1. Exploratory Factor Analysis:

Factor analysis is a statistical method used to describe variability among observed, correlated variables. The goal of performing factor analysis is to search for some unobserved variables called factors.

The correlation values (between different variables in study) varies from a minimum -0.911 to 0.928. These differences suggests the variables could be reduced down to at least two underlying variables or factors.

1.1. Sampling Adequacy

Exploratory factor analysis is only useful if the matrix of population correlation is statistically different from the identity matrix. If these are equal, the variables are few interrelated, i.e., the specific factors explain the greater proportion of the variance and the common factors are unimportant. Therefore, it should be defined when the correlations between the original variables are sufficiently high. Thus, the factor analysis is useful in estimation of common factors. With this in mind, the Bartlett Sphericity test can be used. The hypotheses are:

H0: the matrix of population correlations is equal to the identity matrix

H1: the matrix of population correlations is different from the identity matrix.

1.1.1. Bartlett Test:

Chi_Squared	pvalue	df
438.815	0.000	55.000

Based on the results, it is possible to verify that p-value is 0 ($p<0.05$), which allow us to conclude the null hypothesis is rejected. Thus, the matrix of population correlations is different from the identity matrix. This difference suggests that factorial analysis is appropriate to our data.

1.1.2. KMO Test

KMO (Kaiser-Meyer-Olkin) is a widely method to measure the adequacy of sampling. KMO checks if it is possible to factorize the primary variables efficiently. For reference, Kaiser suggested the following classification of the results:

- 0 to 0.49 unacceptable
- 0.50 to 0.59 miserable

- 0.60 to 0.69 mediocre
- 0.70 to 0.79 middling
- 0.80 to 0.89 meritorious
- 0.90 to 1 marvelous

KMO_value
0.864

Variable	KMO_value_per_variable
mpg	0.915
cyl	0.925
disp	0.918
hp	0.883
drat	0.940
wt	0.885
qsec	0.733
vs	0.819
am	0.836
gear	0.779
carb	0.742

The analysis show a KMO equal to 0.864. This results suggest that the degree of common variance in our dataset is "meritorious". Additionally, all variables have KMO higher than 0.5, and therefore, the factor analysis is appropriate to this data.

1.2. Retained Factors

After deciding about the adequacy or not of the model, the next step is to decide the number of factors to retain.

1.2.1. Results Eigen Values:

Variable	Eigen_values
mpg	7.028
cyl	2.430
disp	0.593
hp	0.260
drat	0.204
wt	0.159
qsec	0.122

Variable	Eigen_values
vs	0.069
am	0.053
gear	0.044
carb	0.038

The Kaisen criteria considers that only factors with eigenvalues greater than one should be retained for interpretation. Thus, the table shows that the analyzed data have 2 eigenvalues higher than 1, and therefore, 2 factors should be retained.

1.2.2. Results Summary:

.rownames	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Standard deviation	136.533	38.148	3.071	1.307	0.906	0.664	0.309	0.286	0.251	0.211	0.198
Proportion of Variance	0.927	0.072	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cumulative Proportion	0.927	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Other method oftentimes used is the variance explained criteria. This is a method based on to retain the number of factors that account for a certain percent of extracted variance. The literature varies on how much variance should be explained before the number of factors is sufficient. However, there is a consensus that should be more than 50%. Considering the minimum acceptable, at least 1 factors should be retained according the variance explained criteria.

1.3. PCA

Principal Component Analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that still contains most of the information in the large set. PCA is a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components.

1.3.1. Communalities Values

The following table show the values of communalities. Communality is the proportion of each variables variance that can be explained by the factors.

names	x
mpg	0.934
cyl	0.952
disp	0.919
hp	0.913
drat	0.814
wt	0.943
qsec	0.938

names	x
vs	0.911
am	0.887
gear	0.909
carb	0.931

Analyzing the communality values, it is possible to verify that all values are higher than 0.5. This means that the percentage of the variance of each variable explained by common factors is greater than 50% and all of them could be considered in the model.

1.3.2. Results PCA:

Variable	Loadings.PC1	Loadings.PC2	Loadings.PC3
mpg	-0.955	-0.031	-0.149
cyl	0.969	0.067	-0.094
disp	0.957	-0.045	0.023
hp	0.907	0.285	0.096
drat	-0.739	0.460	0.236
wt	0.922	-0.239	0.190
qsec	-0.521	-0.763	0.290
vs	-0.795	-0.410	0.333
am	-0.617	0.683	-0.200
gear	-0.611	0.700	0.215
carb	0.621	0.605	0.423

Analyzing the weights of each variable in the factors it is possible to decide the factor with greater weight for the variable. This is the factor that a specific variable belongs to.

Component 1 has variables cyl, disp, hp, wt, carb.

Component 2 has variables mpg, drat, am, gear.

Component 3 has variables qsec, vs.

1.3.3. Results for PCA VARIMAX:

If some doubts regarding the previous model remains, the results should be analyzed after a factor rotation (Varimax rotation, the most popular rotation method due to its simplicity).

Variable	Loadings.Varimax.RC2	Loadings.Varimax.RC3	Loadings.Varimax.RC1
mpg	0.597	0.424	-0.631
cyl	-0.611	-0.603	0.463
disp	-0.670	-0.459	0.508

Variable	Loadings.Varimax.RC2	Loadings.Varimax.RC3	Loadings.Varimax.RC1
hp	-0.386	-0.581	0.653
drat	0.862	0.263	-0.049
wt	-0.767	-0.224	0.551
qsec	-0.173	0.899	-0.317
vs	0.275	0.862	-0.305
am	0.887	-0.202	-0.244
gear	0.948	0.042	0.086
carb	0.082	-0.417	0.866

Component 1 has variables mpg, drat, am, gear.

Component 2 has variables qsec, vs.

Component 3 has variables cyl, disp, hp, wt, carb.

Now, the analyst should pay attention to two suggestions (without and with rotation) and decide which one makes the most sense for their data.