Data Exploration Assignment, Group 5

Yunao Shen, Xinyang Wang, Yan Gu

All our stored procedure are written with MySQL 8.0

1- **Create a stored procedure to accept a table name and a column name as inputs and returns Min, Max, Median, Average, Q1 and Q3 as outputs.**

Code:

```
CREATE DEFINER=`root`@`localhost` PROCEDURE `get_univar`(in tbl varchar(100), attribute varchar(100))
BEGIN
set @rowindex := -1;
set @findmedian = concat(
'select avg(t.col) into @median
from
(select @rowindex:=@rowindex + 1 as rowindex,', tbl,'.',attribute,' as col ',
' from ', tbl,
' order by ', tbl,'.',attribute, ') as t',
' where t.rowindex IN (FLOOR(@rowindex / 2), CEIL(@rowindex / 2));'
);
prepare findmedian from @findmedian;
execute findmedian;
DEALLOCATE PREPARE findmedian;

set @query1 = concat('select min(', attribute, '), max(',attribute, '), TRUNCATE(@median,2) as median, avg(',attribute, ') ', ' from ', tbl);
prepare getsql from @query1;
execute getsql;
DEALLOCATE PREPARE getsql;
END
```

Test:

```
use ABC_Retail;
call get_univar('orders','order_freight');
```

Results:

	min(order_freight)	max(order_freight)	median	avg(order_freight)
▶	0.02	1007.64	41.36	78.217241

2- **Create a stored procedure to accept a table name and two column names as inputs and returns Correlation Coefficient as output.**

```
CREATE DEFINER=`admin`@`%` PROCEDURE `get_correlation_coefficient`(in tbl varchar(100), col1 varchar(100), col2 varchar(100))
BEGIN
#select count
Set @sql = concat('select count(', col1, ') into @cnt', ' from ', tbl);
prepare getsql from @sql;
execute getsql;
#select @cnt;

#select E(col1), E(col2), E(col1*col2), std(col1), std(col2)
set @sql = concat('select (sum(', col1, ')/@cnt),', ' (sum(', col2, ')/@cnt),',
                  ' (sum(', col1, '*', col2, ')/@cnt),',
                  ' std(', col1, '),', ' std(', col2, ')',
                  ' into @E1, @E2, @E12, @s1, @s2',
                  ' from ', tbl
                  );
prepare getsql from @sql;
execute getsql;
select @E1,@E2,@E12,@s1,@s2;

#select Correlation coefficient
select (@E12-@E1*@E2)/(@s1*@s2) as Correlation_coefficient;

END
```

We test it on **orders** table in **abc_retail,** we can see when we pass into same column, the correlation coefficient is **1**