



Guia LADATA de ferramentas para Ciência de Dados

Uma aventura por (quase) tudo o que há de bom em Data Science





Os "n" contextos dos dados



%#(///

##((*,

%%#/*%(

###(/,

##%%%%%#### (/

%%%%%%%%%%%###(((((%%%%%%%%%%%####(((((*

,%%%%%%%#((((#((

%#%%%%%%%%###(######

,&%(,/*(%#%%%#%%%%%#%#((#(&#(,/.

,/*///,(%%###%##%%%%%%###((%#(##(. (*, ./.###/ ###%%##%#%#%#(#%#######((//

,,/(..,...*.//, %%%%%%%#&#%#(&

#(,./..((((((**%.(/,* &%%##%%%%%%%%#(/

,%%%% %#%%%

#,,*##%#

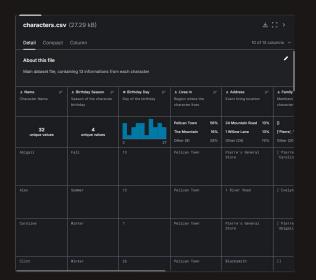
%** .%//,/

#*/(.*,

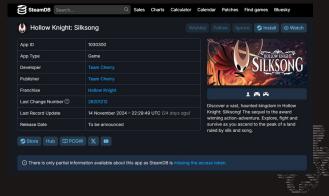
Os "n" contextos da Ciência de dados

- Os dados no contexto Acadêmico
- Uso no mundo empresarial
- Uso de dados para entretenimento





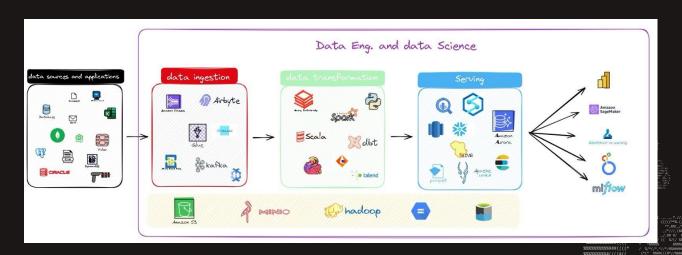




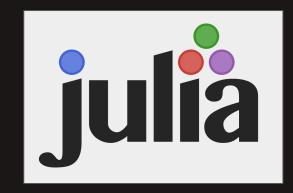


• Que tipo de Ferramentas compõem os projetos de dados?

- Versionamento de código
- Codificação
- Visualização
- Aquisição
- Aprendizagem
- Bibliotecas
- Linguagens





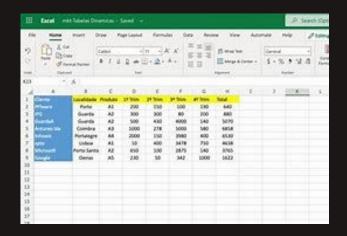


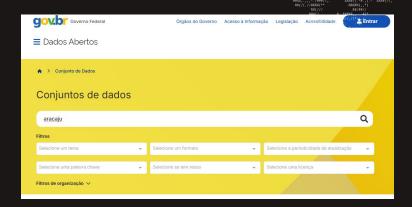


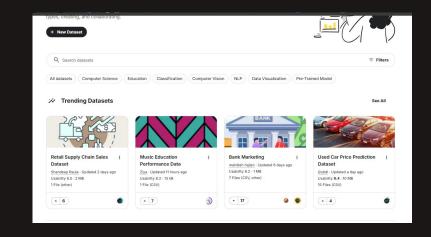
Linguagens e Bibliotecas



- Conjuntos de dados abertos
- Bases de dados da Comunidade
- Tabelas em formatos xls
- Bancos de dados









Codificação além do VS Code

(1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (1997) | ### (

- Uso do Modelo "Notebook"
- Grandes vantagens
- Colaboração "embutida"
- Plataformas Integradas
- Comunidade de dados









Exemplos

A parte da oficina chegou



```
##%%%%%####(/
                           %%%%%%%%%%%##((((
                          %%%%%%%%%%%%###((((*
                                                 %/*(/*,*//*/#%
                          ,%%%%%%%%%#((((#((
                                                (*(* %%##(((#*/
                                                 /%#(**,/// (#
                         %#%%%%%%%%###(######
                       .((##(%%%##%((####%#
                       ,/(..,..,*.//, %%%%%%%%#&#%#(&
   #(,./..((((((**%.(/,* &%%##%%%%%%#(/
           **,&%(,/*(%#%%%#%%%%%%#%#((#( &#(,**/.
 .#(,/
           ./*///,(&%###%##%%%%%%###((%#(##(. (*,
         .,/,## #/ ###%%##%#%#(#%##(##%#####((/
  &%%%(#####%%%%#((((#%%%#/(////**%#/*/,
/ %/*(/*,*//*/#<del>%#######</del>%%#(/*.%*/**(/,/ ///%&#(*./@(.
(*(* %%##(((#*//%&&&#(*(,*(,&#(%,.%#,(/* %&##*..,*. **
 /%%#(**,/// (#*/#.**,.&#(((##(** %.,.,@%#*, &%(.,.&
            //,# (&%%#(((%###...
           *,** &%##(/*(/&####.
 *##/.//.
                                   //,**(#** .%#*.&
          (**,/%#(//*
                         %%##(**
                                   (*/,/@#(/*.(#,*
                         %%%#/(.*#.,(.* %%%#(/(,
  ##%%,.,,,*./###//,.
   ##(/(.//#&%#(**
                           .%&%#%(,,*(
          %#(///
                              &#/##//
          ###(/.
                      #.,*##%#., #(*
```

%** .%//,/(*/,,/,

%%%%%%%%%%####

##((*,

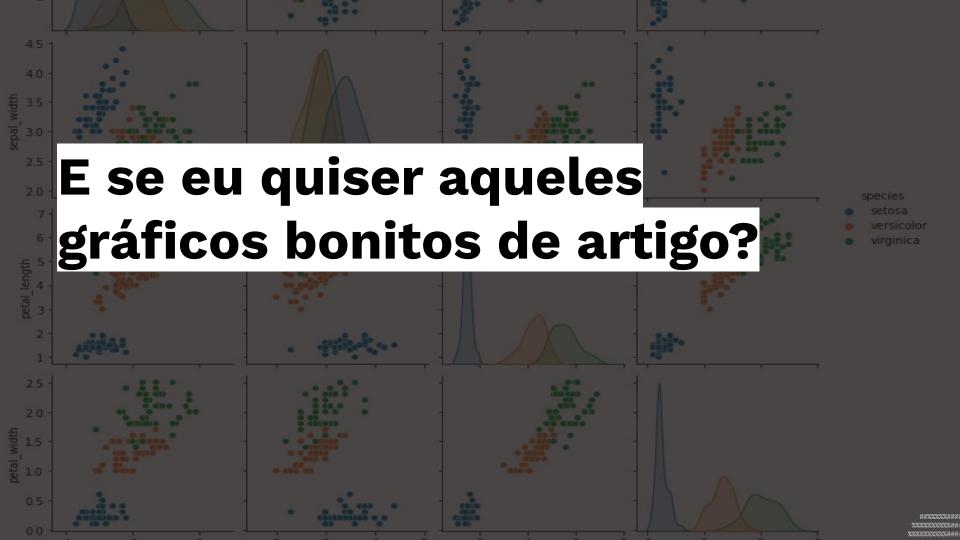
##%%%%%#####(/

%%%%%%%%%%%##((((

%%%%%%%%%%%###(((((*

,/(..,..*.//, %%%%%%%#&#%#(&

#(,./..(((((**%.(/,* &%%##%%%%%%%#(/





Ferramentas próprias



Ferramentas embutidas





Visualização de dados

Ferramentas de ETL e Engenharia de Dados

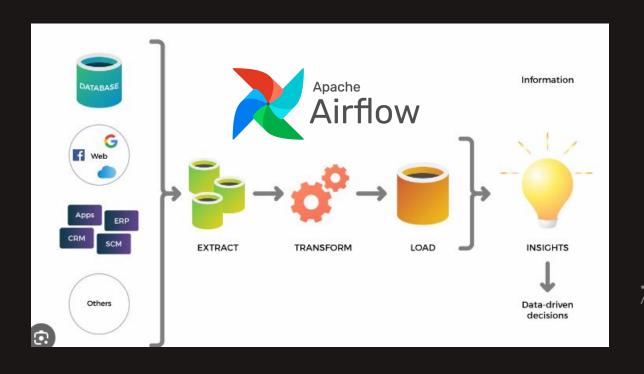
Afinal, o que é ETL?

 Processos que envolvem desde a extração dos dados de uma fonte "crua" de dados até a entrega tratada para seus objetivos finais

E onde entra a Engenharia de dados?

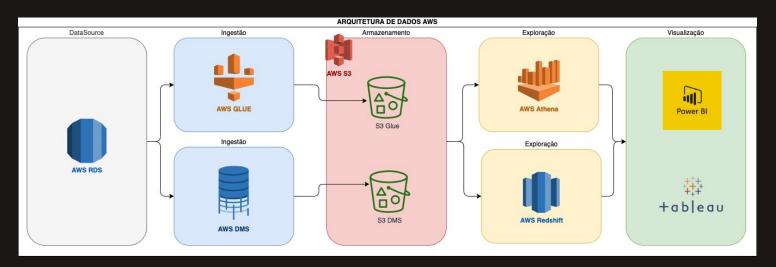
- Trabalham diretamente na aplicação de pipelines de ETL para o dia a dia dos dados
- Mais utilizados no contexto de grandes bases de dados

Ferramentas de ETL e Engenharia de Dados



```
%%%%
                            %#%%%
                          .((##(%%
         .,/,## #/ ###%%##%#%#%#%
%/*(/*,*//*/#<del>%#######</del>%%#(/*.%*<sub>.</sub>
(*(* %%##(((#*//%&&&#(*(,*(,&#(
 /%%#(**,/// (#*/#.**,.&#(((##(
             //,# (&%%#(((%###
            *,** &%##(/*(/&####
 ,*,(*, (**,/%;#(//*
 ##%%,.,,,*./###//,.
  ##(/(.//#&%#(**
           %#(///
          ###(/,
                         #,,*##%#
           ##((*,
                       %** .%//,/
           %%#/*%(
                       #*/(.*,
```

O ecossistema AWS para dados



#(,./..(((((**%.(/,* &%%##%%%%%%#(/





, %%%%%%%% %#%%%%%% . ((##(%%%##%

&%%%(#####%%%%#((((#

Versionamento de Código

- Gerenciamento de projetos
- Controle por "issues"
- Colaboração em times
- Comunidade





Plataformas de Aprendizagem e Materiais

Cursos Pagos				
Plataforma	Curso	Descrição	Preço (aproximado)	
Coursera	IBM Data Science Professional Certificate	Abrange Python, pandas, visualização e machine learning.	R\$ 200/mês (gratuito 7 dias)	
Udemy	Python para Data Science e Machine Learning	Foco em bibliotecas Python como pandas, NumPy, seaborn e scikit-learn.	R\$ 30 a R\$ 150 (promoções)	
DataCamp	Data Scientist in Python	Programa interativo com foco em prática (Python, SQL e machine learning).	R\$ 120/mês	
Alura	Formação Cientista de Dados	Série de cursos cobrindo desde análise exploratória até machine learning e big data.	R\$ 99/mês (assinatura)	

Cursos Gratuitos				
Plataforma	Curso	Descrição	Certificação	
Kaggle	Learn Data Science	Mini-cursos interativos sobre Python, visualização e machine learning.	Sim	
edX	Data Science Essentials (Microsoft)	Tópicos fundamentais de ciência de dados (estatística, Python, visualização).	Opcional (pago)	
Google	Curso de Machine Learning (Google Developers)	Introdutório com exemplos práticos em TensorFlow.	Não	
LinkedIn Learning	Ciência de Dados: Fundamentos	Curso gratuito com princípios básicos para iniciantes.	Sim #	
DIO	Bootcamps de Ciência de Dados	Cursos gratuitos com foco prático em Python, SQL e machine learning.	Sim #	

.((##(%%%##%((####%#

,/(..,...*.//, %%%%%%%#&#%#(& #(,./..(((((**%.(/,* &%%##%%%%%%##(/

Plataformas de Aprendizagem e Materiais



DataLoveLib

```
##(/(.//#&%#(**
                                  .%%#%(,,*(
                                    &#/##//
                             #,,*##%#,, #(*
                 ##((*,
                  %%#/*%(
                           #*/(.*, .*,/
###((/
((/,*/* #/(
                     %%%#####% ((((*
///**%#/*/,
//%&#(*./@(.
%&##*..,*. **
%#*, &%(.,.&
&(. &/#*,,*
(#** .%#*,&
@#(/*.(#,*
%%%#(/(,
```

Obrigado !!

//,# (&%%#(((%###

*,** &%##(/*(/&####

#,,*##%#

%** .%//,/

#*/(.*,

/ %/*(/*,*//*/#%#######%##(/*.%*) (*(* %%##(((#*//%&&&#(*(,*(,&#(! /%%#(**,/// (#*/#.**,.&#(((##(!

(**,/%#(//*

##%%,.,,,*./###//,.

###(/,

##((*,

%%#/*%(

%#(/*

##(/(.//#&%#(**

*##/.//,

ر*) ر*ر

##%%%%%#### (/

%%%%%%%%%%%% %%%%%%%%%%%%%% ####((((**

,%%%%%%%%#((((#((

%#%%%%%%%###(######

.((##(%%%##%((####%#

**, &%(, /*(%#%%#%%%%%#%#((#(&#(, **/

,/*///,(%%###%%%%%###((%#(##(. (*, ../,###/

,/(..,..,*.//, %%%%%%%%#&#%#(&

#(,./..((((((**%.(/,* &%%##%%%%%%%#(/

##%%,.,,,*./###//,. %%%#/(.*#.,(.* %%%#(/(, &#(,**/. #,,*##%#,, #(* (##(. (*, |%####((/ (((((/,*/* #/(/(////**/#/*/, ///%&#(*./@(. %&##*..,*. ** ,@%#*, &%(.,.& &/#*,,* * %%%#(/(,

Siga a LADATA nas redes sociais

LinkedIn: LADATA

Instagram: ladata.ufs

Github: ladata-ufs

Kaggle: ladataufs

/ %/*(/*,*//*/#%##### (*(* %%##(((#*//%&&8

* ** 8%# (**,/%#(/ ##%%,.,,,*./###//,.

%#(///

**, 8%(, /* ,/*//,(&%

e as referências (25-28)

unicu

IMP

