DATA EXPLORATION

Dev Shah -ds1560
Kartik Rattan -kr642
Gautam Sikka -gs805

PROBLEM:

1. *Write a Stored Procedure called usp_univariate*
2. *usp_univariate accepts two parameters, @table_name and @column_name*
3. *usp_univariate returns Count, Average, Median, Mode, Rane, Variance, Standard Deviation and Coefficient of Variation*

CODE:

```
create procedure usp_univariate @table_name nvarchar(30), @column_name nvarchar(500)
as
declare @sql nvarchar(max)
declare @average nvarchar(max)
set @sql = 'select count (' +@column_name + ') as count ,
                  avg('+ @column_name  + ') as average,
                  ( SELECT TOP 1' +  @column_name + ' from ' +  @table_name + ' GROUP  BY '+ @column_name + ' ORDER  BY COUNT(*) DESC) as mode,
                  max (' + @column_name + ') - min (' + @column_name +' ) as range,
                  var(' +@column_name +') as variance,
                  sqrt( var(' + @column_name +') ) as standard_deviation,
                  sqrt( var(' + @column_name +') ) /  avg(' + @column_name  + ') as CV,
                  (((select max(' + @column_name +') from (select top 50 percent '+@column_name+' from '+@table_name+' order by '+@column_name+') as a)+
                  (select min('+ @column_name +') from (select top 50 percent '+@column_name +' from '+ @table_name +' order by '+ @column_name +' desc) as c))/2)
                  as median from '+@table_name
EXECUTE sp_executesql @sql

exec usp_univariate 'orders_denorm','order_dow'
drop procedure usp_univariate
```

RESULT:

	count	average	mode	range	variance	standard_deviation	CV	median
1	1384617	2	0	6	4.69868761293821	2.16764563823015	1.08382281911508	3

PROBLEM:

4. *Write a Stored Procedure called usp_bivariate*
5. *usp_bivariate accepts three parameters, @table_name, @target_colname and @predictor_colname*
6. *usp_bivariate returns Z value*

CODE:

```
create procedure usp_bivariate @table_name nvarchar(max),@target_colname nvarchar(max),@predictor_colname nvarchar(max)
as
declare @cmd nvarchar(max)

create table temp( target varchar(max),count int,avg float,std float, var float)
select @cmd ='insert into temp
select '+@target_colname+', count('+@predictor_colname+') , avg(1.0*'+@predictor_colname+'),stdev('+@predictor_colname+'),var('+@predictor_colname+')
from '+@table_name+' group by '+@target_colname
exec(@cmd)
select * from temp
declare @z_value float
declare @sd1 float
declare @sd2 float
declare @var1 float
declare @var2 float
declare @c1 int
declare @c2 int
declare @avg1 float
declare @avg2 float

select @avg1 = (select top 1 avg from temp order by avg desc)
select @sd1 = (select top 1 std from temp order by avg desc)
select @var1 = (select top 1 var from temp order by avg desc)
select @c1 = (select top 1 count from temp order by avg desc)

select @avg2 = (select top 1 avg from temp order by avg )
select @sd2 = (select top 1 std from temp order by avg )
select @var2 = (select top 1 var from temp order by avg)
select @c2 = (select top 1 count from temp order by avg)
drop table temp
set @z_value = (@avg1 - @avg2)/sqrt((@var1/@c1)+(@var2/@c2))
select @z_value as Z_Value

exec usp_bivariate 'challengerORing', 'ORingFailure', 'LaunchTemperature'
drop procedure usp_bivariate
```

RESULT:

	target	count	avg	std	var
1	N	17	72.294117	4.74031520421308	22.4705882352937
2	Y	6	61.833333	7.08284312029191	50.1666666666664

	Z_Value
1	3.36171798517282