

Seazone Data Scientist Challenge

Report

Pedro Afornali

Index

1. Understanding the data.
2. Challenge questions.
3. About the challenge.

1. Understanding the data.

- Analysis framework.
- First insights.
- Expectations for 2022.

Analysis framework.

The first step to answer this challenge was to understand the database and its structure, for that I imported the two files into Jupyter Notebook and made the first codes in order to start an exploratory analysis.

Immediately the daily_revenue database seemed strange to me, and soon I got confirmation that the data was wrong.

Later, with the new database, I redid the exploratory analysis.

After understanding the meaning of the available data and the challenge questions, I started to put together the framework that I would deliver as an answer.

The main points taken into consideration for this structure were:

- Usability of results in the real world.
- Ease of handling throughout the year.
- Possibility to generate a dashboard of questions 2, 3 and 4 as daily performance monitoring.

The software chosen for this were: Excel and Power Bi.

For the report I chose Power Point.

First Insights.

During my exploratory analysis I tested various ways of drawing conclusions from the data, plotting graphs and testing hypotheses.

The best results I turned into a dashboard in Power Bi and imported to .ppt for sharing (my current account doesn't allow direct sharing).

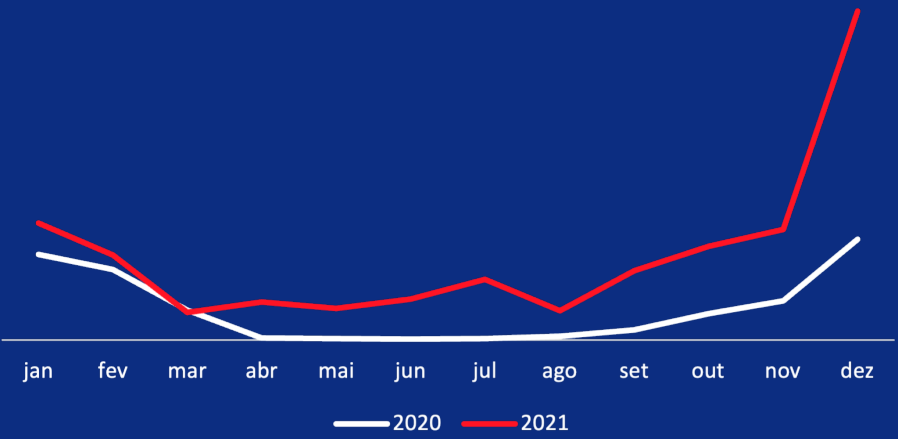
This file is on GitHub as Seazone Insights.

If the company has access to this software, or a similar one, it will be possible to reconstruct the model.

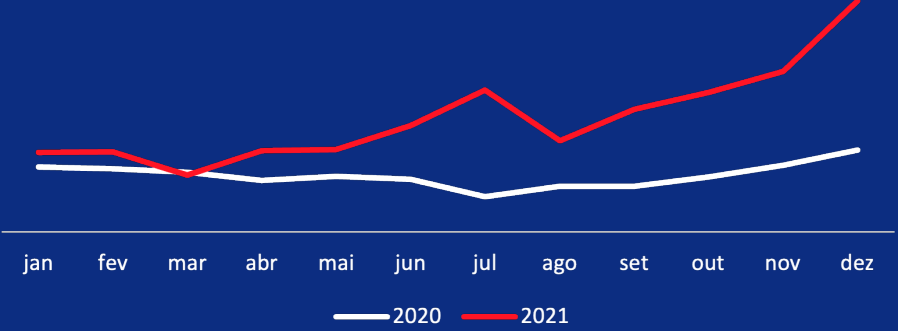
One of the first things I noticed was the company's seasonal periods, its best results come from the holidays, and there is a small increase in the mid-year holidays.

This is valid for revenue and bookings.

Revenue

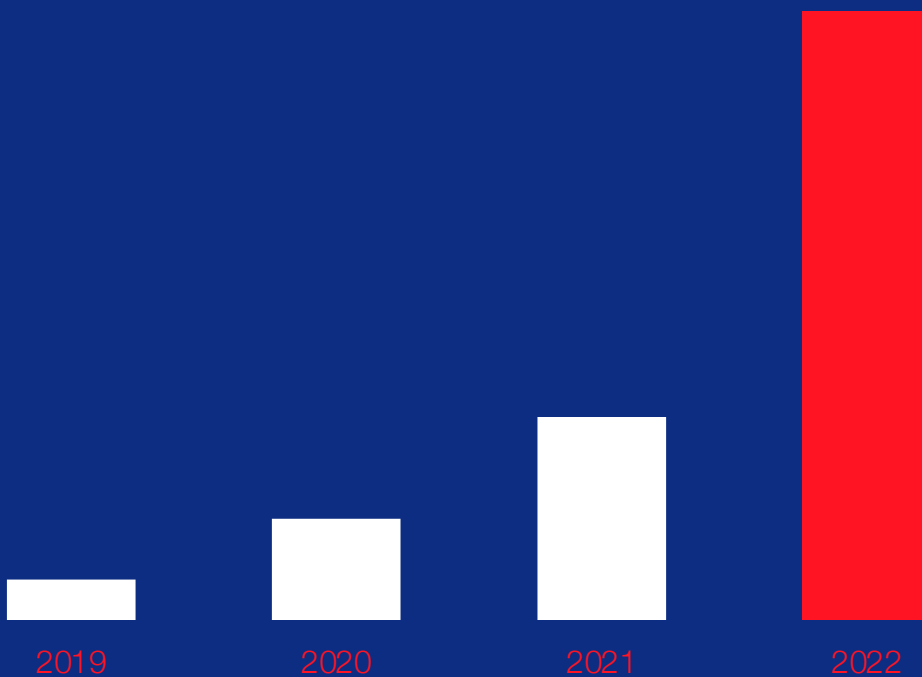


Bookings



Expectations for 2022.

The first **real year** for **Seazone**.



After analyzing all the data provide, 2022 will be the first real year for Seazone, and here is why.

Our data start on final 2019, the company survived covid-19 in 2020, increased your revenue in 2021 and in January of 2022 have your best month in history.



And from a data analysis aspect, this is worrying.

We have a few months of data before covid, an atypical year and a year in which the restrictions and care of the population eased but did not end.

Only now in 2022 are we seeing an official easing of restrictions against the pandemic and a real chance of what would be a return to normality.

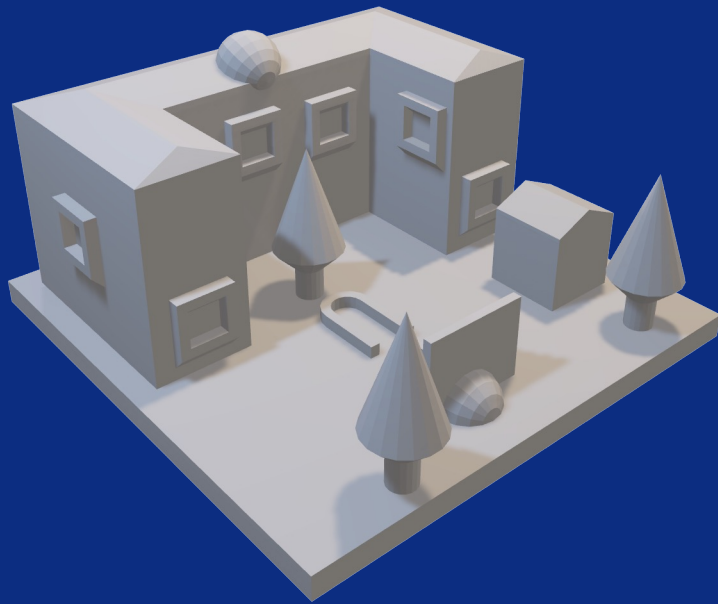
Considering this context, we would not have reliable enough data to measure the company's real potential, a fact that was evident during this analysis.

So, the answers found here should be viewed with care, always thinking about how external factors may have influenced this result or a future projection.

Challenge Questions.



1. What is the expected price and revenue for a listing tagged as JUR MASTER 2Q in march?



For solve this problem we have two possible ways:

1 - Use data from similar properties in same region to estimate the rental price and revenue for march.

2 – Use the property's features to build a price prediction machine learning model.

For a real-world solution, the second choice requires more data then was informed. Features like neighborhood, house facilities, distance for touristic attractions, etc... have a real impact on price rental and revenue.

Is possible to create a machine learning model with the data provide, but the results will not be realistic. For this reason, I chose to use the first option.

Using data from similar properties in same region.

Steps:

1. Filter the data for find MASTER2Q listings in JUR.
2. Filter the data on Daily_Revenue for March 2021, only year with data for this month non affected by pandemic.
3. Use a boxplot graph to capture the price range.

Using data from similar properties in same region.

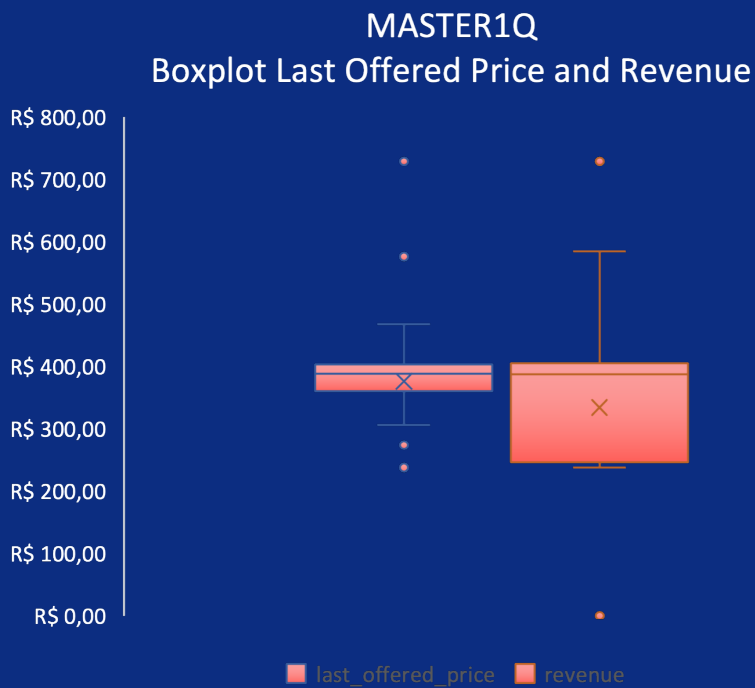
1. Filter the data for find MASTER2Q listings in JUR.

After this step we realized that there are no properties with these characteristics, the closest ones are MASTER1Q and MASTER3Q.

As our category is between these two, we can analyze MASTER Q1 and Q3 separately and put an intermediate price for MASTER2Q.

Steps 2 and 3 remains the same.

Master 1Q.



- Last offered price:

Min: 360,00

Max: 403,00

Mean: 375,00

- Revenue:

Min: 0

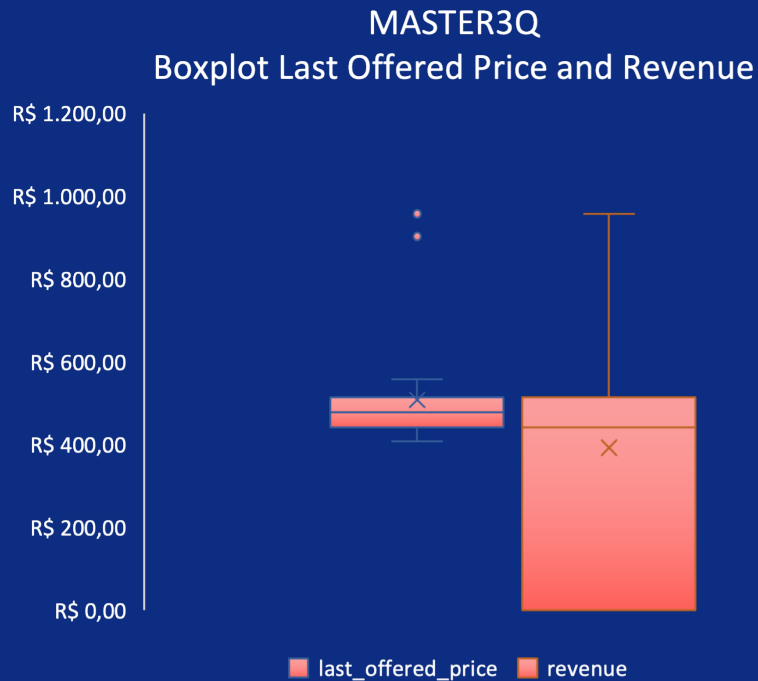
Max: 405,00

Mean: 333,00

* The upper limits and outliers were recorded on March 1st, the last day of the seasonal period.

** The 0 recorded on revenue is for days not sold.

Master 3Q.



- Last offered price:

Min: 442,00

Max: 515,00

Mean: 508,00

- Revenue:

Min: 0

Max: 515,00

Mean: 392,00

* The upper limits and outliers were recorded on March 1st, the last day of the seasonal period.

** The 0 recorded on revenue is for days not sold.

Results Master 2Q JUR - March.

We can estimate the sales price and revenue through the average, positioning the MASTER2Q category between the other two.

	Average Listing Price	Average Revenue per Day
Master 1Q	R\$ 375,00	R\$ 333,00
Master 3Q	R\$ 508,00	R\$ 392,00
Master 2Q	R\$ 441,50	R\$362,50

2. What is Seazone's expected revenue for 2022? Why?



To solve this problem, we can analyze the annual growth rate and project the 2022 revenue through the average of past years.

However, we don't have enough data for that. Few months of data in 2019.

The year 2020 was affected by covid-19, leaving revenues momentarily lower than normal, which results in very high percentages of growth when compared to 2021, and this leaves our projection unrealistic.

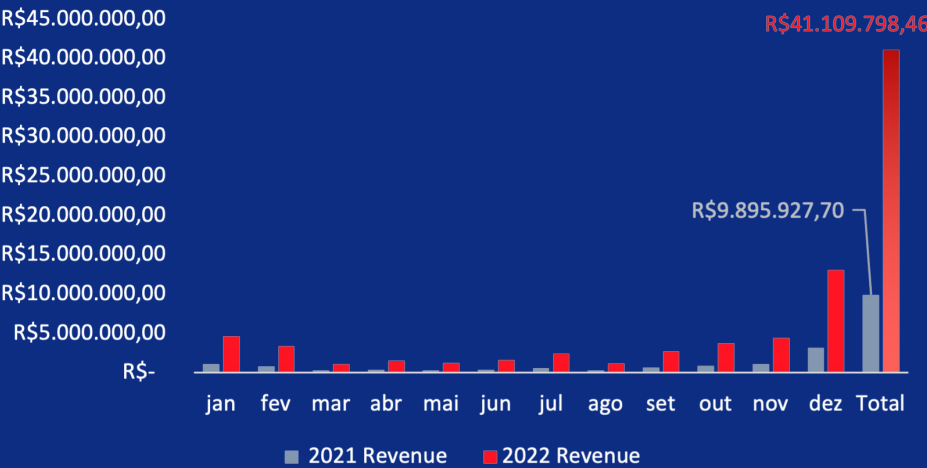
Thus, my choice to determine the 2022 revenue as accurately as possible is to analyze only the month of January 2021 and 2022 and project this growth for the other months.

After the calculus, we have this result:

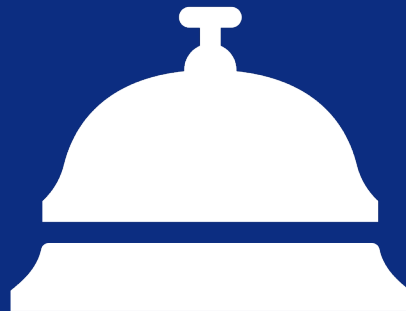
Month	Revenue
Jan 2021	R\$ 1.122.295,79
Jan 2022	R\$ 4.662.256,55
Growth	415%

Month	2021 Revenue	2022 Revenue
jan	R\$ 1.122.295,79	R\$ 4.662.256,55
fev	R\$ 812.533,64	R\$ 3.375.438,38
mar	R\$ 266.841,37	R\$ 1.108.516,08
abr	R\$ 366.162,75	R\$ 1.521.118,31
mai	R\$ 302.491,08	R\$ 1.256.612,59
jun	R\$ 390.310,13	R\$ 1.621.431,69
jul	R\$ 579.209,03	R\$ 2.406.158,09
ago	R\$ 281.171,37	R\$ 1.168.045,96
set	R\$ 663.012,11	R\$ 2.754.293,99
out	R\$ 900.389,25	R\$ 3.740.409,36
nov	R\$ 1.061.541,59	R\$ 4.409.870,62
dez	R\$ 3.149.969,59	R\$ 13.085.646,84
Total	R\$ 9.895.927,70	R\$ 41.109.798,46

2021 – 2022 Revenue



3. How many reservations should we expect to sell per day? Why?



For this problem we must use the data produced in the previous problem and the number of reservations per day to calculate the average ticket in 2021, then, estimate the number of reservations for 2022 with the average ticket and estimated revenue.

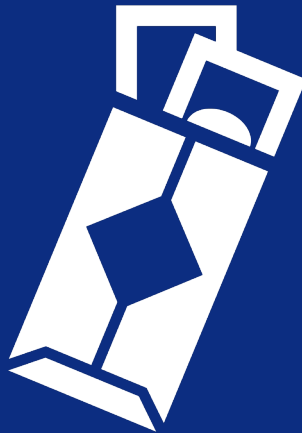
That way:

Month	2021 Revenue	Reservations	Average ticket
jan	R\$ 1.122.295,79	1877	R\$ 597,92
fev	R\$ 812.533,64	1881	R\$ 431,97
mar	R\$ 266.841,37	1346	R\$ 198,25
abr	R\$ 366.162,75	1916	R\$ 191,11
mai	R\$ 302.491,08	1946	R\$ 155,44
jun	R\$ 390.310,13	2503	R\$ 155,94
jul	R\$ 579.209,03	3335	R\$ 173,68
ago	R\$ 281.171,37	2142	R\$ 131,27
set	R\$ 663.012,11	2883	R\$ 229,97
out	R\$ 900.389,25	3281	R\$ 274,43
nov	R\$ 1.061.541,59	3776	R\$ 281,13
dez	R\$ 3.149.969,59	5431	R\$ 580,00

With the average ticket we can determine the reservations per day:

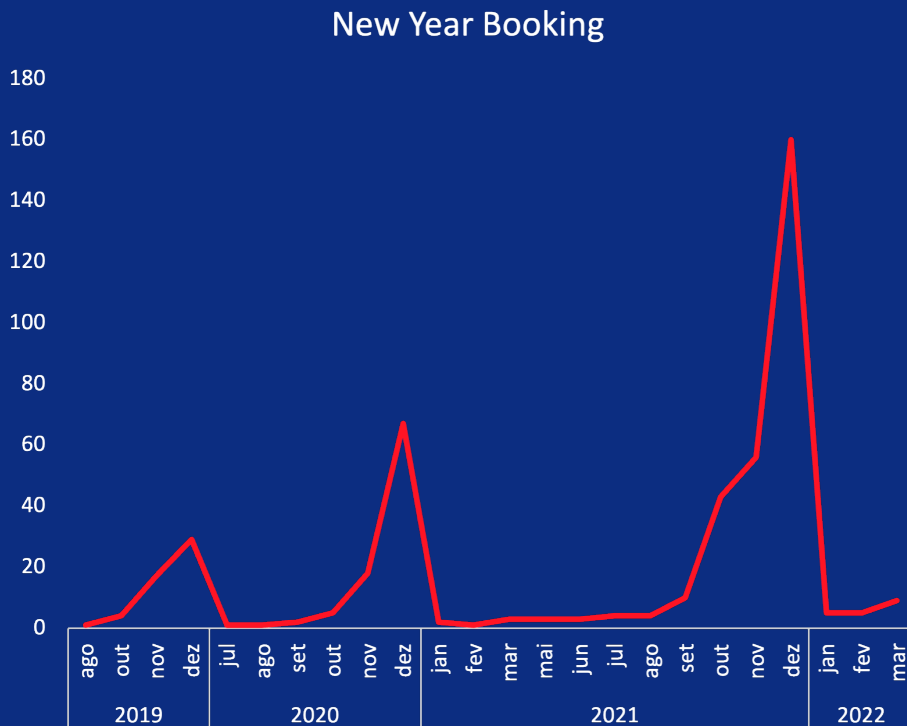
Month	2022 Revenue	Average Ticket	Reservations	Reservations per Day
jan	R\$ 4.662.256,55	R\$ 597,92	7797	260
fev	R\$ 3.375.438,38	R\$ 431,97	7814	260
mar	R\$ 1.108.516,08	R\$ 198,25	5592	186
abr	R\$ 1.521.118,31	R\$ 191,11	7959	265
mai	R\$ 1.256.612,59	R\$ 155,44	8084	269
jun	R\$ 1.621.431,69	R\$ 155,94	10398	347
jul	R\$ 2.406.158,09	R\$ 173,68	13854	462
ago	R\$ 1.168.045,96	R\$ 131,27	8898	297
set	R\$ 2.754.293,99	R\$ 229,97	11977	399
out	R\$ 3.740.409,36	R\$ 274,43	13630	454
nov	R\$ 4.409.870,62	R\$ 281,13	15686	523
dez	R\$ 13.085.646,84	R\$ 580,00	22562	752

4. At what time of the year should we expect to have sold 10% of our new year's nights? And 50%? And 80%?
5. - How can this information be useful for pricing our listings?

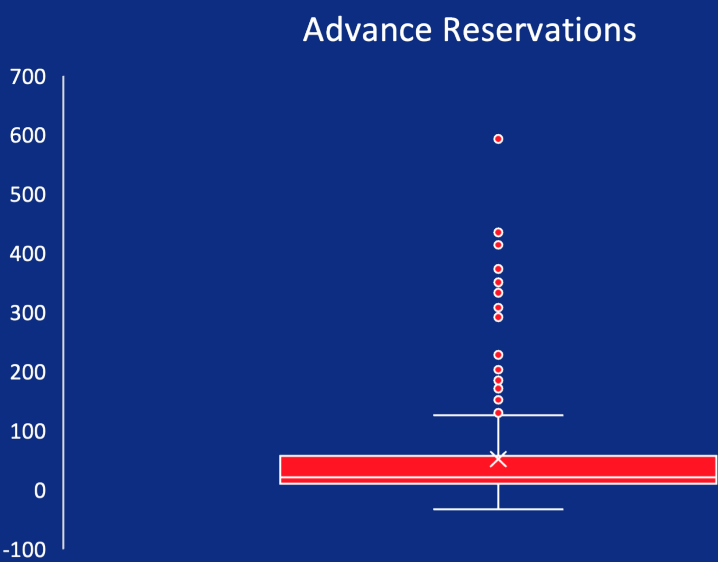


At what time of the year should we expect
to have sold 10% of our new year's nights?
And 50%? And 80%?

New Year's bookings start to gain strength in
October but only peak during December.



Observing the advance reservations, we can see that there are several outliers, however the main range is between 10 and 57 days before the new year.

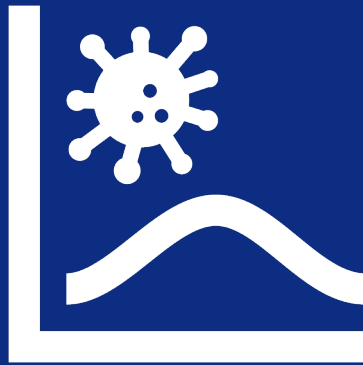


** This analysis needs to be reviewed since I was not able to extract more concrete information to fully respond the question.*

How can this information be useful for pricing our listings?

Knowing the moments of high demand for properties generates a competitive advantage for the company because it allows generating more assertive marketing campaigns and using the moment of high demand to work with better price ranges.

Covid-19 Impact.



Covid-19 Impact.

To understand the pandemic impacts on company's results, we can analyze the revenue difference from 2020 and 2021.

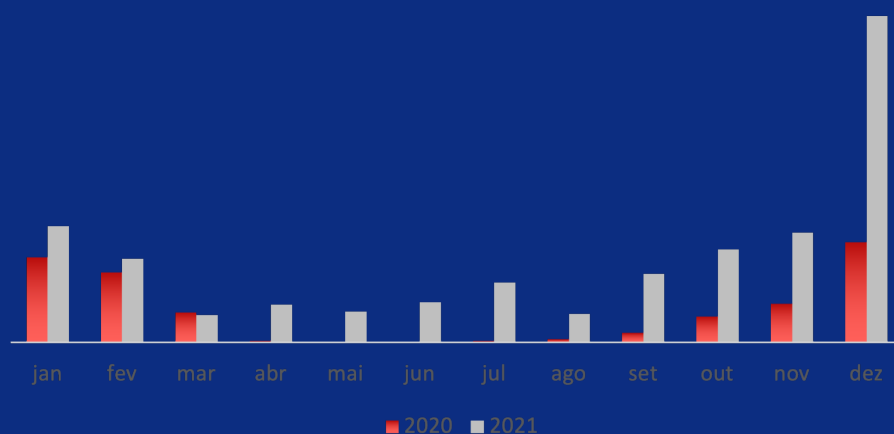
Although 2021 is still a year of restrictions, they have been more lenient, and we could see a significant improvement with the vaccine.

Mês	2020	2021	Revenue Diference
jan	R\$ 821.842,43	R\$ 1.122.295,79	27%
fev	R\$ 675.480,85	R\$ 812.533,64	17%
mar	R\$ 287.703,13	R\$ 266.841,37	-8%
abr	R\$ 18.477,91	R\$ 366.162,75	95%
mai	R\$ 13.627,19	R\$ 302.491,08	95%
jun	R\$ 10.201,71	R\$ 390.310,13	97%
jul	R\$ 14.669,83	R\$ 579.209,03	97%
ago	R\$ 34.281,85	R\$ 281.171,37	88%
set	R\$ 96.198,60	R\$ 663.012,11	85%
out	R\$ 251.116,34	R\$ 900.389,25	72%
nov	R\$ 374.996,12	R\$ 1.061.541,59	65%
dez	R\$ 966.296,50	R\$ 3.149.969,59	69%
Total Geral	R\$ 3.564.892,46	R\$ 9.895.927,70	64%

As we can see, from April, when the pandemic intensified across the country, the company's revenue was 95% lower when compared to the following year, also during the pandemic, but in a more controlled way.

The revenue decreased reached the maximum of 97%, starting a recovery only in August, with 88%.

When we put this data in a graph, we can see that the April to July revenue practically disappeared from the plot.

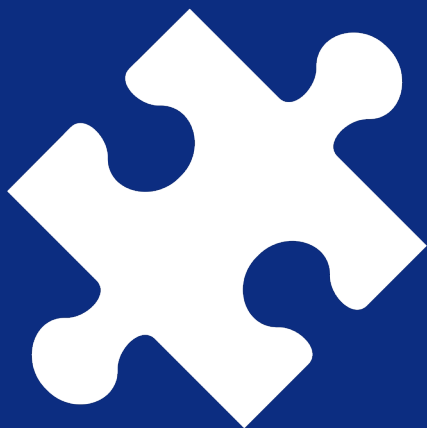


Since the company doesn't have enough data before 2020, it's impossible to say exactly how much the pandemic has affected the business.

What we can conclude after comparing the annual results is that Seazone was indeed affected by Covid-19. But the company managed to recover from the end of 2020 and it's had good results since then.

It is also possible to observe that the company is highly affected by new year seasonality, having year after year its best performance in the months of December and January. Which explains part of his recovery in 2020.

About the challenge.



About the challenge.

This challenge showed an interesting level of difficulty and complexity, mainly due to the database context.

During the analyzed period, there were several variables that impacted the company's results and were outside internal control.

Even so, the company shows incredible growth potential and high expectations for the current year, since its best period occurs in the last months of the year and January 2022 was already the best month in the company's history.

Extracting as much information as possible while planning what would be the best way to use it in the company's day-to-day activities was a fun and challenging task.

Finally, I believe that using the data found to establish the 2022 key performance indicators would be the best way to take advantage of this analysis and I am sure that my knowledge and experiences will be of great value to Seazone.

Pedro Afornali