What are the popular choices of booking rates for different segments of customers?

To perform all tasks, I used Keboola and Snowflake workspace to transform data, and Power BI for data visualization.

I also transform data by using 2 slightly different methods. 1st method is the simple measurement of all criteria such as age, gender and nationality separately.

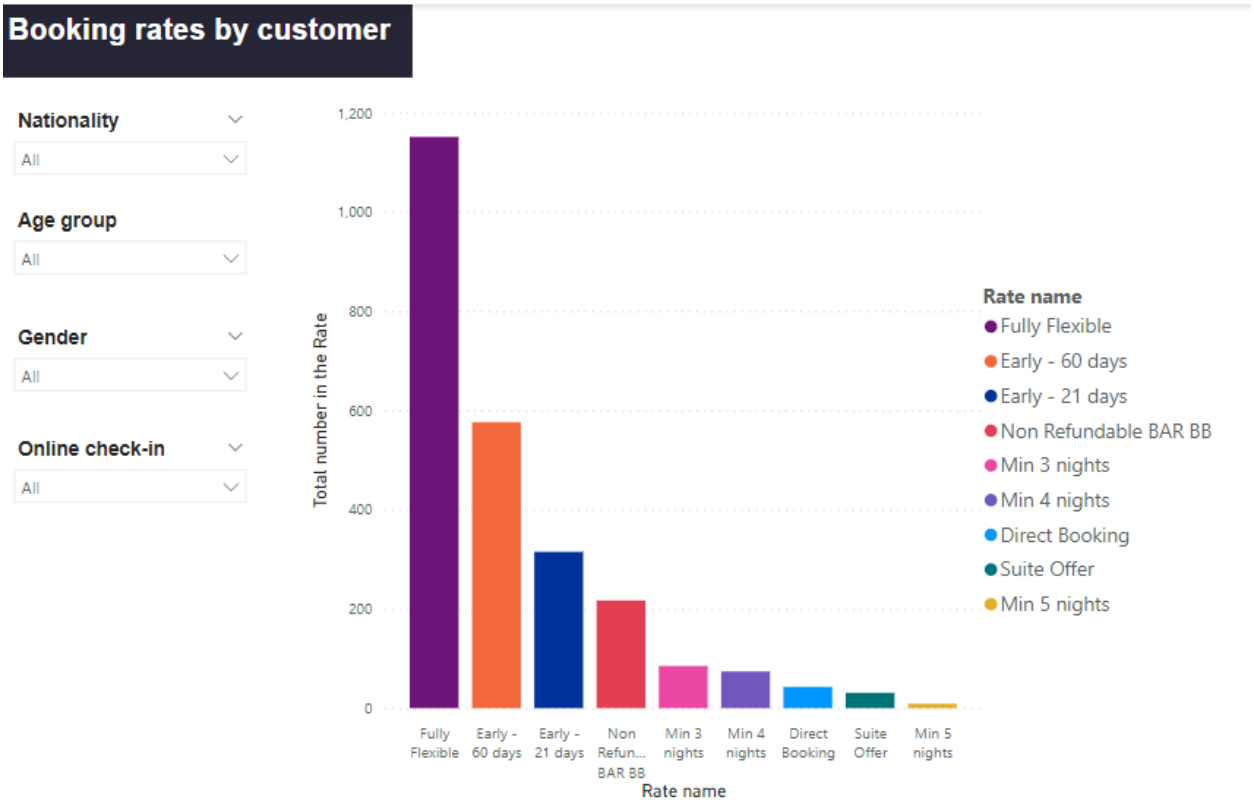
```
-- Join tables + days of week

create or replace table "Reservations with rates" as
select R1.*, R2."RateName", R2."SettlementAction", R2."SettlementTrigger", R2."SettlementValue", CASE
DAYOFWEEK(TO_TIMESTAMP("CreatedUtc"))
   WHEN 0 THEN 'Sunday'
   WHEN 1 THEN 'Monday'
   WHEN 2 THEN 'Tuesday'
   WHEN 3 THEN 'Wednesday'
   WHEN 4 THEN 'Thursday'
   WHEN 5 THEN 'Friday'
   WHEN 6 THEN 'Saturday'
end as "DayOfWeek"
from "Reservations" as R1
left join "Rates" as R2 on R1."RateId" = R2."RateId";

select * from "Reservations with rates";

-- task 1

SELECT t1."RateName", t1."AgeGroup", t2."Gender", t3."NationalityCode"
FROM
(SELECT "RateName", "AgeGroup"
FROM (
   SELECT "RateName", "AgeGroup",
        ROW_NUMBER() OVER (PARTITION BY "RateName" ORDER BY COUNT("AgeGroup") DESC) as rank
   FROM "Reservations with rates"
   where "AgeGroup" != 0
   GROUP BY "RateName", "AgeGroup"
)
WHERE rank = 1) AS t1
JOIN
(SELECT "RateName", "Gender"
FROM (
   SELECT "RateName", "Gender",
        ROW_NUMBER() OVER (PARTITION BY "RateName" ORDER BY COUNT("Gender") DESC) as rank
   FROM "Reservations with rates"
   where "Gender" != 0
   GROUP BY "RateName", "Gender"
)
WHERE rank = 1) AS t2 on t1."RateName" = t2."RateName"
JOIN
(SELECT "RateName", "NationalityCode"
FROM (
   SELECT "RateName", "NationalityCode",
        ROW_NUMBER() OVER (PARTITION BY "RateName" ORDER BY COUNT("NationalityCode") DESC)
as rank
   FROM "Reservations with rates"
   where "NationalityCode" != 'NULL'
   GROUP BY "RateName", "NationalityCode"
)
WHERE rank = 1) AS t3 on t2."RateName" = t3."RateName";
```

Result for the first method

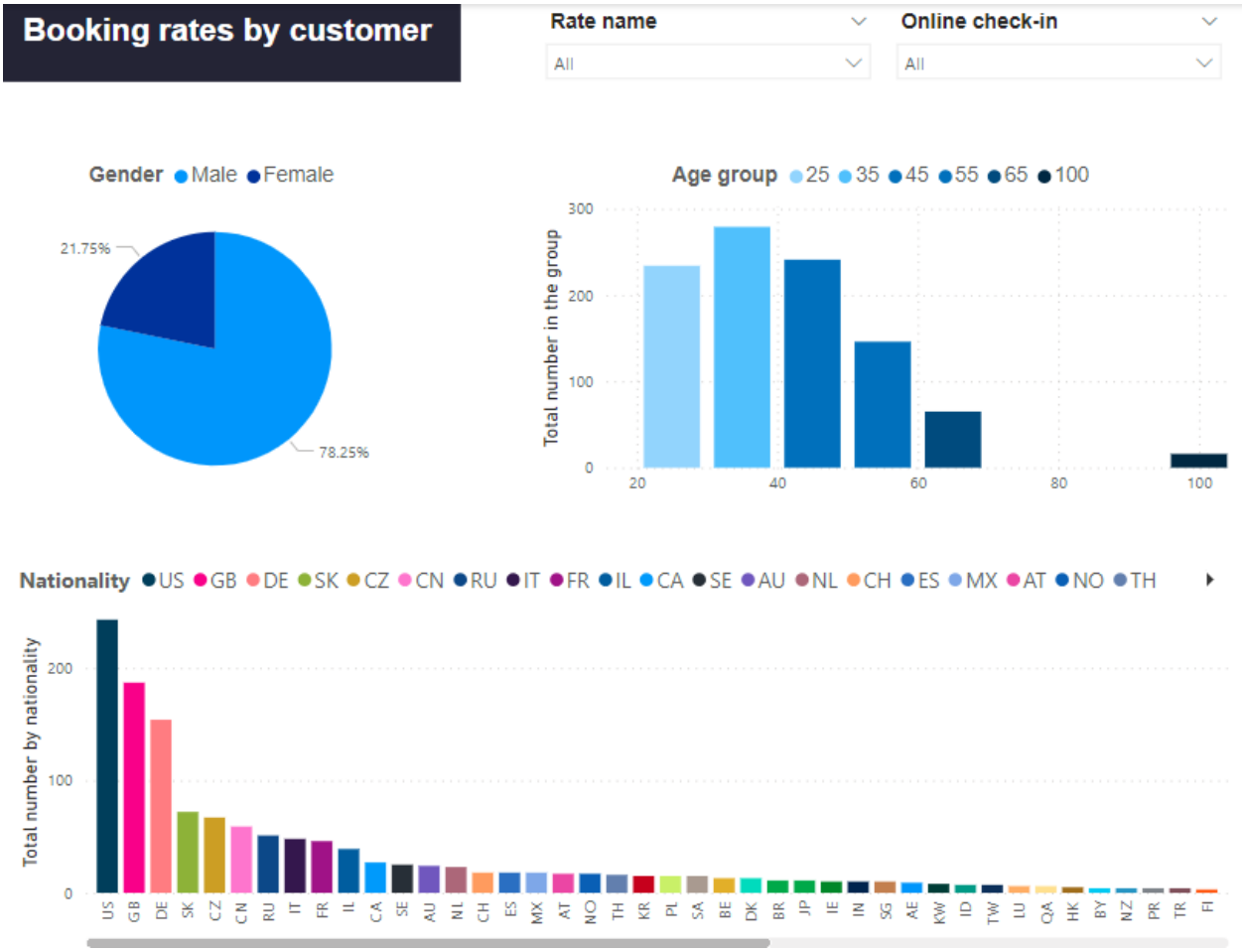
	RateName	AgeGroup	Gender	NationalityCode
1	Direct Booking	55	1	US
2	Min 3 nights	25	1	US
3	Early - 60 days	35	1	GB
4	Fully Flexible	35	1	GB
5	Min 4 nights	25	1	US
6	Suite Offer	45	2	US
7	Min 5 nights	25	1	US
8	Non Refundable BAR BB	35	1	US
9	Early - 21 days	35	1	US

I created 2 types of visualization with different approaches.

The 1st approach allows you to filter for the particular criteria.



The 2nd approach allows you to filter by Rate type and see for whom it is more popular.



The 2nd method consists in creating segments based on all 3 criteria and finds rows with matched criteria, as it is in typical marketing strategy (marketing segmentation). Segmentation from wider criteria (nationality) to narrow criteria (gender). Unfortunately, during this method there appears to be too many different segments, so I decided not to continue with this method, even if for some projects it could be more precis measurement, then separate criteria.

```
-- Join tables + days of week

create or replace table "Reservations with rates" as
select R1.*, R2."RateName", R2."SettlementAction", R2."SettlementTrigger", R2."SettlementValue", CASE
DAYOFWEEK(TO_TIMESTAMP("CreatedUtc"))
   WHEN 0 THEN 'Sunday'
   WHEN 1 THEN 'Monday'
   WHEN 2 THEN 'Tuesday'
   WHEN 3 THEN 'Wednesday'
   WHEN 4 THEN 'Thursday'
   WHEN 5 THEN 'Friday'
   WHEN 6 THEN 'Saturday'
end as "DayOfWeek"
from "Reservations" as R1
left join "Rates" as R2 on R1."RateId" = R2."RateId";

select * from "Reservations with rates";

-- task 1

select "RateName", "NationalityCode", "Gender", "AgeGroup", count(*) as Count1,
concat("NationalityCode",',', "Gender",',', "AgeGroup") as "Segment"
from "Reservations with rates"
where "Gender" != 0 and "NationalityCode" != 'NULL' and "AgeGroup" != 0
group by "RateName", "NationalityCode", "Gender", "AgeGroup"
order by Count1 desc;
```

Result for the second method

	RateName	NationalityCode	Gender	AgeGroup	COUNT1	Segment
1	Fully Flexible	US	1	35	25	US,1,35
2	Fully Flexible	SK	1	65	14	SK,1,65
3	Fully Flexible	US	1	45	14	US,1,45
4	Fully Flexible	GB	1	45	13	GB,1,45
5	Fully Flexible	GB	1	35	13	GB,1,35
6	Non Refundable BAR BB	US	1	25	13	US,1,25
7	Fully Flexible	GB	1	25	11	GB,1,25
8	Fully Flexible	US	1	55	10	US,1,55
9	Fully Flexible	GB	1	55	10	GB,1,55
10	Non Refundable BAR BB	US	1	35	9	US,1,35
11	Fully Flexible	DE	2	45	9	DE,2,45
12	Fully Flexible	GB	2	35	9	GB,2,35
13	Fully Flexible	SK	1	45	9	SK,1,45
14	Fully Flexible	CZ	1	45	8	CZ,1,45
15	Fully Flexible	US	1	25	8	US,1,25
16	Fully Flexible	US	2	25	7	US,2,25
17	Fully Flexible	US	1	65	7	US,1,65

What are the typical guests who do online check-in? Is it somehow different when you compare reservations created across different weekdays?

For the whole week.

```
-- task 2

SELECT
   (SELECT top 1 "AgeGroup"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1
    GROUP BY "AgeGroup"
    ORDER BY COUNT(*) DESC) as "Age",
   (SELECT TOP 1 "Gender"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1
    GROUP BY "Gender"
    ORDER BY COUNT(*) DESC) as "Gender",
   (SELECT TOP 1 "NationalityCode"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1
    GROUP BY "NationalityCode"
    ORDER BY COUNT(*) DESC) as "Nationality";
```

Results:

	Age	Gender	Nationality
1	35	1	US

For each day of the week.

```
-- task 2

SELECT
   'Monday' as "DayOfWeek",
   (
      SELECT top 1 "AgeGroup"
      FROM "Reservations with rates"
      WHERE "IsOnlineCheckin" = 1 and "DayOfWeek" = 'Monday'
      GROUP BY "AgeGroup"
      ORDER BY COUNT(*) DESC
   ) as "Age",
   (
      SELECT TOP 1 "Gender"
      FROM "Reservations with rates"
      WHERE "IsOnlineCheckin" = 1 and "DayOfWeek" = 'Monday'
      GROUP BY "Gender"
      ORDER BY COUNT(*) DESC
   ) as "Gender",
```

```
  (
    SELECT TOP 1 "NationalityCode"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1 and "DayOfWeek" = 'Monday'
    GROUP BY "NationalityCode"
    ORDER BY COUNT(*) DESC
  ) as "Nationality"
UNION ALL
SELECT
  'Tuesday' as "DayOfWeek",
  (
    SELECT top 1 "AgeGroup"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1 and "DayOfWeek" = 'Tuesday'
    GROUP BY "AgeGroup"
    ORDER BY COUNT(*) DESC
  ) as "Age",
  (
    SELECT TOP 1 "Gender"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1 and "DayOfWeek" = 'Tuesday'
    GROUP BY "Gender"
    ORDER BY COUNT(*) DESC
  ) as "Gender",
  (
    SELECT TOP 1 "NationalityCode"
    FROM "Reservations with rates"
    WHERE "IsOnlineCheckin" = 1 and "DayOfWeek" = 'Tuesday'
    GROUP BY "NationalityCode"
    ORDER BY COUNT(*) DESC
  ) as "Nationality"
UNION ALL …….
```

Results:

	DayOfWeek	Age	Gender	Nationality
1	All week	35	1	US
2	Monday	35	1	GB
3	Tuesday	35	1	US
4	Wednesday	55	1	US
5	Thursday	35	1	DE
6	Friday	45	1	RU
7	Saturday	25	1	US
8	Sunday	55	1	GB

Visualization:

Typical guest

Day of the week

Monday

Monday	GB	35	1
Day of the week	Nationality	Age	Gender

Legend

Gender:
1 - Male
2 - Female

Validation

NationalityCode	Count of NationalityCode
GB	6
IL	5
US	4
AT	2
DE	2
DK	2
BE	1
FR	1
IT	1
RU	1
SK	1
Total	**26**

AgeGroup	Count of AgeGroup
35	11
45	9
25	2
55	2
65	2
Total	**26**

Gender	Count of Gender
1	19
2	7
Total	**26**

Look at the average night revenue per single occupied capacity. What guest segment is the most profitable per occupied space unit? And what guest segment is the least profitable?

General:

```
-- task 3

create or replace table "Revenue per single occupied capacity" as
select "NightCount", "NightCost_Sum", "OccupiedSpace_Sum", ("NightCost_Sum"/"OccupiedSpace_Sum")
as "RevenuePerCapacity", "GuestCount_Sum", "IsOnlineCheckin", "NationalityCode", "Gender",
"AgeGroup"
from "Reservations"
where "CancellationReason" = 'NULL' and "OccupiedSpace_Sum" != 0;

-- per gender
select "Gender", avg("RevenuePerCapacity") as "RevenuePerCapacityG"
from "Revenue per single occupied capacity"
where "Gender" != 0
group by "Gender"
order by "RevenuePerCapacityG" desc;

-- per Age
select "AgeGroup", avg("RevenuePerCapacity") as "RevenuePerCapacityA"
from "Revenue per single occupied capacity"
where "AgeGroup" != 0
group by "AgeGroup"
order by "RevenuePerCapacityA" desc;

-- per NationalityCode
select "NationalityCode", avg("RevenuePerCapacity") as "RevenuePerCapacityN"
from "Revenue per single occupied capacity"
where "NationalityCode" != 'NULL'
group by "NationalityCode"
order by "RevenuePerCapacityN" desc;
```

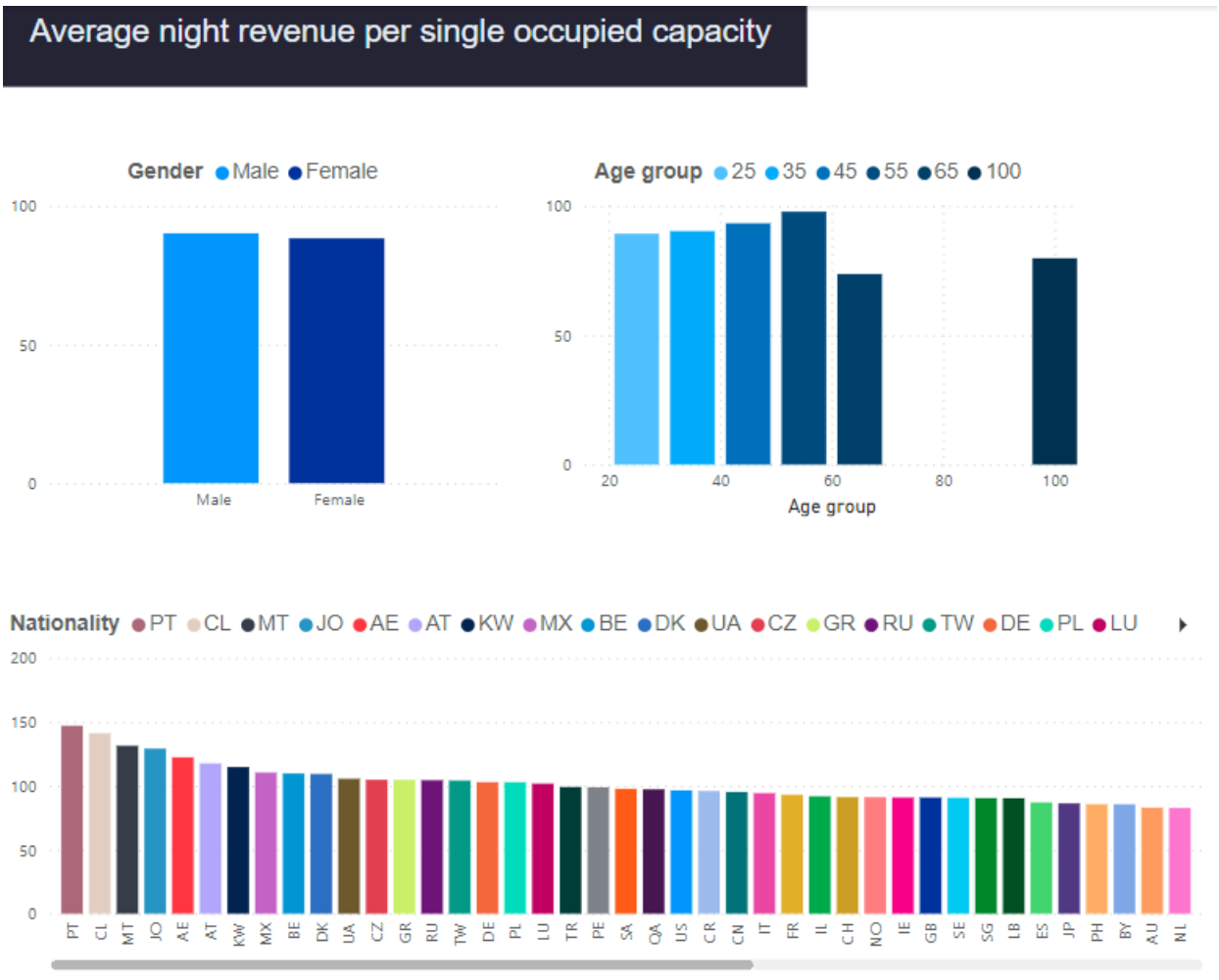
Results:

	Gender	···	RevenuePerCapacityG
1	1		90.176423486
2	2		88.353584383

	AgeGroup	···	RevenuePerCapacityA
1	55		97.815432723
2	45		93.308488614
3	35		90.291638591
4	25		89.333669999
5	100		79.805213705
6	65		73.702322773

	NationalityCode	RevenuePerCapacityN
1	PT	147.018701
2	CL	141.134056
3	MT	131.3951632
4	JO	129.158602545
5	AE	122.294348347
6	AT	117.555858361
7	KW	114.652729668
8	MX	110.443004862
9	BE	109.689407046
10	DK	109.294027803
11	UA	105.558986024

Visualization:



The most and the least profitable guest segment:

```
-- task 3

create or replace table "Revenue per single occupied capacity" as
select "NightCount", "NightCost_Sum", "OccupiedSpace_Sum", ("NightCost_Sum"/"OccupiedSpace_Sum")
as "RevenuePerCapacity", "GuestCount_Sum", "IsOnlineCheckin", "NationalityCode", "Gender",
"AgeGroup"
from "Reservations"
where "CancellationReason" = 'NULL' and "OccupiedSpace_Sum" != 0;

-- per gender Top 1
select TOP 1 "Gender", avg("RevenuePerCapacity") as "RevenuePerCapacityG"
from "Revenue per single occupied capacity"
where "Gender" != 0
group by "Gender"
order by "RevenuePerCapacityG" desc;

-- per gender Bottom 1
select TOP 1 "Gender", avg("RevenuePerCapacity") as "RevenuePerCapacityG"
from "Revenue per single occupied capacity"
where "Gender" != 0
group by "Gender"
order by "RevenuePerCapacityG" asc;

-- per Age Top 1
select TOP 1 „AgeGroup", avg("RevenuePerCapacity") as "RevenuePerCapacityA"
from "Revenue per single occupied capacity"
where "AgeGroup" != 0
group by "AgeGroup"
order by "RevenuePerCapacityA" desc;

-- per Age Bottom 1
select TOP 1 „AgeGroup", avg("RevenuePerCapacity") as "RevenuePerCapacityA"
from "Revenue per single occupied capacity"
where "AgeGroup" != 0
group by "AgeGroup"
order by "RevenuePerCapacityA" asc;

-- per NationalityCode Top 1
select TOP 1  "NationalityCode", avg("RevenuePerCapacity") as "RevenuePerCapacityN"
from "Revenue per single occupied capacity"
where "NationalityCode" != 'NULL'
group by "NationalityCode"
order by "RevenuePerCapacityN" desc;

-- per NationalityCode Bottom 1
select TOP 1  "NationalityCode", avg("RevenuePerCapacity") as "RevenuePerCapacityN"
from "Revenue per single occupied capacity"
where "NationalityCode" != 'NULL'
group by "NationalityCode"
order by "RevenuePerCapacityN" asc;
```

Results:

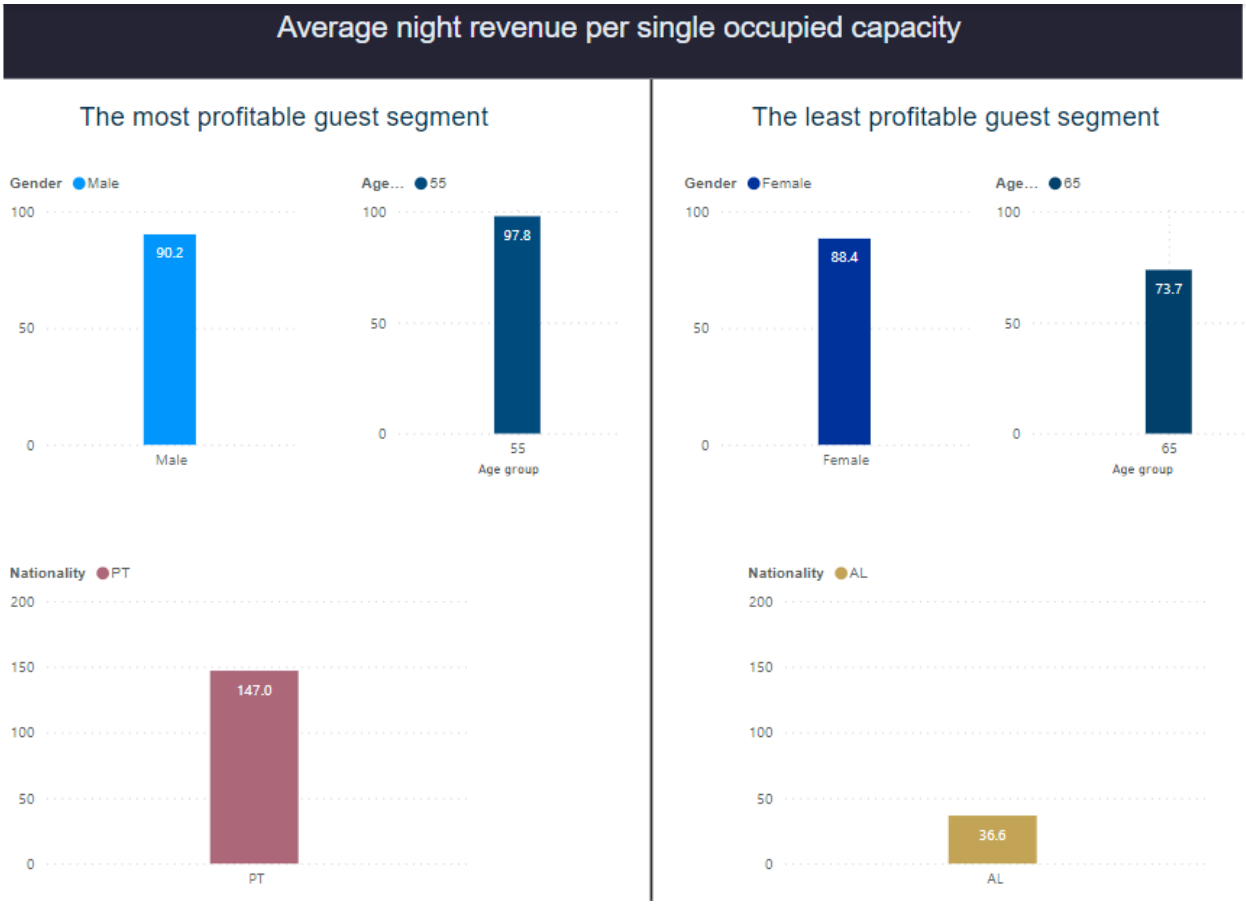
	Gender ···	RevenuePerCapacityG
1	1	90.176423486

	Gender	RevenuePerCapacityG
1	2	88.353584383

	AgeGroup	RevenuePerCapacityA
1	55	97.815432723

	AgeGroup	RevenuePerCapacityA
1	65	73.702322773

	NationalityCode	RevenuePerCapacityN
1	PT	147.018701

	NationalityCode	RevenuePerCapacityN
1	AL	36.613405

Visualization: