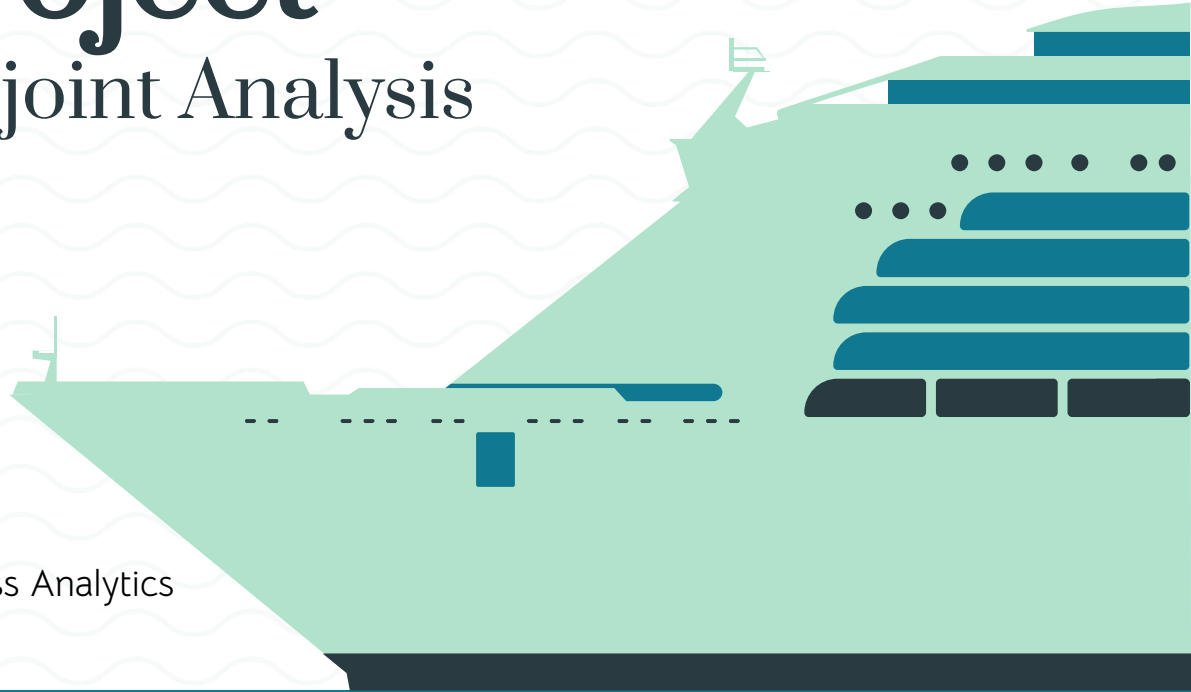


Cruise Project

Traditional Conjoint Analysis

Annamaria Culpo
Milena Atanasova

Laboratory of Customer and Business Analytics
a.y. 2022/2023



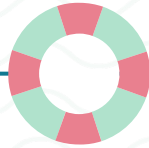
Analysis Phases

Step 01



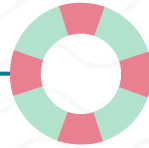
Exploratory
Data Analysis

Step 02



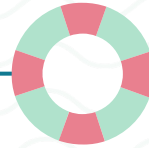
Random Intercept
Model

Step 03



Random Slope
Model

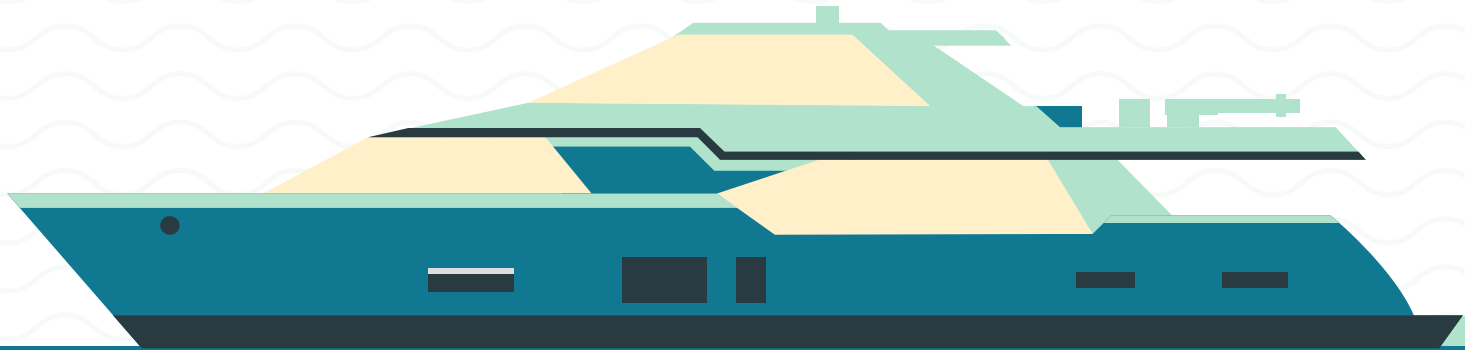
Step 04



Conclusions

01

Exploratory Analysis



Dataset

A cruise company (e.g. Costa) would like to add a new cruise line proposal for the summer season of 2023. It is considering making a customer survey and conducting a traditional conjoint analysis of potential service profiles

1

Brand

Costa
MSC
Royal Caribbean

2

Destination

Egypt
Greece
Sicily
Spain

3

Price

€ 300
€ 700
€ 1200

4

Duration

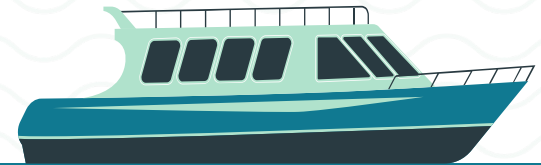
7 days
14 days
21 days

5

Amenities

Animation
Spa
Sport

300 respondents
3 different age groups (young, adult, old)
20 product profiles per respondent



Linear Regression Model

At first, we fit a classical linear regression model for the cruise ratings to later compare it with the multilevel linear model

$$y_i = \beta_0 + \beta X_i + \varepsilon_i$$

- All coefficients (apart from Brand Royal Caribbean) different from zero
- Customers on average prefer longer voyages (Duration 14 and 21 are positive)
- The highest price (€ 1200) is preferred to the lowest (€ 300), which is not the case for the medium price (€ 700)

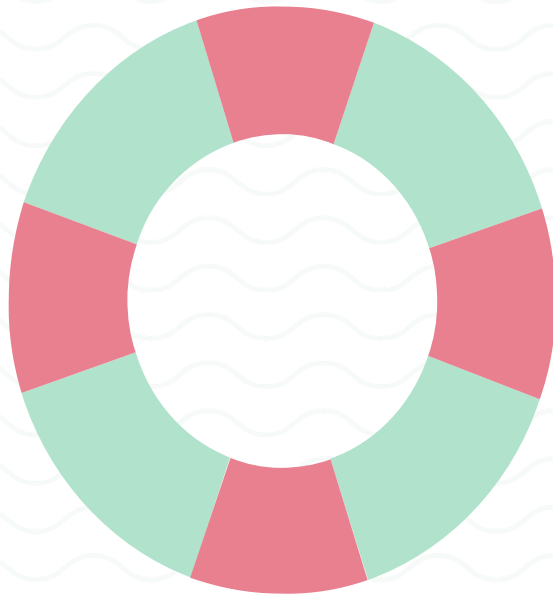
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.31299	0.10246	12.814	< 2e-16	***
BrandMSC	-2.61420	0.05981	-43.712	< 2e-16	***
BrandRoyalCaribbean	0.05220	0.09088	0.574	0.566	
DestinationEgypt	1.93307	0.10187	18.976	< 2e-16	***
DestinationGreece	0.66281	0.08283	8.002	1.46e-15	***
DestinationSpain	1.60500	0.07630	21.036	< 2e-16	***
Price700	-0.85033	0.06328	-13.438	< 2e-16	***
Price1200	1.57241	0.06304	24.943	< 2e-16	***
Duration14	2.56912	0.06793	37.822	< 2e-16	***
Duration21	2.95860	0.07773	38.064	< 2e-16	***
AmenitiesSpa	3.00193	0.08171	36.737	< 2e-16	***
AmenitiesSport	1.25490	0.08781	14.292	< 2e-16	***



02

Random Intercept Model



Estimates of the coefficients



Fitting a random intercept model, we allow the intercept to vary according to each respondent (j) and we can trust s.d.

$$y_{ij} = \beta_{0j} + \beta X_{ij} + \varepsilon_{ij}$$

The estimates of the coefficients are the same of the linear regression model, but the one of the intercept represents the **average** of the distribution of the **intercepts** that vary across the respondents.

Fixed effects:

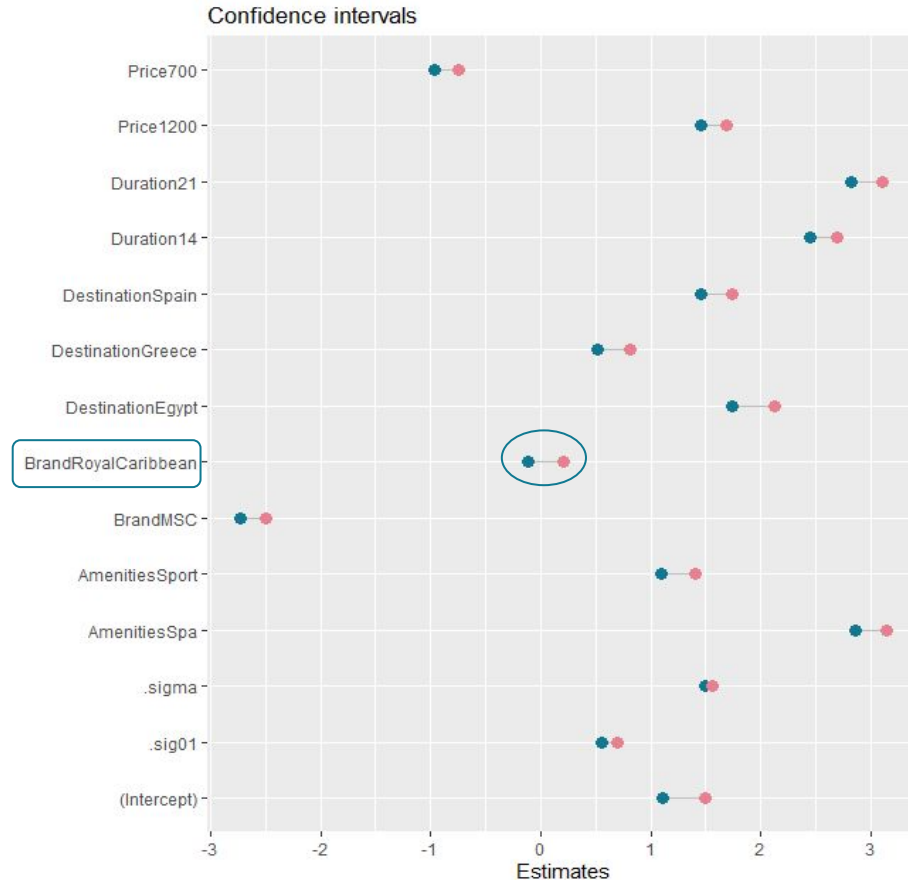
	Estimate	Std. Error	t value
(Intercept)	1.31299	0.10149	12.937
BrandMSC	-2.61420	0.05539	-47.195
BrandRoyalCaribbean	0.05220	0.08417	0.620
DestinationEgypt	1.93307	0.09435	20.488
DestinationGreece	0.66281	0.07672	8.639
DestinationSpain	1.60500	0.07067	22.712
Price700	-0.85033	0.05861	-14.509
Price1200	1.57241	0.05839	26.931
Duration14	2.56912	0.06291	40.836
Duration21	2.95860	0.07199	41.098
AmenitiesSpa	3.00193	0.07568	39.665
AmenitiesSport	1.25490	0.08132	15.431

A **restricted model** without the Destination attribute has been tested, but it has a lower goodness of fit

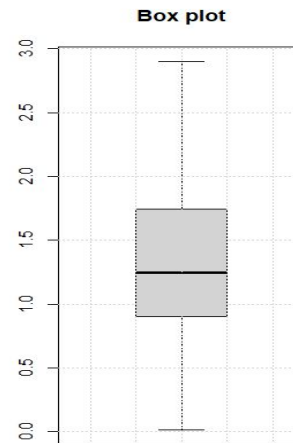
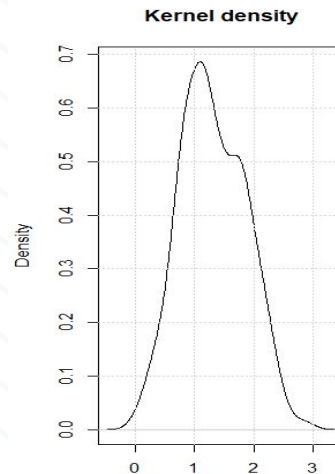
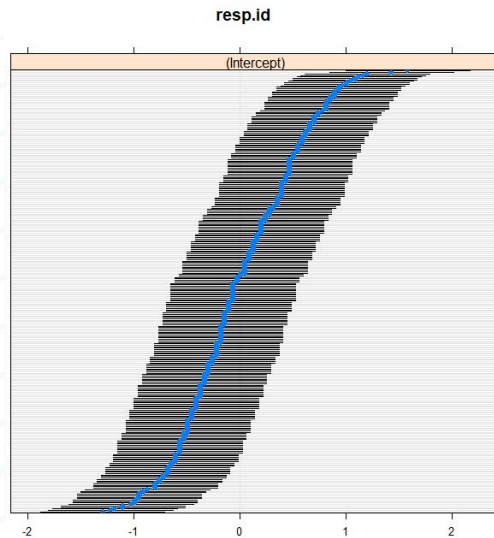
```
Restrm1mRI: rating ~ Brand + Price + Duration + Amenities + (1 | resp.id)
m1mRI: rating ~ Brand + Destination + Price + Duration + Amenities + (1 | resp.id)
npar  AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
Restrm1mRI  11 23384 23458 -11681  23362
m1mRI      14 22581 22675 -11276  22553 809.4  3 < 2.2e-16 ***
```


Confidence Intervals

- The only confidence interval which includes **zero** is the one related to the brand **Royal Caribbean**
- **Small intervals** → Estimates quite precise
- All the estimates are positive, except for the ones of the Brand MSC and the Price of 700 euros
- **Costa** can appreciate the **strength** of its brand with respect to MSC



Random Effects



- Random effects of the intercept with their confidence intervals
- They vary from -1 to 1 → there is a non negligible between-group variation

- The highest frequency is in correspondence of 1.2 - 1.3 (estimate of the average intercept)
- The median is similar to the mean
- There is a second bump with a lower frequency than the first

Model Performance



Intraclass
Correlation
Coefficient

0.142

The across-groups variability (heterogeneity between groups) explains 14% of the overall variability of the model



The grouping structure improves the estimates

Ad-hoc R
Squared
0.612

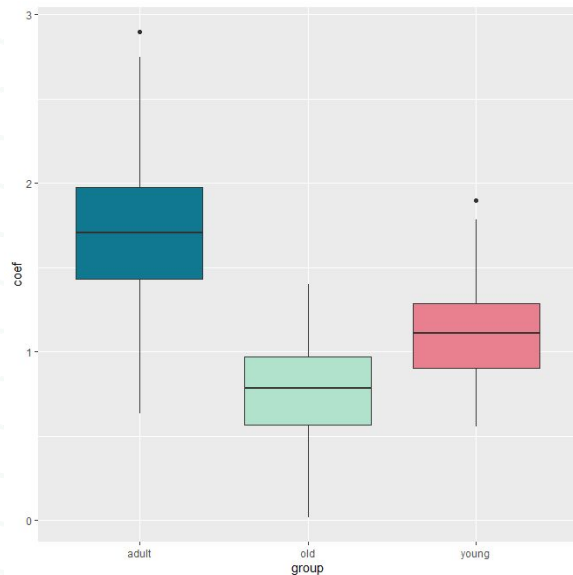


The model explains about 61% of the variance of the ratings above that accounted for in the null model

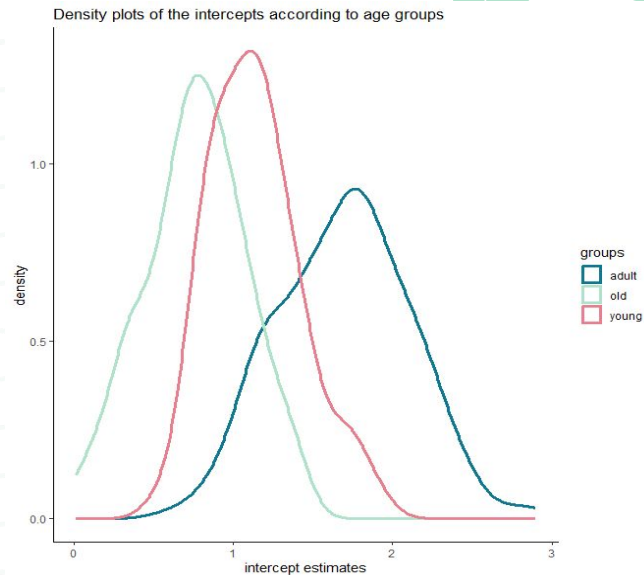


Satisfying goodness-of-fit

Intercepts for different age groups



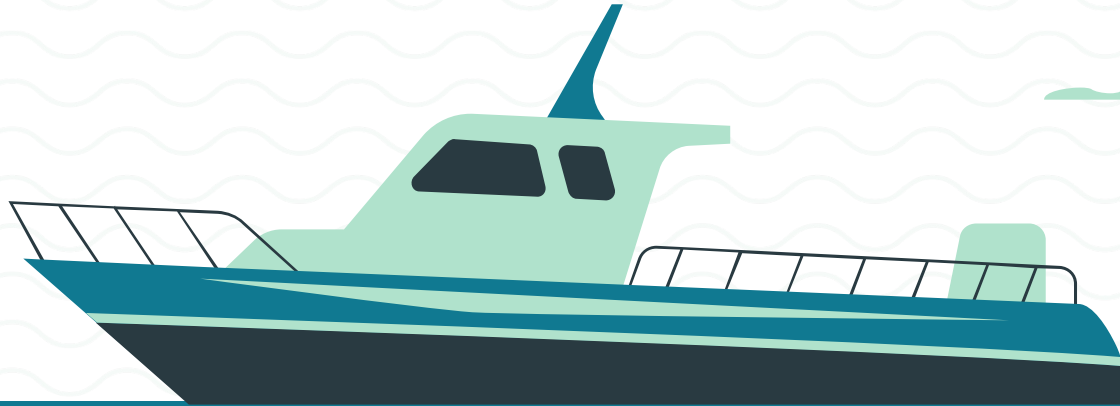
Adults (blue) assign the **highest ratings** to the cruise experience, followed by young respondents (red) and, at last, the olders (green)



The mean and the variance of the density distributions of the intercepts are different. In particular, **adults** present the **greatest heterogeneity**

03 - Random Slope Model

Random Effect for Each Model Parameter



Model Summary

$$y_{ij} = \beta_{0j} + \beta_j X_{ij} + \varepsilon_{ij}$$

Linear mixed model fit by REML ['lmerMod']

Formula: rating ~ Brand + Destination + Price + Duration + Amenities +
(Brand + Destination + Price + Duration + Amenities | resp.id)

Data: cruise

Fixed effects:

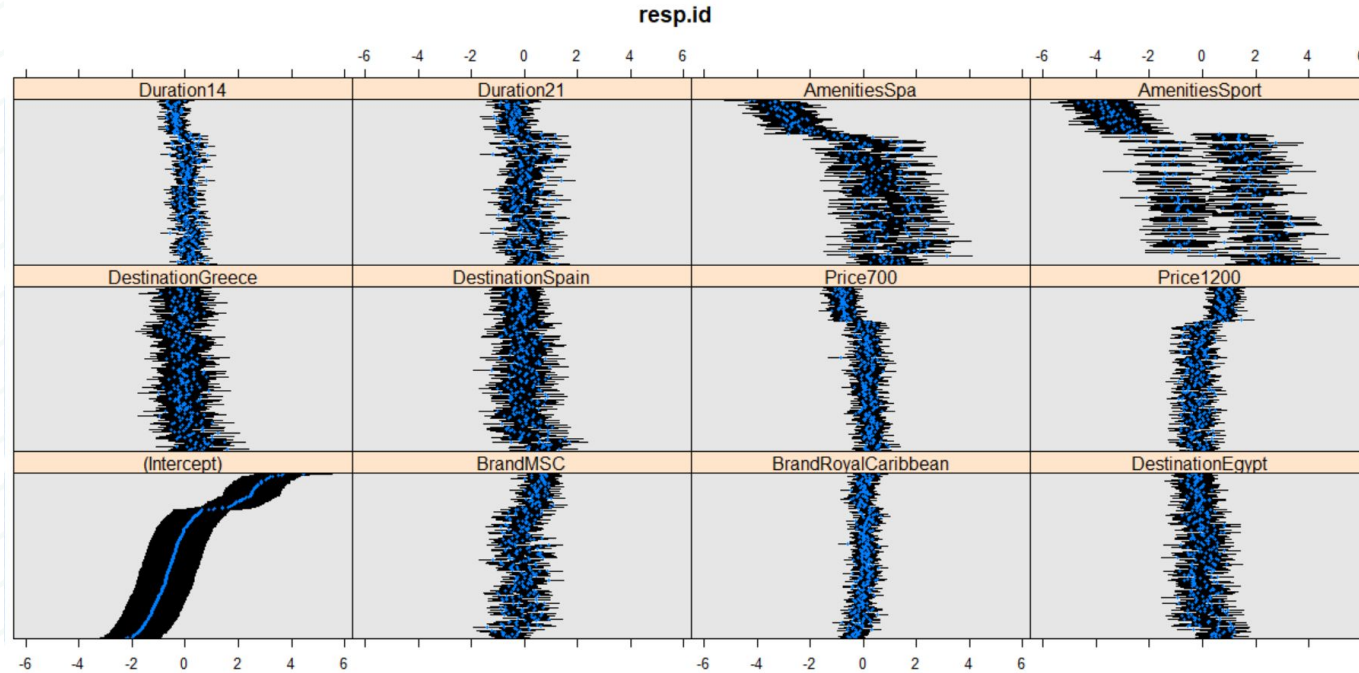
	Estimate	Std. Error	t value
(Intercept)	1.31299	0.10047	13.068
BrandMSC	-2.61420	0.04262	-61.333
BrandRoyalCaribbean	0.05220	0.04344	1.202
DestinationEgypt	1.93307	0.05440	35.532
DestinationGreece	0.66281	0.04968	13.342
DestinationSpain	1.60500	0.04816	33.326
Price700	-0.85033	0.04018	-21.164
Price1200	1.57241	0.04296	36.600
Duration14	2.56912	0.03657	70.257
Duration21	2.95860	0.04775	61.958
AmenitiesSpa	3.00193	0.10657	28.168
AmenitiesSport	1.25490	0.13787	9.102

Random effects:

Groups	Name	Variance	Std.Dev.
resp.id	(Intercept)	2.4254	1.5574
	BrandMSC	0.3396	0.5827
	BrandRoyalCaribbean	0.0917	0.3028
	DestinationEgypt	0.2918	0.5402
	DestinationGreece	0.3463	0.5884
	DestinationSpain	0.3614	0.6012
	Price700	0.2543	0.5043
	Price1200	0.3254	0.5705
	Duration14	0.1361	0.3690
	Duration21	0.3371	0.5806
	AmenitiesSpa	3.0238	1.7389
	AmenitiesSport	5.2593	2.2933
Residual		0.5221	0.7226
Number of obs: 6000, groups: resp.id, 300			

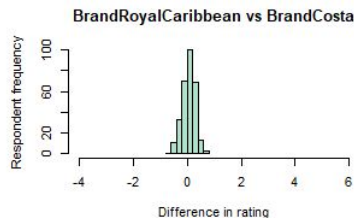
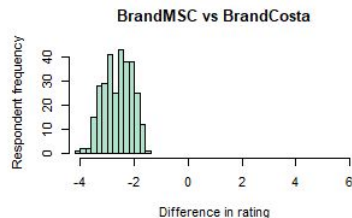
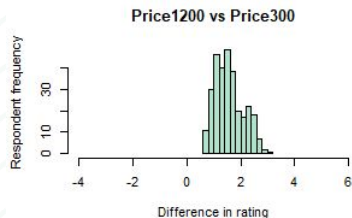
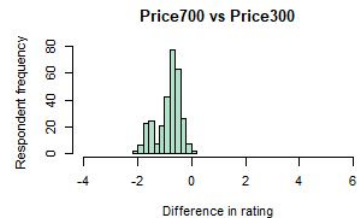
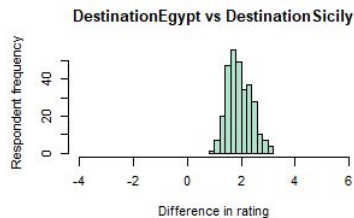
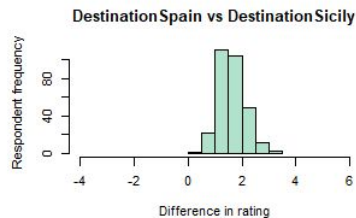
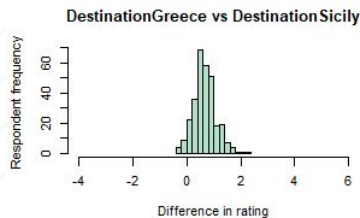
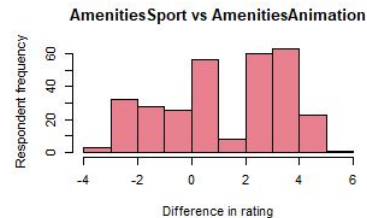
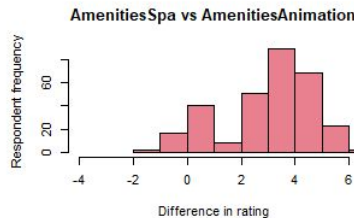
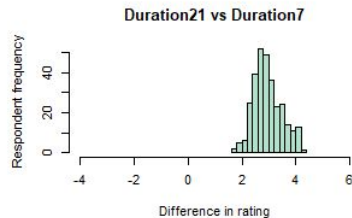
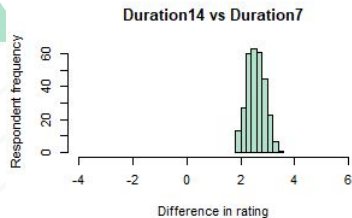
Estimates of the parameters here are the mean values of the distributions of the variables across respondents

Variability of Parameters



Posterior estimates of random effect for each respondent

Histograms for Coefficient Estimates



Comparison with Simpler Models

ANOVA test. Are simpler models reducing data fitting?

Destination and
Duration having only
fixed effects

```
Data: cruise
Models:
m1mRS_dur_des: rating ~ Brand + Destination + Price + Duration + Amenities + (Brand + Price + Amenities | resp.id)
m1mRS: rating ~ Brand + Destination + Price + Duration + Amenities + (Brand + Destination + Price + Duration + Amenities | resp.id)
      npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
m1mRS_dur_des   41 17472 17747 -8695.1    17390
m1mRS           91 17045 17654 -8431.3    16863 527.55 50 < 2.2e-16 ***
```

Duration having only
fixed effects

```
Data: cruise
Models:
m1mRS_dur: rating ~ Brand + Destination + Price + Duration + Amenities + (Brand + Price + Amenities + Destination | resp.id)
m1mRS: rating ~ Brand + Destination + Price + Duration + Amenities + (Brand + Destination + Price + Duration + Amenities | resp.id)
      npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
m1mRS_dur    68 17238 17694 -8551.2    17102
m1mRS        91 17045 17654 -8431.3    16863 239.7 23 < 2.2e-16 ***
```

Destination having
only fixed effects

```
Data: cruise
Models:
m1mRS_des: rating ~ Brand + Destination + Price + Duration + Amenities + (Brand + Price + Amenities + Duration | resp.id)
m1mRS: rating ~ Brand + Destination + Price + Duration + Amenities + (Brand + Destination + Price + Duration + Amenities | resp.id)
      npar    AIC    BIC logLik deviance Chisq Df Pr(>Chisq)
m1mRS_des    58 17261 17650 -8572.5    17145
m1mRS        91 17045 17654 -8431.3    16863 282.31 33 < 2.2e-16 ***
```

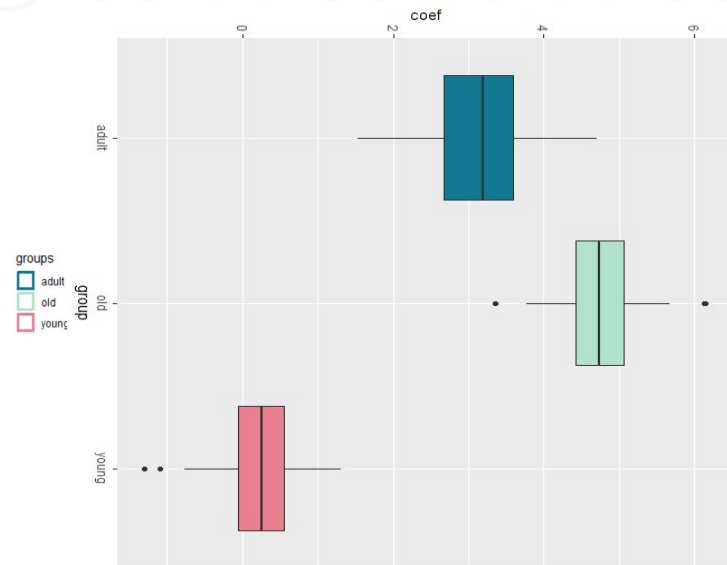
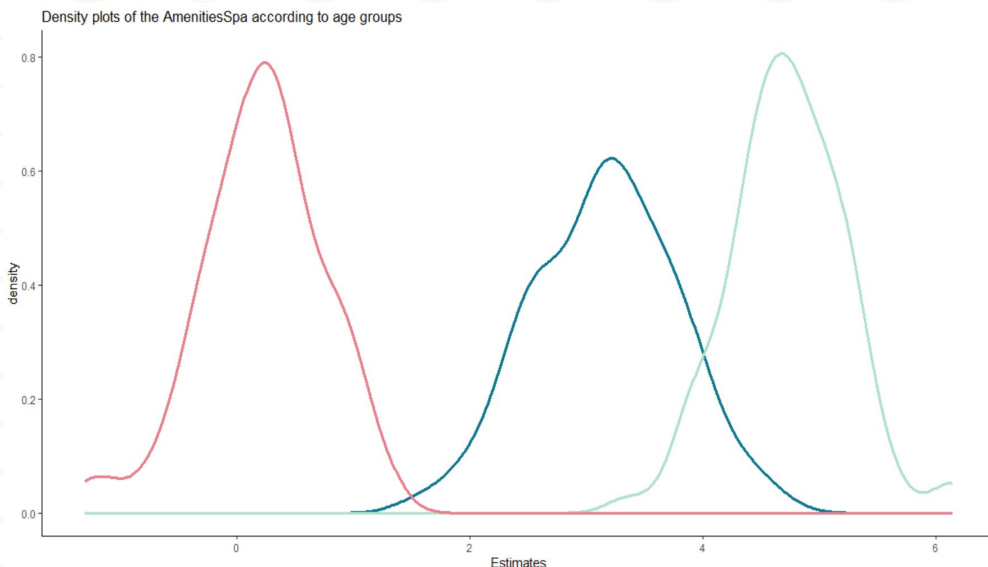


Keep the full model. Reject the null hypothesis of equivalence between the restricted and the full models

Respondent-level Attribute Age

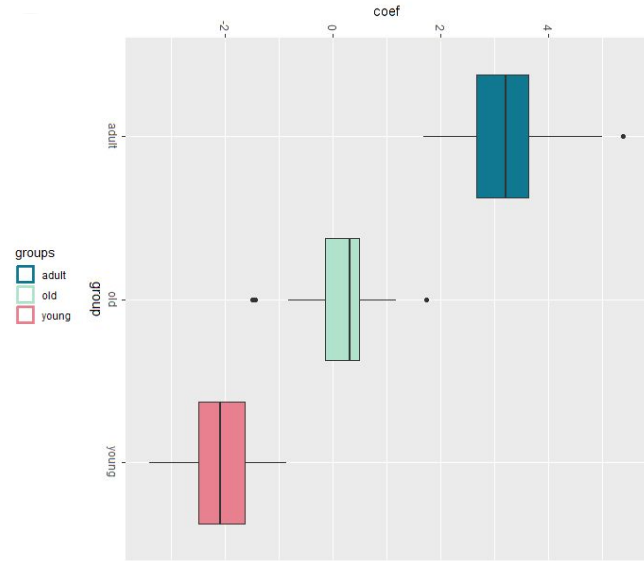
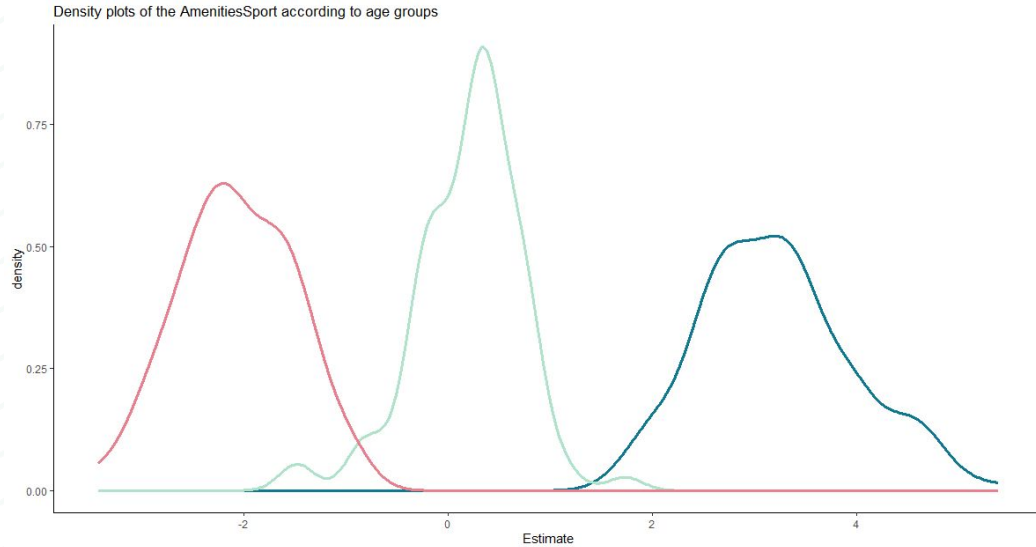
Examine such relationship for the coefficients for which we saw have proven to have an effect on the group-level structure, having variation on a respondent-level.

AmenitiesSpa. Great difference between young and old. Only young consumers exhibit some coefficients lower than zero, which means there is a higher and not-uniform variability in their preferences. There is differentiation in groups' variance, the one of adult being higher, further explained by the fact that the adult group is much more represented in the sample of respondents.



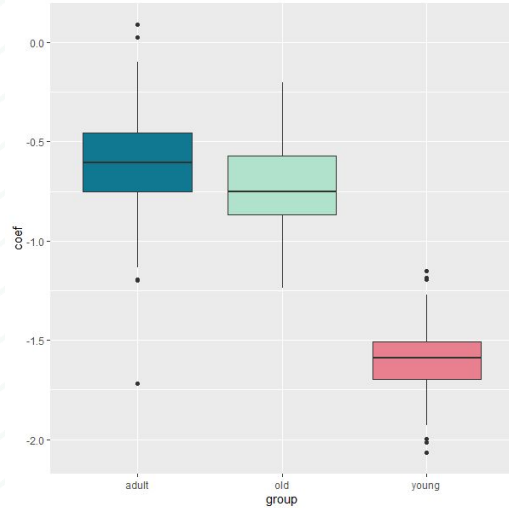
Respondent-level Attribute Age

AmenitiesSport. For the group of young respondents, all the coefficients are below zero which means that such kind of customers actually prefer animation than sport facilities on their ship. Adults value sport a lot and almost all adults have a coefficient for this variable larger than the coefficients for old and young consumers. For a firm to gain more customers therefore, would better consider adding a cruise line where sport amenities are included in order to gather more middle-aged customers.



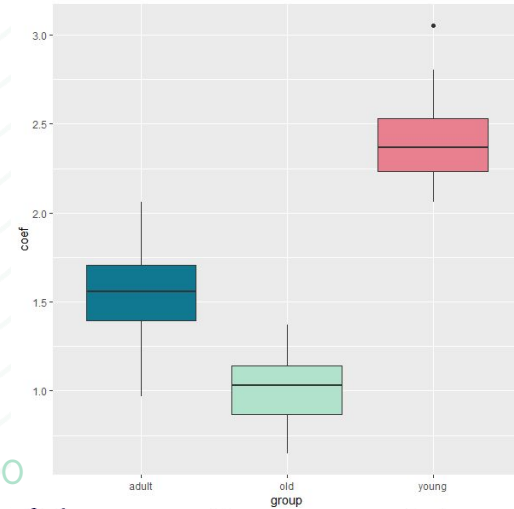
Respondent-level Attribute Age

Price. In case of price of 700, almost all the respondent-level coefficients estimated are negative. On the contrary, they are all positive if we consider price of 1200 (w.r.t. the base 300). Further, we can notice that adults and seniors exhibit quite similar preferences, whereas youngsters' distributions are much more polarized.



Price700

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.06550	-1.03615	-0.74408	-0.85033	-0.54189	0.08987



Price1200

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.648	1.143	1.513	1.572	1.899	3.053



04 - Conclusions



Insights from Analysis



1

Grouping structure

can explain a non negligible share of the variability of preferences

2

Brand MSC much less preferred, Royal Caribbean seemed as similar to Costa

so if we assume the point of view of Costa, we can be quite confident about the strength of our brand, without the necessity to invest in its reinforcement, but just in its maintenance

3

Duration

has a strong effect in shaping preferences, and usually the longer periods are preferred

4

The destination

does not seem to play a role, and neither whether it is a local cruise or to a more exotic place



Insights from Analysis

5

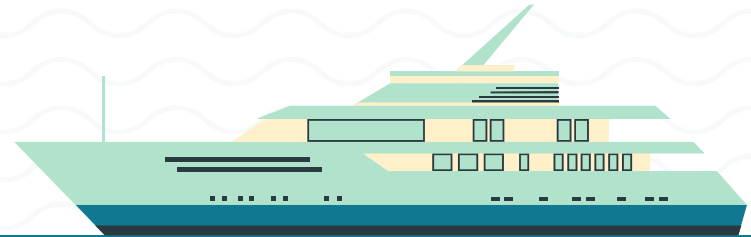
The amenities play a crucial role

and are highly variable depending on respondent, and also on age groups, so this issue could be exploited in the decision-making process of the firm

6

Price plays a particular role

since on average 300 is preferred to 700, but 1200 is preferred to 300. The lowest estimates for the level Price700 are registered for young customers, who are also the age group with the highest estimates for Price1200



Potential Cruise Line Proposals

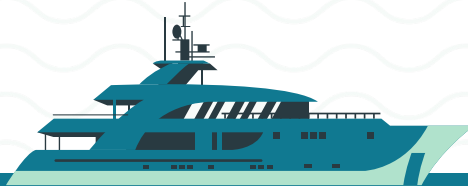
Line 1

Target customers: Young
Price: High (1200)
Amenities: Luxury Animation
Destination: Decide
Duration: 21 days



Line 2

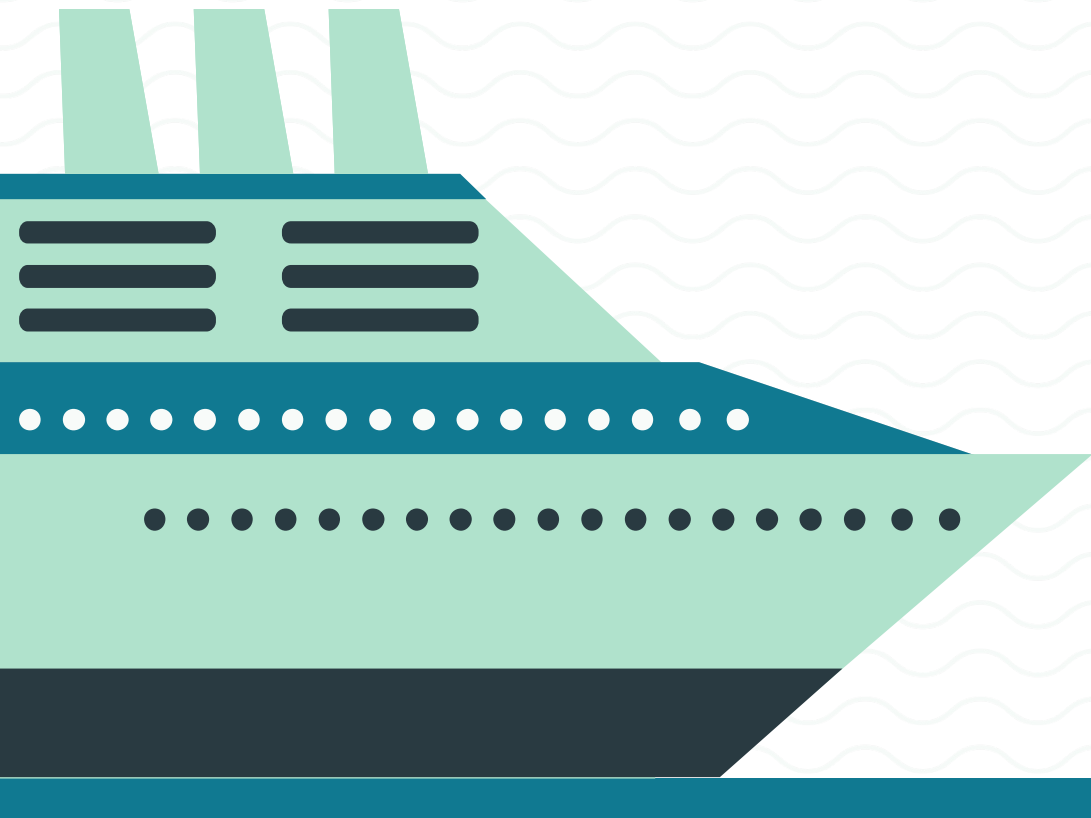
Target customers: Adults
Price: High (1200)
Amenities: Sport
Destination: Decide
Duration: 21 days



Line 3

Target customers: Adults and older
Price: Low (300)
Amenities: Spa
Destination: Decide
Duration: 14 days





Thanks!

Do you have any questions?