

Data Curation: Towards a Tool for All

José Dias¹, Jácome Cunha^{1,2}, Rui Pereira²

1.University of Minho, 2.HASLab/INESC Tec, Portugal

a78494@alunos.uminho.pt, {jacome, ruipereira}@di.uminho.pt



Problems:

- Repeating mundane tasks
- Identifying solutions
- Interpreting data
- Understand transformations

Solutions:

- Data based suggestions
- Automatic statistic calculations
- Spreadsheet-like view of data
- Separate data table to preview



Statistical cards used to help the user understand the data presented in the preview table.

Count	
Nougat	18
Marshmallow	10
Lollipop	5
MARSHMALLOW	3

Replace similar values?

Marshmallow to MARSHMALLOW

MARSHMALLOW to Marshmallow

Both to:

APPLY **APP**

Suggestion cards used to help the user find the right solution for a given problem. Some are specific to the data others more generic.

Data table used to preview any kind of transformation before they are applied. In this table it is also possible to directly alter the data in each cell and use visual filters directly.

	Model	Brand	OS_Version	Codename
1	'VS500PP'	'lge'	'6.0.1'	Marshmallow
2	'AO5510'	'YU'	'5.1.1'	Lollipop
3	'ASUS_Z017D'	'asus'	'7.0'	Nougat
4	'ASUS_X014D'	'asus'	'5.1.1'	Lollipop
5	'Nexus 5'	'google'	'6.0.1'	MARSHMALLOW

	Model	Brand	OS_Version	Codename
1	'VS500PP'	'lge'	'6.0.1'	Marshmallow
2	'AO5510'	'YU'	'5.1.1'	Lollipop
3	'ASUS_Z017D'	'asus'	'7.0'	Nougat
4	'ASUS_X014D'	'asus'	'5.1.1'	Lollipop
5	'Nexus 5'	'google'	'6.0.1'	MARSHMALLOW

A data table that is used to represent the actual state of the data. This table is can't be directly altered but visual filters can also be applied as they do not change the data.

Possible Column Filters:

Filter by condition:

None

Filter by value:

Search

Select all Clear

(Blank cells)

KitKat

MARSHMALLOW

OK

Cancel

Filter by condition:

None

Is empty

Is not empty

Is equal to

Is not equal to

Begins with

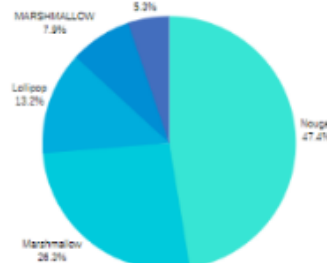
Ends with

Contains

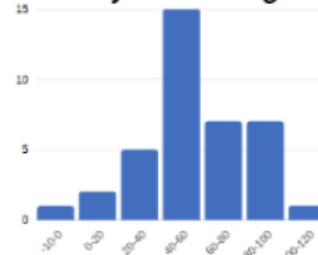
Does not contain

Alternative Statistical Cards:

OS_Version Pie Chart



Battery Level Histogram



This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project UIDB/50014/2020.