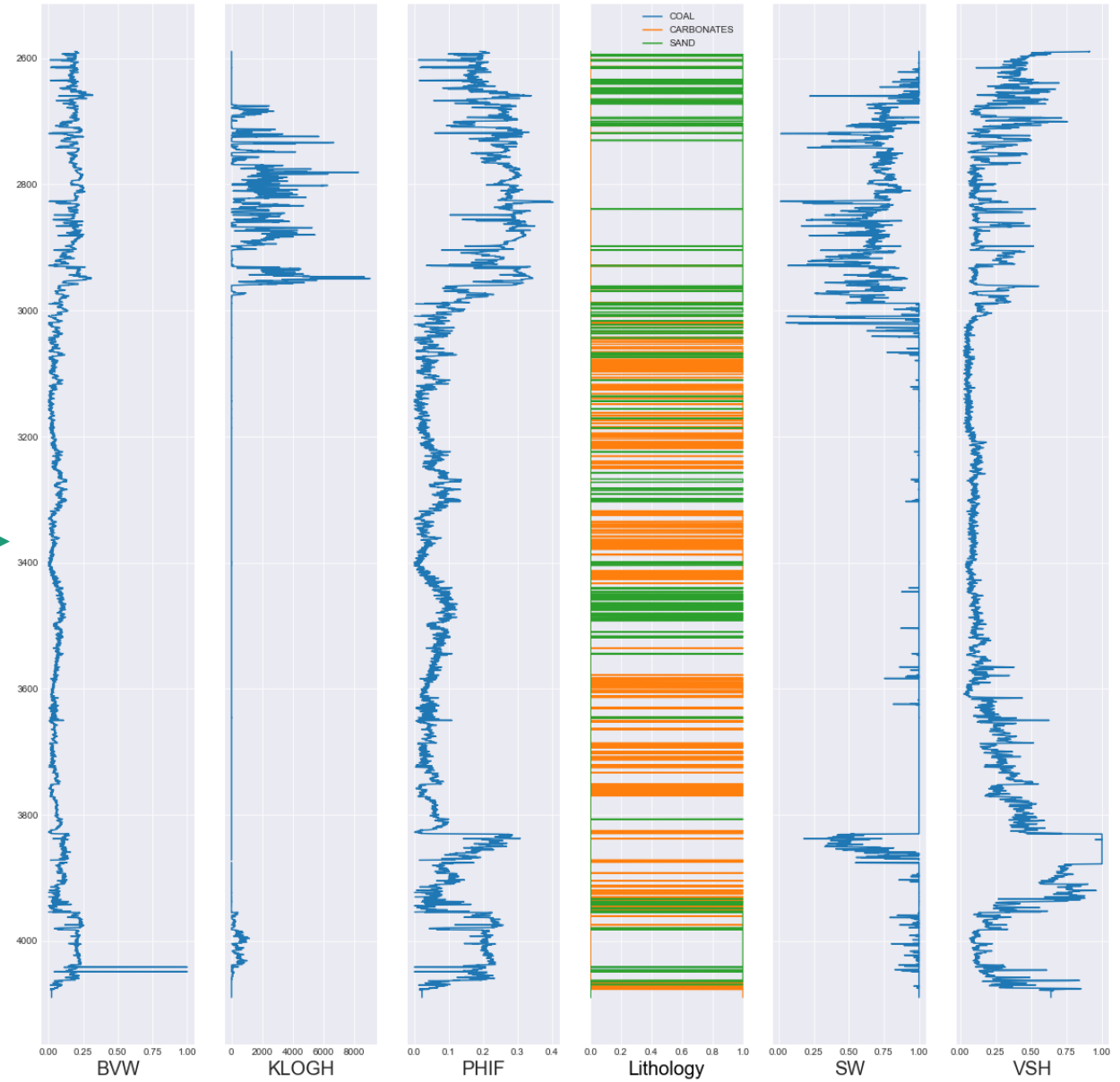
dlis.load(*args)

	BVW	CARB_FLAG	COAL_FLAG	DEPTH	KLOGH	PHIF	SAND_FLAG	SW	VSH
0	0.1902	0.0	NaN	2588.9712	0.001	0.190204	0.0	1.0	0.9136
1	0.1955	0.0	NaN	2589.1236	0.001	0.195471	0.0	1.0	0.9048
2	0.1997	0.0	NaN	2589.2760	0.001	0.199674	0.0	1.0	0.9117
3	0.2021	0.0	NaN	2589.4284	0.001	0.202071	0.0	1.0	0.8773
4	0.1984	0.0	NaN	2589.5808	0.001	0.198369	0.0	1.0	0.8431
...	...	...	...	...	...	...	...	...	...
9845	0.0229	1.0	0.0	4089.3492	0.001	0.022857	0.0	1.0	0.6401
9846	0.0229	1.0	0.0	4089.5016	0.001	0.022857	0.0	1.0	0.6401
9847	0.0229	1.0	0.0	4089.6540	0.001	0.022857	0.0	1.0	0.6401
9848	0.0229	1.0	0.0	4089.8064	0.001	0.022857	0.0	1.0	0.6401
9849	NaN	NaN	NaN	4089.9588	NaN	NaN	NaN	NaN	NaN

9850 rows × 9 columns

Librería : matplotlib.pyplot

plt.subplots(*args)



Archivo Separado por comas (.csv)

```
DEPT,ROP5,BPOS,BVEL,SWOB,HKLD,TQA,RPM,Stick_RT,TFLO,SPPA,GR,ECD,an_Temp,DateTime,Well,Run
226.0035,9.356,13.595,0.003,7.56,103.31700000000001,3.3810000000000002,39.0,69.0,3456.143,
226.0345,9.339,13.564,0.003,7.632999999999999,103.244,1.9240000000000002,39.0,69.0,3456.14
226.0575,9.339,13.540999999999999,0.003,7.422000000000001,103.455,2.7760000000000002,39.0,
226.081,9.339,13.517999999999999,0.003,7.5329999999999995,103.345,-1.8119999999999998,40.0
226.1111,9.339,13.487,0.003,7.507000000000001,103.37,-0.349,39.0,114.0,3456.143,88.796,2.4
226.1336,9.339,13.465,0.002,7.689,103.18799999999999,-0.349,39.0,114.0,3456.143,89.281,2.4
226.1565,9.339,13.442,0.002,7.097,103.78,0.705,39.0,72.0,3456.143,86.052,2.417,0.06,15.0,2
```

Librería : Pandas

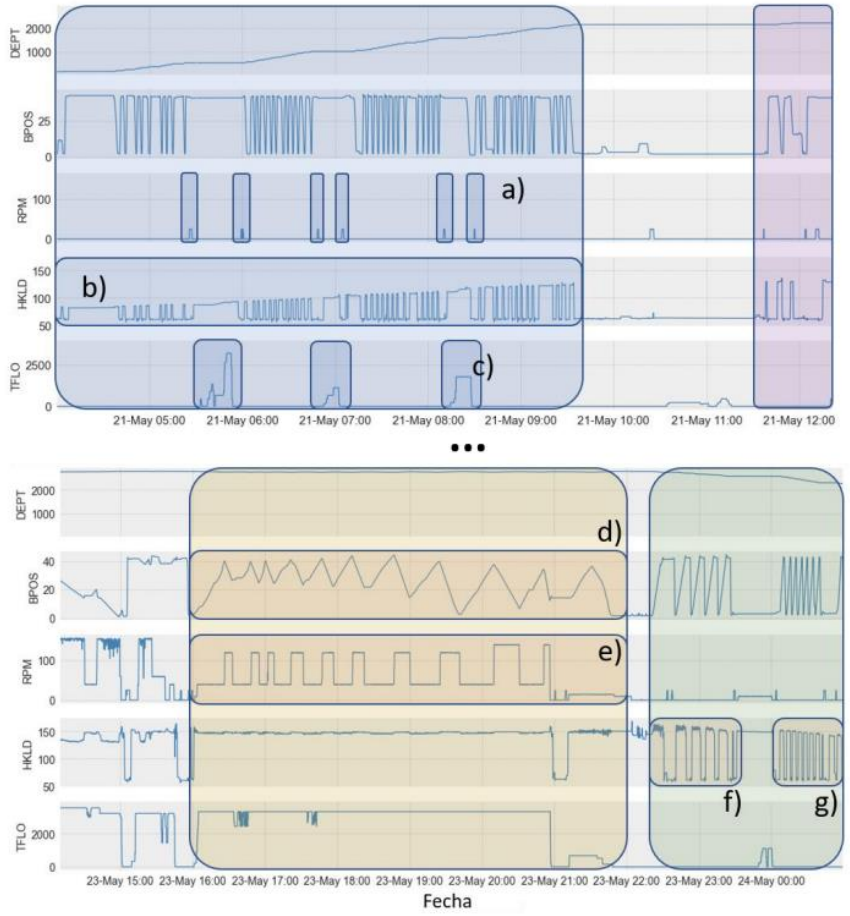
pd.read_csv(*args)

	DEPT	ROP5	BPOS	BVEL	SWOB	HKLD	TQA	RPM	Stick_RT	TFLO	SPPA	GR	ECD	an_Temp	DateTime	Well	Run
0	226.0035	9.3560	13.5950	0.0030	7.5600	103.3170	3.3810	39.0	69.0	3456.1430	89.0010	7.2500	0.0350	15.0	2007-12-18 02:13:26	F-5	run_2
1	226.0345	9.3390	13.5640	0.0030	7.6330	103.2440	1.9240	39.0	69.0	3456.1430	89.6530	4.8330	0.0480	15.0	2007-12-18 02:13:36	F-5	run_2
2	226.0575	9.3390	13.5410	0.0030	7.4220	103.4550	2.7760	39.0	60.0	3456.1430	92.3380	4.8330	0.0480	15.0	2007-12-18 02:13:46	F-5	run_2
3	226.0810	9.3390	13.5180	0.0030	7.5330	103.3450	-1.8120	40.0	60.0	3456.1430	87.1240	2.4170	0.0350	15.0	2007-12-18 02:13:56	F-5	run_2
4	226.1111	9.3390	13.4870	0.0030	7.5070	103.3700	-0.3490	39.0	114.0	3456.1430	88.7960	2.4170	0.0350	15.0	2007-12-18 02:14:06	F-5	run_2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
42359	3791.7510	7.3932	33.5655	-0.0456	0.1851	117.7876	20.6663	130.0	45.0	2104.7023	209.7894	56.1877	1.5863	102.0	2008-07-29 22:08:12	F-5	run_5
42360	3791.7510	7.3932	33.5655	-0.0456	0.1851	117.7876	20.6663	130.0	45.0	2104.7023	209.7894	56.1877	1.5863	102.0	2008-07-29 22:08:22	F-5	run_5
42361	3791.7510	7.3932	33.5655	-0.0456	0.1851	117.7876	20.6663	130.0	45.0	2104.7023	209.7894	56.1877	1.5863	102.0	2008-07-29 22:08:32	F-5	run_5
42362	3791.7510	7.3932	33.5655	-0.0456	0.1851	117.7876	20.6663	130.0	45.0	2104.7023	209.7894	56.1877	1.5863	102.0	2008-07-29 22:08:42	F-5	run_5
42363	3791.7510	7.3932	33.5655	-0.0456	0.1851	117.7876	20.6663	130.0	45.0	2104.7023	209.7894	56.1877	1.5863	102.0	2008-07-29 22:08:52	F-5	run_5

42364 rows × 17 columns

Librería : matplotlib.pyplot

plt.subplots(*args)



Survey de perforación *(.xml)*


```
▼<logs xmlns="http://www.witsml.org/schemas/1series" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
 instance" version="1.4.1.1">
  ▼<log uidWell="W-924688" uidWellbore="B-924688" uid="L-957090-MD">
    <nameWell>15/9-F-10</nameWell>
    <nameWellbore>15/9-F-10 - Main Wellbore</nameWellbore>
    <name>8.5 in Section - MD Log</name>
    <objectGrowing>false</objectGrowing>
    <serviceCompany>Schlumberger</serviceCompany>
    <runNumber>9,10</runNumber>
    <pass>Drilling</pass>
    <creationDate>2009-05-17T06:09:03.000Z</creationDate>
    <indexType>measured depth</indexType>
    <startIndex uom="m">4802.124</startIndex>
    <endIndex uom="m">5062.765</endIndex>
    <direction>increasing</direction>
    <indexCurve>DEPTH</indexCurve>
```


Librería : bs4

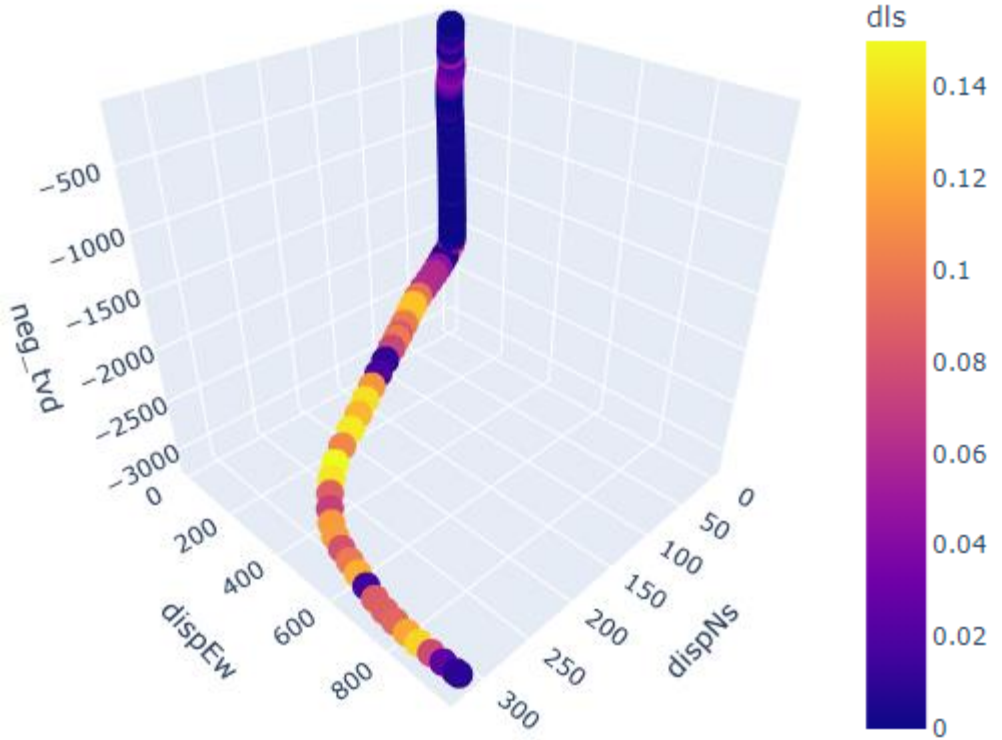
bs4.BeautifulSoup(*args)

	md	tvd	dispNs	dispEw	incl	azi	dls	neg_tvd
0	145.899835	145.899835	-1.210056	2.420112	0.00	0.00	0.000000	-145.899835
1	155.089860	155.089860	-1.195685	2.421476	0.18	5.42	0.019587	-155.089860
2	195.439894	195.439589	-1.040746	2.498650	0.33	37.78	0.005016	-195.439589
3	235.799986	235.798157	-0.835734	2.752488	0.61	58.22	0.007979	-235.798157
4	276.200006	276.190558	-0.470167	3.419976	1.55	62.50	0.023337	-276.190558
...	...	...	...	...	...	...	...	...
87	3596.234921	3059.530279	320.556822	861.660786	51.69	92.76	0.136588	-3059.530279
88	3638.070550	3086.324028	318.656939	893.727148	48.65	94.05	0.076424	-3086.324028
89	3678.386446	3112.682827	316.609552	924.162966	49.69	93.65	0.026867	-3112.682827
90	3718.835234	3138.720672	314.598419	955.051238	50.17	93.80	0.012201	-3138.720672
91	3729.857717	3145.774049	314.041492	963.503146	50.26	93.74	0.009173	-3145.774049

92 rows × 8 columns

Librería : plotly.express

px. scatter(*args)



Datasets de Ing.Petrolera

- Live Coding
  - Visualizar
    - Excel
    - DLIS
    - LAS
    - WITSML
    - CSV

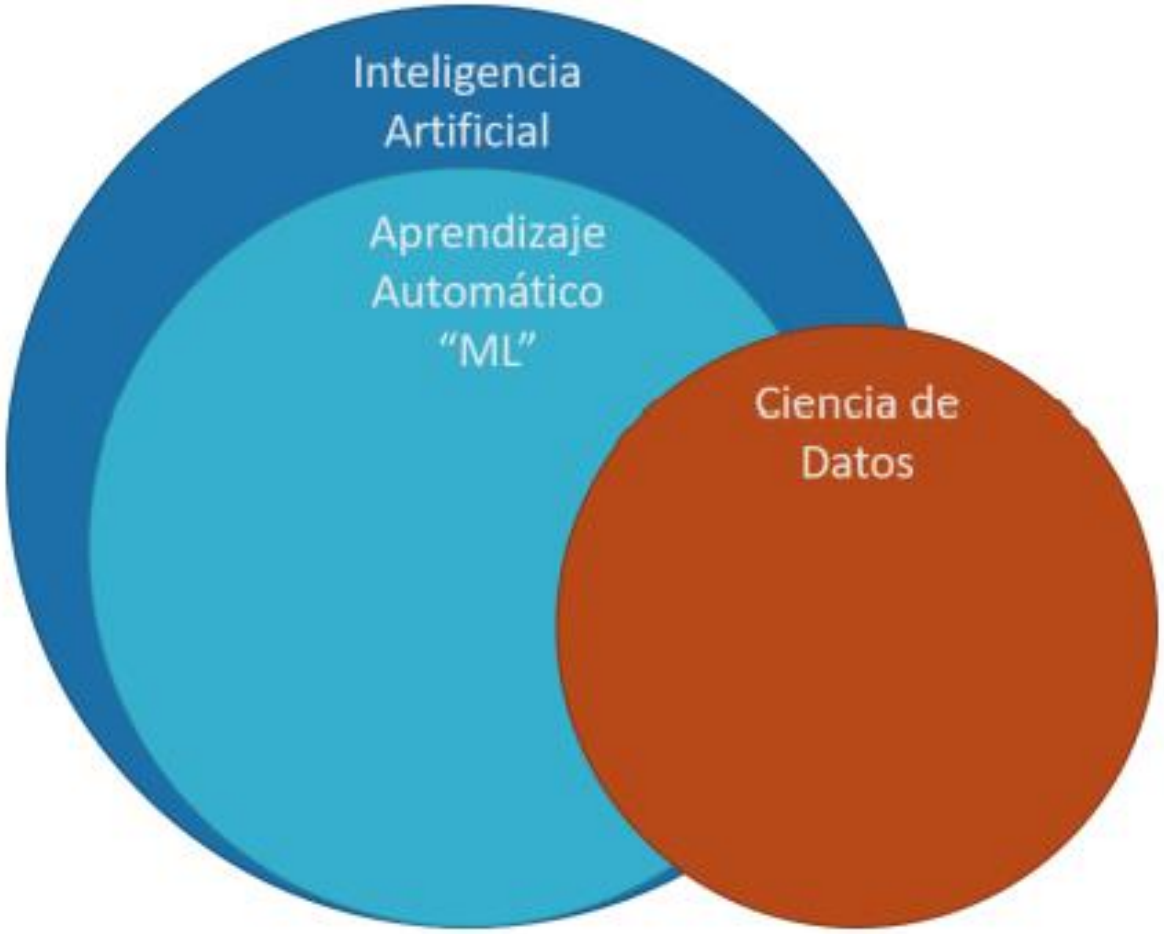
Ciencia de Datos

Objetivos :

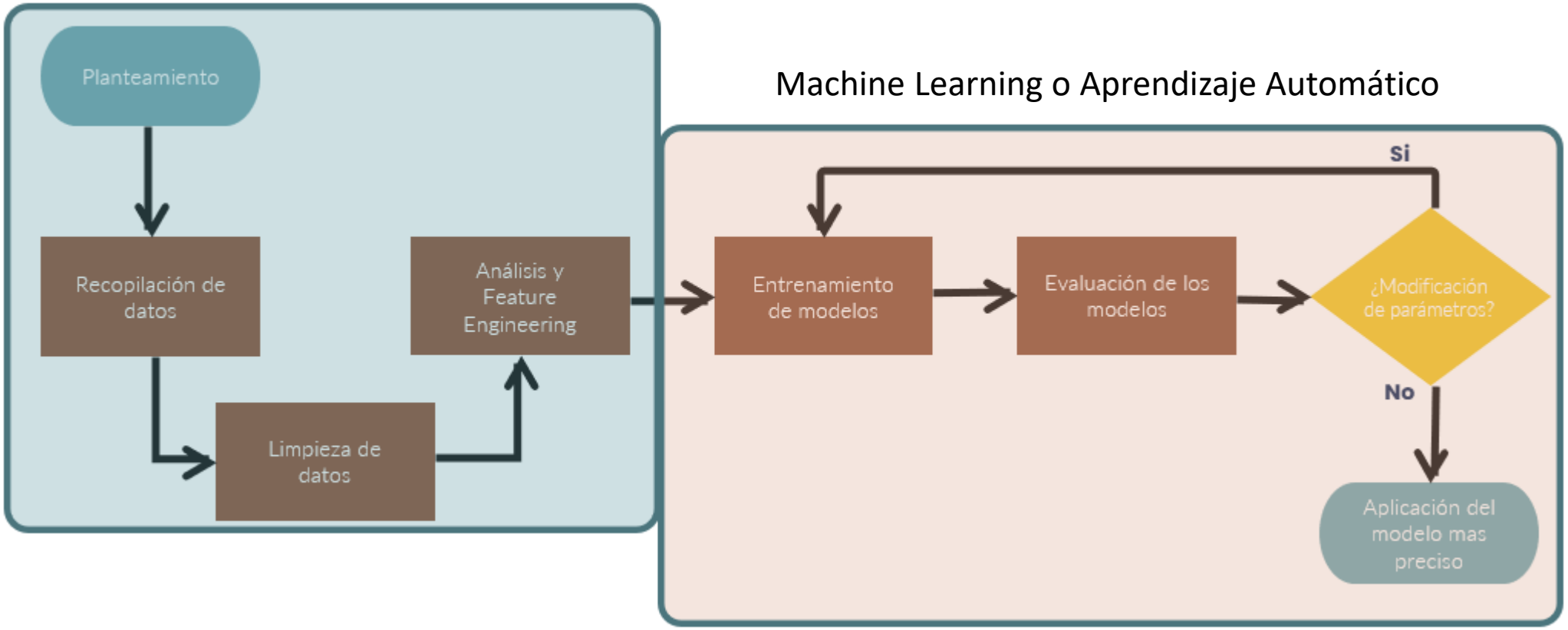
-Analizar datos “crudos”

- Limpieza
- Visualización de comportamiento
- Determinar relaciones

Importancia



Importancia



Ciencia de Datos

- Live Coding
  - Producción Volve
    Manejo de DataFrames
    Plots de producción
    Análisis de variables

  - Registro de perforación sin filtrar
    Plots estadísticos principales

Gracias por su atención

Contacto:

- Luis Enrique Navarro Morales | LinkedIn
- LuisNavarro93 (github.com)

- 93lenm@gmail.com
- https://t.me/+TIVLACKN0FTuIGM8