

通过 MySQL 进行数据清洗

● 分为三步：

A. 删除重复值

B. 对数据进行标准化操作

C. 处理空白值

① 删除重复值：

源数据

	company	location	industry	total_on_boarding	percentage_on_boarding	date	stage	country	funds_raised_millions
▶	Casper	New York City	Retail	NULL	NULL	9/14/2021	Post-IPO	United States	339
	Casper	New York City	Retail	78	0.21	4/21/2020	Post-IPO	United States	339
	Casper	New York City	Retail	NULL	NULL	9/14/2021	Post-IPO	United States	339

处理后数据

	company	locations	industry	total_on_boarding	percentage_on_boarding	date	stage	country	funds_raised_millions
▶	Casper	New York City	Retail	NULL	NULL	2021-09-14	Post-IPO	United States	339
	Casper	New York City	Retail	78	0.21	2020-04-21	Post-IPO	United States	339

② 对数据进行标准化操作（处理数据不一致、日期标准化）

源数据

	company	location	industry	total_on_boarding	percentage_on_boarding	date	stage	country	funds_raised_millions
	BitGo	SF Bay Area	Crypto	NULL	0.12	4/17/2020	Series B	United States	69
	Purse	SF Bay Area	Crypto	NULL	1	4/16/2020	Seed	United States	1
	CipherTrace	SF Bay Area	Crypto	NULL	NULL	4/9/2020	Unknown	United States	18
	Elliptic	London	Crypto	NULL	0.3	4/9/2020	Series B	United Kingdom	40
	Bitfarms	Quebec	Crypto	NULL	NULL	4/6/2020	Post-IPO	Canada	25
	Unstoppable Dom...	SF Bay Area	Crypto Currency	42	0.25	7/14/2022	Series B	United States	7
	GSR	Hong Kong	Crypto Currency	NULL	NULL	10/11/2022	Unknown	Hong Kong	NULL
	Gemini	New York City	CryptoCurrency	68	0.07	7/18/2022	Unknown	United States	423

处理后数据

	company	locations	industry	total_on_boarding	percentage_on_boarding	date	stage	country	funds_raised_millions
▶	Gemini	New York City	Crypto	NULL	0.1	2023-01-23	Unknown	United States	423
	Gemini	New York City	Crypto	100	0.1	2022-06-02	Unknown	United States	423
	Gemini	New York City	Crypto	68	0.07	2022-07-18	Unknown	United States	423
	GSR	Hong Kong	Crypto	NULL	NULL	2022-10-11	Unknown	Hong Kong	NULL
	Unstoppable Domains	SF Bay Area	Crypto	42	0.25	2022-07-14	Series B	United States	7

③ 处理空白值

源数据

	company	location	industry	total_on_boarding	percentage_on_boarding	date	stage	country	funds_raised_millions
▶	Airbnb	SF Bay Area		30	NULL	3/3/2023	Post-IPO	United States	6400
	Airbnb	SF Bay Area	Travel	1900	0.25	5/5/2020	Private Equity	United States	5400

处理后数据

	company	locations	industry	total_on_onboarding	percentage_on_onboarding	date	stage	country	funds_raised_millions
▶	Airbnb	SF Bay Area	Travel	30	NULL	2023-03-03	Post-IPO	United States	6400
	Airbnb	SF Bay Area	Travel	1900	0.25	2020-05-05	Private Equity	United States	5400