Sales Analysis

Maeva Braeckevelt

During the Covid19 pandemic, every company had to adapt to the new measures. Restaurants were particularly challenged in order to stay profitable. Managers of restaurants struggle to staff their restaurant accordingly. I decided to analyze how certain variables affect the sales during this special period. As this situation is new,I collected the data from the first day of Belgium new lockdown regulation (20/10/2020) until november 30th, day I gather the data. As a robustness check, more data should be added, as the study serve only as a basis. The result for this particular restaurant are the following : first, the weather has little impact on the sales, but can be measured. Secondly, one additional productive hour corresponds to an average of 82.1 euro extra earnings when all the variables remain the same and thirdly, the sales change in function of the day of the week. Finally, I chose two models to predict the sales. However, with the two models chosen, I observed very high residuals. I would investigate those residuals and see if there were external problems, if not, I will wait to gather more data before giving those guidelines to the manager.

Introduction

What factors could impact the sales of a restaurant during the Covid-19 pandemic? That's the question I want to answer. Restaurants are living a rough period : in Belgium, since october 20th of 2020, new sanitarian measures, among others,appeared : the restaurant can not welcome sitting clients anymore (only take-away meals) and there is a curfew at 10pm. The restaurants of my company are struggling to be profitable and their managers need as much information as possible to be able to take the right decision in a unstable environment. A way of staying profitable is to staff the restaurant according to the demand. Maybe knowing how the productive hours or the day of the week (and other variables) are correlated to the sales, in this particular period, could be insightful for this decision. I have taken the data of one of our restaurants and I will use it for my analysis. There is not such analysis done under "normal" cirucumstances yet but it would be interesting to compare those in the future.Due to the recentness of the situation I only have 42 observations, from october 20th to november 3Oth. Thus, the unlikely circumstances and the lack of observations may be problematic for significant result during the analysis and for external validity. Unfortunately, the actual circumstance will remain so the analysis is worth being done even with few observations. Moreover, once the model is set, adding observations and see the evolution will be interesting as well.

Data

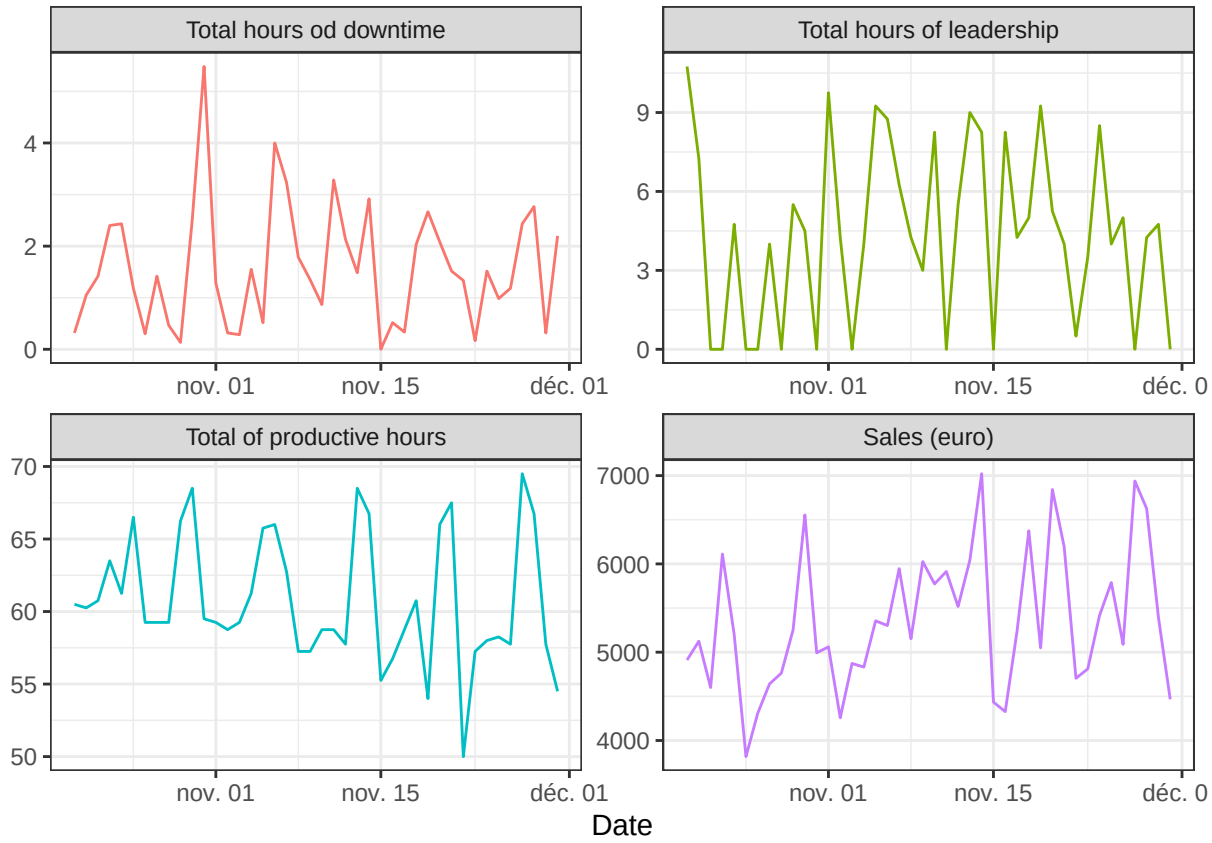
Staffing a restaurant according to the demand is a difficult decision. Providing statistical directives could be an advantage to the manager.There is a lot of variables that could affect the sales, I chose the variable in a way that either the manager has control on the variables or that he can know it in advance to take it into consideration when he is doing the schedule.

Manager's decision : Working hours and hours of downtime

The manager can choose how many hours will be used per day and if he will be working on the shift as well. So, my goal is to observe how increasing or decreasing the productive hours could impact the sales and if

the presence of the manager has an impact on the sales and how. I used the Sales in EURO (HT) per day (y). The productive hours (hours of kitchen's employee and waiters) as a variable and the leadership hours (hours of the managers working either in the kitchen or as a waiter) as another variable. The restaurant is using an external delivery platform. When the team feels swamped by the number of clients on-site ordering in the restaurant, the team can chose to pause the delivery platform (downtime hours) so that no client can order on the platform during a given period. How does this method impact the sales?

variable	n	mean	median	min	max	sd	skew
Sales	42	5358.265835	5217.607440	3817.94	7021.000704	789.843909	0.4048451
Productive_hours	42	60.750000	59.250000	50.00	69.500000	4.479969	0.2248577
Leadership_hours	42	4.375000	4.250000	0.00	10.750000	3.366464	0.0809803
hours_of_downtime	42	1.575397	1.416667	0.00	5.483333	1.176833	0.9955381



I observe that for every variable the median and the mean are fairly close for each other. So the distribution is close to symmetric. I chose not to transform any of ma variable. This analysis is aim at helping manager, so it is essential to keep it the most simple as possible. There is two extrem values, the 31/10/2020 there is a high total of hours of downtime and the 25/10/202 I have a fery few of observations and we are in a special period so drop or control some variables will loose the sense this analysis so I chose to keep all my data but keep that in mind during the analysis.

Out of the control variable : day of the week and weather

In this analysis I choose to proceed as a stationnary time serie analysis.Indeed, there is a probability that the day of the week have a seasonality effect but with the few variable that I have, that will made my model

more complicated and difficult to interpret and the difference will not be significative enough. (The PACF model help me to take that decision). However, with more observation it will be imperative to take lags and verify if there is a seasonality or if it was repressed by the pandemic. This two variables are nominal : weather can be sunny, cloudy and rainy and the day of the week can either Monday, Tuesday etc.

Source of the data and transformation

I chose not to transform my data. This analysis aims at helping the managers so over-complicated interpretation could diminish its usefulnesst.The sales are coming from our POS (technology provider to the European hospitality market).The productive hours and the leadership hours are coming from our scheduler program. The downtime of the delivery platfrom is coming from the website of that supplier.The weather was compiled by myself looking into https://www.timeanddate.com/weather/belgium/brussels/historic website. I will chose a significative level of 0.2 because it is enough for guidelines.

Model choice and interpretation

After taking a look at the scatterplot, I realised that the leadership hours don't have a correlation with the Sales. I also did not keep the downtime hours for the following reason : by knowing the process of the pause of the delivery platform and my personal experience, the relation between sales and downtime hours is mediated by the number of client during the day and the time of the day, both variable that I did not include in my data. From example, on a busy Friday evening, the restaurant needs to close the delivery platform for a few hours while they deal with the amount of customers on site. The number of orders and the time of the downtime could be found so, for the next analysis, it would be useful to include it. While I was looking at all the different models, I realized that even if there were no high correlation between the variables, the day of the week and the productive hours are related. It makes much sense because the manager changes the schedule according to the day of the week, taking the tendency pre-covid. It would be interesting to test a week with the same amount of productive hours everyday to detect a difference in the sales.I chose to keep two models, because I think they can both be valuable. I decided for the reason explained before to split days of the week and productive hours. The models are linear regression, indeed the following graph show a linear correlation.

Model 1 : **Formula : Sales = 248.54 + 82.1 * Productive hours + 371.24* Sunny weather - 336.27* rainy weather**

B0 : The interpretation is meaningless **Beta 1** : One additional productive hour corresponds to an average of 82.1 euro extra earnings when all the variable remain the same. **Beta 2** : On a sunny day, the sales change by 371 (euro) on average per day when the total hours of productivity remains the same. **Beta_3** : The confidence for this interpretation is less than 80% so I won't interpret it.

What information can I use for the manager ?

The sales are a little bit higher on a sunny day, that's an interesting information to use in general. Knowing how the productive hours impact the sales is very valuable. In the past, We already calculated how much a productive hours need to add if we want to stay profitable (no overstaffing) and keep the team in good working condition (not understaffed) and the result was 65 euro per hour. The new regulation, specially the take-away obligation, might allow the employees to be more productive (less cleaning, serving at the table, etc). It could be interesting to calculate again the ideal number in the new conditions and compare it to the 82.1 euro. Now with that information, the manager could estimate the sales with the hours of staffing and vice versa.

Model 2 : **Formula : Sales = 6115.9 - 1407.9 * Monday - 301.9 * Saturday - 1392.3 * Sunday - 1054.7 * Thursday -857.8 * Tuesday -732.7 * Wednesday -324 * Rainy + 274* Sunny**

B0 : The average Sales for a cloudy friday is 6115,9 euro. **Beta 1** : On a Monday the average sales is 1407.9 euro less on average when the weather is cloudy. **Beta 2** : The confidence for this interpretation is

less than 80% so I wont interpret it. **Beta 3** : 0n a Thursday the sales is 1392.3 euro less on average when the weather is cloudy. **Beta 4** : On a Tuesday the sales is 857.8 euro less on average when the weather is cloudy. **Beta 5** : On a Wednesday the sales is 732.7 euro less on average when the weather is cloudy. **Beta 6** : The confidence for this interpretation is less than 80% so I won't interpret it. **Beta 7** : On a sunny Friday, the sales changes by 274 (euro) on average per day.

What information can I provide to the manager ?

Same interpretation for the weather. It would have make more sens to transform the variable and talk into percentage. The reason I didn't do it is operational : to create the schedule of the week, the manager has to insert the projected sales into our scheduler program and it calculates automatically the ratio of hours to use. So it is a lot easier for them to talk about average sales that about percentage. Thanks to this formula I could create a pre-defined excel, to help them evaluate the sale on a sunny monday for example.

Analysis of the residuals

Table 2: Modele 1 & 2 : List of the 5 sales with the largest negative errors

Date	Sales	reg2_y_pred	reg2_res	Date	Sales	reg4_y_pred	reg4_res
2020-10-25	3817.940	5729.090	-1911.1503	2020-10-25	3817.940	4723.591	-905.6510
2020-10-26	4310.046	5131.587	-821.5413	2020-10-31	4992.995	6088.012	-1095.0175
2020-11-04	4831.693	5667.655	-835.9613	2020-11-03	4872.546	5532.052	-659.5061
2020-11-05	5355.682	6038.519	-682.8370	2020-11-04	4831.693	5657.251	-825.5575
2020-11-06	5301.545	6059.122	-757.5778	2020-11-06	5301.545	6389.901	-1088.3565

For both of the model, I observe high negative residuals, it means that the sales have been overestimated. This dates could be worth investigated for: if some external factors could have been a reason for a low sales on that day. Broken machines, a strike, an IT malfunction for example. I can also see that there are the two extreme value. By keeping them, I knew it will probably found it in the residuals.

Table 3: Modele 1 & 2 : List of the 5 sales with the largest positive errors

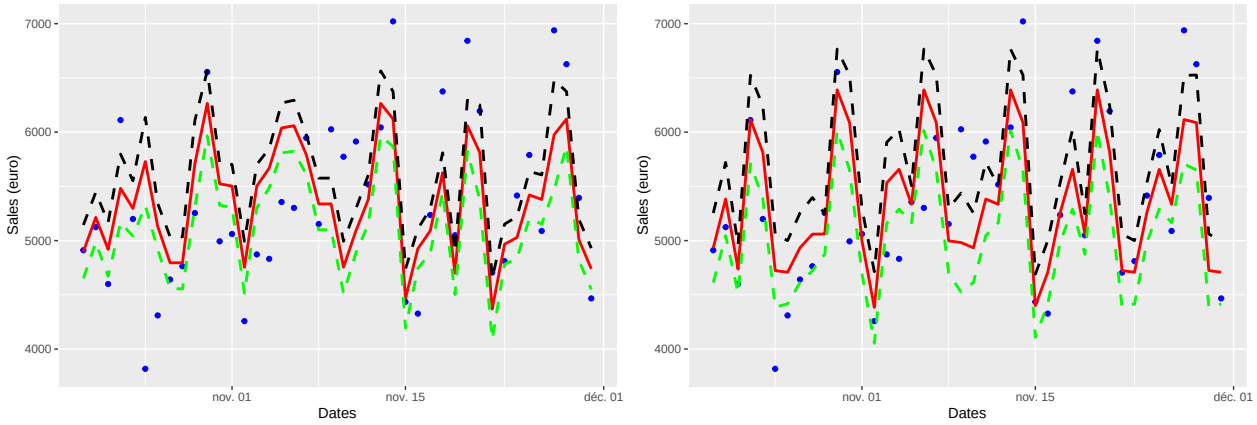
Date	Sales	reg2_y_pred	reg2_res	Date	Sales	reg4_y_pred	reg4_res
2020-11-10	5773.381	4754.110	1019.2712	2020-11-09	6025.951	4982.019	1043.9319
2020-11-11	5913.556	5090.380	823.1763	2020-11-10	5773.381	4933.990	839.3906
2020-11-14	7021.001	6120.933	900.0677	2020-11-14	7021.001	6088.012	932.9884
2020-11-20	6842.264	6059.122	783.1418	2020-11-18	6375.254	5657.251	718.0033
2020-11-27	6938.410	5976.333	962.0772	2020-11-27	6938.410	6115.868	822.5417

For both model, I observe high positive residuals, it means that the sales have been underestimated. With the same logic, I would investigate those dates: if a big catering order has been made, if another restaurant had a problem and sent the client to this one, or if it was a evening before a bank holiday for example.

Visualization of fit of the predicted value

The predicted values are not exactly on the line, but I can see that the tendency is no too far from it. I can also say that the predicted value is more accurate when the sales are small for both of the models. The more observations we will add to the model, the more it will become accurate.

Uncertainties in my prediction and plot: confidence interval of predicted values and prediction interval.



From my personal experience, a shift could be very difficult if the projection of sales was off by 500 (euro). Between -500 and 500 some solutionS are provided : during the shift when there is less client than expected, we let people go home early, and deduct the work hours. If there are more clients than expected, employees know they are likely to work a bit longer. The step I would do is investigate those residuals and see if there were any external problems, if not, I will wait to gather more data before giving those guidelines to the manager.

External validity

As I mentionned before, because of the low amount of variable, I am not able to test my model. However, next month, with more data, I will be able to test this model and add lags for example if necessary. I will keep the structure of this analysis for a basis for the other restaurant but I will definitely not apply those

results to another restaurant of the company before testing and I know already, from personal experience that I will have to change the models.

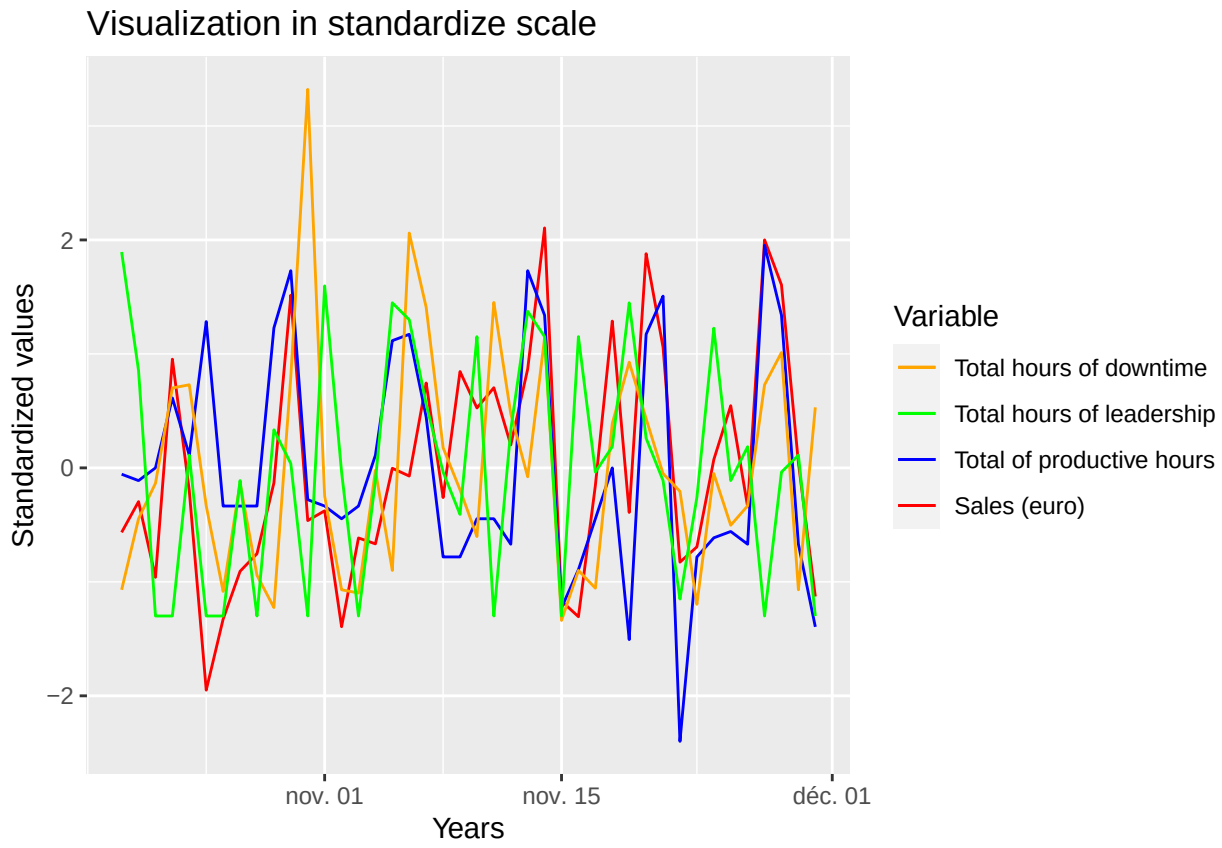
Summary

During the Covid19 pandemic, every company had to adapt to the new measures. Restaurants were particularly challenged in order to stay profitable. Managers of restaurants struggle to staff their restaurant accordingly. I decided to analyze how certain variables affect the sales during this special period. As this situation is new,I collected the data from the first day of Belgium new lockdown regulation (20/10/2020) until november 30th, day I gather the data. As a robustness check, more data should be added, as the study serve only as a basis. The result for this particular restaurant are the following : first, the weather has little impact on the sales, but can be measured. Secondly, one additional productive hour corresponds to an average of 82.1 euro extra earnings when all the variables remain the same and thirdly, the sales change in function of the day of the week. Finally, I chose two models to predict the sales. However, with the two models chosen, I observed very high residuals. I would investigate those residuals and see if there were external problems, if not, I will wait to gather more data before giving those guidelines to the manager.

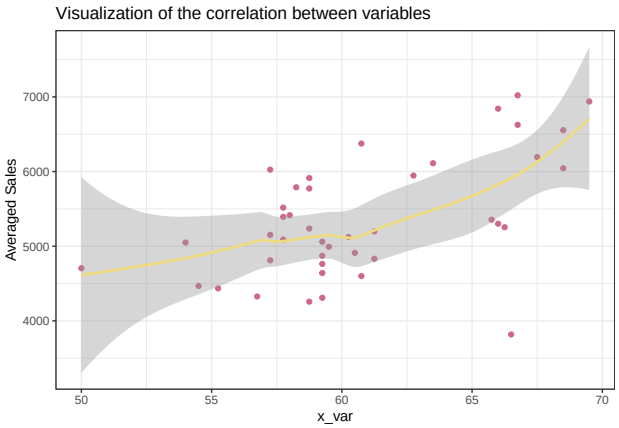
Appendix

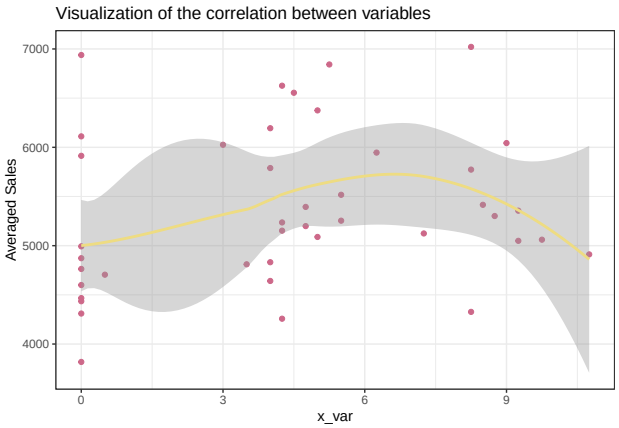
This appendix contains the documentation of the analysis.

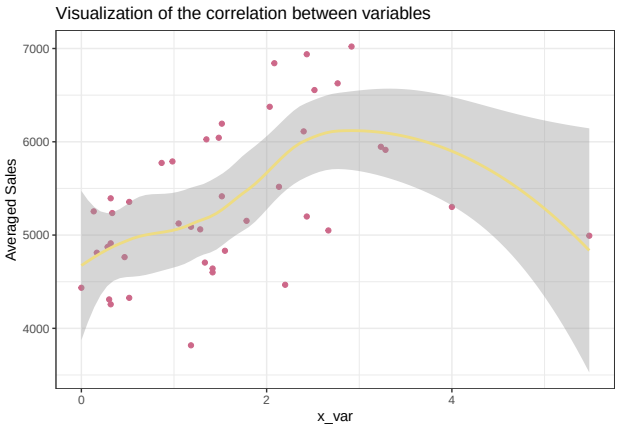
1. Standardization - good to compare the (co)-movement



2. Correlation between variables

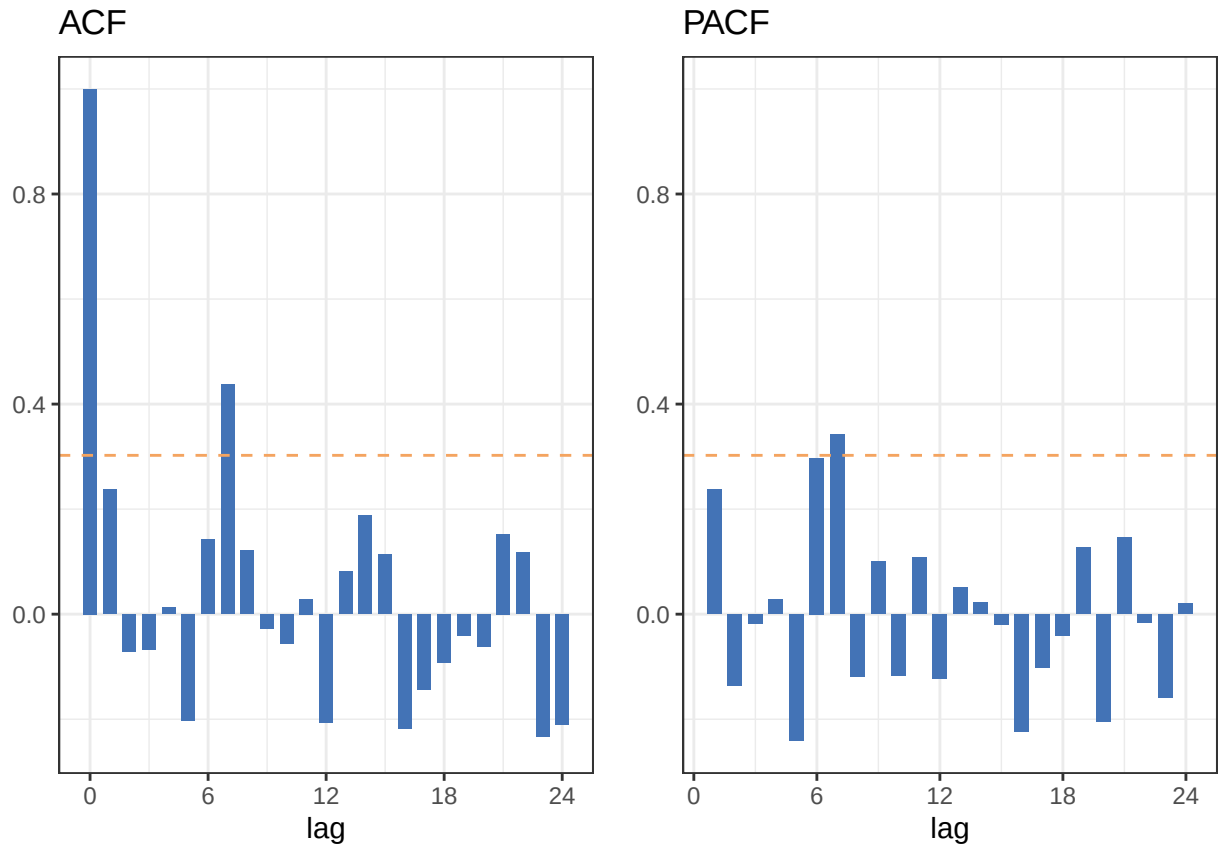




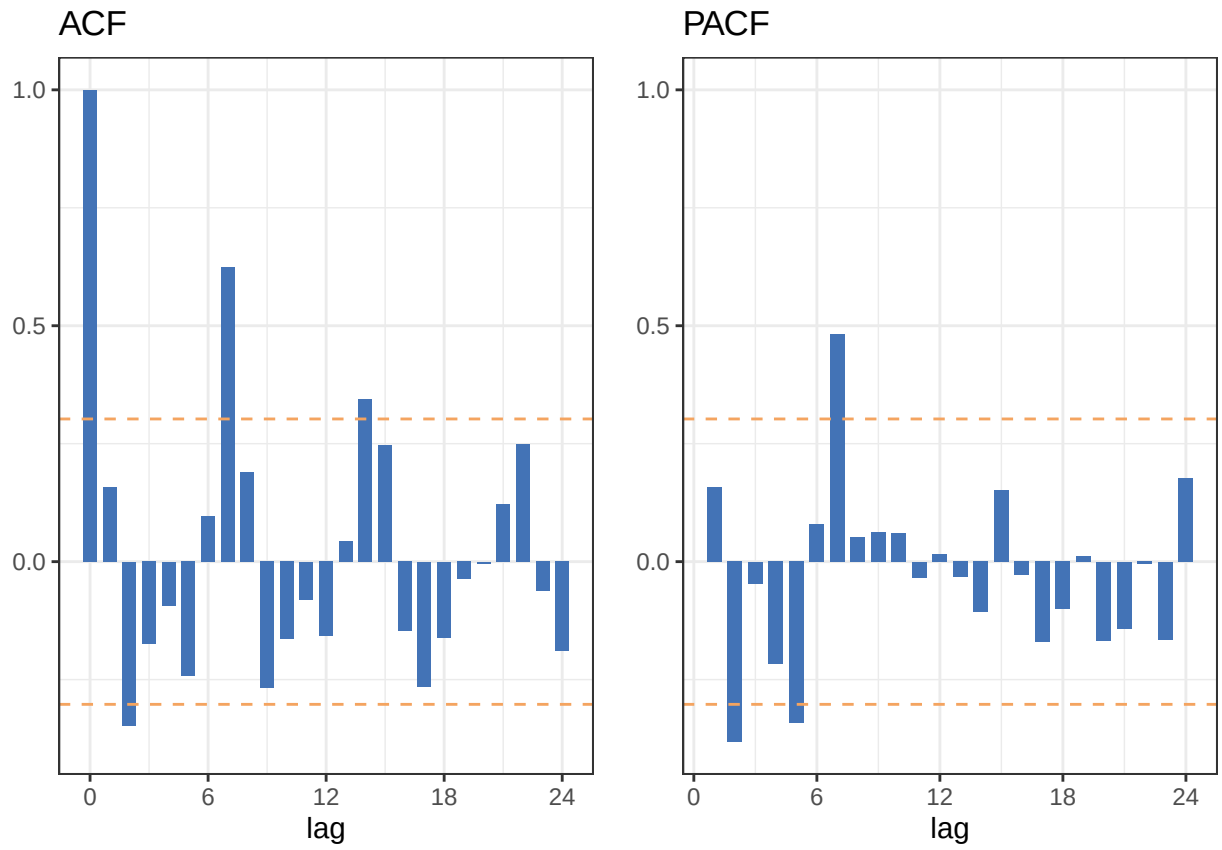


3. ACF PACF Serial-correlation (a.k.a. Auto-correlation)

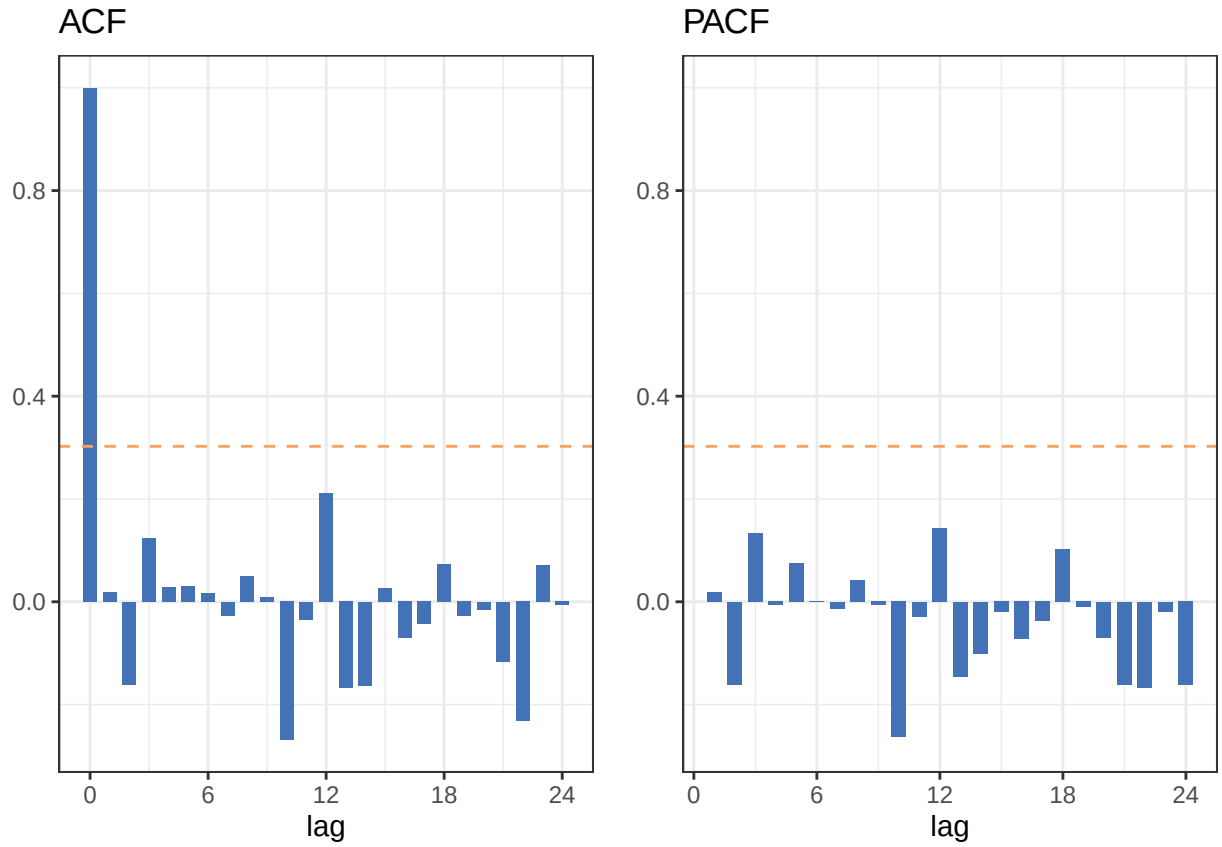
Sales



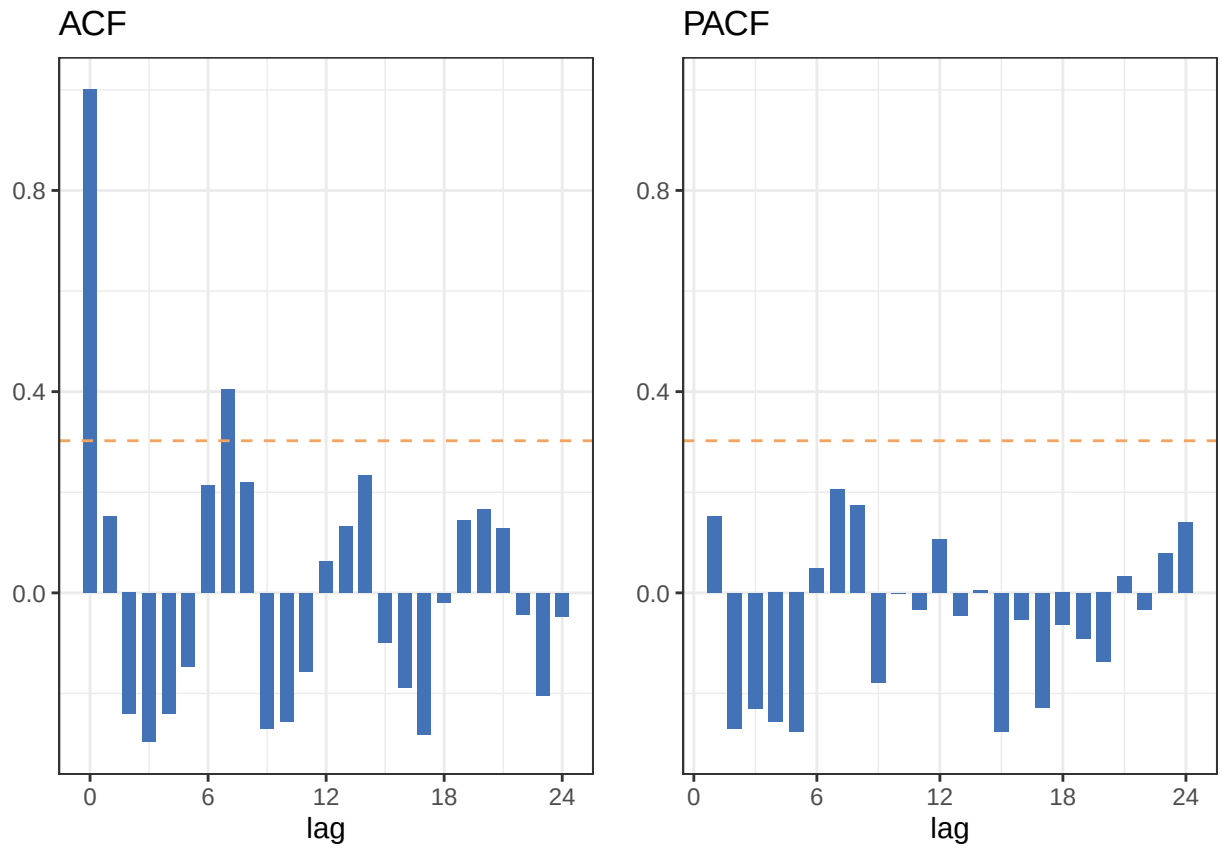
Productive hours



Leadership hours

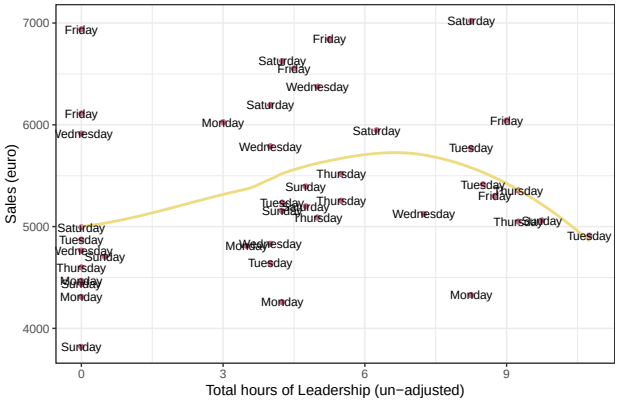


Hours of downtime

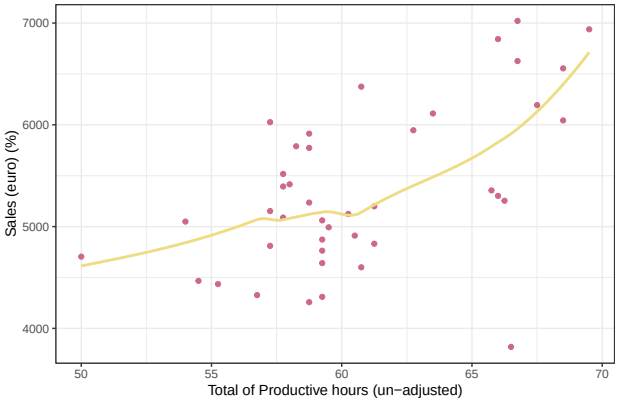


4. Scatterplot for investigation

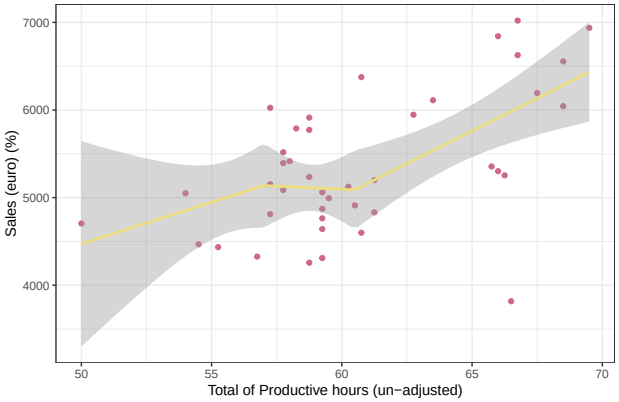
Scatterplot Sales − Leadership hours per day of the week



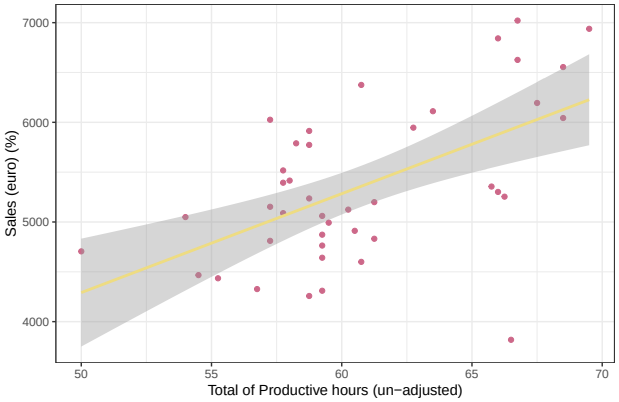
Scatterplot Sales − productive hours



PLS : Sales and Productive hours



Linear regression : Sales and Productive hours



Scatterplot : Sales and downtime hours per day of the week