

New MySQL functions

Abstract

In this report, I explore new MySQL functions. Specifically, I try string manipulation, extracting day of the week, getting variance and standard deviation, and manually replacing null values using coalesce()

For the first query I experiment with 2 string functions, specifically CONCAT_WS and LEFT. CONCAT_WS concatenates two strings with a separator of one's choice. LEFT select a certain number of characters from a string starting from the left.

QUERY 1:

```
select
title,
collcode,
itemtype,
CONCAT_WS("_", collcode,itemtype) as combo_class,
LEFT(itemtype,1) as adult_or_juvenile
from spl_2016.inraw limit 10
```

RESULT:

The result is stored in concat_and_left.csv.

title	collcode	itemtype	combo_class	adult_or_juvenile
Portrait of Jennie	nalpfc	acbkc	nalpfc_acbkc	a
Big Jimmys Kum Kau Chinese take out	ncpic	jcbkc	ncpic_jcbkc	j
Seminoles	ccnf	jcbkc	ccnf_jcbkc	j
Too many pumpkins	ncpic	jcbkc	ncpic_jcbkc	j
art of the handwritten note a guide to reclaiming civilized communications	nanf	acbkc	nanf_acbkc	a
forgotten a Peter Decker Rina Lazarus novel	namys	acbkc	namys_acbkc	a
Youth league baseball coaching and playing	ncnf	jcbkc	ncnf_jcbkc	j
Ready set go	ncrdrc	jcbkc	ncrdrc_jcbkc	j
story about me	ncpic	jcbkc	ncpic_jcbkc	j
Five senses	cceck	jckit	cceck_jckit	j

Combo_class is created using the CONCAT_WS() command by combining collection code and item type. The result is a classification that includes both variables.

Adult or juvenile is created by using LEFT() to select the first letter of the itemtype, since in this classification adult items start with “ac” and juvenile items start with “jc”. In the end result, the first letter “a” stands for adult, and the first letter “j” stands for juvenile.

For the second query, I will use the command DAYNAME(). This command applies to a date and returns the day of the week of that date. As an example, I will get aggregate counts per days of the week in March 2015, to see which day is typically the busiest in the library.

QUERY 2:

```
select
DAYNAME(cout) as day_of_the_week,
count(*) as counts
from spl_2016.inraw
where date_format(cout, '%Y-%m-%d') between '2015-03-01' and '2015-03-31'
group by 1
order by 2
```

RESULT: the result is stored in dayname.csv. As seen in the table below, Monday is typically the busiest day in SPL.

day_of_the_week	counts
Monday	83411
Tuesday	72426
Saturday	72388
Wednesday	60123
Thursday	57292
Sunday	54700
Friday	42596

In the next query I will use some statistical functions, namely sample standard deviation and sample variance, both being measures of spread. I will use STDDEV_SAMP() and VAR_SAMP() commands, to get both statistics for daily checkout counts in March 2015. In addition, I will use ROUND() command to round both to 2 decimal places.

QUERY 3:

```
select
```

```

avg(counts) as mean,
round(stddev_samp(counts),2) as standard_deviation,
round(var_samp(counts),2) as variance
from(
select
day(cout) as days,
count(*) as counts
from spl_2016.inraw x
where date_format(cout, '%Y-%m-%d') between '2015-03-01' and '2015-03-31'
group by 1) y

```

RESULT: the result is stored in sd_and_var.csv. and displayed below. Relative to the mean, the standard deviation is pretty high, meaning that there was noticeable variation in book counts in March 2015.

mean	standard_deviation	variance
14288.26	2641.08	6975282

In the next query I will use COALESCE() command to replace null values in a table with something else. This command has a syntax: COALESCE(expression, value), where “expression” can be a column to scan, and “value” is what I will replace null values with.

QUERY 4:

```

select
title,
subj,
COALESCE(subj,"some subject") as coalesce_example
from spl_2016.inraw
where subj is null
limit 10

```

RESULT: the result is stored in “coalesce.csv”. As seen in the table below, coalesce took “subj” that is null and replaced it with an arbitrary name that I gave it. The benefit of this command is that it can be used to manually fill in null values in the database based on a certain condition.

title	subj	coalesce_example
Portrait of Jennie	NULL	some subject
Big Jimmys Kum Kau Chinese take out	NULL	some subject

Seminoles	NULL	some subject
Too many pumpkins	NULL	some subject
art of the handwritten note a guide to reclaiming civilized communications	NULL	some subject
forgotten a Peter Decker Rina Lazarus novel	NULL	some subject
Youth league baseball coaching and playing	NULL	some subject
Ready set go	NULL	some subject
story about me	NULL	some subject
Five senses	NULL	some subject