```
/*Creating dataset*/
proc sql;
create table cars2 as
select make,normalized_losses,fuel_type,body_style,horsepower,price,height,width,highway_mpg,engine_size
from saspro.cars1
;
run;

/*Description of the dataset*/
proc contents data=cars2;
run;


/*Checking null values*/
title 'null values';
proc means data=cars2 N Nmiss;
run;

/* 2 missing record in horsepower so we calculating mean of horsepower to update null values  mean value */
proc means data=cars2 mean maxdec=2;
var horsepower;
run;

data saspro.carsinfo;
set cars2;
if horsepower='.' then horsepower='104.26';
run;

/*update missing value to of normalized_losses */
data saspro.carsinfo;
set cars2;
if normalized_losses='?' then normalized_losses='124';
run;

/*Description of the dataset*/
proc contents data=saspro.carsinfo;
run;

/*plotting histogram of horsepower*/
proc univariate data = saspro.carsinfo;
   histogram horsepower
   /
   normal (
   mu = est
   sigma = est
   color = blue
   w = 2.5
   )
barlabel = percent
midpoints =176 to 330 by 50;
run;

/*
Correlation between 3 varaibles
*/
proc corr data=saspro.carsinfo;
title 'correlation between variables';
var horsepower price highway_mpg;
run;

/*Correlation between engine_size and price*/
PROC SGPLOT DATA=saspro.carsinfo;
 reg X =engine_size  Y = price / lineattrs=(color=red thickness=2);
 TITLE 'Correlation between engine_size and price';
RUN;

/*
checking the VIF and tolerance  of data
*/
proc reg data=saspro.carsinfo;
title 'checking the mulicolinarty and tolerance  of data';
model  price= highway_mpg horsepower engine_size/ vif tol;
run;
```

```
/*
There is no multi colinarity in dataset
*/

/*model of linear regression */
proc reg data=saspro.carsinfo;
title 'model of linear regression';
model price=engine_size;
run;

/*
This is right model to predict the price of car with value of R sqaured is 0.74
*/
```