|  |  |
| --- | --- |
| **How to use regular expressions to retrieve analytics fields by alpha and numeric characters and combinations** |  |



**Yoel Kortick**

**Senior Librarian**

**Question**

In our institution the primary identifier is either all alphabetic, all numeric, or a mixture of the two. In some cases it is all alphabetic with a certain numeric string at the end. We know the types of users based on the alphabetic or numeric characters and therefore we need a way to retrieve them in analytics as follows:

1. All numbers
2. All letters
3. All numbers followed by all letters
4. All numbers followed by specific letters
5. All letters followed by all numbers
6. All letters followed by specific numbers

First, we will provide a “short” answer and then a “full” answer.

**Short Answer:**

Here we will use field "User Details"."Primary Identifier" from the “Users” subject area. The same formula can be used on any field.

First Example: All numeric characters

This will return "1" if the string is all numbers and 0 if there is at least one alphabetic character (you can then filter on this for = 1)

**evaluate('REGEXP\_INSTR(%1, ''^[0-9]+$'')',"User Details"."Primary Identifier")**

Second Example: All alphabetic characters

This will return "1" if the string is all alphabetic characters and 0 if there is at least one number in the string (you can then filter on this for = 1)

**evaluate('REGEXP\_INSTR(%1, ''^[A-z]+$'')',"User Details"."Primary Identifier")**

Third Example: Starts with numeric characters and ends alphabetic with characters

This will return "1" if the string starts with numbers and ends with characters and 0 if not (you can then filter on this for = 1)

**evaluate('REGEXP\_INSTR(%1, ''^[0-9].\*[A-z]+$'')',"User Details"."Primary Identifier")**

Fourth Example: Starts with numeric characters and ends with specific alphabetic character 'A'

This will return "1" if the string starts with numbers and ends with 'A' and 0 if not (you can then filter on this for = 1)

**evaluate('REGEXP\_INSTR(%1, ''^[0-9].\*A+$'')',"User Details"."Primary Identifier")**

Fifth Example: Starts with alphabetic characters and ends with numeric characters

This will return "1" if the string starts with alphabetic characters and ends with numbers and 0 if not (you can then filter on this for = 1)

**evaluate('REGEXP\_INSTR(%1, ''^[A-z].\*[0-9]+$'')',"User Details"."Primary Identifier")**

Sixth Example: Starts with alphabetic characters and ends with specific numeric characters '613'

This will return "1" if the string starts with alphabetic characters and ends with 613 and 0 if not (you can then filter on this for = 1)

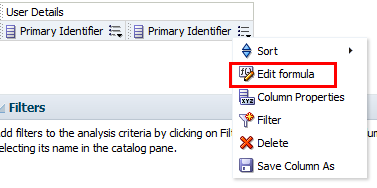
**evaluate('REGEXP\_INSTR(%1, ''^[A-z].\*613+$'')',"User Details"."Primary Identifier")**

After getting a “1” or a “0” filter the field by “1” to get the desired rows from the report.

**Full Answer**

**ONE**

In the ‘Users” subject area create a report and in the criteria tab include the field “Primary Identifier” from the “User Details” folder twice. On one of the fields do “edit formula”.



**TWO**

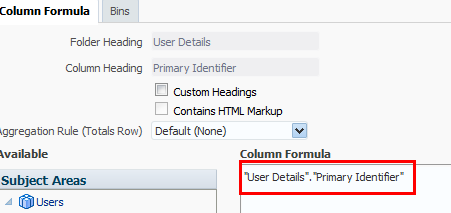
Replace the default text:

**"User Details"."Primary Identifier"**

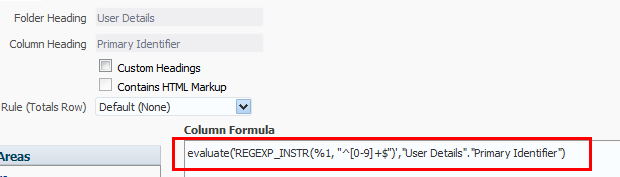
with the text from the first example above:

**evaluate('REGEXP\_INSTR(%1, ''^[0-9]+$'')',"User Details"."Primary Identifier")**

Instead of this



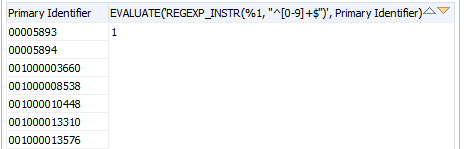
Have this

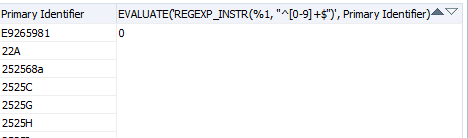


**THREE**

See that there is a "1" if the string is all numbers and 0 if there is at least one alphabetic character.

You can then filter this field by “1” to get only identifiers with all numbers.

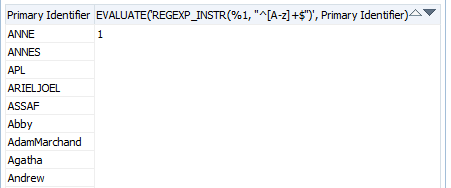


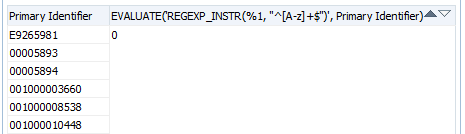


**FOUR**

Now go back to the criteria tab and replace the formula with the second example above.

See that you get "1" if the string is all alphabetic characters and 0 if there is at least one number in the string (you can then filter on this for = 1)

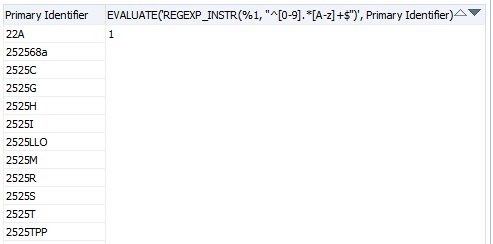


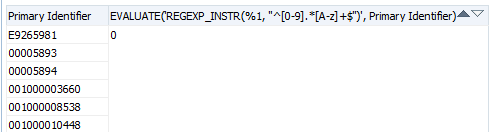


**FIVE**

Now go back to the criteria tab and replace the formula with the third example above.

See that you get "1" if the string starts with numbers and ends with characters and 0 if not (you can then filter on this for = 1)





**SIX**

Continue in the same manner for other examples.

After getting a “1” or a “0” filter the field by “1” to get the desired rows from the report.