AIRLINE PASSENGER SATISFACTION

АНАЛІЗ ГОЛОВНИХ КОМПОНЕНТ

ДАНІ

CODEBOOK

Airline Passenger Satisfaction

Team Four

1 Variable Information

№	Variable	Attribute Type	Label	Units of measurement	Details
1.	id	numerical	id number of the passengers	-	not used in our research
2.	Satisfaction	binary class label	airline satisfaction level	-	2 class labels: “dissatisfied” and “satisfied”
3.	Gender	nominal	gender of the passengers	-	2 terms: “female” and “male”
4.	Customer type	nominal	the customer type	-	2 terms: “loyal customer” and “disloyal customer”
5.	Age	numerical	the actual age of the passengers	year	integer numbers from 7 to 85
6.	Type of travel	nominal	purpose of the flight of the passengers	-	2 terms: “personal travel” and “business travel”
7.	Class	nominal	travel class in the plane of the passengers	-	3 terms: “eco”, “business” and “eco plus”
8.	Flight distance	ordinal	the flight distance of this journey	km	integer from 56 to 4893
9.	Inflight Wi-Fi service	ordinal	*	-	from 0 = least satisfied to 5 = most satisfied
10.	Departure/Arrival time convenient	ordinal	*	-	from 0 = least satisfied to 5 = most satisfied

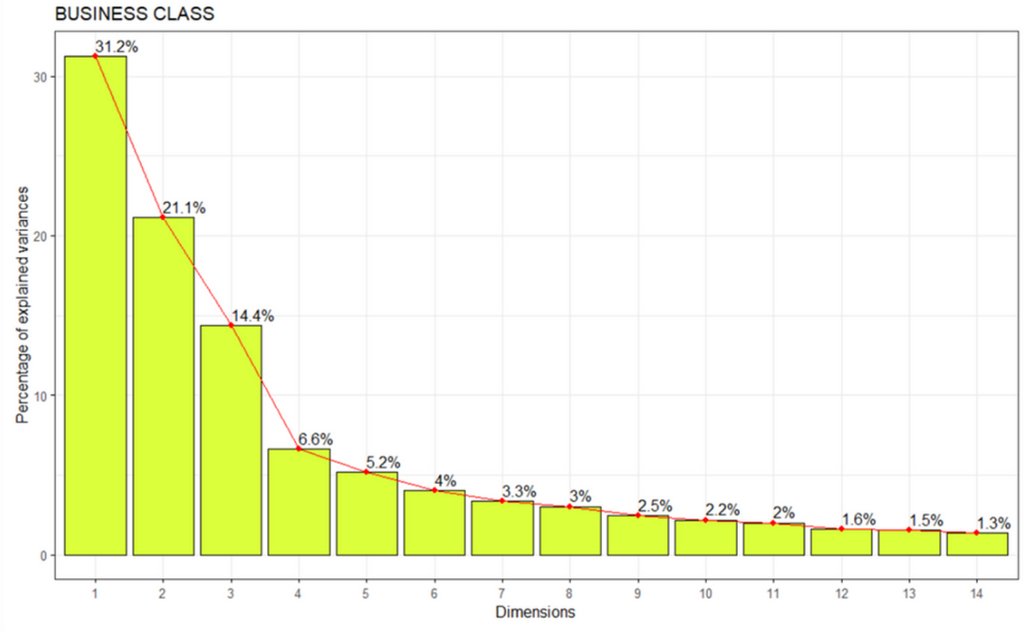
ДОЦІЛЬНІСТЬ ПРОВЕДЕННЯ PCA

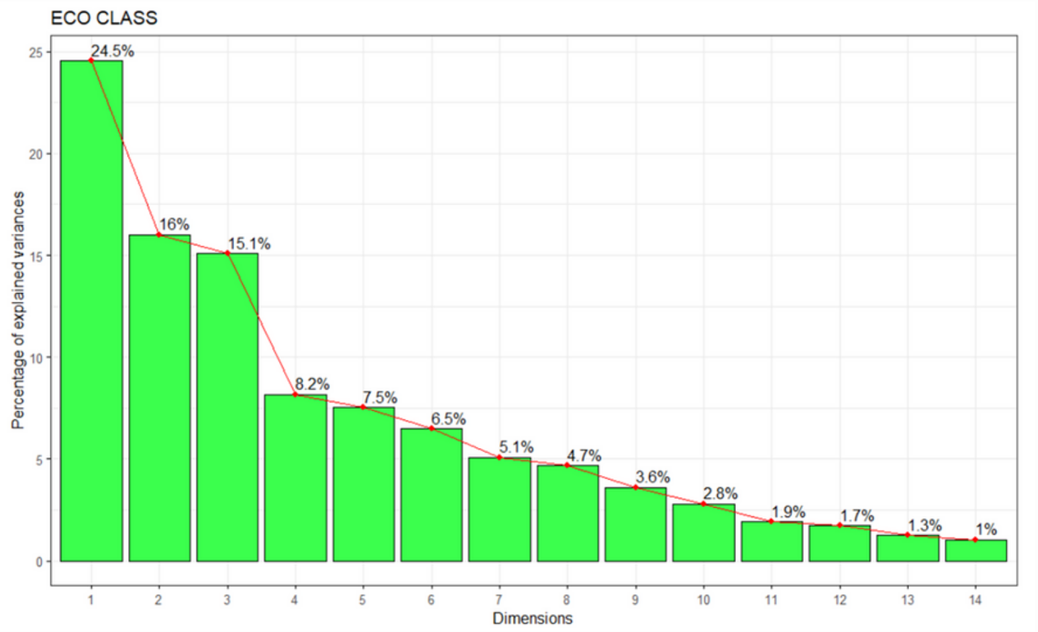
Нульова гіпотеза тесту Бартлетта стверджує, що дисперсії залежної змінної однакові для всіх груп. Якщо *p-value* менше заданого рівня значущості, то ми відхиляємо нульову гіпотезу і стверджуємо, що дисперсії неоднорідні.

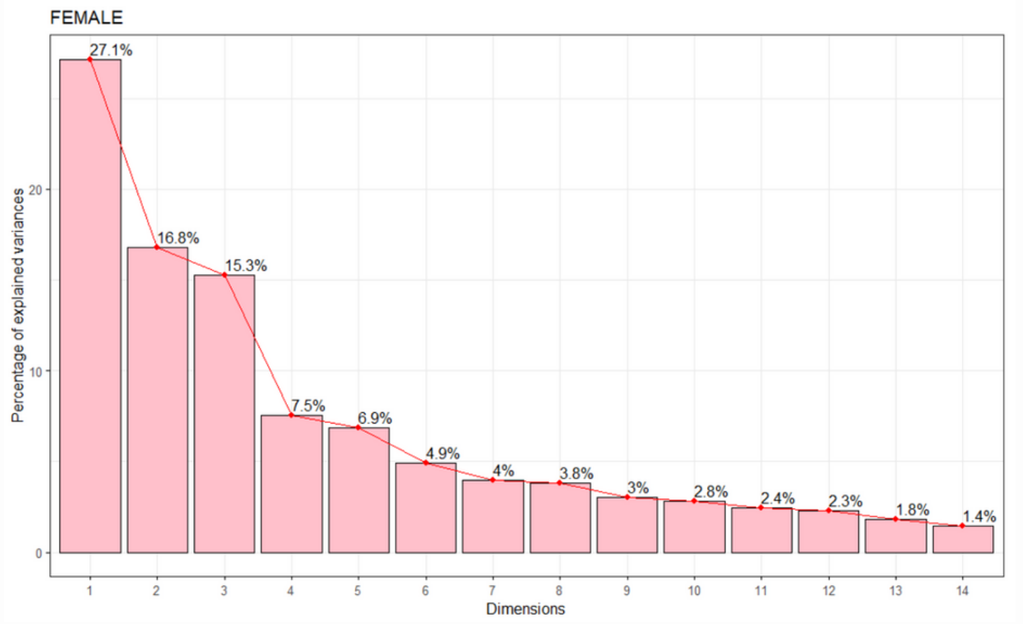
Bartlett's test: p-value = 0

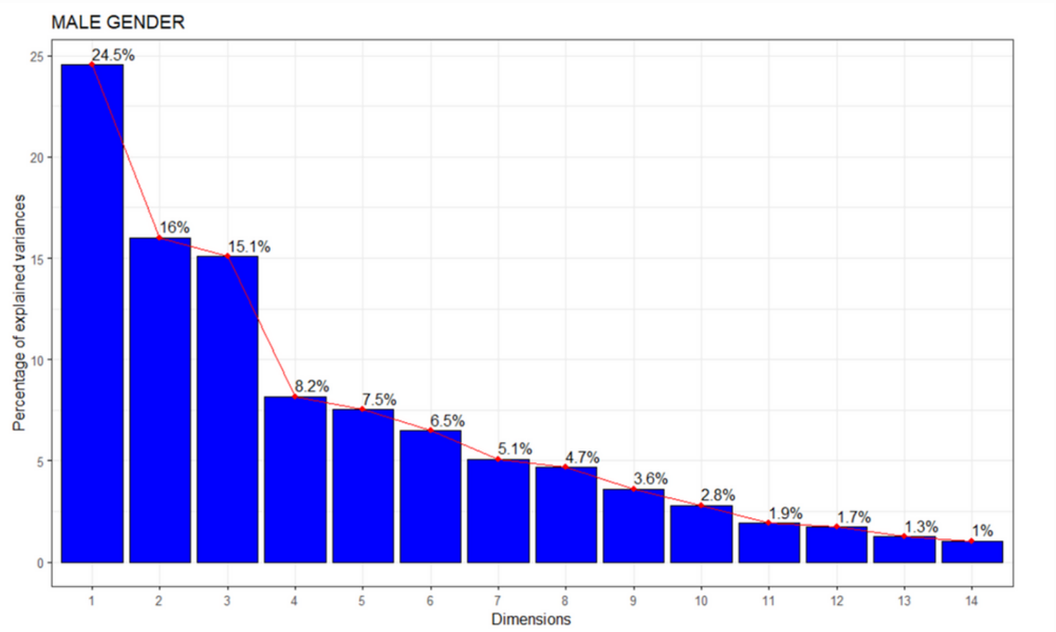
Результати показують, що значення p < 0,001 і є статистично значущим. PCA можна застосувати.

SCREE PLOTS

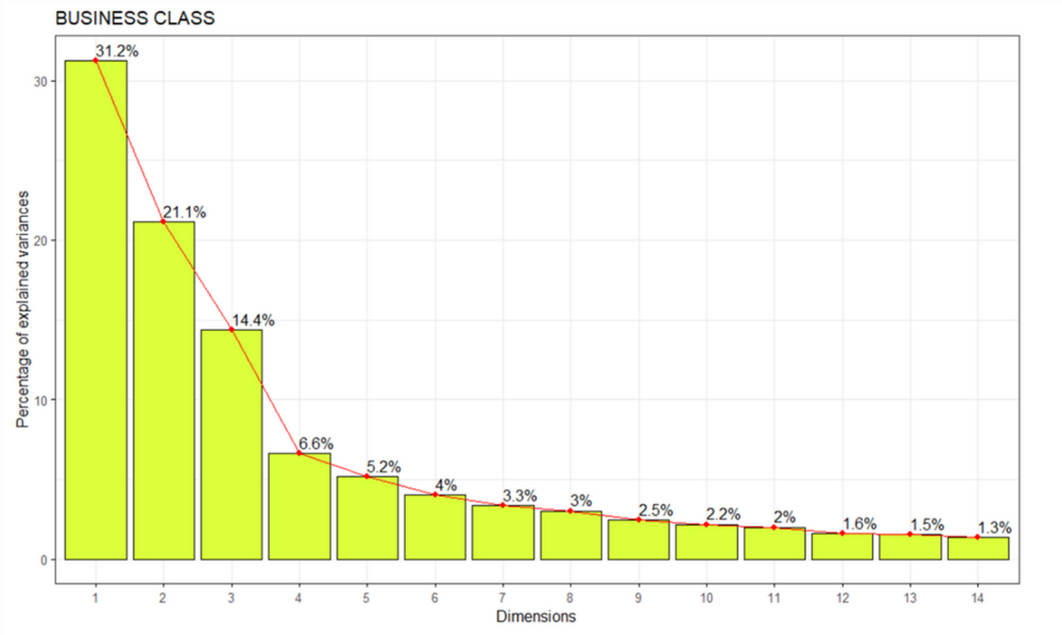

BUSINESS CLASS
Percentage of explained variances
Dimensions
31.2%
21.1%
14.4%
6.6%
5.2%
4%
3.3%
3%
2.5%
2.2%
2%
1.6%
1.5%
1.3%

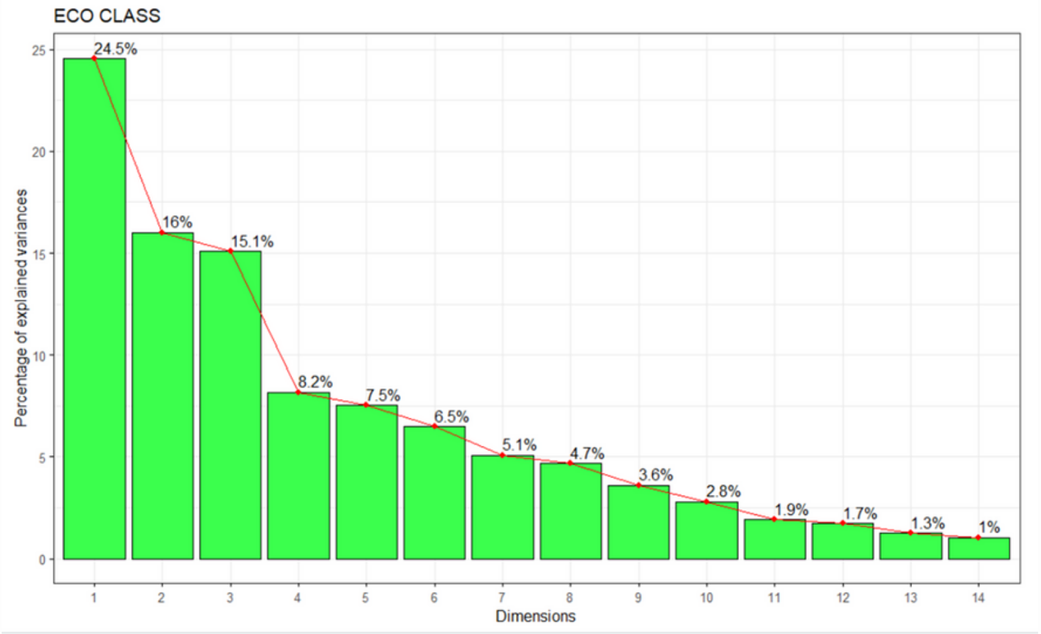

ECO CLASS
Percentage of explained variances
Dimensions
24.5%
16%
15.1%
8.2%
7.5%
6.5%
5.1%
4.7%
3.6%
2.8%
1.9%
1.7%
1.3%
1%


FEMALE
Percentage of explained variances
Dimensions
27.1%
16.8%
15.3%
7.5%
6.9%
4.9%
4%
3.8%
3%
2.8%
2.4%
2.3%
1.8%
1.4%


MALE GENDER
Percentage of explained variances
Dimensions
24.5%
16%
15.1%
8.2%
7.5%
6.5%
5.1%
4.7%
3.6%
2.8%
1.9%
1.7%
1.3%
1%

МЕТОДИ ОЦІНКИ ОПТИМАЛЬНОЇ КІЛЬКОСТІ КОМПОНЕНТ. МЕТОД 1





Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	4.372	2.958	2.011	0.929	0.725	0.562	0.467
% of var.	31.227	21.130	14.362	6.635	5.177	4.014	3.339
Cumulative % of var.	31.227	52.356	66.719	73.354	78.531	82.545	85.884

Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	3.436	2.242	2.114	1.145	1.057	0.906	0.711
% of var.	24.542	16.015	15.102	8.177	7.547	6.474	5.082
Cumulative % of var.	24.542	40.557	55.659	63.837	71.383	77.857	82.939

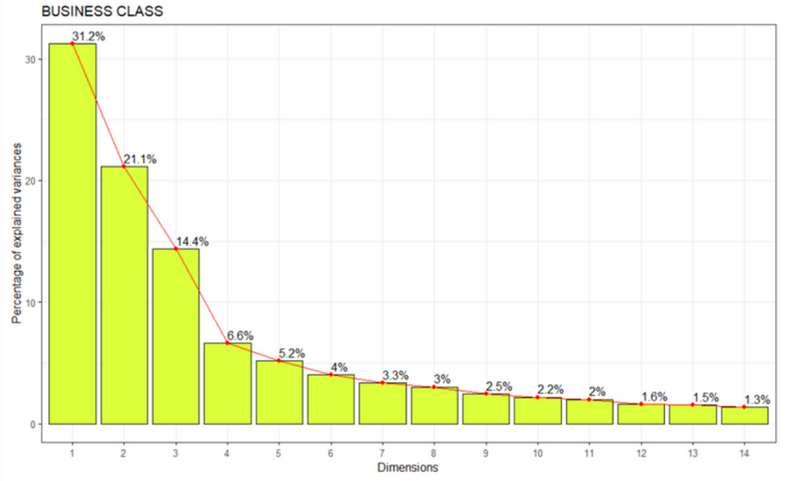
Для **бізнес-класу три** змінні описують кумулятивну дисперсію на рівні 66%.

В **економ-класі** такий самий рівень дисперсії описують **п'ять** змінних.

МЕТОДИ ОЦІНКИ ОПТИМАЛЬНОЇ КІЛЬКОСТІ КОМПОНЕНТ. МЕТОД 2

Parallel Analysis Results

Method: pca
Number of variables: 14
Sample size: 62147
Number of correlation matrices: 300
Seed: 42
Percentile: 0.95



Component <dbl>	Mean <dbl>	0.95 <dbl>
1	1.025	1.030
2	1.019	1.023
3	1.015	1.018
4	1.011	1.014
5	1.008	1.011
6	1.005	1.007
7	1.002	1.004
8	0.998	1.001
9	0.995	0.998
10	0.992	0.995

num [1:14]
4.372
2.958
2.011
0.929
0.725
...

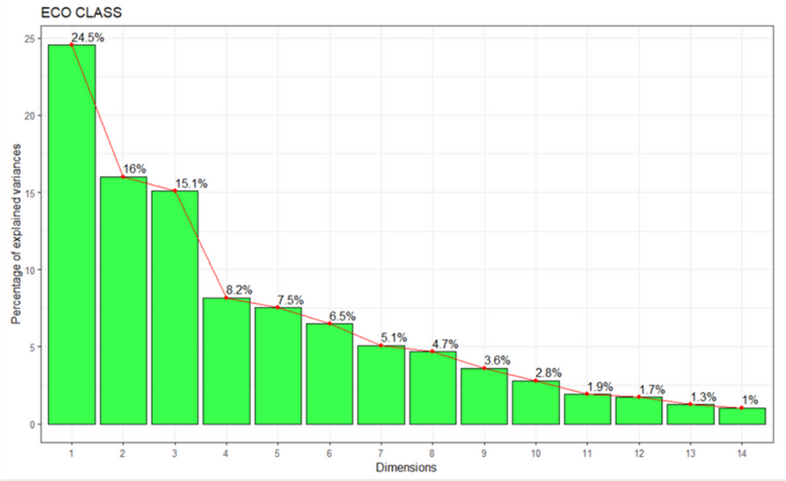
Parallel analysis

Parallel analysis, also known as Horn's parallel analysis, is a statistical method used to determine the number of components to keep in a principal...

Wikipedia / Aug 18, 2022

Parallel Analysis Results

Method: pca
Number of variables: 14
Sample size: 58293
Number of correlation matrices: 300
Seed: 42
Percentile: 0.95

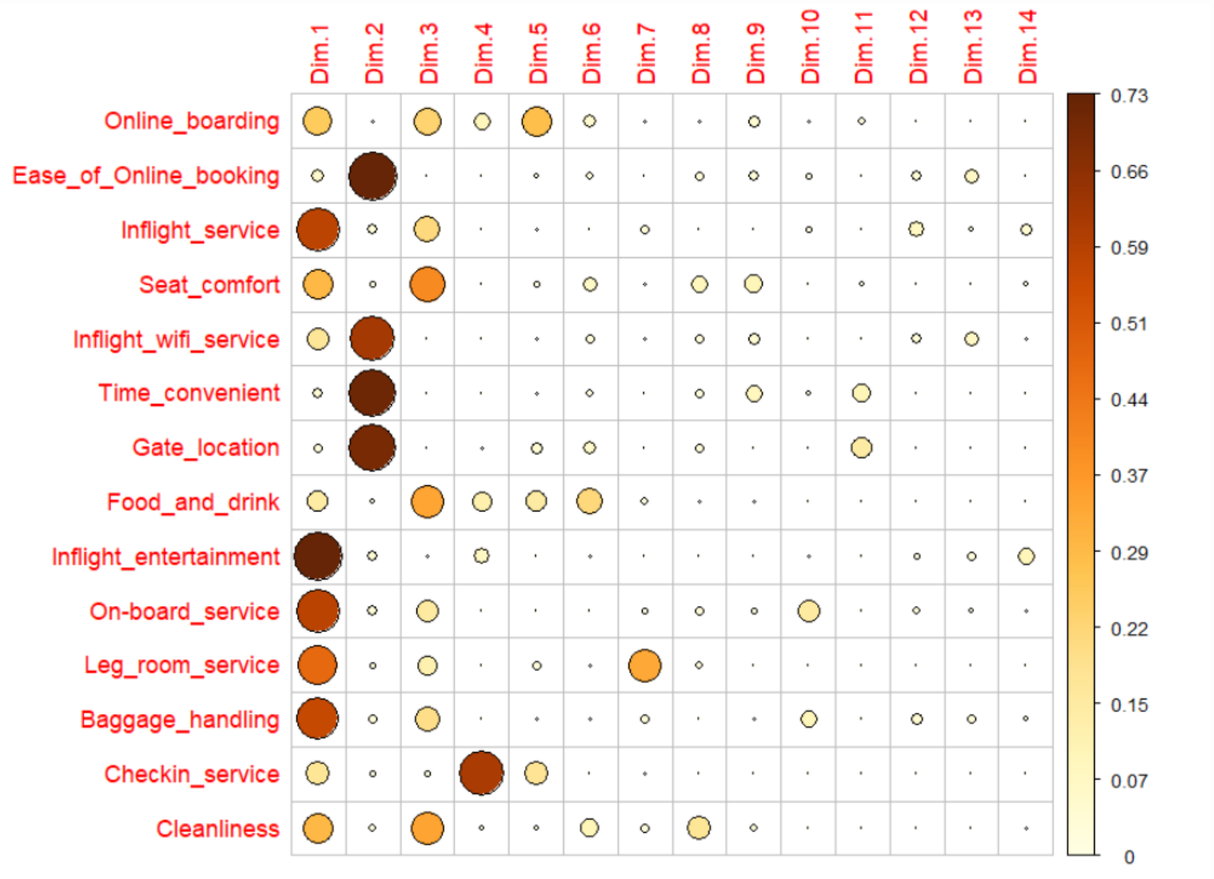


Component <dbl>	Mean <dbl>	0.95 <dbl>
1	1.026	1.030
2	1.020	1.024
3	1.016	1.019
4	1.012	1.015
5	1.008	1.011
6	1.005	1.008
7	1.001	1.004
8	0.998	1.001
9	0.995	0.997
10	0.992	0.994

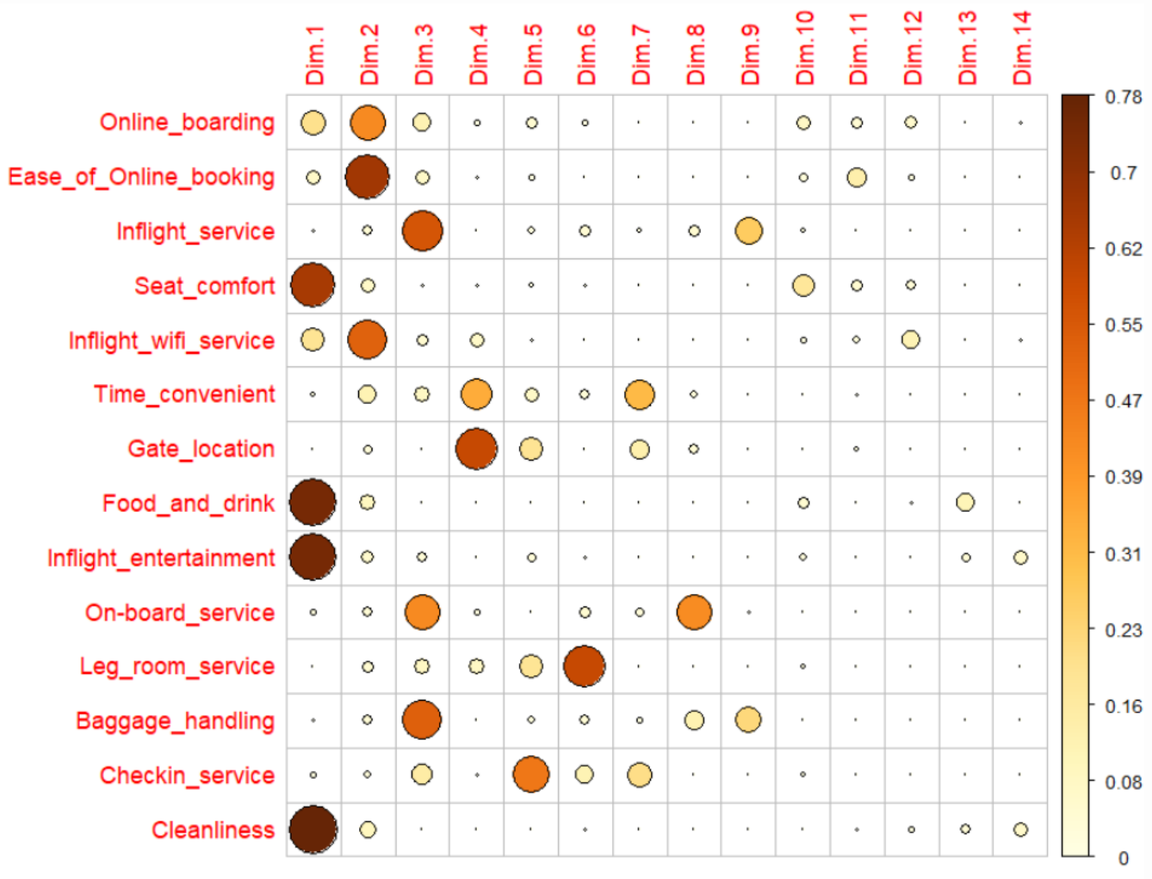
num [1:14]
3.44
2.24
2.11
1.14
1.06
...

REPRESENTATION QUALITY

BUSINESS

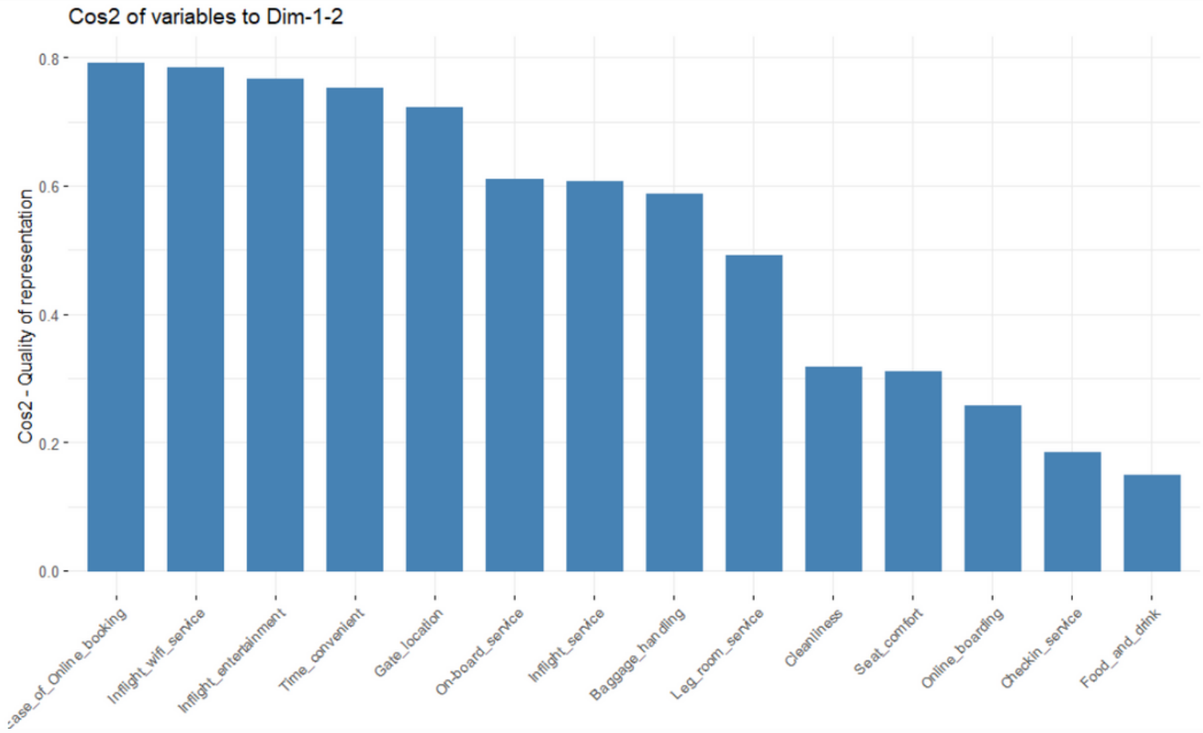


ECO

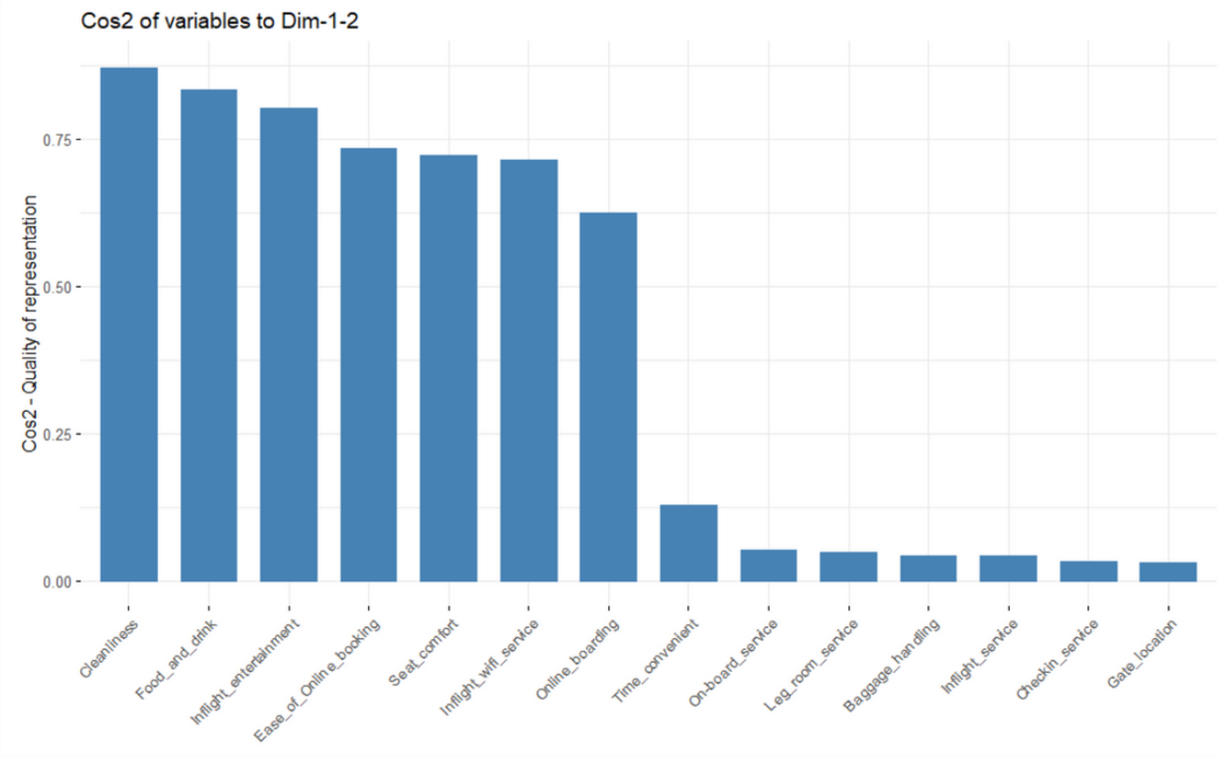


REPRESENTATION QUALITY

BUSINESS

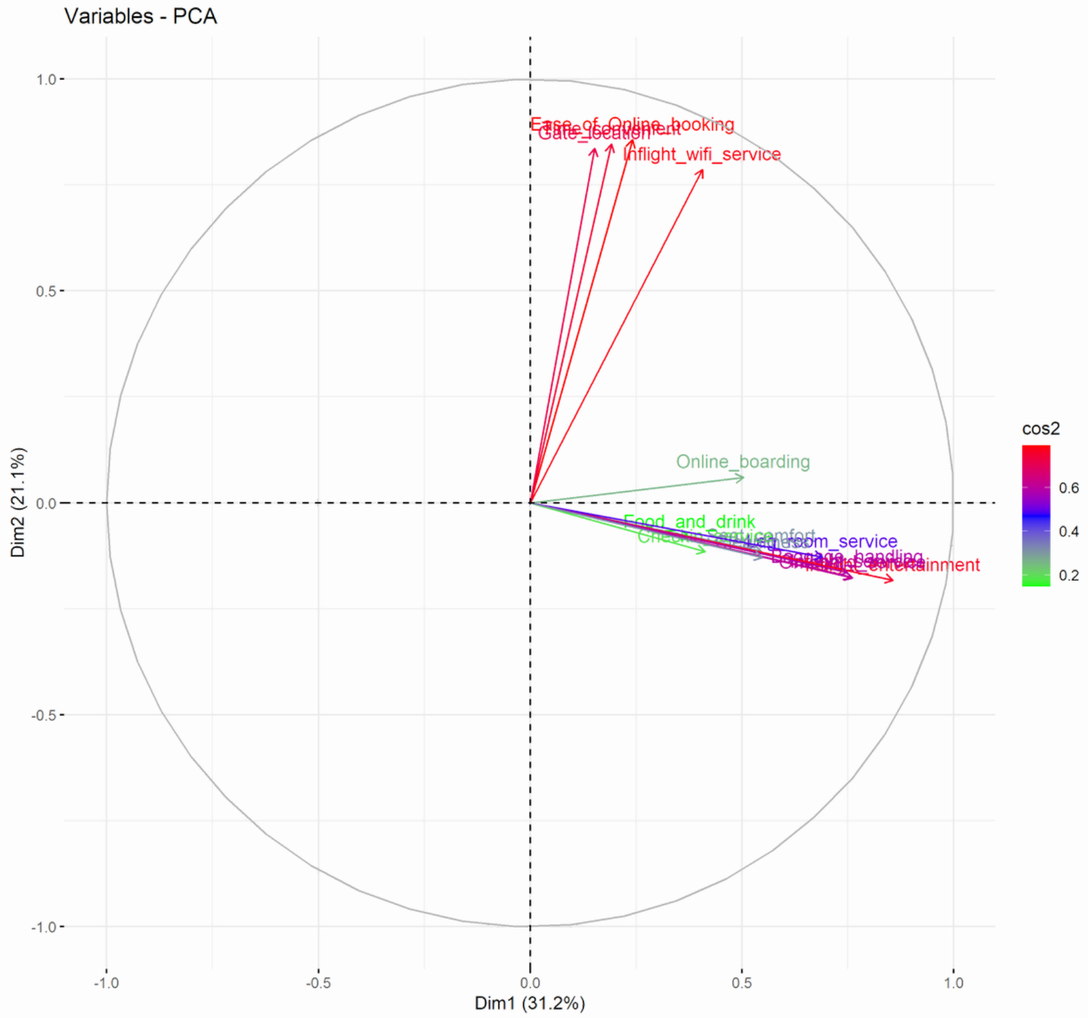


ECO

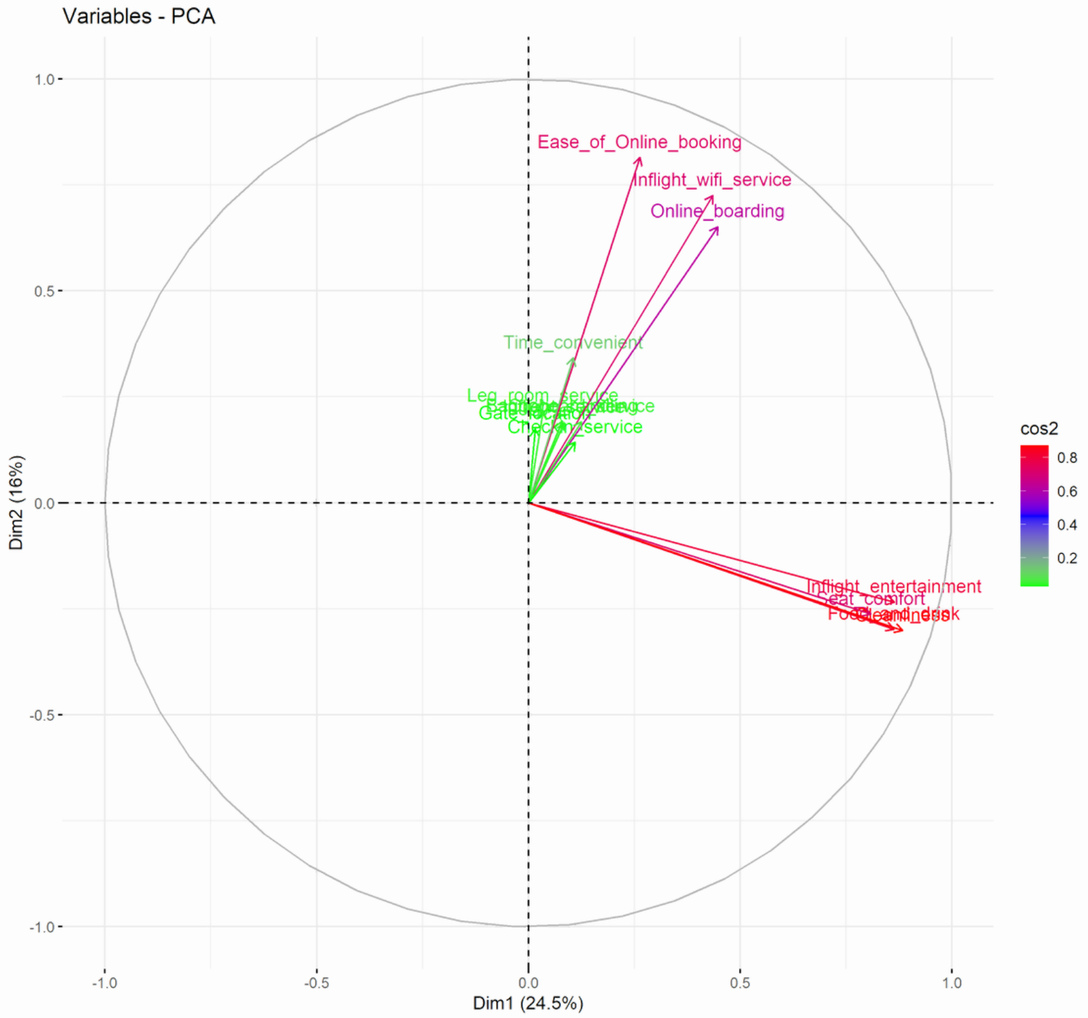


PROJECTION OF VARIABLES

BUSINESS

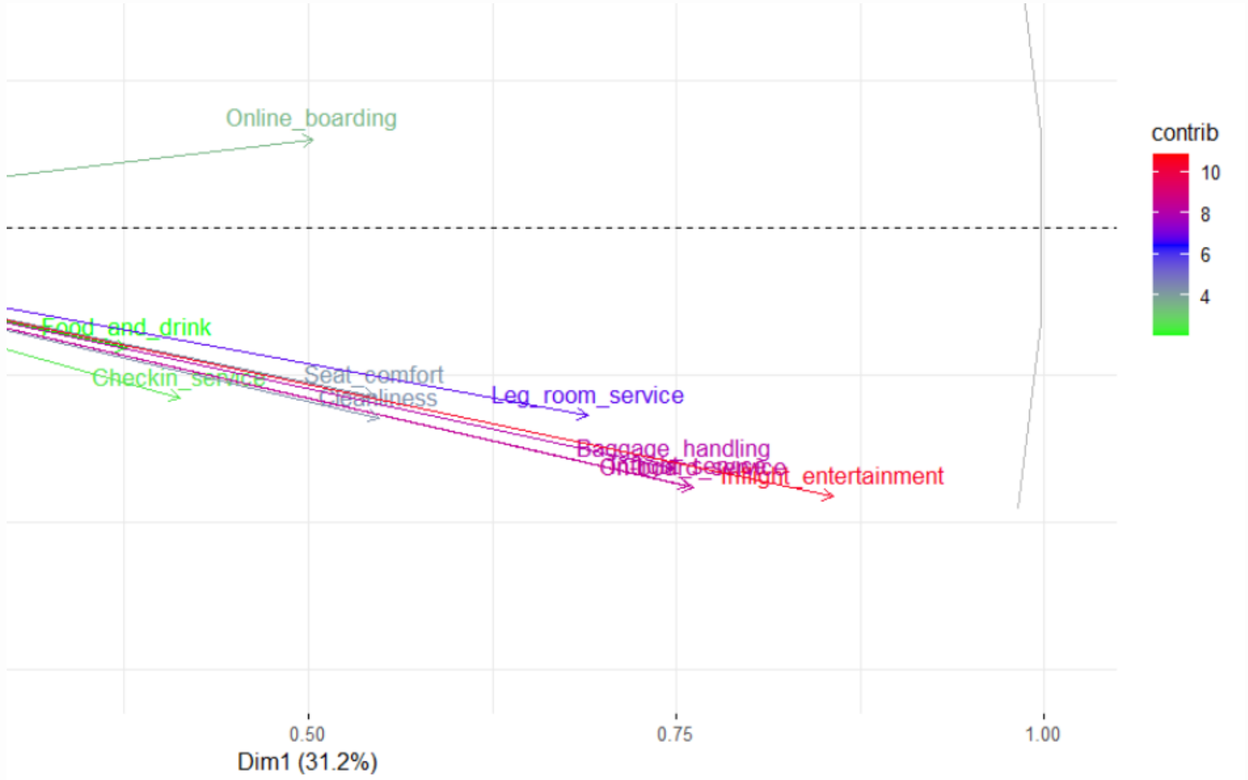


ECO

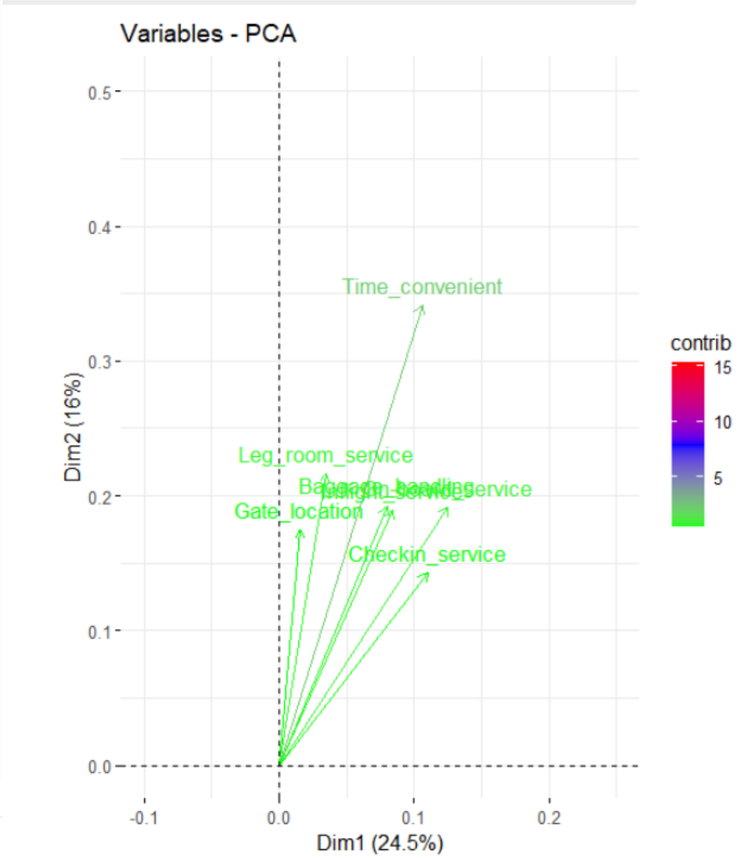


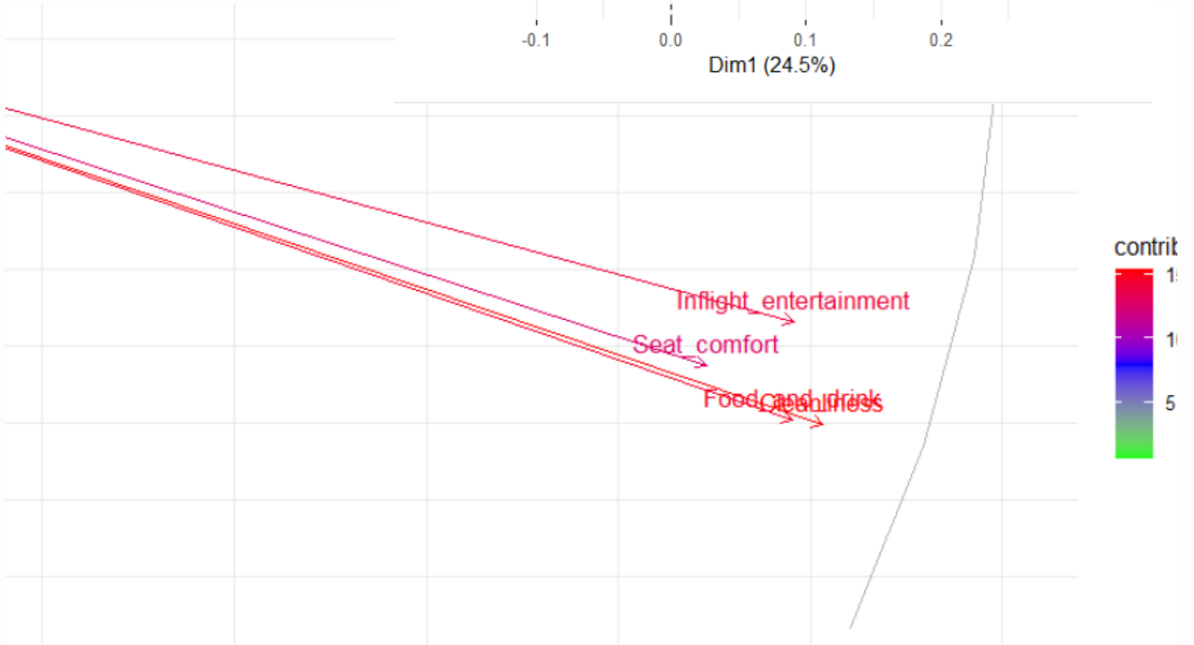
PROJECTION OF VARIABLES

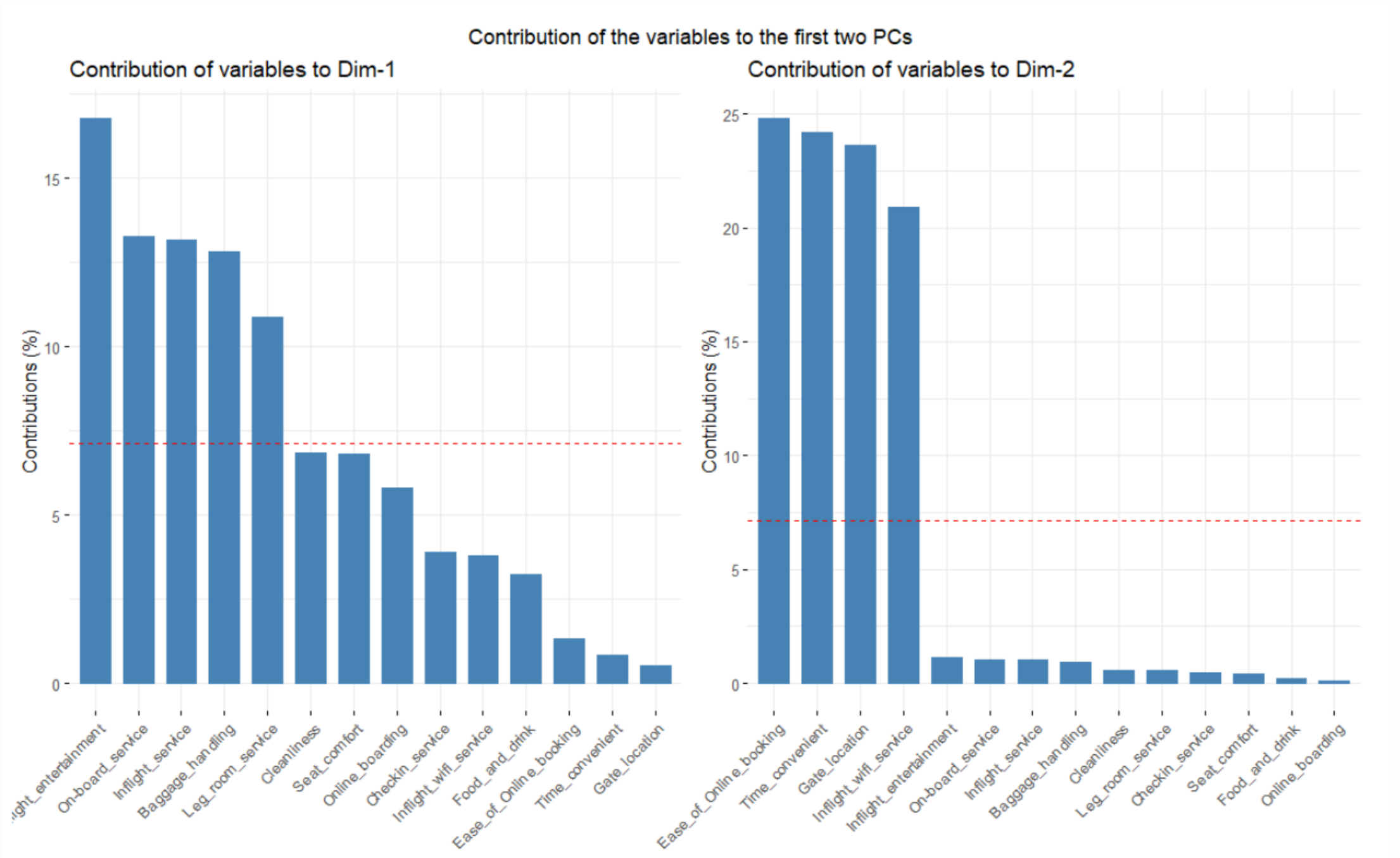
BUSINESS

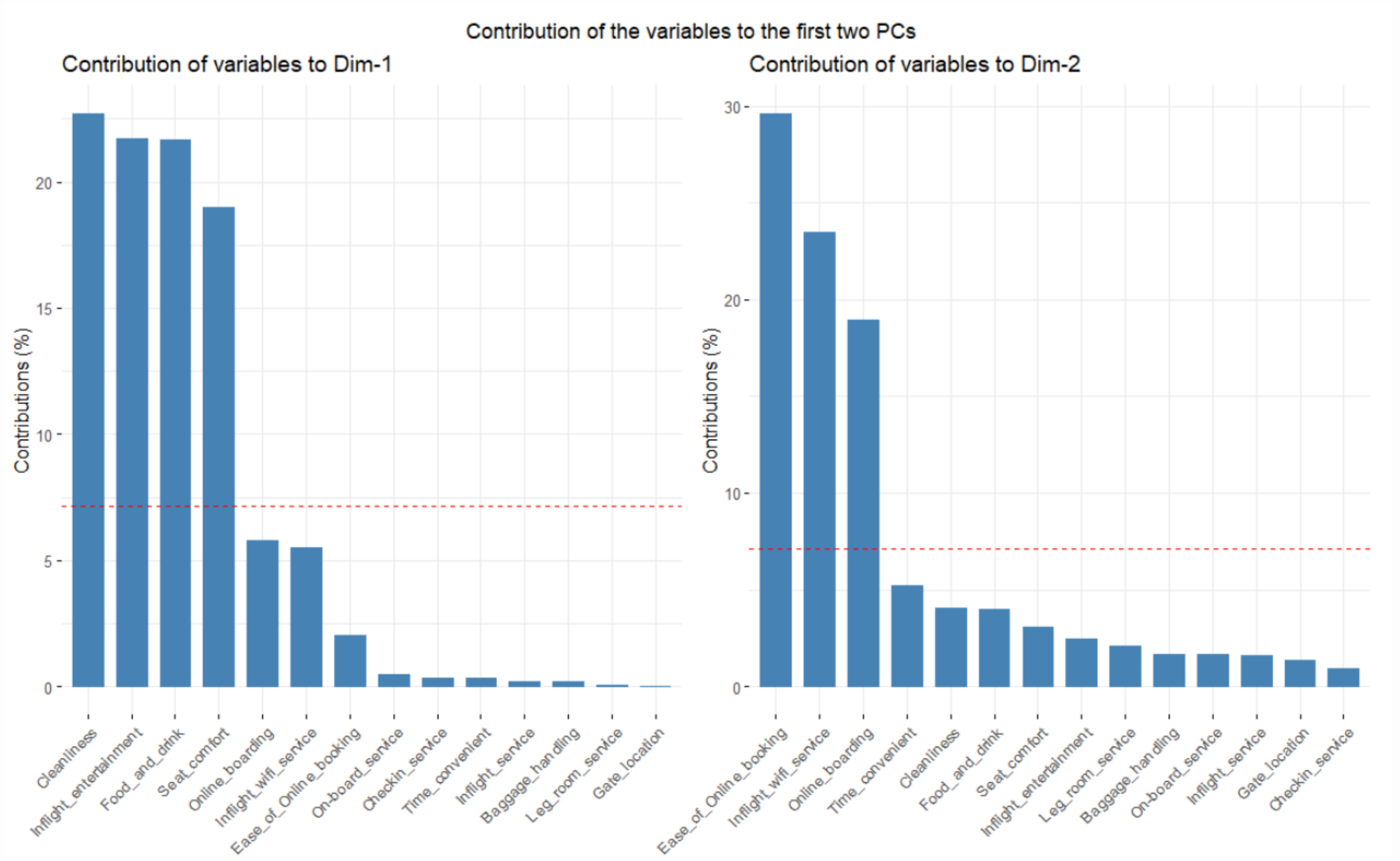


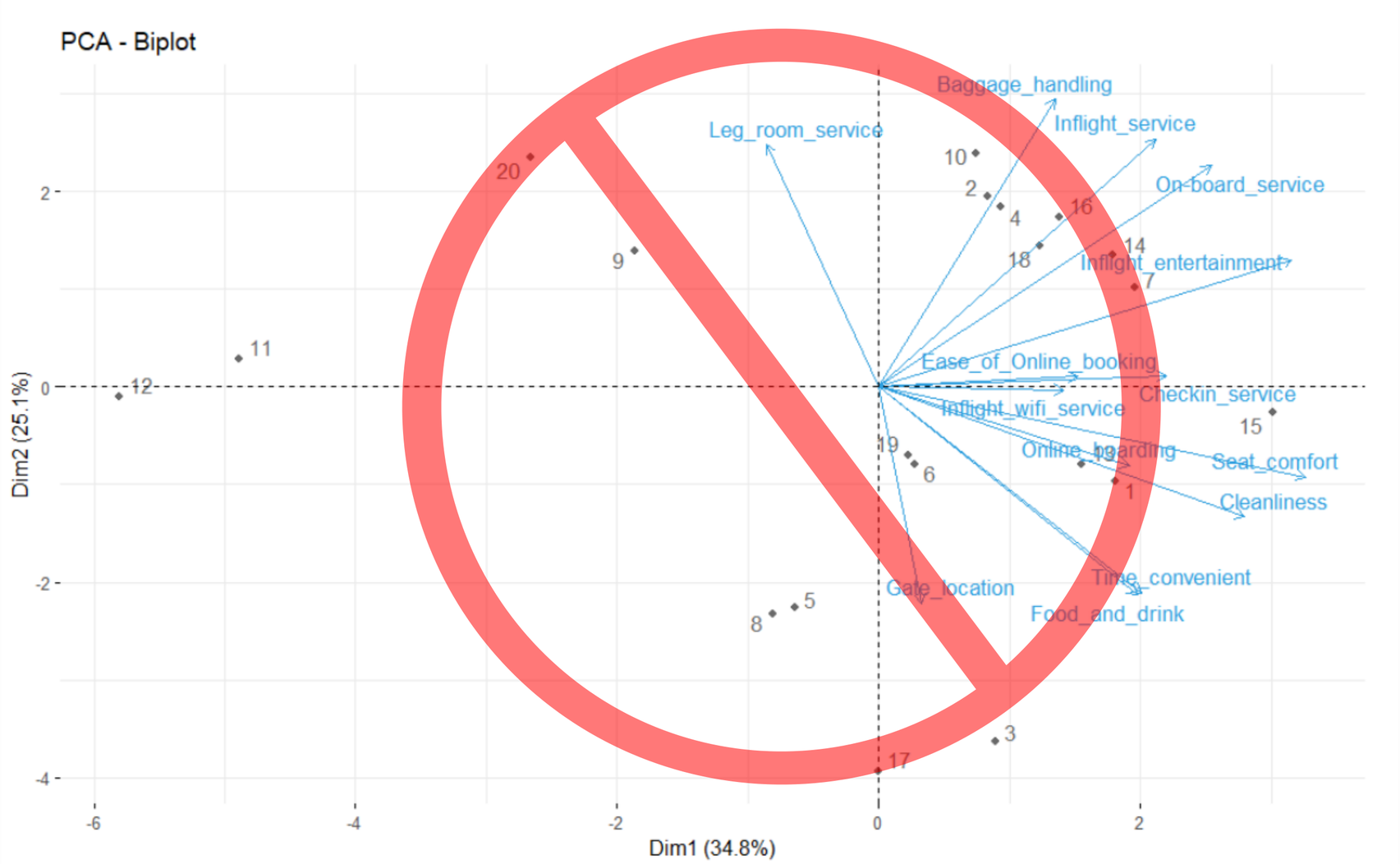
ECO






Contribution of the variables to the first two PCs
Contribution of variables to Dim-1
Contributions (%)
0
5
10
15
ght_entertainment
On-board_service
Inflight_service
Baggage_handling
Leg_room_service
Cleanliness
Seat_comfort
Online_boarding
Checkin_service
Inflight_wifi_service
Food_and_drink
Ease_of_Online_booking
Time_convenient
Gate_location
Contribution of variables to Dim-2
Contributions (%)
0
5
10
15
20
25
Ease_of_Online_booking
Time_convenient
Gate_location
Inflight_wifi_service
Inflight_entertainment
On-board_service
Inflight_service
Baggage_handling
Cleanliness
Leg_room_service
Checkin_service
Seat_comfort
Food_and_drink
Online_boarding


Contribution of the variables to the first two PCs
Contribution of variables to Dim-1
Contributions (%)
0
5
10
15
20
Cleanliness
Inflight_entertainment
Food_and_drink
Seat_comfort
Online_boarding
Inflight_wifi_service
Ease_of_Online_booking
On-board_service
Checkin_service
Time_convenient
Inflight_service
Baggage_handling
Leg_room_service
Gate_location
Contribution of variables to Dim-2
Contributions (%)
0
10
20
30
Ease_of_Online_booking
Inflight_wifi_service
Online_boarding
Time_convenient
Cleanliness
Food_and_drink
Seat_comfort
Inflight_entertainment
Leg_room_service
Baggage_handling
On-board_service
Inflight_service
Gate_location
Checkin_service


PCA - Biplot
Baggage_handling
Inflight_service
Leg_room_service
On-board_service
Inflight_entertainment
Ease_of_Online_booking
Checkin_service
Inflight_wifi_service
Online_boarding
Seat_comfort
Cleanliness
Time_convenient
Gate_location
Food_and_drink
20
10
2
4
16
14
18
9
7
11
12
19
6
15
1
8
5
3
17
Dim2 (25.1%)
Dim1 (34.8%)
2
0
-2
-4
-6
-4
-2
0
2

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

«Business Intelligence in Airline Passenger Satisfaction Study — A Fuzzy-Genetic Approach with Optimized Interpretability-Accuracy Trade-Off» - Marian B. Gorzałczany, Filip Rudzi´nsk, and Jakub Piekoszewski, Department of Electrical and Computer Engineering, Kielce University of Technology, Poland, 2021

2. «Investigating airline passenger satisfaction: Data mining method» - Tri Noviantoro, Jen-Peng Huang, College of Business, Southern Taiwan University of Science and Technology, Taiwan, 2022.

3. «Feature Analysis on Airline Passenger Satisfaction using Orange Tool» - Hannah Susan Mathew,
Department of Computer Science, Rajagiri College of Social Sciences, Kochi, India, 2022.