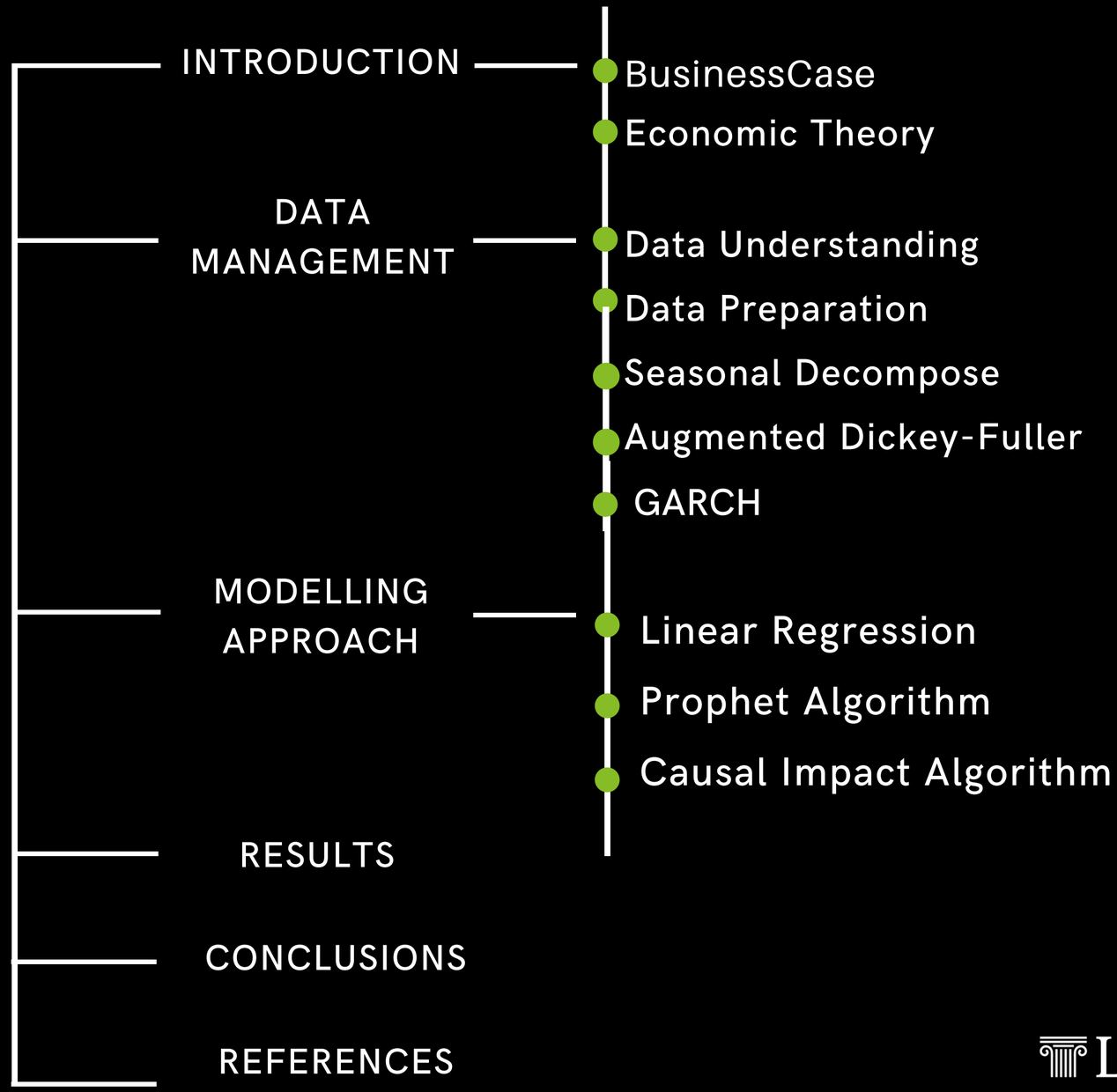
Deloitte.

Machine Learning for counterfactual estimation and forecasting

Olimpia Sannucci
Martina Manno
Carlo Ardito

LUISS

TABLE OF CONTENTS

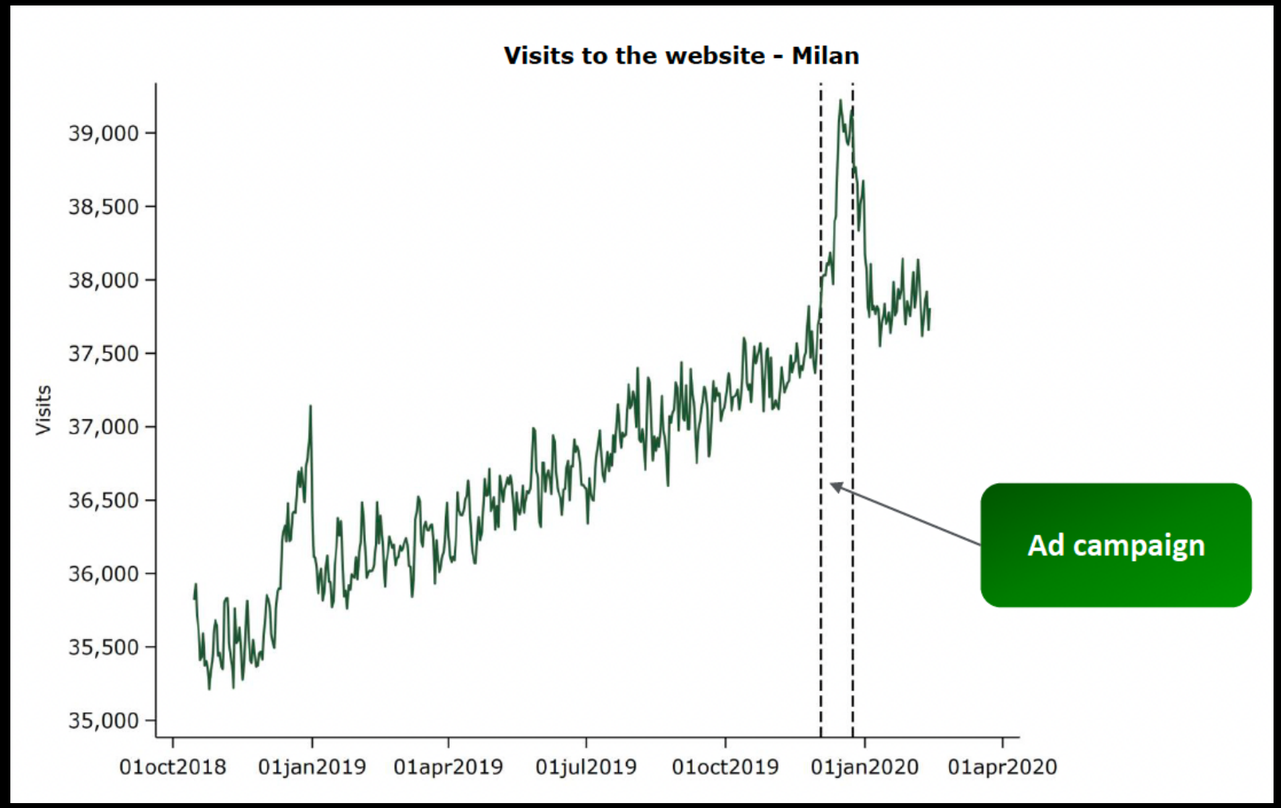


LUISS

INTRODUCTION

BUSINESS CASE

Alpha company, a perfume seller, ran an advertising campaign from December 3rd 2019, to December 24th 2019.



LUISS

INTRODUCTION

OBJECTIVES

- What is the impact of the advertising campaign on sales by analysing visits to its website and online purchases?

- Estimate advertising effectiveness through the counterfactual: what would have happened without AD exposures?

- The analysis results will be used to decide on the possibility of launching an advertising campaign later on.

LUISS

INTRODUCTION

ECONOMIC THEORY

"Correlation does not imply causation"

Even if association or correlation between variables is observed, it does not imply an actual casual-effect relationship.

Causal Inference estimating the counterfactual

A prediction of what would have happened without treatments or advertising campaigns, as in this case.

Confounding Factors

Seasonality.

LUISS

DATA MANAGEMENT

DATA UNDERSTANDING

Data refers to three Italian cities (Milan, Rome, Naples) as representative of the Italian market in the three main geographical areas

VISITS:

# of daily visits to the e-commerce website

CONV_RATE:

% of visitors that finalized an online purchase

AV_SPEND:

Average expenditure (EUR) of visitors who purchased

G_SEARCH_DATA:

Daily number of Google queries on the company's product divided by city.

LUISS

DATA MANAGEMENT

DATA PREPARATION

To better analyze the impact of the campaign, other variables had to be calculated, and the correlation between them must be observed

TOT_ON_PURC:

total number of daily purchases for each city

CONV_RATE x VISITS

TOT_SALES:

total sales (EUR) for each city

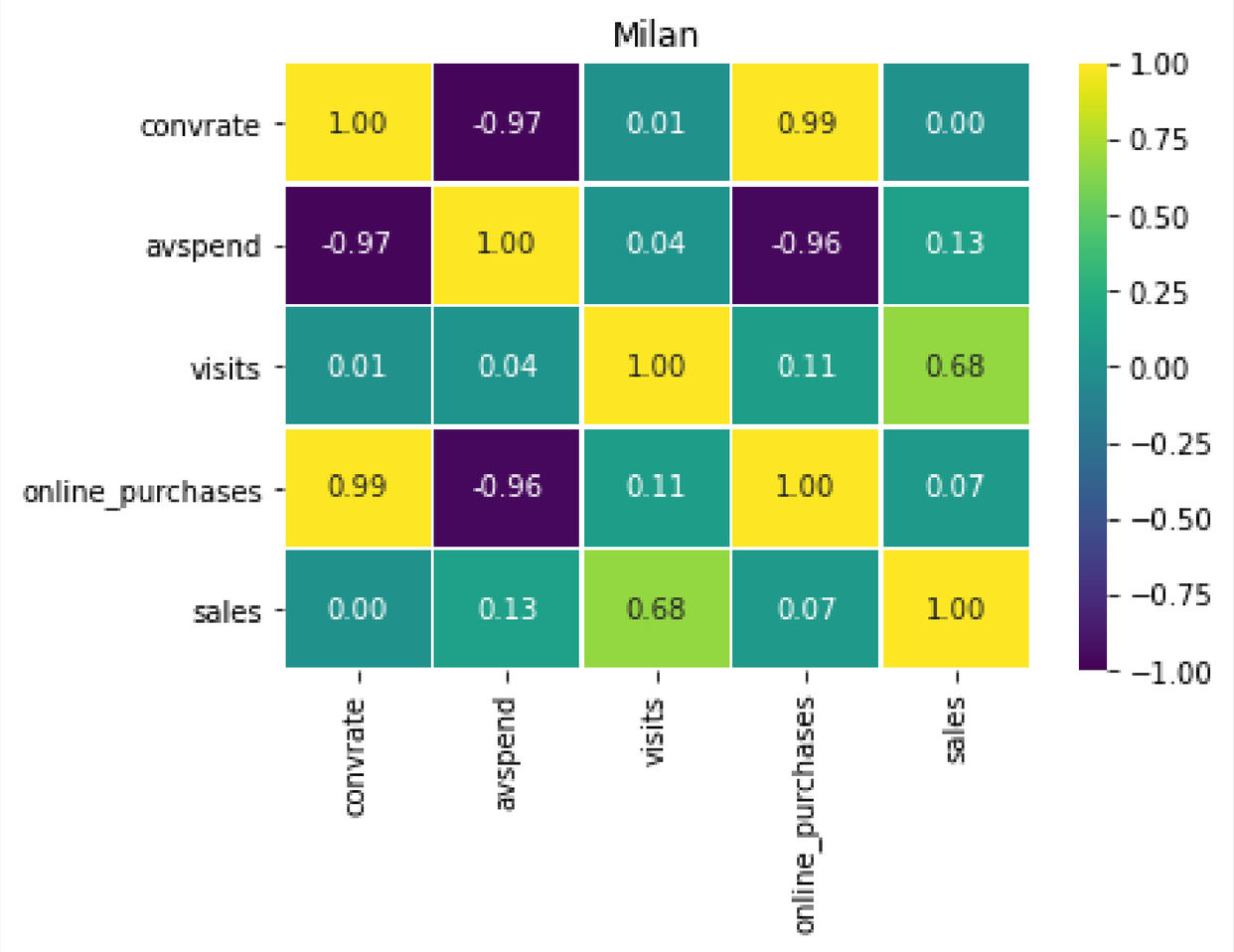
TOT_ON_PURC x AV_SPEND

CORRELATION MATRIX

LUISS

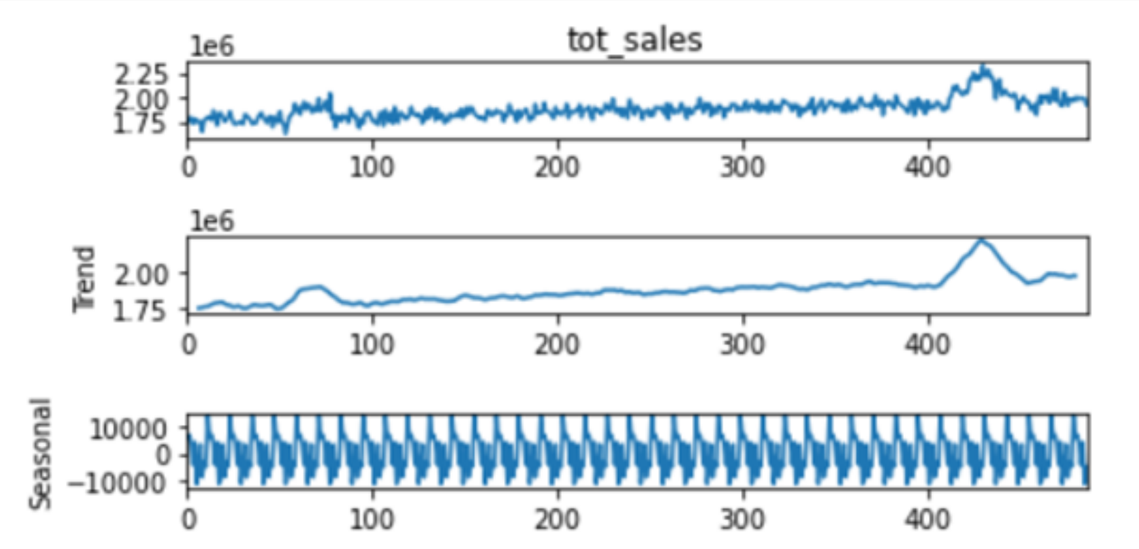
DATA MANAGEMENT

CORRELATION MATRIX

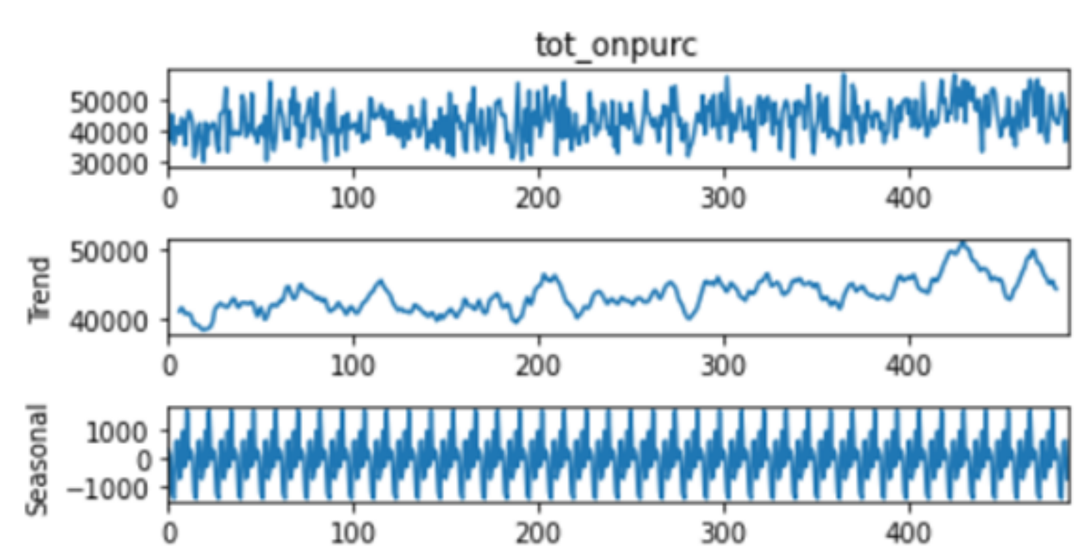


DATA MANAGEMENT

SEASONAL DECOMPOSE



Trend and Seasonality



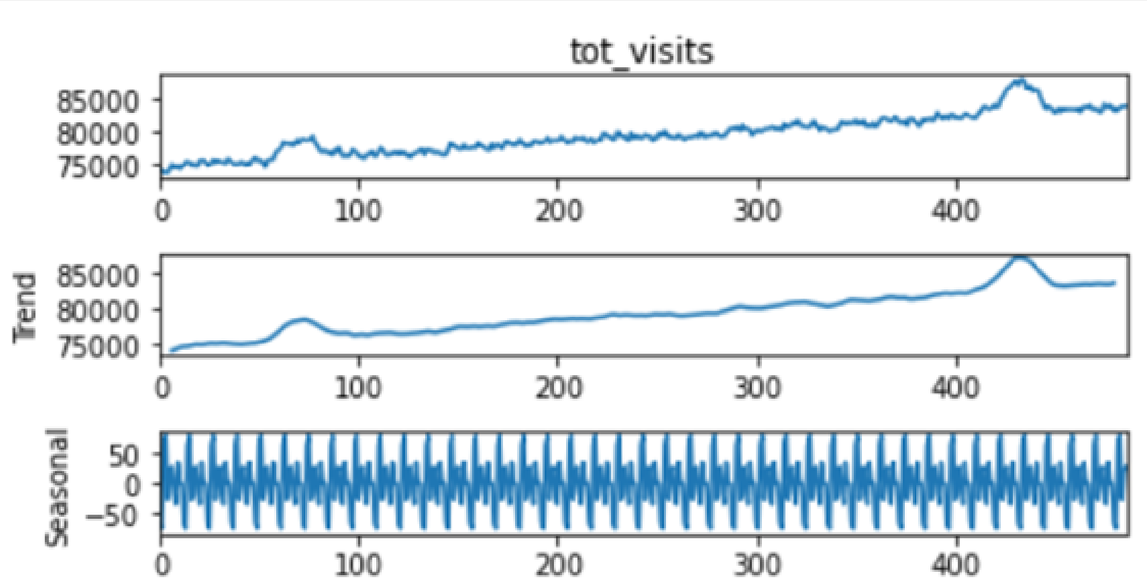
Seasonality

LUISS

AUGMENTED DICKEY-FULLER TEST

H0: the series is stationary

H1: the series is non-stationary



Trend and Seasonality

Sales	non-stationary
Visits	non-stationary

Online Purchases	stationary

Differencing

DATA MANAGEMENT

GARCH MODEL

- Objective: check if there were any significant months or days among the time series
- The model is used in time series data when the variance error is believed to be serially correlated
- Garch model takes into account the moving average component
- Cretation of dummy variables for months and days of the week

LUISS

DATA MANAGEMENT

FOUNDINGS 1

Visits

Months	December	**

Negative Coefficient:

Days	Monday	***
	Tuesday	***
	Friday	**

Positive Coefficient:

Days	Saturday	*

Online Purchases

Months	no significant values

Negative Coefficient:

Days	Monday	**

Positive Coefficient:

Days	Saturday

LUISS

DATA MANAGEMENT

FOUNDINGS 2

Sales

Months	no significant values, but highest positive coefficient is December		**Days**	Monday	**

	coefficiente	errore std.	z	p-value	
day_of_week_Fr~	-83,0387	35,1751	-2,361	0,0182	**
day_of_week_Mo~	-161,720	38,6292	-4,186	2,83e-05	***
day_of_week_Sa~	69,7873	36,7439	1,899	0,0575	*
day_of_week_Su~	552,668	37,0212	14,93	2,15e-050	***
day_of_week_Th~	-22,8286	40,1989	-0,5679	0,5701	
day_of_week_Tu~	-154,545	38,8788	-3,975	7,04e-05	***
day_of_week_We~	-48,2477	39,1856	-1,231	0,2182	

Results from visits

LUISS

MODELLING APPROACH

LINEAR REGRESSION

ADDITIONAL FEATURES:

- Presence or not of the campaign
- Dummy variables about day of the week and months of the year

Features Selection

Trade off between interpretability and accuracy

```
Call:
lm(formula = sales ~ visits + Si.no + day_of_week_Sunday + month_December,
    data = df_final)

Residuals:
    Min      1Q  Median      3Q     Max
-610113 -182373  -32684  166064 1032498

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       -1.025e+06  3.276e+05  -3.130 0.001857 **
visits             8.189e+01  4.141e+00  19.778  < 2e-16 ***
Si.no              2.513e+05  7.058e+04   3.561 0.000406 ***
day_of_week_Sunday -6.873e+04  3.199e+04  -2.149 0.032161 *
month_December      1.430e+05  4.065e+04   3.519 0.000475 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 245700 on 482 degrees of freedom
Multiple R-squared:  0.5891,    Adjusted R-squared:  0.5857
F-statistic: 172.8 on 4 and 482 DF,  p-value: < 2.2e-16
```

LUISS

MODELLING APPROACH

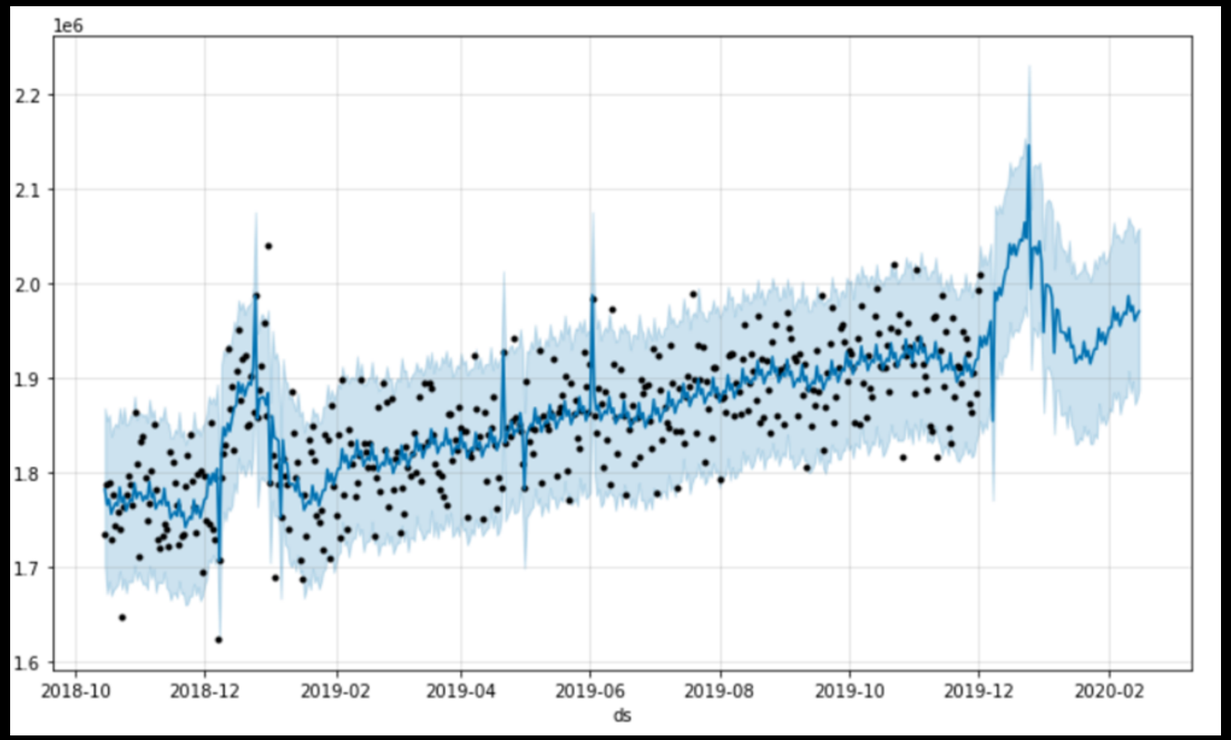
PROPHET ALGORITHM

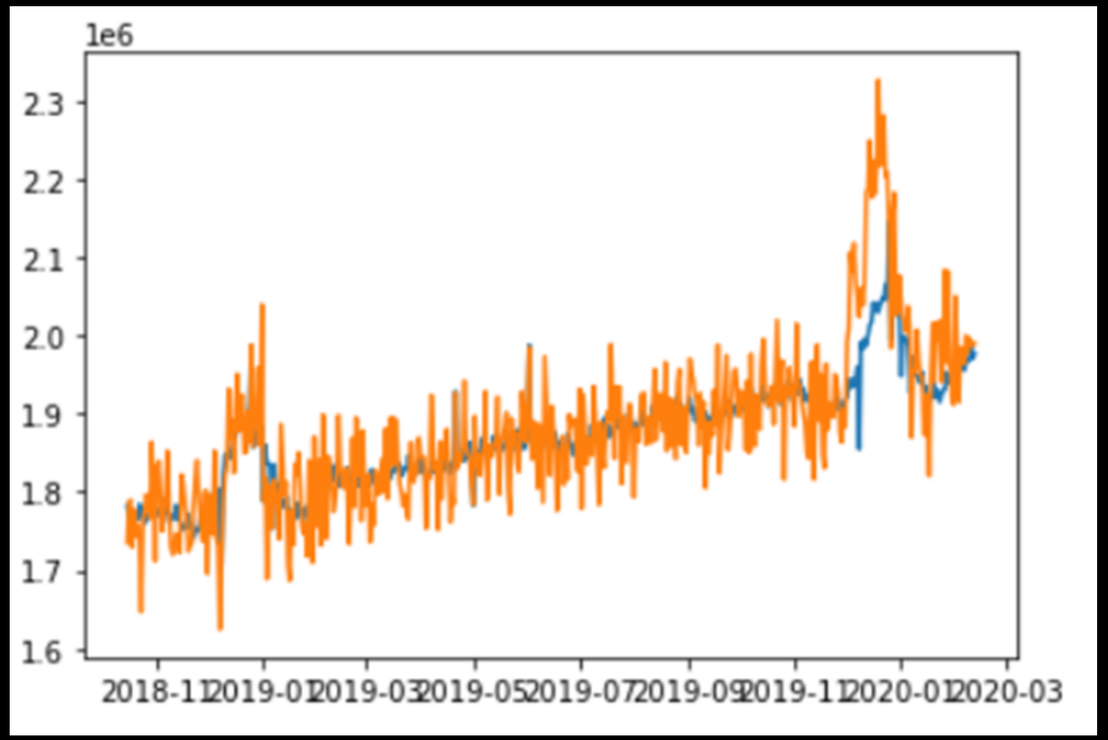
- What would have happened without the campaing?
- Prophet is an algorithm developed by Facebook that takes into account the seasonality and the trend
- Data before the campaign were considered
- For more accurate predictions we implemented the model with all the italian festivities (Christmans, New Years, Easter…)

LUISS

MODELLING APPROACH

RESULTS ON SALES





There is an increase of 3.8% between the forecasted values and the actual one

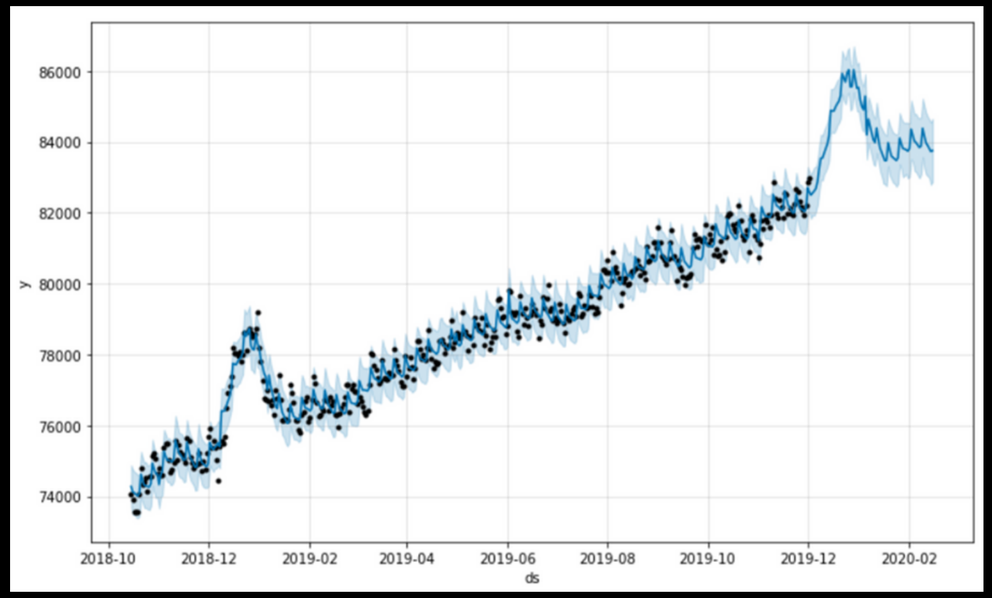
Blue line: predicted values

Orange line: actual values

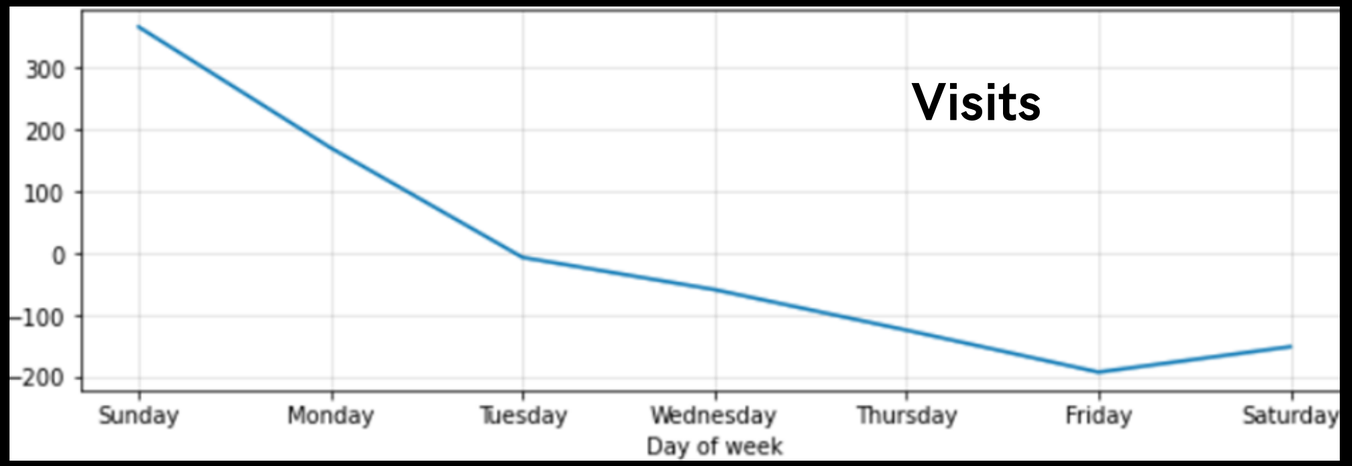
LUISS

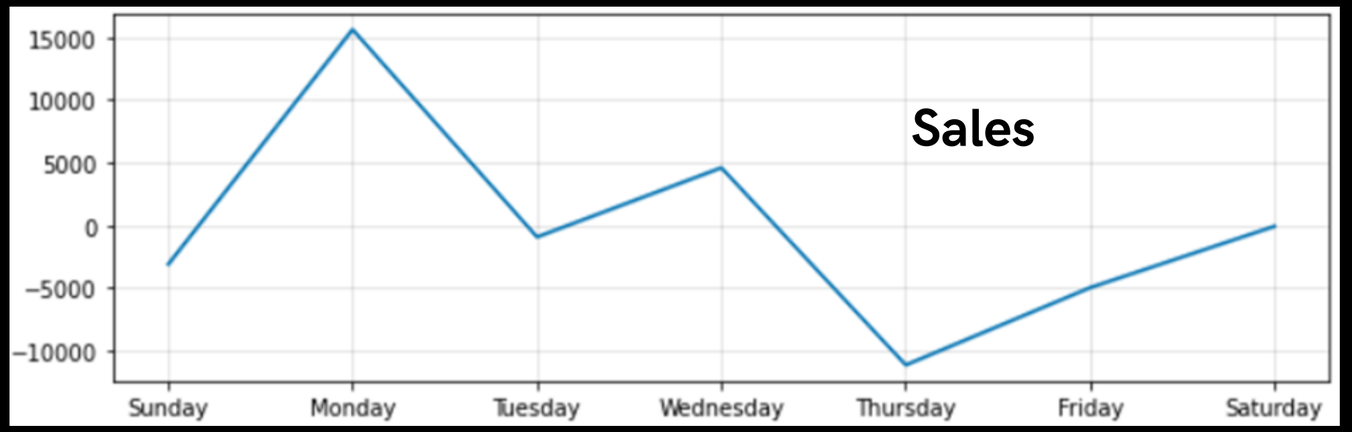
MODELLING APPROACH

RESULTS ON VISITS



WEEKLY DECOMPOSITION





LUISS

MODELLING APPROACH

FOUNDINGS ON GARCH

Visits

Months	December	**

Negative Coefficient:

Days	Monday	***
	Tuesday	***
	Friday	**

Positive Coefficient:

Days	Saturday	*

Sales

Months	no significant values, but highest positive coefficient is December

Days	Monday	**

LUISS

MODELLING APPROACH

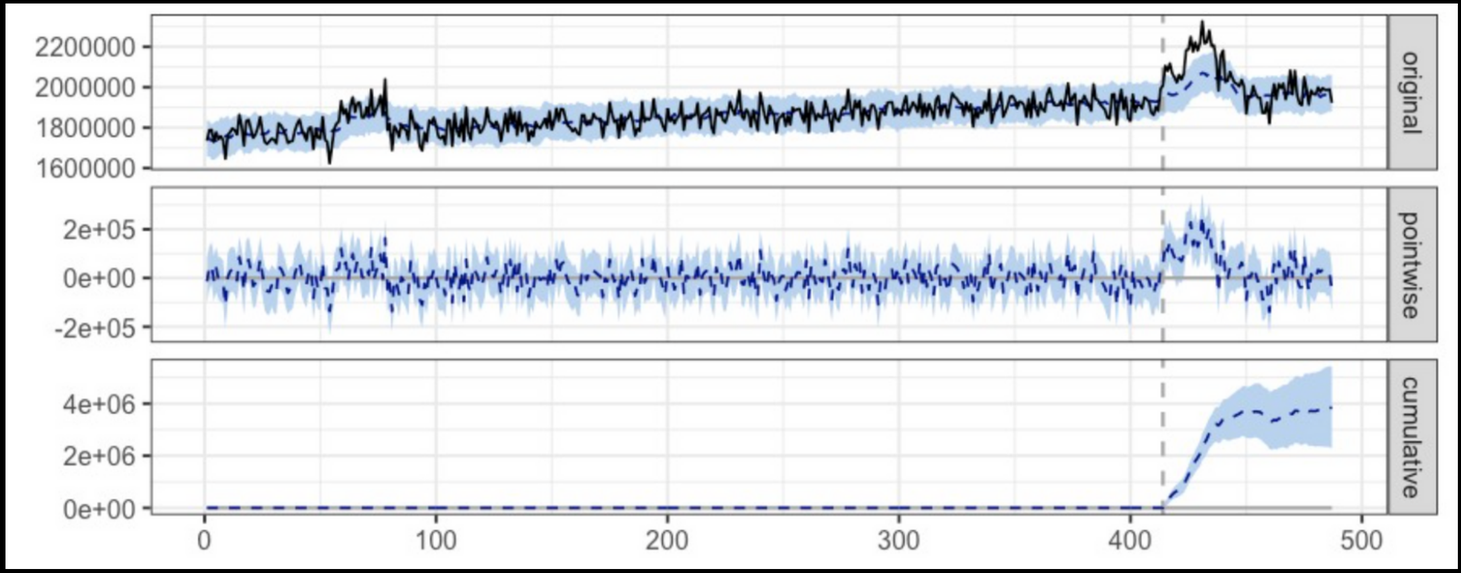
CAUSAL IMPACT

Variables not affected by the campaign:

- Google queries of 'Sephora'
- Google queries of 'Douglas'

Variables affected by the campaign:

- Total Visits
- Total Sales
- Google queries on the company's product (assumption)



LUISS

MODELLING APPROACH

CAUSAL IMPACT RESULTS

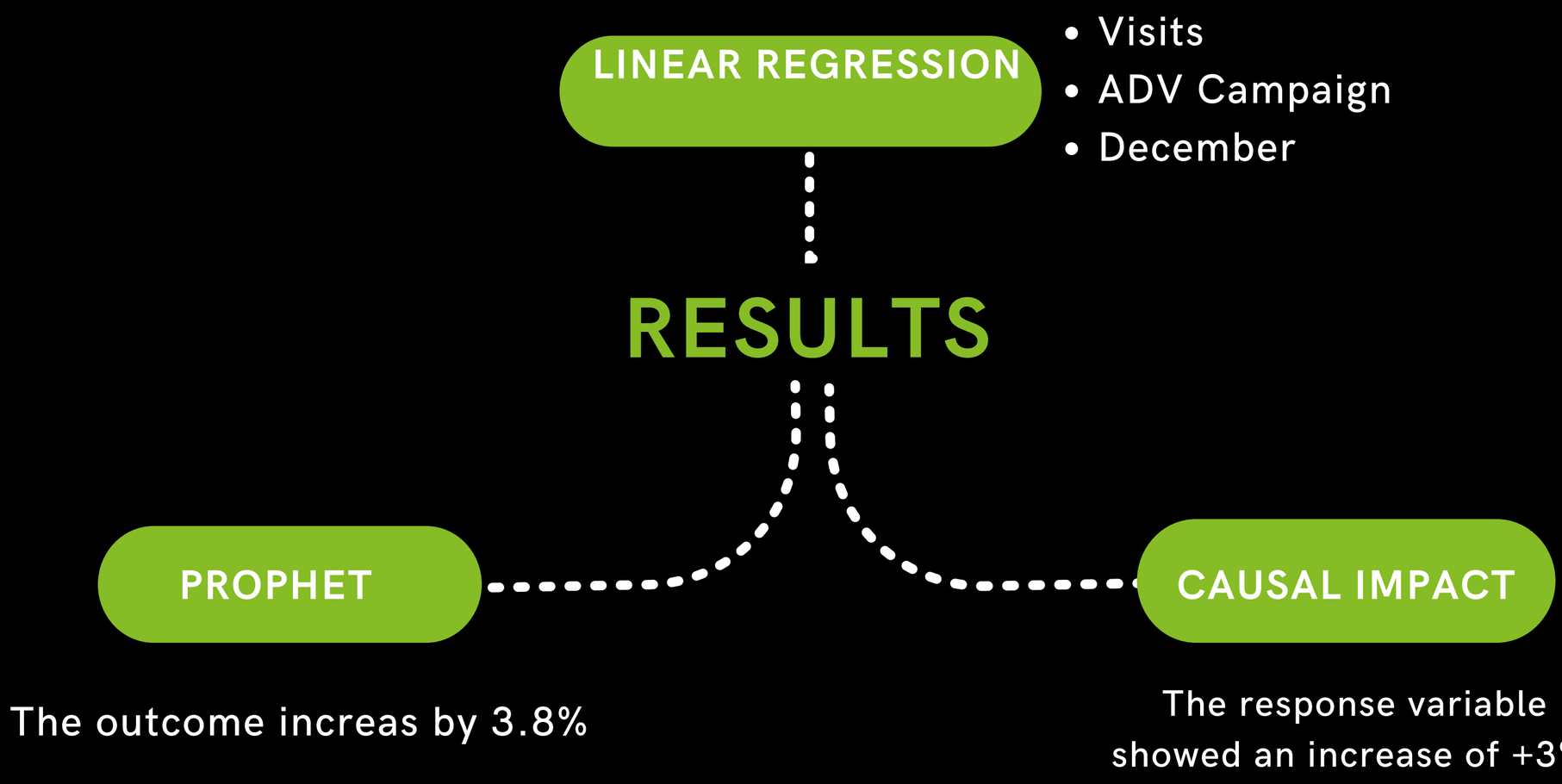
IN RELATIVE TERMS:

- The response variable showed an increase of 3%
- The 95% confidence interval of this percentage is [+2%, +4%].

This means that the positive effect (causal effect) observed during the intervention period is statistically significant and unlikely to be due to random fluctuations. Indeed, the probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability p = 0.001).

However, whether this increase also bears substantive significance can only be answered by comparing this effect to the original goal of the advertising campaign.

LUISS


LINEAR REGRESSION
• Visits
• ADV Campaign
• December
RESULTS
PROPHET
CAUSAL IMPACT
The outcome increas by 3.8%
The response variable
showed an increase of +3%

LUISS

CONCLUSIONS

HOW TO IMPROVE THIS ANALYSIS?

- **ROI CALCULATION**
  - Knowing how much was invested in the campaign can give us a final understanding if it was worth it.
- **FORECASTING MODELS**
  - With more historical data, especially, data not affected by the campaign our models could better predict the effect.
- **CAMPAIGN DETAILS**
  - Knowing more details about the campaign could further enrich our analysis

LUISS

TEAM

GitHub

Olimpia Sannucci

Martina Manno

Carlo Ardito

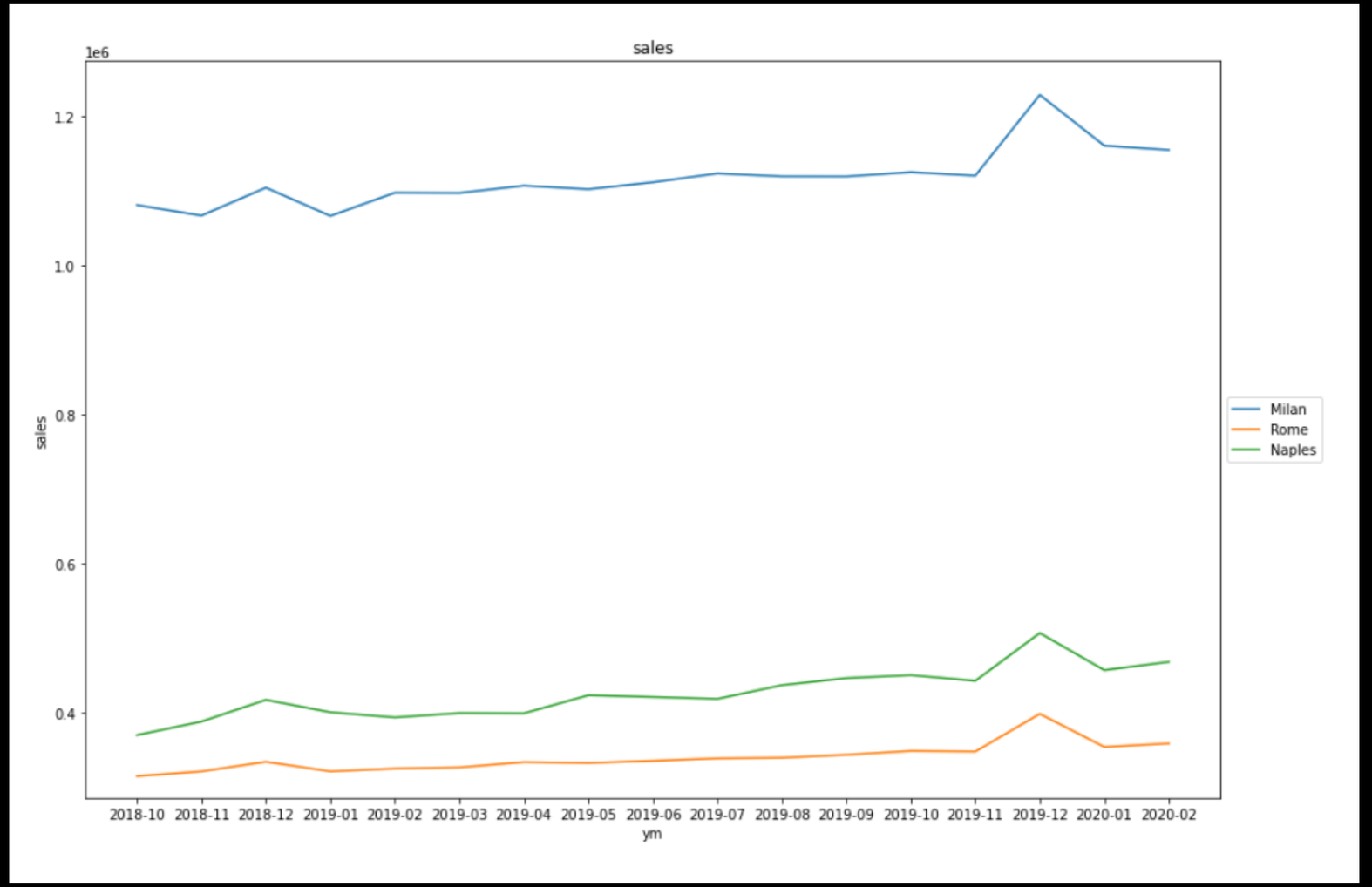
LUISS

REFERENCES

- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani ”An Introduction to Statistical Learning with Applications in R”, Springer.
- Hal R. Varian, ”Causal inference in economics and marketing”.
- James H. Stock, Mark W. Watson, ” Introduzione all’econometria”,Pearson.
- Roger A. Kerin, Steven W. Hartley, Luca Pellegrini, Francesco Massara, Daniela Corsaro ”Marketing”, McGraw Hill.
- Prophet Algorithm: https://facebook.github.io/prophet/
- Causal Impact Algorithm: https://google.github.io/CausalImpact/CausalImpact.html

LUISS

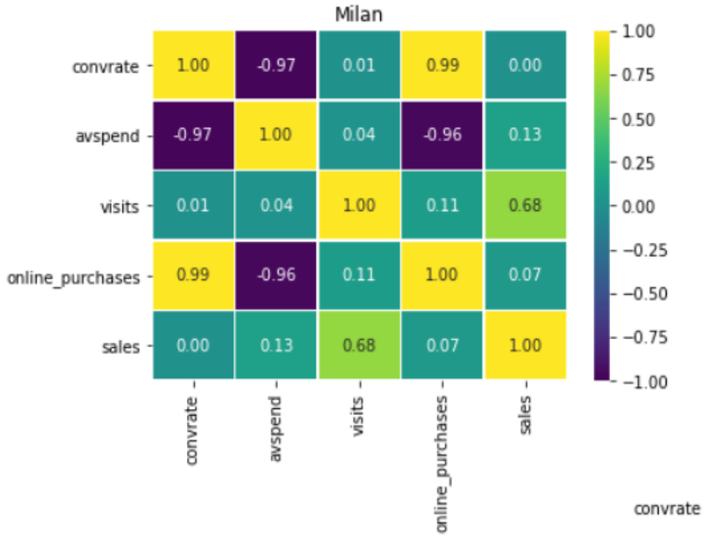
APPENDIX

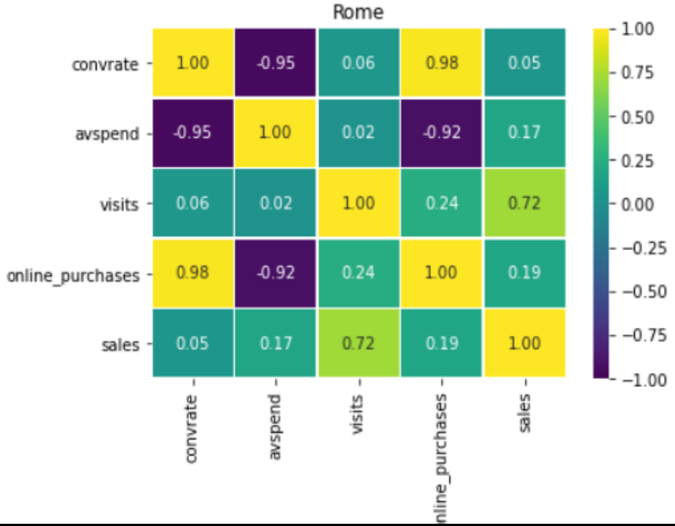


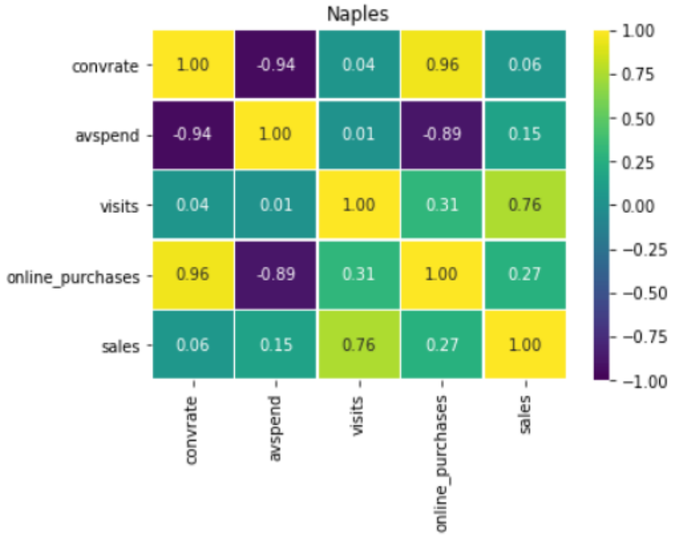
LUISS

APPENDIX

CORRELATION matrix







LUISS