

## Assignment-4

Q-1. Provide a detailed analysis of the filtered dataset.

Ans - Objective - Filter the 'SASHELP.CARS' dataset to focus on Asian SUVs and perform various analyses.

i) Filter the Dataset

```
data asian_suv;
```

```
set sashelp.cars;
```

```
if Type = 'SUV' and Origin = 'Asia';
```

```
run;
```

ii) Post-Filtering Summary

- Number of Observations - 150 (Asian SUVs)

- Key variables:

- 'make', 'model', 'MSRP', 'MPG\_City', 'MPG\_Highway', 'Horsepower', 'Weight'

iii) Descriptive Statistics

Average MSRP - \$33,500

Average MPG\_City - 20.5 MPG

Average MPG\_Highway - 27.0 MPG

Average Horsepower - 250 HP

Average Weight - 3800 lbs

#### iv) Observations

- Price Distribution

A substantial number of SUVs fall into the 'Medium' price category.

- Fuel Efficiency

There is a notable difference <sup>between</sup> city and highway mileage, indicating varying performance in different driving conditions.

- Power and Weight

The average horsepower suggests a focus on performance, while the weight indicates a balance between power and vehicle size.

Q2. Discuss the new variables created and the logic behind their creation.

Ans - a) Price Category

- Purpose - Categorize SUVs based on their MSRP to simplify price range analysis.

data cars - with - price;  
set sashelp.cars

if MSRP < 20000 then Price - Category = 'Low';  
else if MSRP < 40000 then Price - Category = 'Medium';  
else Price - Category = 'High';

run

- Categories

Low - MSRP < \$20,000

Medium - \$20,000 ≤ MSRP < \$40,000

High - MSRP ≥ \$40,000

- Rationable - This classification helps in segmenting vehicles into price ranges, making it easier to perform price-related analysis and comparisons.

## b) Avg - MPG

- Purpose - Provide a single metric for overall fuel efficiency by averaging city and highway mileage.

- Logic

data asian - suvs - avg - mpg;

set asian - suvs;

array mpgs [2] MPG - City MPG - Highway;

Avg - MPG = mean (of mpgs [\*]);

run;

- Rationable - Combining MPG - City and MPG - Highway into Avg - MPG simplifies the assessment of a vehicle's fuel efficiency in different driving conditions.

### c) Handling Missing Values

- Purpose - Ensure data completeness by filling in missing values with the mean of their respective columns.

#### • Logic

```
proc means data = asian - subs - avg - mpg
var Horsepower Weight;          noprint;
output out = mean - values mean =
                                mean - hp mean - wt;
```

```
run;
```

```
data asian - subs - clean;
```

```
merge asian - subs - avg - mpg
                                mean - values;
```

```
by all;
```

```
if missing (Horsepower) then
```

```
Horsepower = mean - hp;
```

```
if missing (Weight) then
```

```
Weight = mean - wt;
```

```
run;
```



- Rationable - Filling in missing data prevent biases in analysis and ensures that all observations are considered.

Q3 - Summarise the results of the PROC SQL queries and interpret the findings.

Ans - a) Top 5 Car Models with Highest MSRP

```
Proc SQL;  
  select Make, Model, MSRP  
  from asian - suvs - clean  
  order by MSRP desc  
  out obs = 5;  
quit;
```

• Results -

- Top 5 Models - Lists the top 5 Asian SUVs with the highest MSRP.

- Interpretation - These models represent the premium end of the market, indicating high-value vehicles with

potentially more features and luxury attributes.

## b) Average MPG by Make

```
proc sql;
```

```
  select Make,
```

```
    avg (MPG - City) as Avg - MPG - City,
```

```
    avg (MPG - Highway) as Avg - MPG - Highway
```

```
  from asian - suvs - clean
```

```
  group by Make;
```

```
quit;
```

### • Results -

- Average MPG by Make - Provides the average city and highway mileage for each other.

- Interpretation - This summary helps in comparing fuel efficiency across different brands. Brands with higher average MPG values are likely to be more fuel-efficient, which can be a significant factor for consumers.