
Analisi trend di guadagni di ristoranti

Lorenzo Lecce (n. matricola 830881) e Andrea Lucini Paioni (n. matricola 826578)

Abstract

Il seguente progetto si focalizza sull'analisi delle dinamiche di vendita di sei ristoranti, situati nelle due regioni di Lombardia ed Emilia-Romagna. L'analisi prende in considerazione i dati di sei ristoranti in un periodo di osservazione che si estende dal gennaio 2018 all'aprile 2022. Ad una prima fase di import, pulizia ed elaborazione delle informazioni ottenute a partire dai due dataset inizialmente forniti, segue una fase di integrazione dei dati con dataset contenenti informazioni inerenti al fenomeno in studio ed una fase di esplorazione dei dati.

La seconda parte del progetto riguarda la creazione di modelli e serie storiche che permettano di rappresentare in modo fedele l'andamento dei guadagni dei sei ristoranti di interesse, col fine di individuare eventuali pattern stagionali, trend significativi ed insight che possano aiutare nella stima delle perdite dovute all'avvento del Covid 19 e ai conseguenti periodi di lockdown e limitazioni.

Parole chiave: Data Science - Analisi Esplorativa - Serie Temporal - Python

Indice

Introduzione	2
Scopo del progetto	2
Import e preprocessing	2
I dati originali	2
Integrazione dati	3
Fase di preprocessing	3
Analisi esplorativa	4
Analisi esplorativa singoli ristoranti	11
Ristorante 1	11
Ristorante 2	12
Ristorante 3	14
Ristorante 4	15
Ristorante 5	17
Ristorante 6	18
Modelli e serie temporali	20
Conclusione ed eventuali sviluppi futuri	23
Indice figure	24
Indice tabelle	25
Riferimenti	26

Introduzione

All'interno della cultura italiana, la cucina e la ristorazione giocano un ruolo di fondamentale importanza, tanto da essere considerate un'autentica espressione dell'identità nazionale italiana. La cucina italiana è tra le più rinomate al mondo per la sua ricchezza, per la particolare cura nelle preparazioni e per la qualità e varietà di ingredienti di prima classe. E l'importanza della cucina italiana non risiede solamente nella preparazione dei piatti, bensì anche nei momenti dei pasti visti come occasione di condivisione e convivialità tra amici e famiglie.

In questo contesto, è subito chiaro quanto sia stato enorme l'impatto che la pandemia da COVID-19 ha avuto sulla cucina e sulla ristorazione italiana: il settore è stato profondamente colpito dalle restrizioni e dalle chiusure, con numerosi ristoranti e aziende alimentari che hanno lottato per sopravvivere. Il concetto di slow food, di ristorante come luogo di incontro tra conoscenti e momento di condivisione andava in forte contrasto con le esigenze cliniche della tragica situazione della pandemia. Tuttavia, questa crisi ha portato ad un'innovazione del settore, con nuove formule fino a qualche anno fa pressoché inesistenti in Italia (se non in qualche realtà, come le metropoli), come la consegna a domicilio e la possibilità di ordinare un pasto online tramite siti e applicazioni.

Scopo del progetto

L'obiettivo di questo progetto consiste nell'analisi dei dati relativi ai 6 ristoranti di interesse, in modo da poter stabilire l'impatto che la pandemia da COVID-19 ha avuto sui guadagni di queste attività, e inoltre la creazione di modelli e serie temporali che possano dare una prospettiva dell'andamento futuro delle vendite nel settore.

Possiamo pensare il seguente progetto come diviso in tre fasi principali:

- Una fase di import dei dati già forniti, e una fase di integrazione dei dati con altre informazioni potenzialmente utili, su cui è stata inoltre compiuta una fase di preprocessing (col fine di rendere i dati maggiormente utilizzabili per i nostri scopi).
- Una fase di esplorazione dei dati completi, in modo tale da avere un quadro sia generale per l'intero dataset, sia un focus sulle singole attività, così da arrivare all'ultima fase del progetto con tutte le possibili informazioni a disposizione.
- Una fase di creazione di serie temporali e modelli di previsione, per utilizzare i dati raccolti e analizzati nelle prime due fasi con fine predittivo.

Import e preprocessing

La prima fase del progetto riguarda operazioni di import del dataset in Python, strumento utilizzato in questa prima fase, col fine di rendere i dati leggibili, puliti e facilmente integrabili con ulteriori informazioni online; al termine di queste operazioni, l'obiettivo è di avere un dataset pulito, completo, informativo e pronto ad essere analizzato nella fase di esplorazione.

I dati originali

Vengono forniti inizialmente due dataset da cui ottenere le informazioni fondamentali.

Il dataset `df_ristorazione`, estratto dal file `.csv Serie-storiche-ristoranti_def.csv`, è formato da 8817 osservazioni e 4 colonne. Il dataset contiene le seguenti features:

- `id_ristorante`: indica a quale ristorante si riferiscono i dati dell'osservazione;
- `data`: indica la data a cui si riferiscono tutte le informazioni contenute in ogni osservazione;
- `scontrini`: indica il numero di scontrini emessi dal ristorante di quell'osservazione in quel determinato giorno;
- `lordototale`: indica il guadagno lordo totale ricavato dagli scontrini emessi dal ristorante d'interesse in quella giornata.

Il secondo dataset si chiama `df_info_ristoranti`, ed è estratto dal file `.csv RistorantiCitta.csv`; è formato da solamente 6 osservazioni e 2 colonne, che definiscono le città in cui si trovano i differenti ristoranti.

Integrazione dati

Il dataset finale è frutto dell'integrazione dei due dataset forniti ad inizio progetto, oltre ad una serie di informazioni ottenute da fonti di dati trovate online. Contiene le seguenti features, formato da un totale di 8811 osservazioni:

- *Codice*: variabile in formato stringa che indica il ristorante tra i sei presenti dei dataset;
- *Location*: la città in cui si trova il ristorante;
- *Provincia*: la provincia dove si trova il ristorante;
- *Regione*: la regione dove si trova il ristorante;
- *Data*: data delle informazioni dell'osservazione, in formato yyyy-mm-dd;
- *Scontrini*: numero intero che indica il numero di scontrini emessi in quel giorno dal ristorante;
- *Lordo_totale*: numero decimale che indica il guadagno lordo del ristorante in quel determinato giorno;
- *Giorno*: giorno della settimana;
- *Mese*: mese dell'anno;
- *Anno*: anno;
- *Stagione*: stagione dell'anno;
- *Festività*: variabile categorica, che indica se si tratta di un giorno festivo, e nel caso di quale festività si tratta (tra *Capodanno*, *Epifania*, *Liberazione*, *Lavoro*, *Repubblica*, *Ferragosto*, *Ognissanti*, *Immacolata*, *Natale*, *Santo Stefano*, *Pasqua*, o *No* in caso di giorno non di festività);
- *Festivo_feriale*: variabile categorica che indica, con due possibili outcome (*Festivo* o *Feriale*), se si tratta di giorno festivo o feriale;
- *Covid*: variabile categorica che indica, con due possibili outcome (*Pre-Covid19* o *Post-Covid19*), se si tratta di un giorno prima dell'inizio della pandemia di Covid-19 in Italia;
- *Colore*: variabile categorica che indica il colore delle restrizioni Covid applicate in quella provincia in quel giorno (*bianco*, *rosso*, *arancione*, *giallo*, o *Pre-Covid19* se nel periodo prima dell'inizio del Covid in Italia);
- *Inflazione*: livello dell'inflazione generale nel mese di interesse (valori positivi o negativi);
- *Inflazione_cibo*: livello dell'inflazione nei prezzi del cibo nel mese di interesse (valori positivi o negativi);
- *Inflazione_combustibili*: livello dell'inflazione nei prezzi dei carburanti nel mese di interesse (valori positivi o negativi);
- *Maschi*: numero di vaccinazioni tra i maschi in quella data;
- *Femmine*: numero di vaccinazioni tra le femmine in quella data;
- *Prima_dose*: numero di vaccinazioni come prima dose in quella data;
- *Seconda_dose*: numero di vaccinazioni come seconda dose in quella data;
- *Contagio_precedente*: numero di vaccinazioni come ulteriore dose dopo un contagio in quella data;
- *Dose_Booster*: numero di vaccinazioni come dose booster in quella data;
- *Seconda_dose_Booster*: numero di vaccinazioni come seconda dose booster in quella data;
- *Totale_casi*: totale di nuovi positivi in quel giorno nella provincia di interesse.
- *Scontrino_Medio*: valore medio dello scontrino nella giornata di interesse per il ristorante d'interesse.

Fase di preprocessing

Una prima fase di preprocessing è necessaria per rendere il dataset, formato dall'unione dei due dataset originali, leggibile ed utilizzabile: infatti, il formato delle date dev'essere reso utilizzabile (viene fornito inizialmente come un intero, e non in formato data con separatori tra giorni, mesi e anni).

Inoltre, vengono completate una serie di operazioni per rendere tutte le features ugualmente leggibili come formato:

- È stata completata la colonna relativa al colore della zona della regione (denominata *Colore*) per completare i dati (erano indicate informazioni relative al periodo compreso tra il 06/11/2020 e il 30/04/2022, dunque è stato completato il resto del dataset);
- C'è stata una fase di imputazione di alcuni valori errati per le variabili *Scontrini* e *Lordo_Totale* (sono state escluse dall'analisi 6 osservazioni che avevano valori *NaN* per una o addirittura entrambe le features), in quanto in alcune osservazioni avevano valori corretti per il numero di scontrini ma indicavano valore pari a 0 come *Lordo_Totale*, e viceversa altre osservazioni avevano valori corretti per il lordo totale ma indicavano 0 scontrini per la giornata: per i dati con 0 scontrini e lordo totale maggiore di 0, il numero di scontrini è stato

arrotondato all'intero più vicino con divisione $\frac{Scontrino_medio}{Lordo_totale}$, mentre per i dati con 0 lordo totale e numero di scontrini maggiore di 0, il calcolo deriva dalla moltiplicazione $Scontrini * Scontrino_medio$.

- Alcuni formati delle colonne sono cambiati (la variabile *Scontrini* era indicata come numero decimale), mentre per altre colonne (come il *Lordo_Totale*) tutti i valori sono stati arrotondati a 2 cifre decimali.
- Sono stati tolti una serie di dati prima delle fasi di analisi esplorativa e di creazione dei modelli, che possono essere considerati outliers, in quanto contenenti i dati di *Scontrini* e *Lordo_Totale* relativi al mese intero (infatti, sono 8 osservazioni per ogni ristorante, tutte coincidenti col primo giorno del mese e seguite dalle osservazioni dei giorni successivi contenenti valori di *Scontrini* e *Lordo_Totale* pari a 0): essendo comunque all'inizio del dataset, dunque non nel periodo di interesse vero e proprio della nostra analisi (ovvero il periodo a cavallo della pandemia da Covid-19), abbiamo deciso di escludere i primi otto mesi dell'osservazione, dunque considerando a partire dal 2018-09-01.

Analisi esplorativa

Per la fase di analisi esplorativa, sono stati analizzati sia tutti i ristoranti nel loro complesso, per cercare eventuali elementi in comune, sia i singoli ristoranti per focalizzarci meglio su alcuni aspetti delle singole attività.

Prima di tutto, è bene notare che il numero di osservazioni a disposizione è differente da un ristorante all'altro: infatti, il ristorante col codice R002 (situato a Piacenza) ha un numero sensibilmente inferiore di osservazioni, con solamente 912 righe rispetto alle circa 1300 degli altri ristoranti (in particolare, il ristorante R000, R001, R004 e R005 con 1338 osservazioni e il ristorante R003 con 1337). In totale, il dataset conta, rispetto alle fasi iniziali, 27 features e 7601 osservazioni.

Ora osserviamo i boxplot relativi al numero totale di scontrini nei singoli ristoranti (variabile *Scontrini*), e al guadagno totale per giornata dei singoli ristoranti (variabile *Lordo_Totale*):

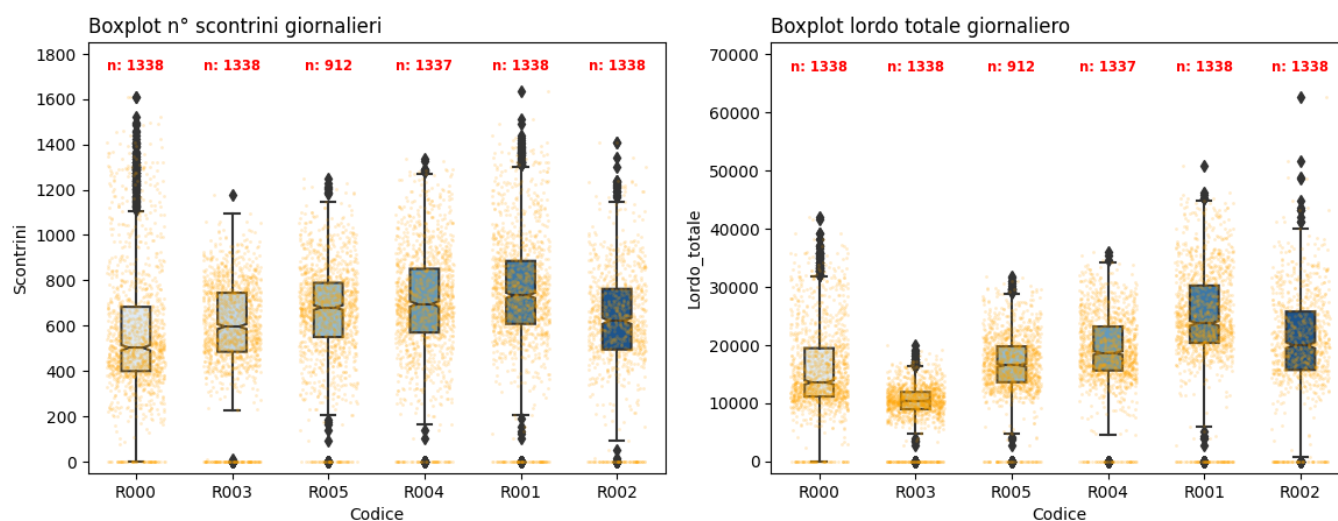


Figura 1: boxplot variabili *Scontrini* e *Lordo_Totale* per ristorante

CODICE	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
R000	1338.0	563.82	296.44	0.0	399.25	502.5	683.0	1611.0
R001	1338.0	733.08	288.12	0.0	608.0	733.5	885.75	1633.0
R002	912.0	608.88	262.93	0.0	494.75	621.0	762.25	1409.0
R003	1337.0	582.79	231.01	0.0	485.0	596.0	745.0	1175.0
R004	1338.0	693.28	258.24	0.0	569.75	695.5	851.0	1335.0
R005	1338.0	654.99	229.17	0.0	550.0	680.0	789.0	1247.0

Tabella 1: statistiche variabile *Scontrini* per ristorante

CODICE	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
R000	1338.0	15533.11	7580.49	0.0	11138.11	13562.32	19440.3	41895.55
R001	1338.0	24358.23	9224.63	0.0	20350.2	23723.25	30208.83	50843.73
R002	912.0	20161.37	9078.93	0.0	15704.09	19937.25	25738.22	62623.4
R003	1337.0	9911.95	3669.78	0.0	8953.42	10365.33	11898.12	19997.85
R004	1338.0	18975.15	6584.44	0.0	15610.86	18611.26	23165.69	35933.65
R005	1338.0	16541.88	5608.85	0.0	13600.05	16504.28	19753.64	31699.78

Tabella 2: statistiche variabile Lordo_Totale per ristorante

Sia dai boxplot che dalle tabelle di statistiche descrittive, è possibile osservare che il numero di scontrini si mantiene costante tra i sei ristoranti, con mediane comprese tra 500 e 750 scontrini al giorno, con distribuzioni simili tra i vari ristoranti (qualche outlier in più per il ristorante R000 e R001 con valori particolarmente elevati, con massimi di 1611 e 1633 rispettivamente).

Al contrario dei valori relativi al numero di scontrini, il lordo totale cambia in maniera significativa tra i vari ristoranti: abbiamo distribuzioni più concentrate, come per il ristorante R003 (con deviazione standard di 3669.78), mentre per gli altri ristoranti i valori sono molto più sparsi, con picchi di 9078.93 e 9224.63 per la deviazione standard rispettivamente del ristorante R002 e R001. In generale, il ristorante R001 è quello con la media del lordo totale maggiore di tutti, pari a 24358.23, e al contrario il ristorante R003 è quello con media del lordo totale minore (9911.95).

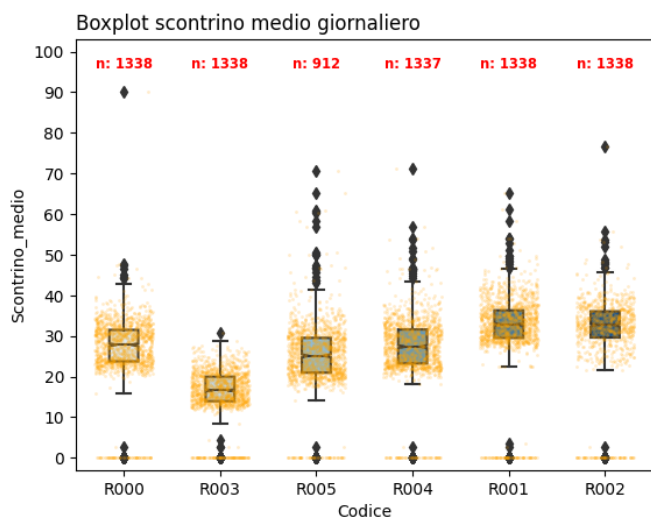


Figura 2: boxplot variabile Scontrino_Medio per ristorante

CODICE	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
R000	1338.0	27.11	9.8	0.0	23.73	27.82	31.48	243.18
R001	1338.0	31.72	9.35	0.0	29.51	32.8	36.31	65.3
R002	912.0	30.76	10.16	0.0	29.6	32.72	36.04	76.74
R003	1337.0	16.14	5.87	0.0	13.93	16.6	19.97	30.87
R004	1338.0	27.02	7.85	0.0	23.22	27.34	31.59	71.08
R005	1338.0	24.97	7.94	0.0	20.96	25.1	29.52	70.51

Tabella 3: statistiche variabile Scontrino_Medio per ristorante

Verificando ora i valori dello scontrino medio dei differenti ristoranti, notiamo che il ristorante R003 è quello meno costoso, con uno scontrino che mediamente fa incassare 16.14€; al contrario, gli altri valori sono tutti di gran lunga superiori, con picco di 30.76 e 31.72 di media per i due ristoranti R002 e R001.

In generale, gli scontrini del ristorante R003 hanno valori molto meno variabili, con deviazione standard pari a 5.87, mentre per gli altri ristoranti gli scontrini presentano valori molto più sparsi, essendo tutti compresi tra valori della deviazione standard di 7.85 e di 10.16.

Verifichiamo ora se ci sono differenze particolari a livello di guadagni dei ristoranti in base al giorno della settimana:

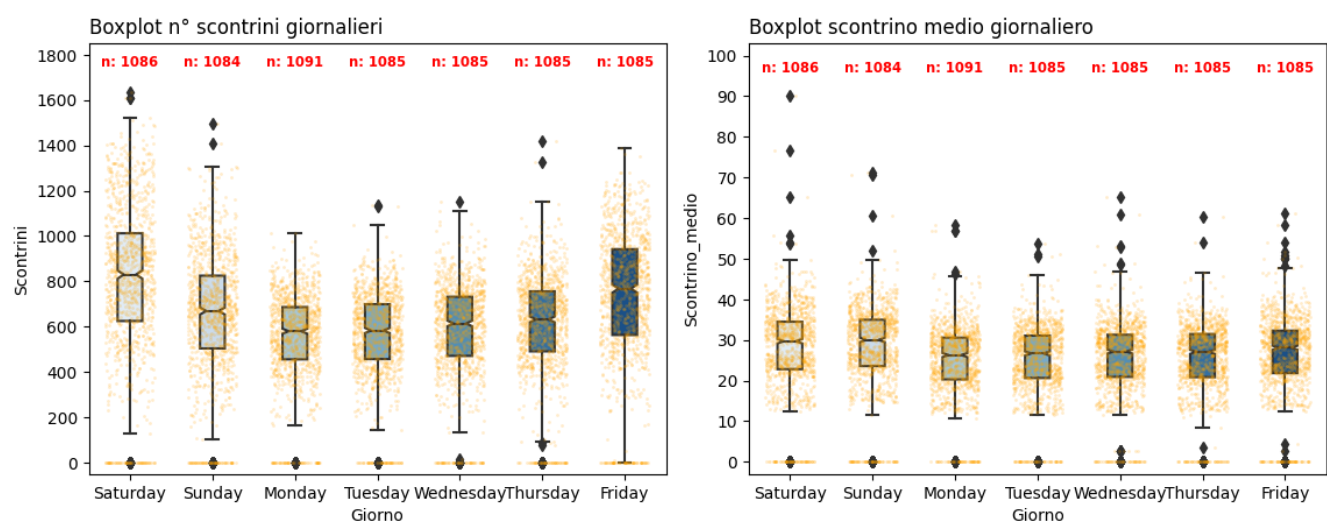


Figura 3: boxplot variabili Scontrini e Scontrino_Medido per ristorante

GIORNO	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
FRIDAY	1086.0	728.51	295.17	0.0	563.25	768.0	941.75	1387.0
MONDAY	1084.0	553.08	198.6	0.0	455.75	580.0	686.25	1015.0
SATURDAY	1091.0	806.01	340.03	0.0	625.0	830.0	1012.0	1633.0
SUNDAY	1085.0	651.79	275.53	0.0	504.0	669.0	825.0	1496.0
THURSDAY	1085.0	605.9	225.06	0.0	491.0	632.0	756.0	1421.0
TUESDAY	1085.0	559.47	198.92	0.0	457.0	583.0	699.0	1134.0
WEDNESDAY	1085.0	582.52	211.91	0.0	471.0	614.0	732.0	1153.0

Tabella 4: statistiche variabile Scontrino per giorno

GIORNO	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
FRIDAY	1086.0	26.15	9.85	0.0	21.83	28.08	32.28	61.02
MONDAY	1084.0	24.64	8.88	0.0	20.28	26.26	30.56	58.42
SATURDAY	1091.0	27.69	10.67	0.0	22.8	29.68	34.47	90.01
SUNDAY	1085.0	28.28	12.2	0.0	23.56	30.0	34.96	243.18
THURSDAY	1085.0	25.21	9.0	0.0	20.84	27.06	31.43	60.22
TUESDAY	1085.0	24.96	8.78	0.0	20.63	26.67	31.05	53.63
WEDNESDAY	1085.0	25.34	9.12	0.0	20.9	27.14	31.26	65.31

Tabella 5: statistiche variabile Scontrino_Medio per giorno

A livello di scontrino medio, notiamo una piccola differenza in generale, con una spesa lievemente più elevata nel weekend (media di 27.69 e di 28.28 il sabato e la domenica, e in parte un valore di 26.15 al venerdì, mentre nel resto della settimana ci si mantiene su medie comprese tra il 24.64 del lunedì e il 25.34 del mercoledì), come indicato nella tabella 5.

La situazione cambia come numero di scontrini: come prevedibile, il numero di scontrini aumenta in modo significativo nel fine settimana, con una media di 806 per il sabato e di 651 per la domenica; anche il venerdì ha valori elevati, con una media di 728 scontrini, mentre per gli altri giorni della settimana non si supera mai i 605 scontrini medi del giovedì.

Cerchiamo ora di distinguere, all'interno dei giorni della settimana, i giorni festivi con quelli feriali (considerando le feste programmate, ad esempio):

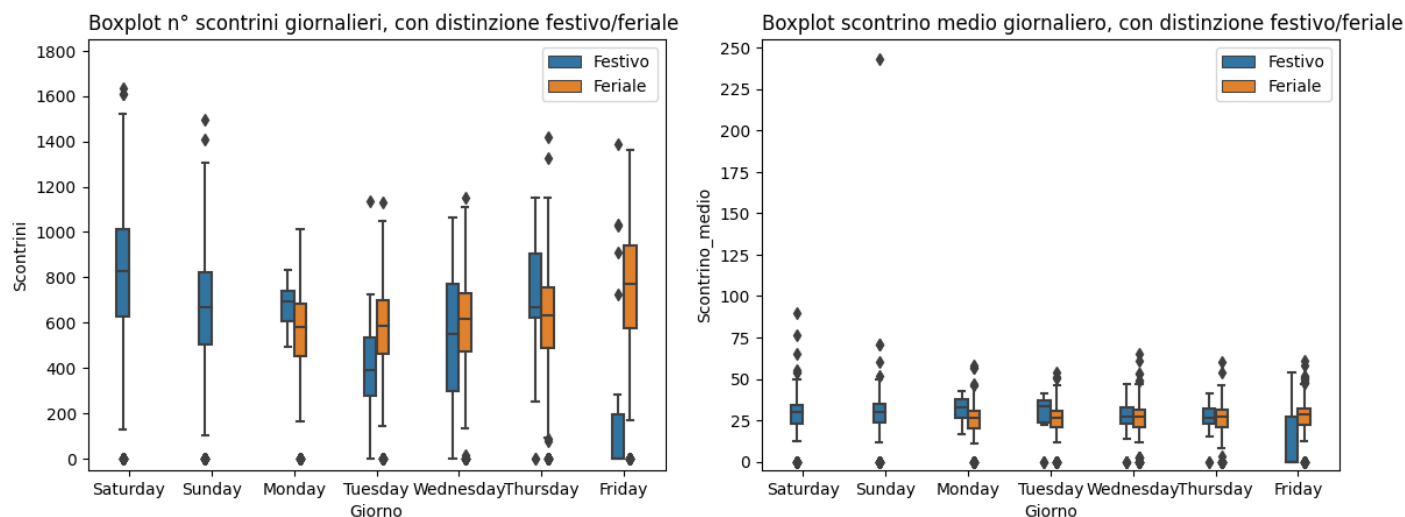


Figura 4: boxplot variabili Scontrini e Scontrino_Medido per giorno, con distinzione festivo/feriale

GIORNO	FESTIVO FERIALE	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
FRIDAY	Feriale	1062.0	739.62	282.1	0.0	575.25	772.0	942.0	1364.0
FRIDAY	Festivo	24.0	237.17	425.92	0.0	0.0	0.0	193.75	1387.0
MONDAY	Feriale	1067.0	551.01	199.19	0.0	453.0	579.0	684.5	1015.0
MONDAY	Festivo	17.0	682.82	89.98	493.0	609.0	695.0	741.0	832.0
SATURDAY	Festivo	1091.0	806.01	340.03	0.0	625.0	830.0	1012.0	1633.0
SUNDAY	Festivo	1085.0	651.79	275.53	0.0	504.0	669.0	825.0	1496.0
THURSDAY	Feriale	1058.0	603.85	222.96	0.0	490.25	630.0	755.75	1421.0
THURSDAY	Festivo	27.0	686.15	289.7	0.0	620.5	666.0	903.5	1153.0
TUESDAY	Feriale	1063.0	562.44	196.4	0.0	461.0	586.0	700.5	1129.0
TUESDAY	Festivo	22.0	415.95	264.43	0.0	275.75	392.5	535.5	1134.0
WEDNESDAY	Feriale	1045.0	584.26	207.11	0.0	473.0	616.0	730.0	1153.0
WEDNESDAY	Festivo	40.0	537.02	312.47	0.0	296.0	549.5	770.5	1065.0

Tabella 6: statistiche variabile Scontrino per giorno, con distinzione festivo/feriale

GIORNO	FESTIVO_FERIALE	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
FRIDAY	Feriale	1062.0	26.43	9.37	0.0	21.98	28.23	32.29	61.02
FRIDAY	Festivo	24.0	13.69	18.99	0.0	0.0	0.0	27.22	53.98
MONDAY	Feriale	1067.0	24.52	8.86	0.0	20.25	26.11	30.48	58.42
MONDAY	Festivo	17.0	32.11	6.75	16.38	26.24	32.95	37.75	42.58
SATURDAY	Festivo	1091.0	27.69	10.67	0.0	22.8	29.68	34.47	90.01
SUNDAY	Festivo	1085.0	28.28	12.2	0.0	23.56	30.0	34.96	243.18
THURSDAY	Feriale	1058.0	25.2	8.98	0.0	20.79	27.11	31.42	60.22
THURSDAY	Festivo	27.0	25.62	9.74	0.0	23.0	26.47	32.06	40.91
TUESDAY	Feriale	1063.0	24.9	8.68	0.0	20.56	26.59	30.94	53.63
TUESDAY	Festivo	22.0	28.07	12.7	0.0	23.8	33.28	36.97	41.08
WEDNESDAY	Feriale	1045.0	25.32	9.0	0.0	20.89	27.16	31.24	65.31
WEDNESDAY	Festivo	40.0	25.69	11.95	0.0	22.98	27.03	32.95	46.69

Tabella 7: statistiche variabile Scontrino_Medio per giorno, con distinzione festivo/feriale

Ovviamente non abbiamo dati feriali relativi al sabato e alla domenica. Per gli altri giorni, invece, notiamo sia dai boxplot (figura 4) che dalle tabelle con statistiche descrittive (Tabella 6 e Tabella 7) che abbiamo un numero esiguo di giorni normalmente feriali ma che sono festa perchè coincidono con feste (Pasqua, Natale...), da un minimo di 17 volte il lunedì fino ad un massimo di 40 volte per il mercoledì.

In generale, il comportamento è molto variegato, con alcuni giorni in cui le feste portano i clienti ad uscire ed andare al ristorante maggiormente (il lunedì ad esempio, dove la media di scontrini passa da 551 a 682, e allo stesso modo si

comporta anche il giovedì), ed altri giorni in cui il numero di clienti cala (il martedì, il mercoledì e soprattutto il venerdì, con un calo drastico da 739 scontrini di media nei venerdì feriali a 237 di media nei venerdì festivi).

Per quanto riguarda invece lo scontrino medio, il mercoledì e il giovedì la differenza è pressoché nulla tra giorni festivi e feriali; il martedì, lo scontrino medio aumenta leggermente dal caso feriale al festivo (da 24.9€ di media a 28.07€ di media), il lunedì aumenta in maniera più significativa, da 24.52€ di media a 32.11€ di media (ma valore da tenere d'occhio, in quanto non ci sono stati lunedì festivi con 0 scontrini, caso unico nell'intero dataset dove sia per giorni festivi che feriali c'è sempre stato almeno un minimo di 0 scontrini). Invece, il venerdì il valore crolla, con una media che si dimezza da 26.43€ di media dello scontrino medio a 13.69€.

Inoltre, verifichiamo se una festa in particolare porta le persone ad andare maggiormente al ristorante, e se questi sono portati a spendere cifre significativamente più alte rispetto al normale:

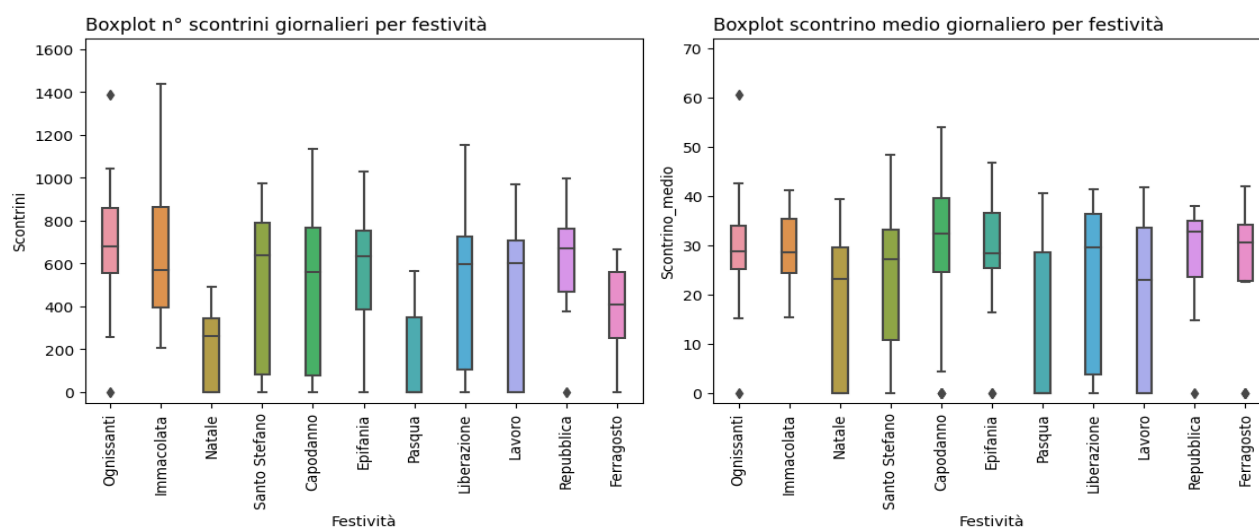


Figura 5: boxplot variabili Scontrini e Scontrino_Medido per festività

FESTIVITÀ	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
CAPODANNO	23.0	494.91	391.48	0.0	76.5	562.0	768.0	1134.0
EPIFANIA	23.0	564.74	270.79	0.0	385.0	634.0	751.5	1030.0
FERRAGOSTO	17.0	377.59	227.82	0.0	254.0	407.0	562.0	666.0
IMMACOLATA	23.0	681.78	343.77	208.0	396.0	570.0	863.0	1437.0
LAVORO	17.0	451.29	363.86	0.0	0.0	599.0	705.0	970.0
LIBERAZIONE	22.0	505.59	358.94	0.0	103.25	597.5	723.25	1153.0
NATALE	23.0	208.26	184.33	0.0	0.0	261.0	343.5	491.0
NO	7367.0	646.38	264.83	0.0	501.0	653.0	795.0	1633.0
OGNISSANTI	23.0	691.09	296.68	0.0	556.0	681.0	861.0	1387.0
PASQUA	23.0	157.83	210.68	0.0	0.0	0.0	349.5	564.0
REPUBBLICA	17.0	628.0	251.07	0.0	470.0	669.0	761.0	998.0
SANTO STEFANO	23.0	503.43	373.73	0.0	83.0	636.0	789.5	976.0

Tabella 8: statistiche variabile Scontrini per festività

FESTIVITÀ	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
CAPODANNO	23.0	28.33	16.67	0.0	24.49	32.44	39.6	53.98
EPIFANIA	23.0	28.7	12.08	0.0	25.39	28.45	36.54	46.69
FERRAGOSTO	17.0	25.61	13.27	0.0	22.67	30.48	34.14	41.95
IMMACOLATA	23.0	29.35	7.19	15.38	24.42	28.61	35.36	41.08
LAVORO	17.0	19.22	15.89	0.0	0.0	23.0	33.49	41.72
LIBERAZIONE	22.0	23.28	15.94	0.0	3.74	29.55	36.38	41.29
NATALE	23.0	17.45	14.95	0.0	0.0	23.22	29.64	39.41
NO	7367.0	26.1	9.74	0.0	21.37	27.47	32.22	243.18
OGNISSANTI	23.0	29.04	11.26	0.0	25.14	28.74	33.94	60.51

PASQUA	23.0	12.11	15.85	0.0	0.0	0.0	28.49	40.65
REPUBBLICA	17.0	28.21	10.29	0.0	23.56	32.83	35.0	37.94
SANTO STEFANO	23.0	23.68	15.91	0.0	10.72	27.14	33.16	48.32

Tabella 9: statistiche variabile Scontrino_Medio per festività

Le medie dei numeri di scontrini sono molto variabili (anche considerate rispetto al 646 di media del numero di scontrini): alcune festività hanno valori medi molto esigui, come Pasqua, Natale e Ferragosto (valori della media di rispettivamente di 157, 208 e 377 scontrini), mentre le altre festività si avvicinano tutte ai valori dei giorni non di festività (tutte con medie comprese tra 451 e 691).

Da notare alcune particolarità, come il fatto che l'Immacolata abbia sempre fatto registrare un numero significativo di scontrini, per cui il minimo numero di scontrini è di 208 (tutte le altre festività hanno avuto almeno un giorno di chiusura, e quindi hanno minimo 0). Inoltre, oltre il 25% delle ricorrenze di Natale, Pasqua e Festa del Lavoro hanno avuto valori pari a 0, probabilmente per le chiusure a causa del Covid19. Anche per questi motivi, i valori delle medie degli scontrini medi risentono particolarmente di questi valori pari a 0, motivo per cui sono le uniche medie degli scontrini medi minori di 20 sono di Pasqua, Natale e Festa del Lavoro, con rispettivamente 12.11€, 17.45€ e 19.22€.

Tutte le altre medie degli scontrini medi si avvicinano alla media dei giorni non di festa (pari a 26.10€), con valori compresi tra 23.28€ e 29.35€.

Vediamo ora se c'è qualche differenza significativa, nei singoli ristoranti, nel cambiamento di prezzi prima e dopo l'inizio della pandemia da Covid-19:

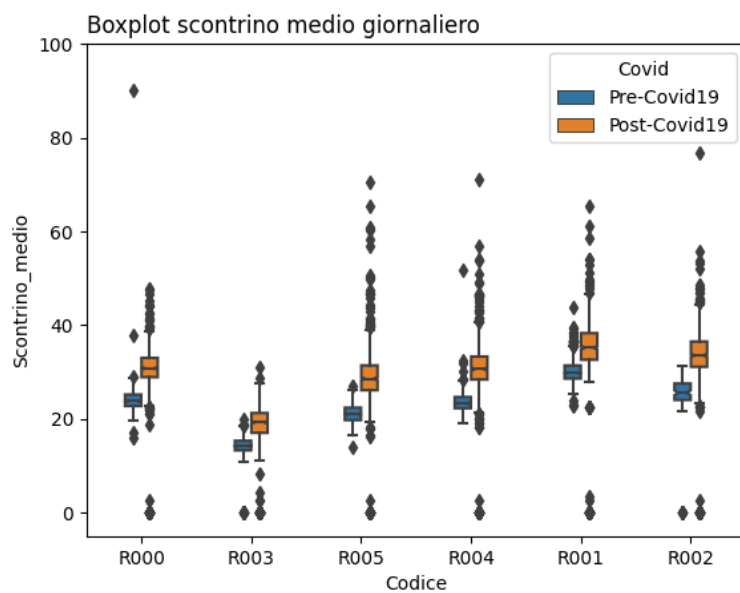


Figura 6: boxplot variabile Scontrino_Medio rispetto a ristorante e pre/post Covid

CODICE	COVID	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
R000	Post-Covid19	800.0	28.91	9.2	0.0	28.89	30.76	33.01	47.71
R000	Pre-Covid19	538.0	24.45	10.05	15.99	22.73	23.8	25.18	243.18
R001	Post-Covid19	800.0	32.69	11.82	0.0	32.6	35.17	38.29	65.3
R001	Pre-Covid19	538.0	30.28	2.5	22.84	28.59	29.82	31.38	43.76
R002	Post-Covid19	800.0	31.64	10.25	0.0	31.1	33.52	36.49	76.74
R002	Pre-Covid19	112.0	24.45	6.69	0.0	24.02	25.55	27.53	31.35
R003	Post-Covid19	799.0	17.47	7.04	0.0	17.01	19.28	21.24	30.87
R003	Pre-Covid19	538.0	14.17	2.35	0.0	13.28	14.21	15.33	19.84
R004	Post-Covid19	800.0	29.34	9.29	0.0	28.38	30.63	33.29	71.08
R004	Pre-Covid19	538.0	23.57	2.2	19.05	22.25	23.27	24.65	51.75
R005	Post-Covid19	800.0	27.59	9.28	0.0	26.1	28.51	31.4	70.51
R005	Pre-Covid19	538.0	21.08	1.94	14.06	19.72	20.99	22.38	26.93

Tabella 10: variabile Scontrino_Medio rispetto a ristorante e pre/post Covid

Si osserva, sia dai boxplot (Figura 7) che dalla tabella con le statistiche descrittive (Tabella 10), che in generale lo scontrino medio è aumentato per tutti e sei i ristoranti prima e dopo l'inizio della pandemia: il ristorante che ha aumentato maggiormente i prezzi è il ristorante R002, dove la media è passata da 24.45€ a 31.64€, un incremento del 22.75%. Al contrario, il ristorante R001 è quello che ha aumentato meno i prezzi, con una differenza di soli 2.41€ tra lo scontrino medio Pre-Covid19 e Post-Covid19.

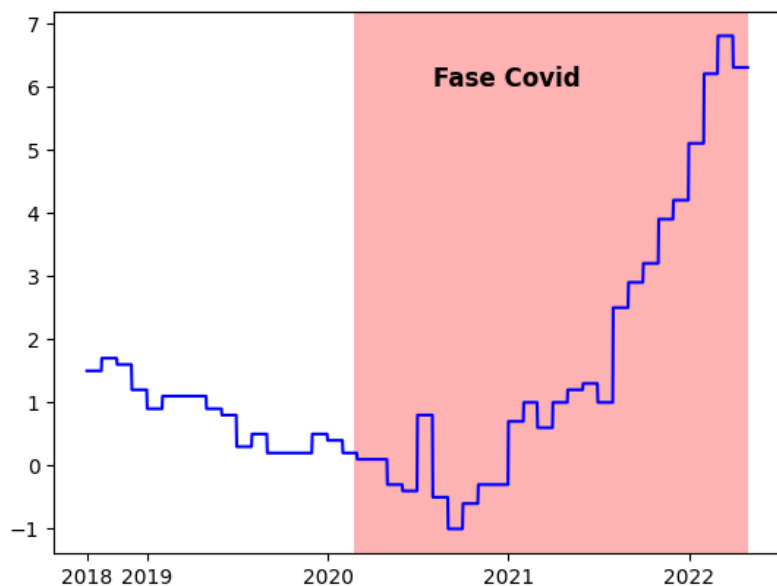


Figura 7: andamento inflazione

ANNO	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
2018	610.0	1.5	0.19	1.2	1.2	1.55	1.7	1.7
2019	1886.0	0.64	0.35	0.2	0.2	0.5	0.9	1.1
2020	2196.0	-0.15	0.47	-1.0	-0.5	-0.3	0.1	0.8
2021	2190.0	1.96	1.25	0.6	1.0	1.3	3.2	4.2
2022	719.0	6.1	0.63	5.1	5.1	6.3	6.8	6.8

Tabella 11: statistiche descrittive variabile Inflazione per anno

ANNO	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
2018	610.0	23.04	10.66	0.0	19.97	23.07	26.02	243.18
2019	1886.0	22.57	5.93	0.0	19.6	22.86	25.92	90.01
2020	2196.0	24.22	12.94	0.0	17.85	27.84	32.54	71.08
2021	2190.0	30.03	7.38	0.0	27.07	31.09	34.38	70.51
2022	719.0	31.09	7.42	0.0	28.6	31.89	35.09	76.74

Tabella 12: variabile Scontrino_medio per anno

Si osserva che generalmente, per tutti i ristoranti, i prezzi si sono mantenuti costanti nel periodo compreso tra il 2018 e il 2020. Il 2021 è stato probabilmente l'anno dove lo scontrino medio è cresciuto in maniera maggiore per tutti i ristoranti, completando l'aumento di valori nel 2022. Tuttavia, va considerata comunque la situazione particolare del 2020, ovvero l'anno con il numero maggiore di giorni di chiusura, che inevitabilmente hanno aumentato la quantità di osservazioni con valori di Scontrino_Medio pari a 0, andando ad incidere in maniera molto significativa sulla media (si può notare, nella Tabella 13, sia una deviazione standard molto più alta degli altri anni, sia valori del primo quartile più bassi rispetto agli anni precedenti).

L'inflazione, in compenso, si è mantenuta a livelli controllati fino al 2021, con una media compresa tra -0.15 e 1.96; in compenso, l'aumento consistente è stato nel 2022, dove il valore minimo è di 5.1 (considerando che dal 2018 al 2020 i valori dell'inflazione non avevano mai superato quota 1.7, e che il massimo nel 2021 era stato di 4.2) e la media di 6.09.

Analisi esplorativa singoli ristoranti

Ristorante 1

Passiamo ora ad un'analisi dei singoli ristoranti, iniziando col primo, il ristorante codice R000 situato a Montebello, provincia di Pavia.

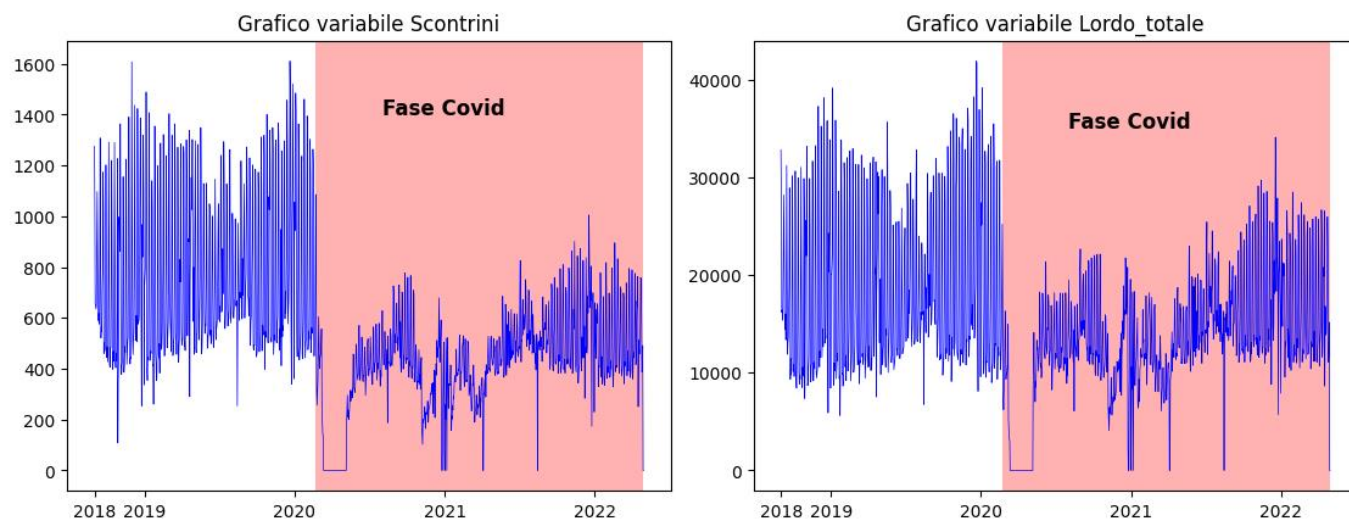


Figura 9: grafici variabili Scontrini e Lordo_totale ristorante 1

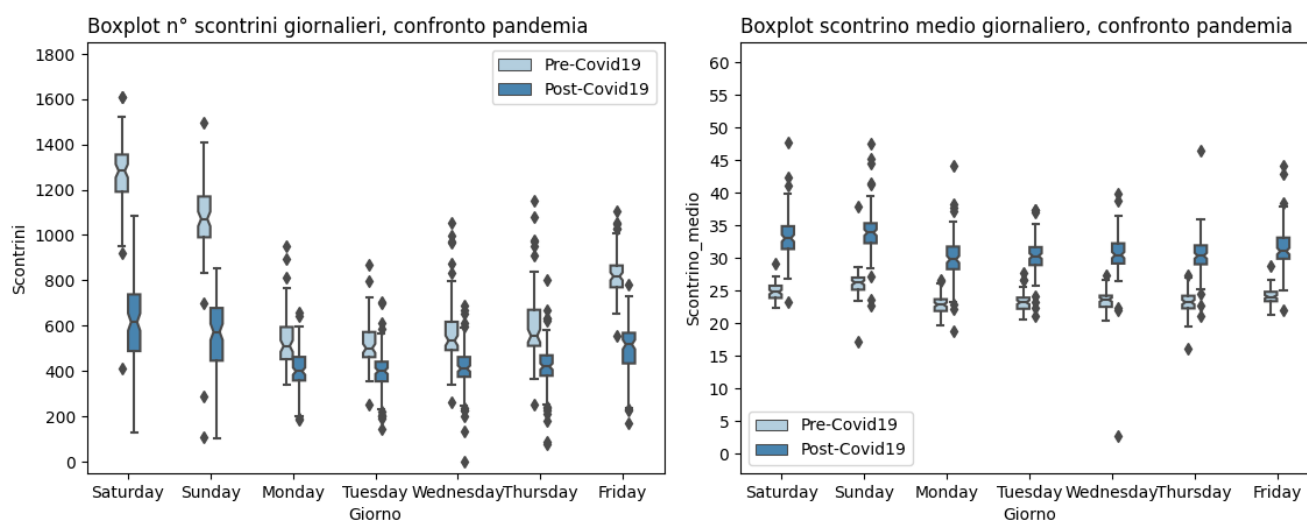


Figura 8: boxplot variabili Scontrini e Scontrino_medio per giorno, con confronto Pre/Post Covid ristorante 1, esclusi giorni con 0 scontrini

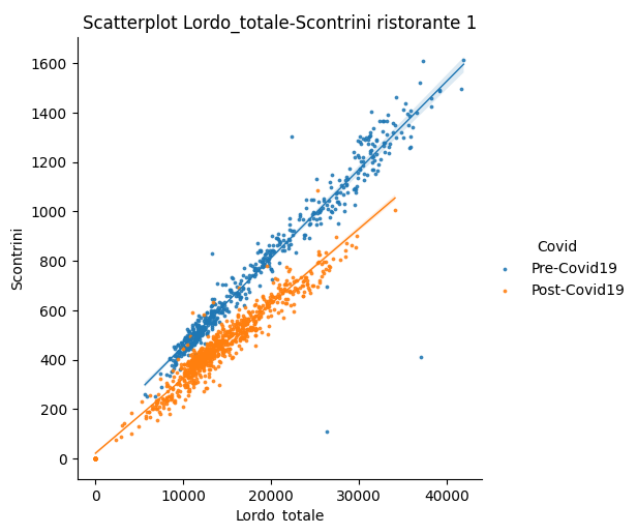


Figura 10: Scatterplot Lordo_totale Scontrini ristorante 1

Dalla *Figura 8* è possibile individuare chiaramente il periodo del primo lockdown, che coincide col crollo dei valori di *Scontrini* e *Lordo_Totale* della prima metà del 2020: anche dopo la fine del primo lockdown, la situazione non sembra più essere quella precedente, con valori che si mantengono molto più bassi rispetto al periodo precedente (soprattutto per quanto riguarda il numero degli scontrini, ma anche per il valore *Lordo_totale*, forse parzialmente compensato dall'aumento dei prezzi).

Per il ristorante 1, osserviamo che sembra essere un locale che lavora molto nel fine settimana, un po' meno nei giorni feriali: il numero medio di scontrini tocca picchi di 1259 circa nei sabati pre-Covid (e numeri importanti anche domenica e venerdì, con medie rispettivamente di 1061 e 825), per calare ad un massimo di 600 scontrini di media il giovedì pre-Covid e un minimo di 519 di media il martedì. La situazione rimane simile anche dall'inizio della pandemia, sempre con numeri superiori il weekend, ma il numero di scontrini medi cala comunque drasticamente (540 scontrini medi il sabato, 486 la domenica e 447 il venerdì) assottigliando la differenza con il resto della settimana.

È possibile notare dalla *Figura 8* come lo scontrino medio sia aumentato in maniera significativa tra il periodo precedente al Covid e dopo l'inizio della pandemia: siccome è poco realistico ipotizzare che le persone dall'inizio della pandemia abbiano iniziato ad uscire a mangiare maggiormente in gruppo (per le restrizioni appunto), si può dedurre invece un significativo aumento dei prezzi, come evidenziato dagli aumenti dei valori di inflazione osservati in precedenza.

Infine, nella *Figura 10* è possibile notare una forte correlazione positiva tra i valori di *Lordo_totale* e *Scontrini*: al crescere dei valori di una variabile, i valori dell'altra tendono a crescere con andamento simile; solo in rari casi vediamo valori lontani dalle distribuzioni (inoltre, è da notare il fatto che la maggior parte di questi è del periodo Pre-Covid).

Ristorante 2

Analizziamo ora il ristorante n°2, situato in Piacenza.

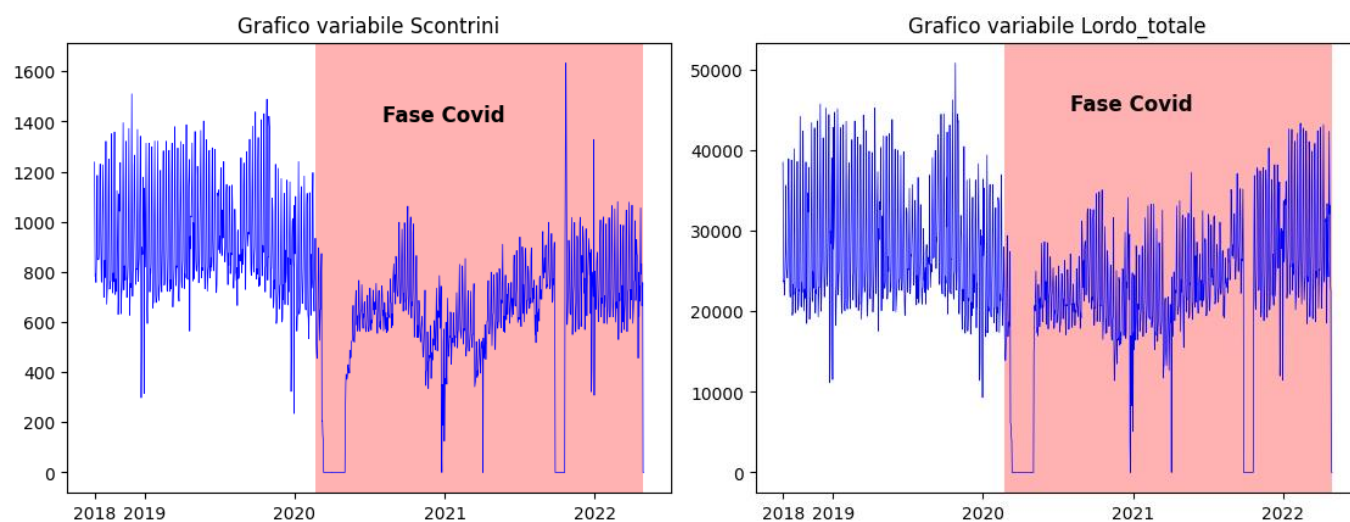


Figura 11: grafici variabili *Scontrini* e *Lordo_totale* ristorante 2

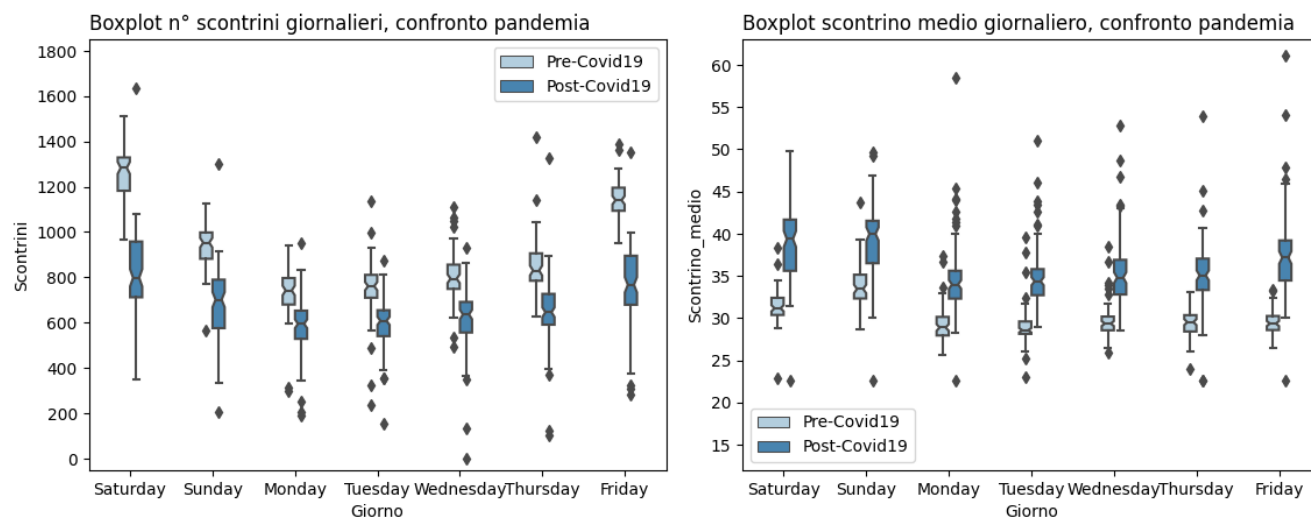


Figura 12: boxplot variabili Scontrini e Scontrino_medio per giorno, con confronto Pre/Post Covid ristorante 2, esclusi giorni con 0 scontrini

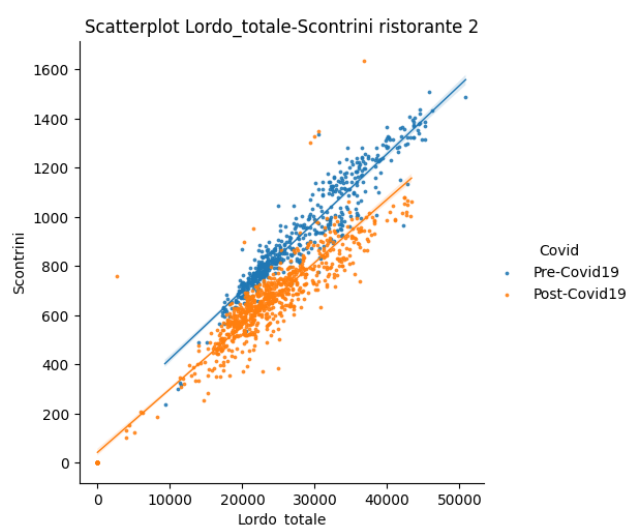


Figura 13: Scatterplot Lordo_totale Scontrini ristorante 2

Come nella situazione del primo ristorante, anche per il ristorante n°2 si può notare (dalla Figura 11) un calo significativo dei valori di *Scontrini* e *Lordo_Totale* della prima metà del 2020, anche se meno significativo rispetto alle perdite del primo ristorante: la situazione non sembra più essere quella precedente, con valori del *Lordo_totale* probabilmente compensati in parte dall'aumento dei prezzi, ma che corrispondono comunque ad un numero di scontrini molto più basso. Rispetto al primo ristorante, notiamo comunque che lo scontrino medio è più elevato, con media di 31.72 per scontrino (come riportato in Tabella 3), rispetto ai 27.11 del primo ristorante.

Per il ristorante 2, sembra essere un locale che lavora molto nel fine settimana, meno nei giorni feriali: il numero medio di scontrini tocca picchi di 1257 circa nei sabati pre-Covid (con numeri importanti più il venerdì della domenica, e medie rispettivamente di 1135 e 940); questi numeri calano molto meno del primo ristorante durante la settimana, con numeri di scontrini nel periodo pre-Covid compresi tra un minimo di 750 il martedì e un massimo di 851 il giovedì. La situazione cambia con l'inizio della pandemia, con numeri che tendono a diventare regolari per tutta la settimana (compresi tra un massimo di 811 scontrini di media il sabato e un minimo di 587 scontrini il lunedì).

Dalla Figura 12 osserviamo che lo scontrino medio è aumentato in maniera significativa tra il periodo precedente al Covid e dopo l'inizio della pandemia, come probabile conseguenza dell'aumento dei prezzi attuato dai ristoratori per combattere gli effetti della pandemia sui loro ricavi. Infine, nella Figura 13 è possibile osservare la forte correlazione positiva che intercorre tra i valori di *Lordo_totale* e *Scontrini*: al crescere dei valori di una variabile, i valori dell'altra tendono a crescere con andamento simile; solo in rari casi vediamo valori lontani dalle distribuzioni (in questo caso, rispetto al primo ristorante, la maggior parte di questi fa parte del periodo dopo l'inizio della pandemia).

Ristorante 3

Ora analizziamo i dati relativi al terzo ristorante, anche questo situato in Piacenza. La particolarità dei dati relativi a quest'attività, rispetto alle altre, sta nel fatto che i dati iniziano in data 08/11/2019, rispetto al 01/09/2018 degli altri ristoranti.

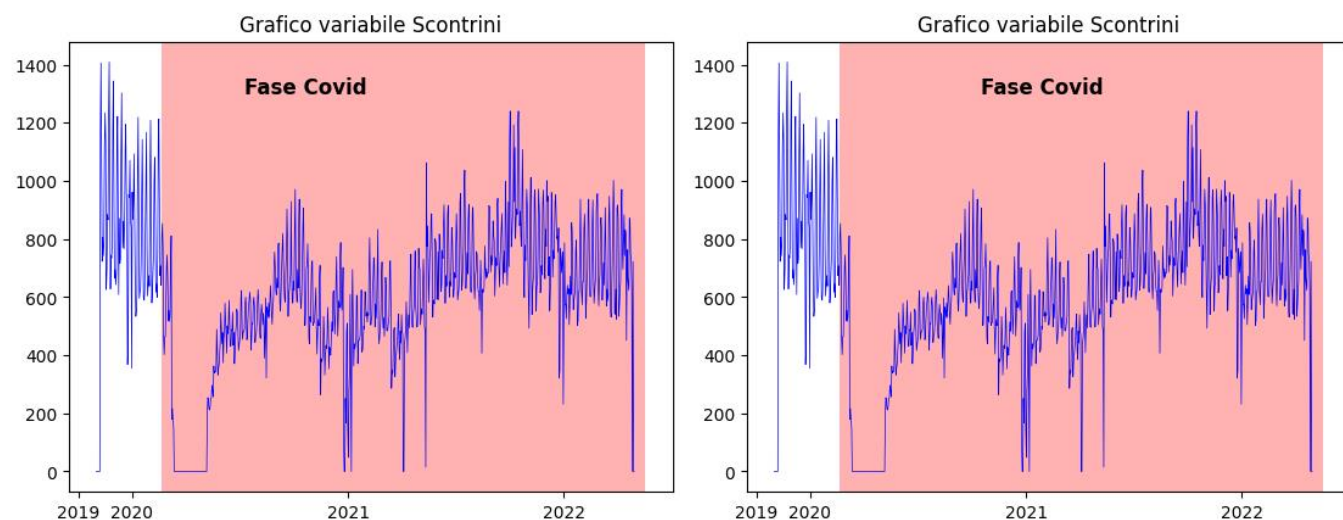


Figura 14: grafici variabili Scontrini e Lordo_totale ristorante 3

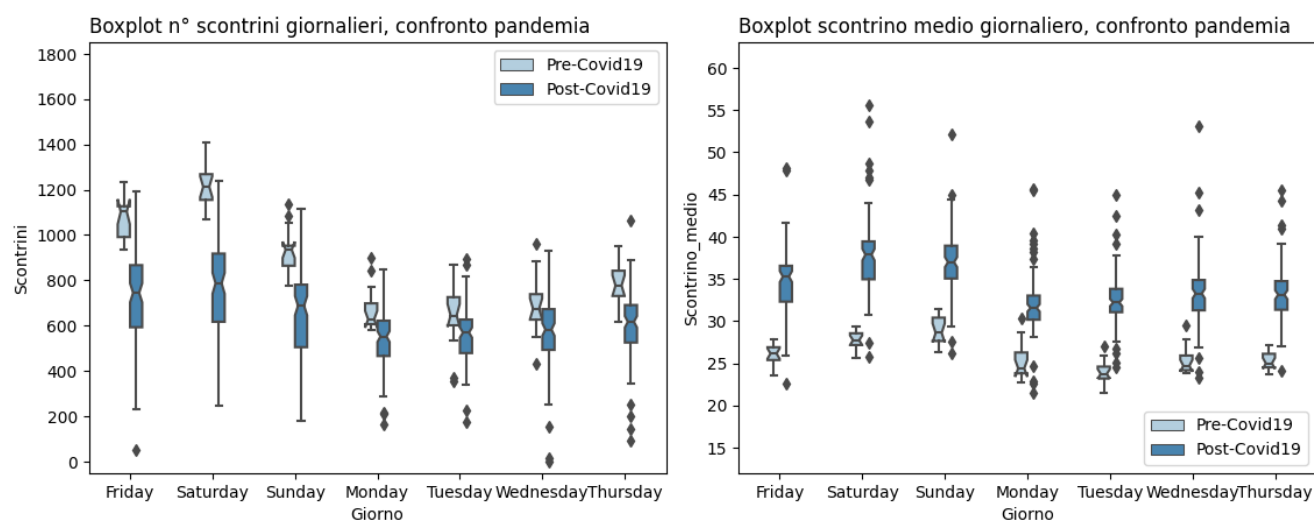


Figura 15: boxplot variabili Scontrini e Scontrino_medio per giorno, con confronto Pre/Post Covid ristorante 3, esclusi giorni con 0 scontrini

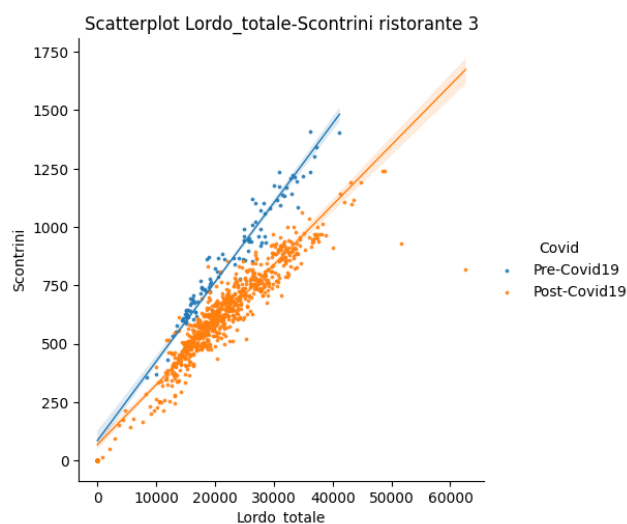


Figura 16: Scatterplot Lordo_totale Scontrini ristorante 3

Rispetto ai primi due ristoranti, è più complesso fare un confronto tra la situazione prima e dopo l'inizio della pandemia del Covid-19, in quanto praticamente ci sono 14 mesi in meno di dati a disposizione per quanto riguarda la situazione pre-Covid: tuttavia, si riconosce comunque dalla *Figura 14* un calo dei valori di *Scontrini* e *Lordo_Totale* dalla prima metà del 2020, seppur più moderato rispetto soprattutto al primo ristorante: la situazione non sembra più essere quella precedente, con valori del *Lordo_totale* probabilmente compensati in parte dall'aumento dei prezzi, ma che corrispondono comunque ad un numero di scontrini molto più basso. Da notare, inoltre, che i valori dello scontrino medio si avvicinano più a quelli del ristorante n°2 che al primo.

Come per il ristorante n°2, anche il n°3 sembra essere un locale che lavora molto nel fine settimana, e che riduce il numero di coperti nei giorni feriali: si passa da picchi di 1144 circa nei sabati pre-Covid (con domeniche da 879 scontrini di media, e venerdì da 1000 scontrini di media), a numeri più ridotti durante la settimana (608 scontrini di media il martedì come valore minimo). Con l'inizio della pandemia, il numero di scontrini diventa molto regolare durante tutta la settimana (con valori compresi tra un massimo di 695 scontrini di media il sabato e un minimo di 502 scontrini il lunedì). Come negli altri casi, lo scontrino medio è aumentato in maniera significativa tra il periodo precedente al Covid e dopo l'inizio della pandemia (come indicato in *Figura 15*), come probabile conseguenza dell'aumento dei prezzi e dell'inflazione.

Infine, la *Figura 15* mostra una forte correlazione positiva che intercorre tra i valori di *Lordo_totale* e *Scontrini*: al crescere dei valori di una variabile, i valori dell'altra tendono a crescere con andamento simile; rispetto alle precedenti situazioni, da notare una correlazione maggiormente positiva nella situazione Pre-Covid (il *Lordo_totale* aumenta maggiormente con l'aumentare del numero di scontrini nel periodo pre-Covid, rispetto che dopo l'inizio della pandemia).

Ristorante 4

Analizziamo ora il ristorante n°4, situato anch'esso a Piacenza.

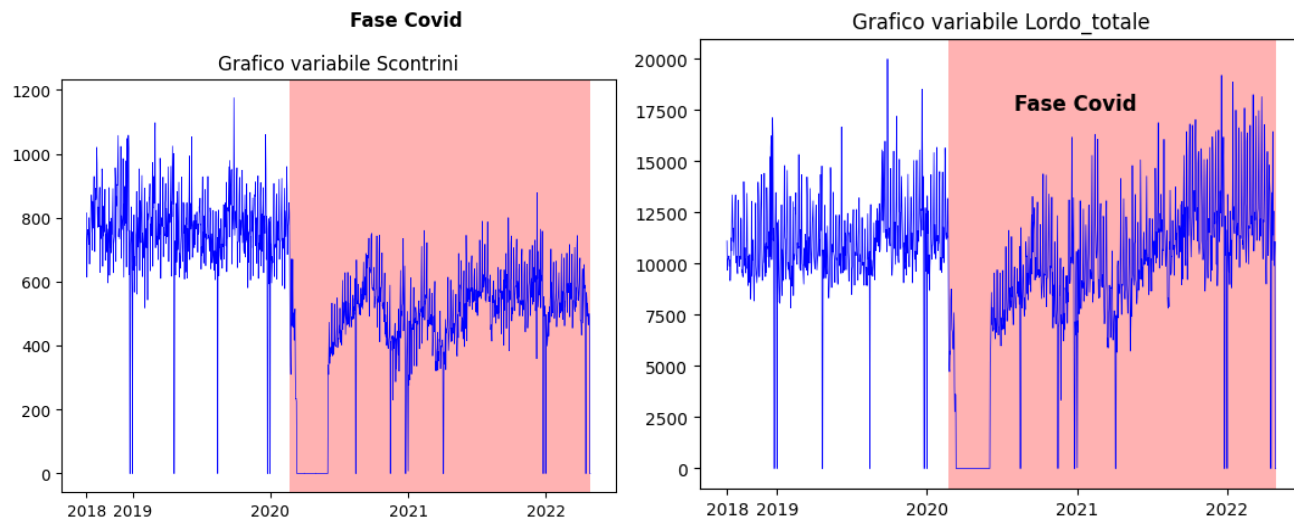


Figura 17: grafici variabili *Scontrini* e *Lordo_totale* ristorante 4

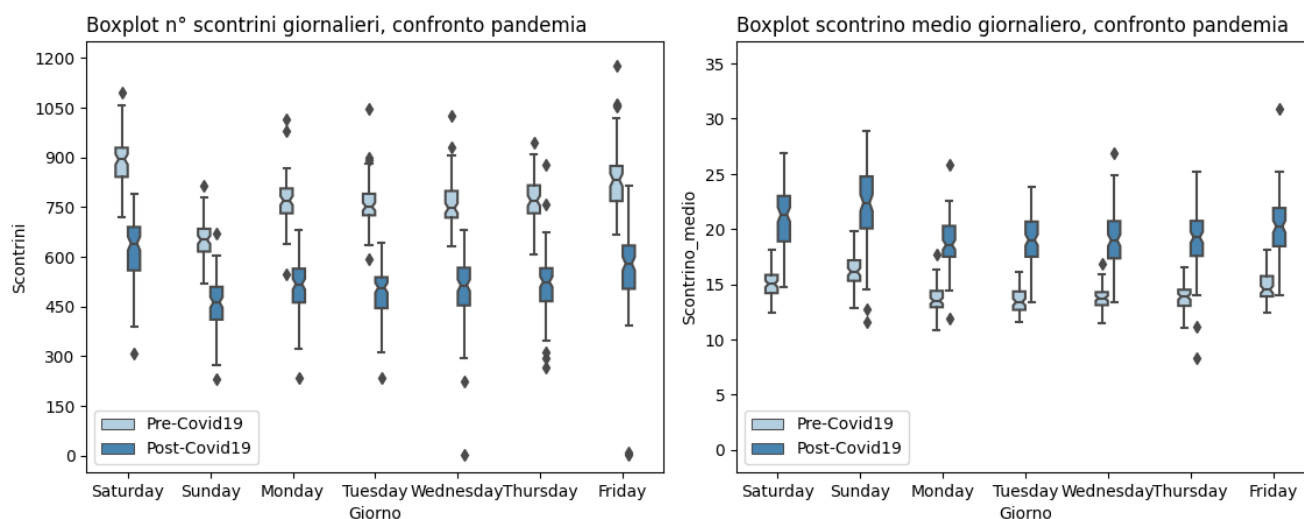


Figura 18: boxplot variabili Scontrini e Scontrino_medio per giorno, con confronto Pre/Post Covid ristorante 4, esclusi giorni con 0 scontrini

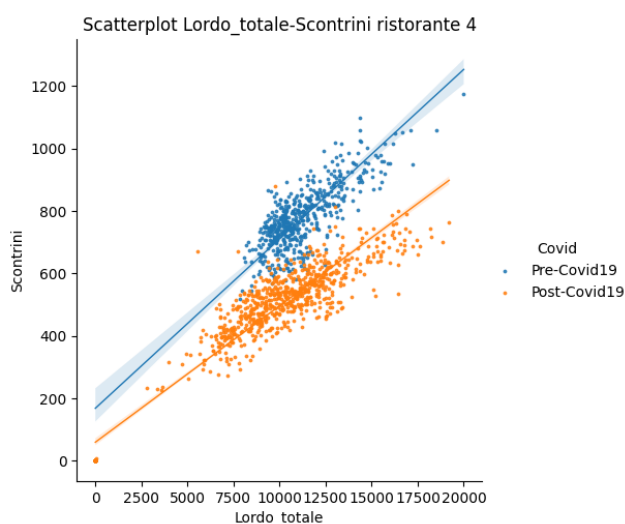


Figura 19: Scatterplot Lordo_totale Scontrini ristorante 4

Il ristorante n° 4 è un caso particolare rispetto a quelli visti in precedenza, per alcune caratteristiche: probabilmente è un ristorante più piccolo (da cui un numero significativamente minore di scontrini giornalieri, confrontando sia il periodo pre-Covid che post-Covid), e con prezzi decisamente minori (dalla Tabella 3 notiamo uno scontrino medio di 16.14, rispetto agli altri ristoranti con medie attorno ai 25/30€). Inoltre, a differenza degli altri locali, sembra essere un ristorante che lavora parecchio durante la settimana, e mantiene un numero simile di clientela anche nei weekend (come è facile notare dalla Figura 18).

Come per tutti i precedenti ristoranti, anche in questo caso si nota (Figura 17) un calo significativo dei valori di Scontrini della prima metà del 2020, tuttavia (esclusi i periodi di chiusura) a livello di Lordo_totale il ristorante sembra mantenere la situazione precedente all'inizio della pandemia, come conseguenza di diversi fattori (verosimilmente l'aumento dei prezzi, l'inflazione, ma magari anche il fatto di avere meno coperti, dunque di riuscire comunque a riempire il locale e di raggiungere comunque i guadagni sperati).

Come già detto, il ristorante n°4 sembra essere un locale che lavora molto durante la settimana, con un numero medio di scontrini che tocca picchi di 892 scontrini circa nei sabati pre-Covid, e valori minimi di 645 nelle domeniche pre-Covid; con l'inizio della pandemia, la situazione rimane simile, sempre con massimi di 530 scontrini il sabato e minimi di 387 scontrini la domenica, e tutti gli altri valori compresi tra 440 e 500 scontrini circa. Dalla Figura 18 osserviamo che lo scontrino medio è aumentato in maniera significativa tra il periodo pre-Covid e post-Covid, tuttavia

si tratta comunque di aumenti contenuti tenendo conto dei valori di scontrino medio molto più bassi di tutte le altre attività.

Infine, come per gli altri ristoranti, nella *Figura 19* è possibile osservare la forte correlazione positiva tra i valori di *Lordo_totale* e *Scontrini*, per cui al crescere dei valori di una variabile, i valori dell'altra tendono a crescere con andamento simile.

Ristorante 5

Viene analizzato ora il ristorante n°5, con sede a Stradella, in provincia di Pavia.

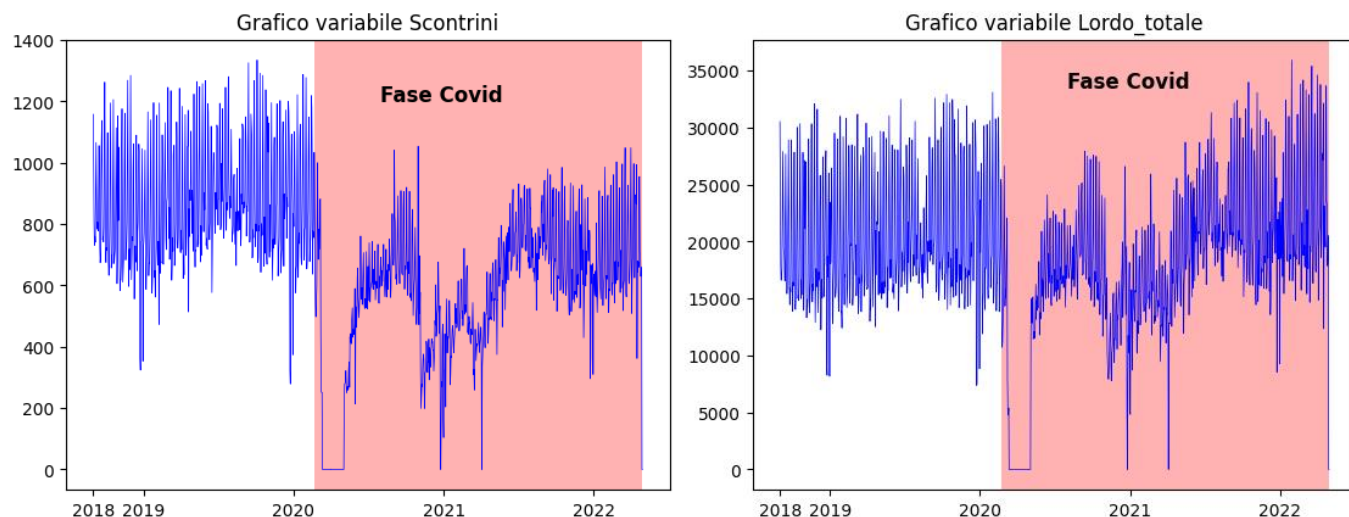


Figura 21: grafici variabili *Scontrini* e *Lordo_totale* ristorante 5

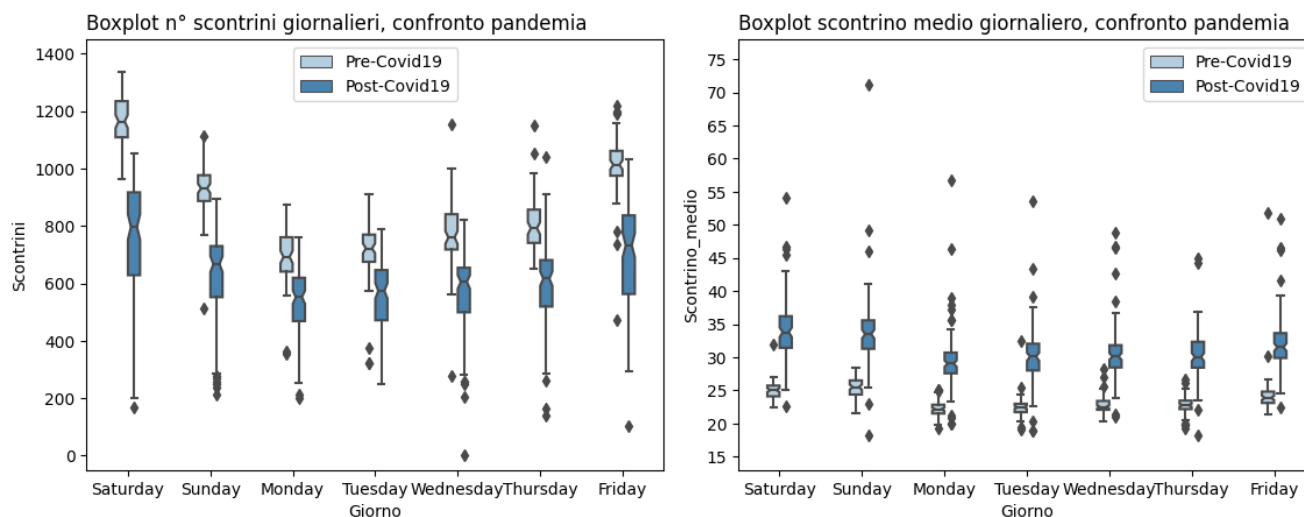


Figura 20: boxplot variabili *Scontrini* e *Scontrino_medio* per giorno, con confronto Pre/Post Covid ristorante 5, esclusi giorni con 0 scontrini

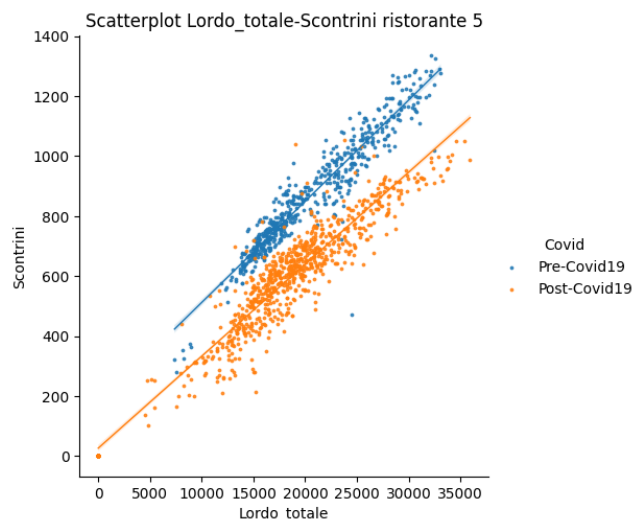


Figura 22: Scatterplot Lordo_totale Scontrini ristorante 5

La situazione del ristorante n°5 è simile a quella dei primi due ristoranti: ristoranti con un gran traffico di clienti soprattutto nel fine settimana, che ha visto un netto calo del numero di scontrini e nei valori di Lordo_totale con l'inizio della pandemia, e dunque per limitare le perdite si è visto costretto ad aumentare i prezzi (come delineato dalla Figura 21). Il valore di scontrino medio è elevato, con valori simili al ristorante n°1 e n°2, con media di 27.02 per scontrino (come riportato in Tabella 3).

Il ristorante 5, prima dell'inizio della pandemia, lavorava parecchio nei fine settimana, meno nei giorni feriali: il numero medio di scontrini tocca picchi di 1164 circa nei sabati pre-Covid (con numeri importanti il venerdì e la domenica, con medie rispettivamente di 1035 e 933); i numeri calavano durante la settimana, con numeri di scontrini nel periodo pre-Covid compresi tra un minimo di 698 il lunedì e un massimo di 803 il giovedì. La situazione con la pandemia è cambiata parecchio, con numeri più regolari in tutta la settimana (con massimi di 691 scontrini di media il sabato e un minimo di 503 scontrini il lunedì).

Dalla Figura 21 osserviamo che lo scontrino medio è aumentato in maniera significativa tra il periodo precedente al Covid e dopo l'inizio della pandemia, probabile conseguenza dell'inflazione e dell'aumento di prezzi per combattere gli effetti della pandemia sui loro ricavi. Da notare, in particolare, che rispetto agli altri ristoranti sembrano esserci molti più outliers nelle distribuzioni degli scontrini medi, soprattutto per valori elevati nel periodo post-Covid: potrebbe trattarsi di cerimonie, o feste, con numeri di invitati elevati e dunque scontrini medi con valori molto più elevati. La Figura 22 indica la forte correlazione positiva tra i valori di Lordo_totale e Scontrini, come per tutte le altre distribuzioni degli altri ristoranti: al crescere dei valori di una variabile, i valori dell'altra tendono a crescere con andamento simile.

Ristorante 6

Infine, analizziamo l'ultimo ristorante, situato in Voghera, provincia di Pavia.

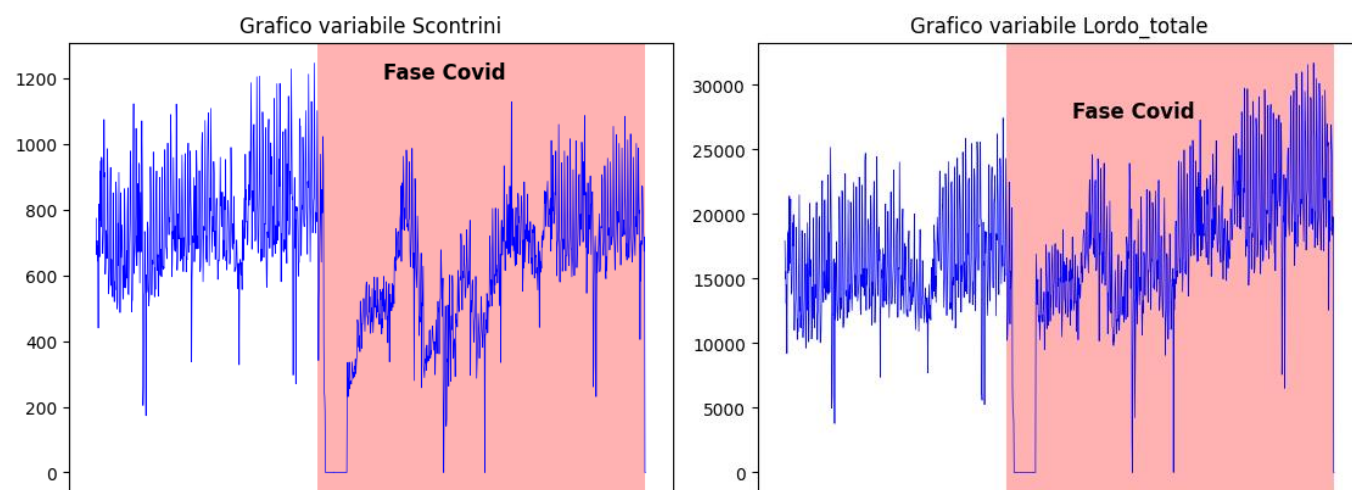


Figura 23: grafici variabili Scontrini e Lordo_totale ristorante 6

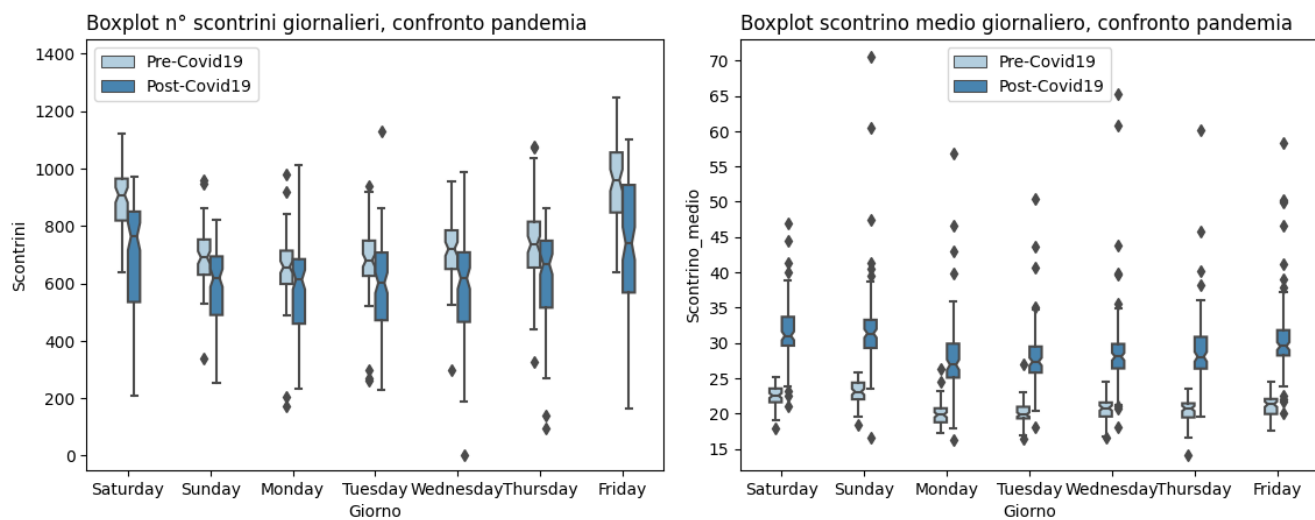


Figura 24: boxplot variabili Scontrini e Scontrino_medio per giorno, con confronto Pre/Post Covid ristorante 6, esclusi giorni con 0 scontrini

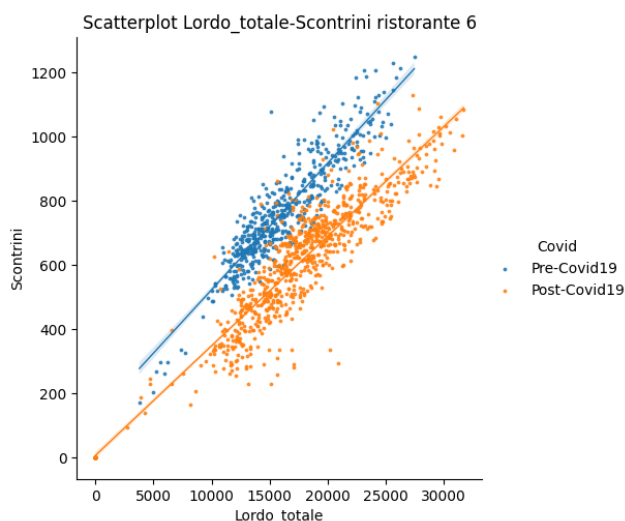


Figura 25: Scatterplot Lordo_totale Scontrini ristorante 6

La situazione del ristorante n°6 è una via di mezzo tra il ristorante 1,2 e 5 e il ristorante n°4: si tratta probabilmente di un'attività con un numero minore di coperti rispetto agli altri ristoranti (proprio come il ristorante n°4), tuttavia ha prezzi più elevati del quarto ristorante, ed ha non solo mantenuto comunque numeri simili come scontrini nel confronto tra prima dell'inizio del Covid e dopo, ma ha addirittura aumentato il lordo totale a partire da metà del 2021.

Prima dell'inizio della pandemia, il ristorante n°6 lavorava maggiormente nei fine settimana, mentre con la pandemia la situazione è cambiata, con numeri molto più regolari in tutta la settimana. Da notare, rispetto alle altre attività, che i valori medi più elevati come numero di scontrini sono il venerdì nel periodo pre-Covid, rispetto al sabato per tutti gli altri ristoranti.

Dalla Figura 24 osserviamo che lo scontrino medio è aumentato in maniera significativa tra il periodo precedente al Covid e dopo l'inizio della pandemia, come probabile conseguenza dell'aumento dei prezzi attuato dal proprietario per combattere gli effetti della pandemia sui ricavi. Come nel caso del ristorante n°5, sembrano esserci molti più outliers nelle distribuzioni degli scontrini medi, soprattutto per valori elevati nel periodo post-Covid: potrebbe trattarsi di cerimonie, o feste, con valori elevati.

Infine, la *Figura 25* indica la forte correlazione positiva tra i valori di *Lordo_totale* e *Scontrini*, come per tutte le altre distribuzioni degli altri ristoranti: al crescere dei valori di una variabile, i valori dell'altra tendono a crescere con andamento simile.

Modelli e serie temporali

Vengono ora creati alcuni modelli di serie temporali, con l'obiettivo di studiare ancora meglio l'andamento e i trend dei differenti. Questi sono i dati originali a cui si fa riferimento:

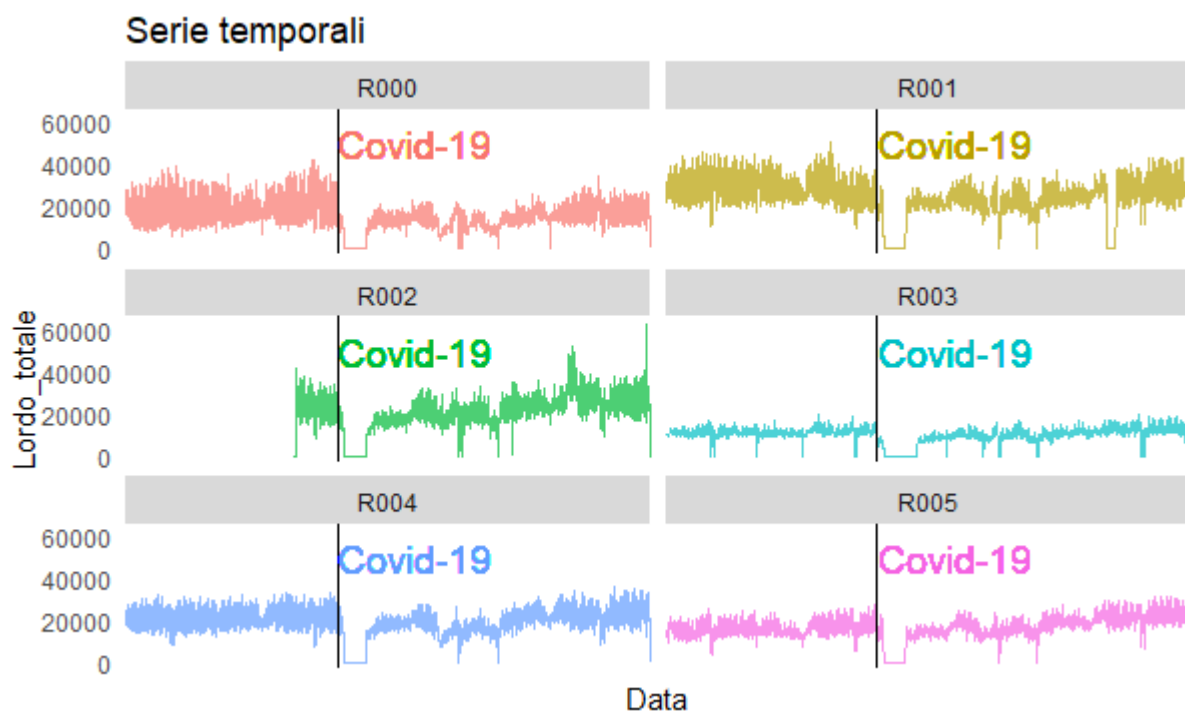


Figura 26: raccolta variabile *Lordo_totale* serie temporali di tutti i ristoranti

Osserviamo ora, in particolare, le serie temporali ottenute dai valori di *Lordo_totale* dei differenti ristoranti, per il periodo precedente allo scoppio della pandemia:

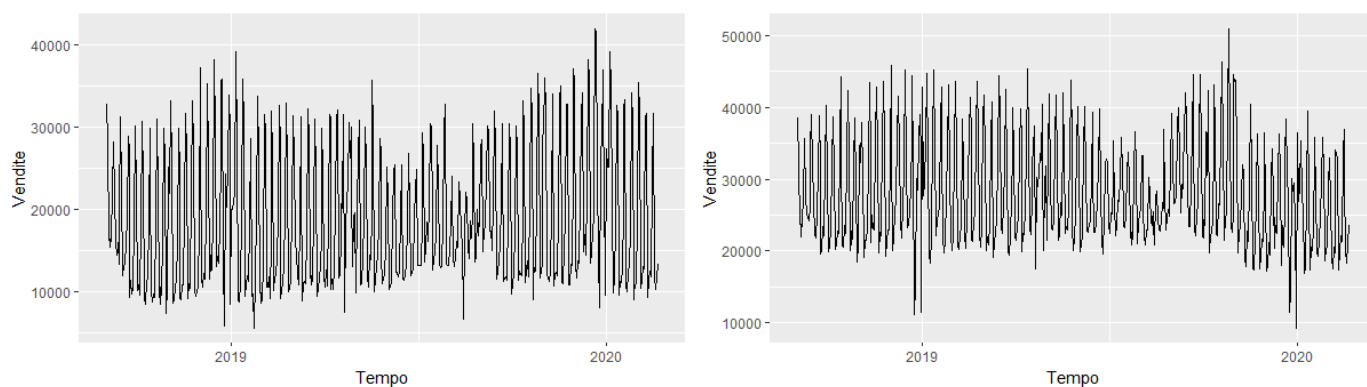


Figura 27: serie temporale *Lordo_totale* ristorante 1 e 2

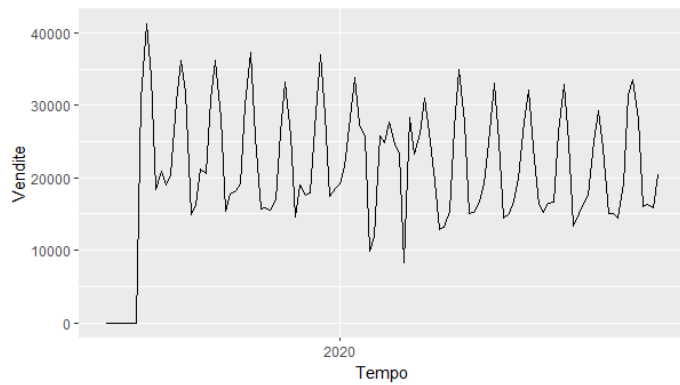


Figura 29: serie temporale Lordo_totale ristorante 3 e 4

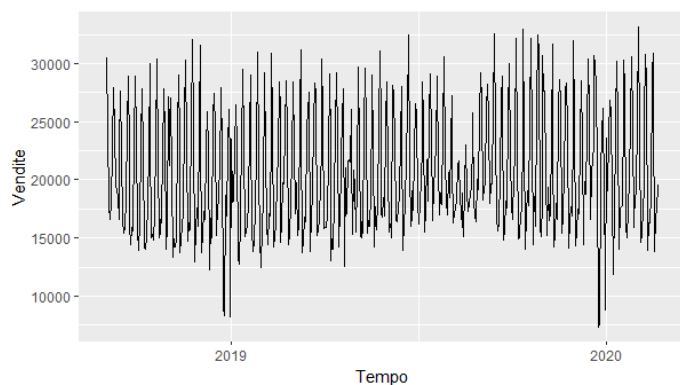
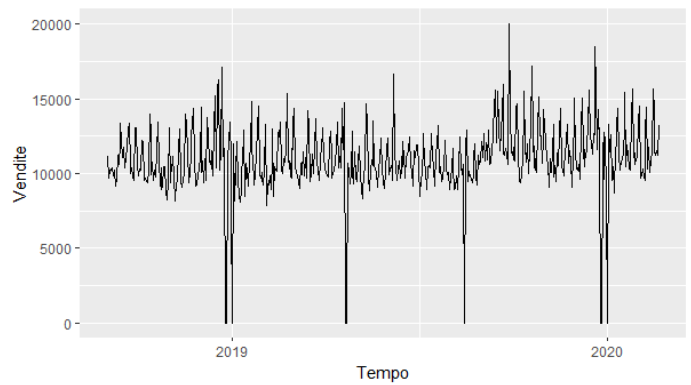
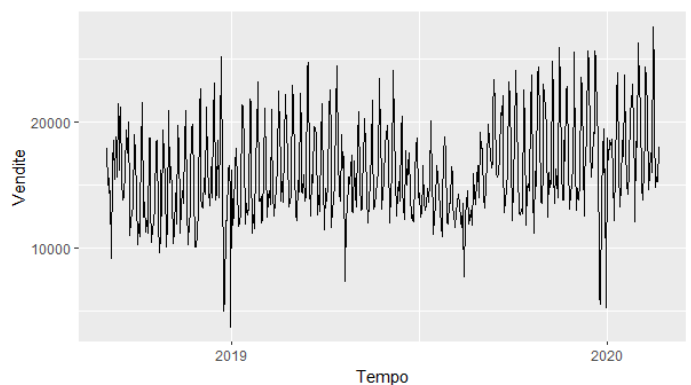


Figura 28: serie temporale Lordo_totale ristorante 5 e 6



È possibile notare, una certa irregolarità all'interno delle serie: si tratta di dati giornalieri, ma in questi influisce la presenza di stagionalità all'interno dei dati, sia stagionalità settimanale, come indicato nella sezione di analisi esplorativa (ad esempio, i giorni del weekend con molti più clienti e molto più guadagno per alcune attività), sia stagionalità annuali (come per le feste programmate, o in generale per periodi dell'anno con più o meno movimento; questa è comunque più difficilmente individuabile, per mancanza di due periodi completi). La stagionalità annuale non è visibile ovviamente per il ristorante 3, avendo troppi pochi dati, tuttavia la stagionalità settimanale è ben visibile.

Proviamo a differenziare le serie temporali, per andare a stabilizzare la media di una serie temporale in modo tale da poter osservare serie stazionarie:

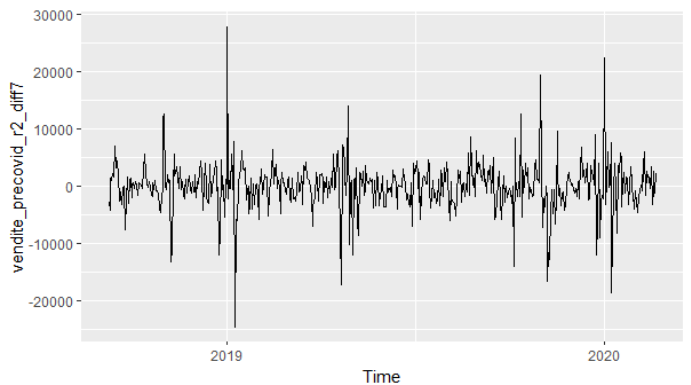
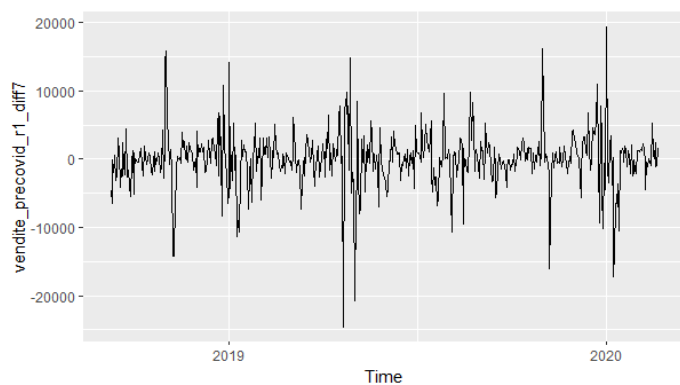


Figura 30: serie temporale differenziata del Lordo_totale ristorante 1 e 2

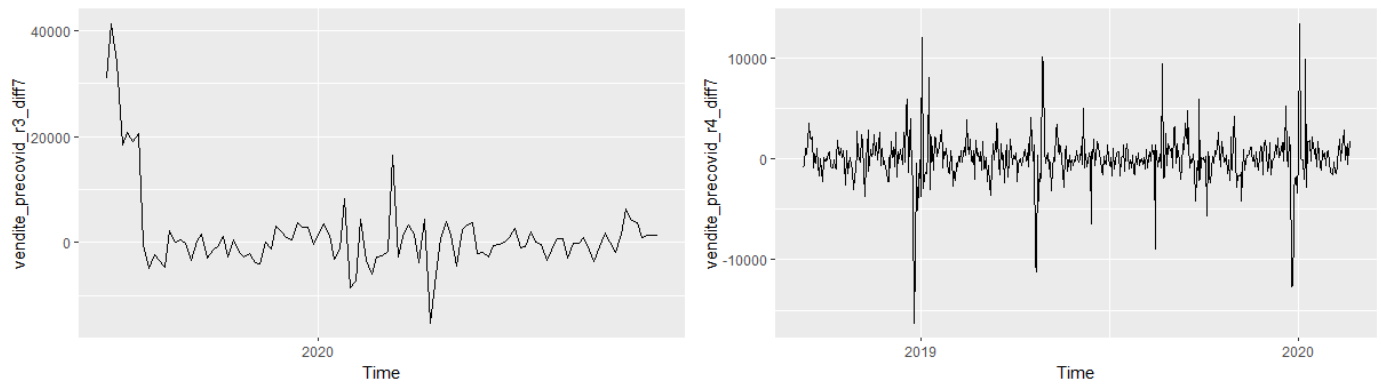


Figura 31: serie temporale differenziata del Lordo_totale ristorante 3 e 4

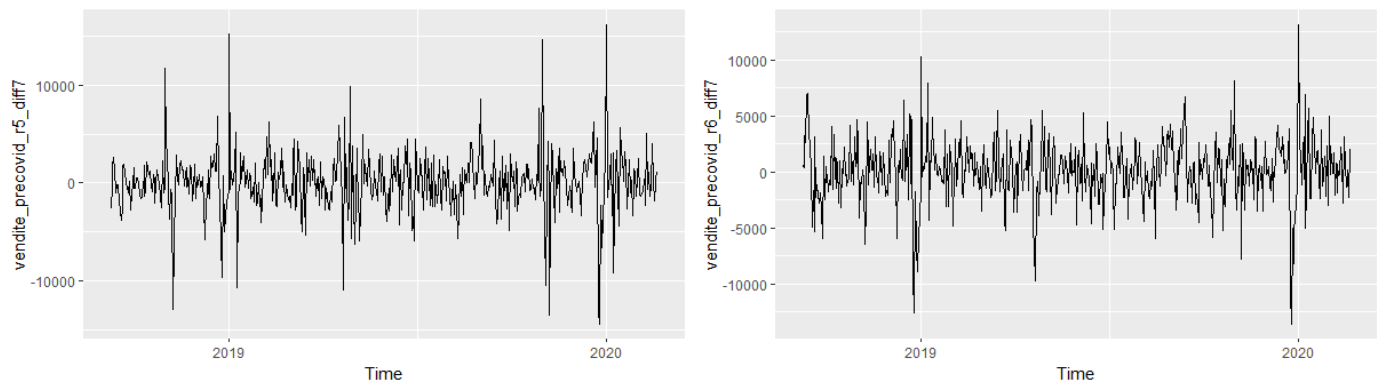


Figura 32: serie temporale differenziata del Lordo_totale ristorante 5 e 6

Con una differenziazione settimanale, abbiamo eliminato l'elemento di trend, o stagionalità, che era presente all'interno della serie temporale originale: ora non si presentano più trend particolari legati al giorno della settimana in cui viene considerata la serie. Valutiamo, inoltre, se tramite una trasformazione di BoxCox possiamo andare a risolvere il problema della non stazionarietà della serie. I risultati delle stime di lambda sono differenti per ogni distribuzione della variabile Lordo_totale dei sei ristoranti: con -0.34 viene suggerita una trasformazione dell'inverso del quadrato per la distribuzione del ristorante 1, con valori di 1.06 , 1.12 e 0.86 per le distribuzioni del ristorante n° 2, 5 e 6 non è necessaria alcuna trasformazione, con valori di 1.71 e 1.99 per le distribuzioni dei ristoranti 4 e 3 viene suggerita trasformazione quadratica. Effettuate queste operazioni, è possibile differenziare ulteriormente: in questo modo l'ACF non decade a zero lentamente e non presenta pattern stagionali per tutte le serie temporali.

Possiamo, a questo punto, stimare i modelli ARIMA delle differenti distribuzioni:

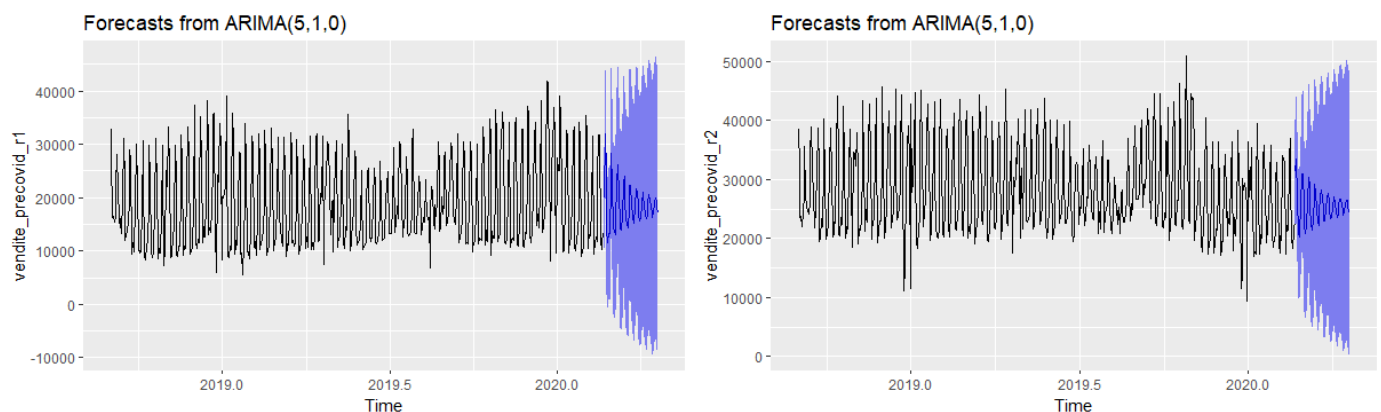


Figura 33: previsioni Lordo_totale ristoranti 1 e 2, tramite modello ARIMA

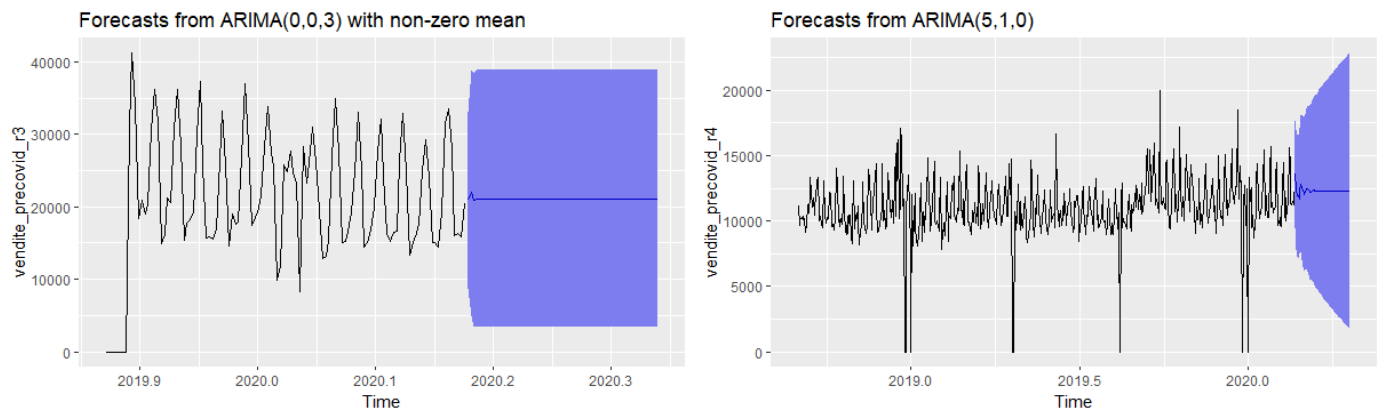


Figura 35: previsioni Lordo_totale ristoranti 3 e 4, tramite modello ARIMA

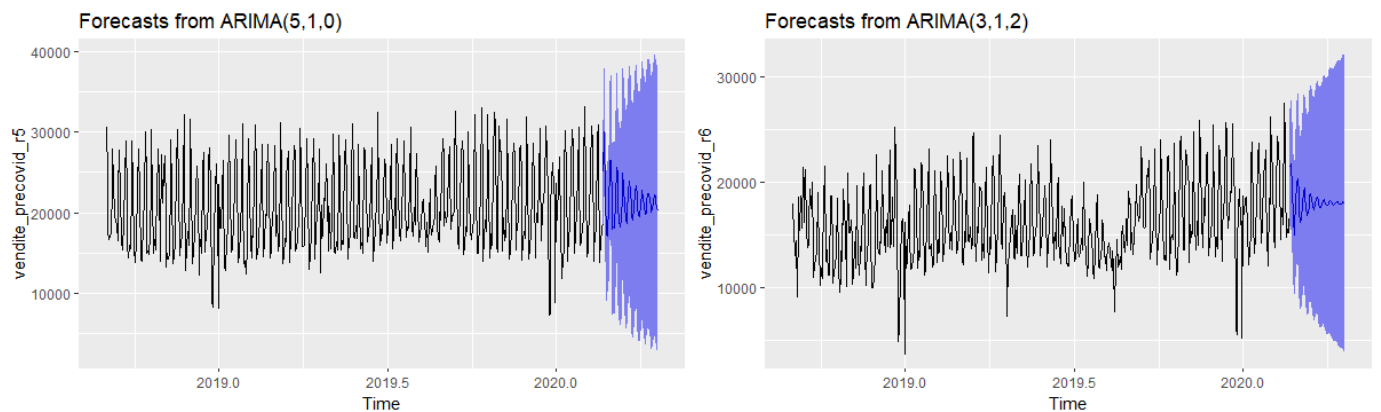


Figura 34: previsioni Lordo_totale ristoranti 5 e 6, tramite modello ARIMA

Queste sono le previsioni ottenute a partire dai modelli ARIMA, in modo da stimare i guadagni previsti se non ci fosse stata la pandemia e il periodo di lockdown da marzo 2020. Infatti per questi ristoranti i mesi di marzo e aprile 2020 sono stati segnati da un crollo delle vendite, solo parzialmente coperto dagli aiuti e dalle sovvenzioni stanziati dal governo, ed è facile immaginare che anche i mesi successivi al primo lockdown siano stati condizionati da quei mesi...

Conclusione ed eventuali sviluppi futuri

Con questo progetto è stato possibile osservare lo sviluppo di sei attività in un periodo molto difficile, caratterizzato da situazioni impensabili fino a 4/5 anni fa. La fase di analisi esplorativa ha permesso di conoscere al meglio queste attività, non solo per le loro caratteristiche (numeri di scontrini, lo scontrino medio pagato dai clienti...), ma anche per il contesto in cui hanno lavorato in questi anni (come la città, dunque legata alle restrizioni in base al momento della pandemia, i numeri di casi e vaccini, le feste programmate...). La fase di previsioni ha cercato di spiegare l'impatto economico che ha avuto la pandemia su queste attività, utilizzando un modello adatto alla situazione, al contesto e alle variabili a nostra disposizione.

Il seguente lavoro può essere ulteriormente ampliato in diversi modi:

- Creazione di modelli di serie temporali differenti (come i modelli UCM, o modelli di Machine Learning), in modo da individuare i modelli che permettono le previsioni migliori;
- Ampliamento delle informazioni a disposizione per ogni locale (come numero di dipendenti e di coperti, tipologia di ristorante, se legato ad altre attività come hotel o b&b) in modo da avere un quadro ancora più completo.
- Utilizzo dei modelli di serie temporali per fare previsioni anche sui guadagni nel periodo post-Covid;
- Ampliamento dei dati a disposizione, con l'integrazione del dataset con i dati successivi al 30 aprile 2022, anche per verificare l'eventuale efficacia dei modelli predittivi implementati;
- Ampliamento dei dati a disposizione aggiungendo informazioni relative ad altri ristoranti, magari anche di zone differenti d'Italia, per analizzare l'impatto della pandemia anche in contesti molto differenti da questo;

Indice figure

FIGURA 1: BOXPLOT VARIABILI SCONTRINI E LORDO_TOTALE PER RISTORANTE	4
FIGURA 2: BOXPLOT VARIABILE SCONTRINO_MEDIO PER RISTORANTE	5
FIGURA 3: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER RISTORANTE	6
FIGURA 4: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON DISTINZIONE FESTIVO/FERIALE	7
FIGURA 5: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER FESTIVITÀ	8
FIGURA 6: BOXPLOT VARIABILE SCONTRINO_MEDIO RISPETTO A RISTORANTE E PRE/POST COVID	9
FIGURA 7: ANDAMENTO INFLAZIONE	10
FIGURA 8: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON CONFRONTO PRE/POST COVID RISTORANTE 1, ESCLUSI GIORNI CON 0 SCONTRINI	11
FIGURA 9: GRAFICI VARIABILI SCONTRINI E LORDO_TOTALE RISTORANTE 1	11
FIGURA 10: SCATTERPLOT LORDO_TOTALE SCONTRINI RISTORANTE 1	11
FIGURA 11: GRAFICI VARIABILI SCONTRINI E LORDO_TOTALE RISTORANTE 2	12
FIGURA 12: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON CONFRONTO PRE/POST COVID RISTORANTE 2, ESCLUSI GIORNI CON 0 SCONTRINI	13
FIGURA 13: SCATTERPLOT LORDO_TOTALE SCONTRINI RISTORANTE 2	13
FIGURA 14: GRAFICI VARIABILI SCONTRINI E LORDO_TOTALE RISTORANTE 3	14
FIGURA 15: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON CONFRONTO PRE/POST COVID RISTORANTE 3, ESCLUSI GIORNI CON 0 SCONTRINI	14
FIGURA 16: SCATTERPLOT LORDO_TOTALE SCONTRINI RISTORANTE 3	14
FIGURA 17: GRAFICI VARIABILI SCONTRINI E LORDO_TOTALE RISTORANTE 4	15
FIGURA 18: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON CONFRONTO PRE/POST COVID RISTORANTE 4, ESCLUSI GIORNI CON 0 SCONTRINI	16
FIGURA 19: SCATTERPLOT LORDO_TOTALE SCONTRINI RISTORANTE 4	16
FIGURA 20: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON CONFRONTO PRE/POST COVID RISTORANTE 5, ESCLUSI GIORNI CON 0 SCONTRINI	17
FIGURA 21: GRAFICI VARIABILI SCONTRINI E LORDO_TOTALE RISTORANTE 5	17
FIGURA 22: SCATTERPLOT LORDO_TOTALE SCONTRINI RISTORANTE 5	18
FIGURA 23: GRAFICI VARIABILI SCONTRINI E LORDO_TOTALE RISTORANTE 6	18
FIGURA 24: BOXPLOT VARIABILI SCONTRINI E SCONTRINO_MEDIO PER GIORNO, CON CONFRONTO PRE/POST COVID RISTORANTE 6, ESCLUSI GIORNI CON 0 SCONTRINI	19
FIGURA 25: SCATTERPLOT LORDO_TOTALE SCONTRINI RISTORANTE 6	19
FIGURA 26: RACCOLTA VARIABILE LORDO_TOTALE SERIE TEMPORALI DI TUTTI I RISTORANTI	20
FIGURA 28: SERIE TEMPORALE LORDO_TOTALE RISTORANTE 1 E 2	20
FIGURA 29: SERIE TEMPORALE LORDO_TOTALE RISTORANTE 5 E 6	21
FIGURA 27: SERIE TEMPORALE LORDO_TOTALE RISTORANTE 3 E 4	21
FIGURA 30: SERIE TEMPORALE DIFFERENZIATA DEL LORDO_TOTALE RISTORANTE 1 E 2	21
FIGURA 31: SERIE TEMPORALE DIFFERENZIATA DEL LORDO_TOTALE RISTORANTE 3 E 4	22
FIGURA 32: SERIE TEMPORALE DIFFERENZIATA DEL LORDO_TOTALE RISTORANTE 5 E 6	22
FIGURA 34: PREVISIONI LORDO_TOTALE RISTORANTI 1 E 2, TRAMITE MODELLO ARIMA	22
FIGURA 35: PREVISIONI LORDO_TOTALE RISTORANTI 5 E 6, TRAMITE MODELLO ARIMA	23
FIGURA 33: PREVISIONI LORDO_TOTALE RISTORANTI 3 E 4, TRAMITE MODELLO ARIMA	23

Indice tabelle

TABELLA 1: STATISTICHE VARIABILE SCONTRINI PER RISTORANTE	4
TABELLA 2: STATISTICHE VARIABILE LORDO_TOTALE PER RISTORANTE	5
TABELLA 3: STATISTICHE VARIABILE SCONTRINO_MEDIO PER RISTORANTE	5
TABELLA 4: STATISTICHE VARIABILE SCONTRINO PER GIORNO	6
TABELLA 5: STATISTICHE VARIABILE SCONTRINO_MEDIO PER GIORNO	6
TABELLA 6: STATISTICHE VARIABILE SCONTRINO PER GIORNO, CON DISTINZIONE FESTIVO/FERIALE	7
TABELLA 7: STATISTICHE VARIABILE SCONTRINO_MEDIO PER GIORNO, CON DISTINZIONE FESTIVO/FERIALE	7
TABELLA 8: STATISTICHE VARIABILE SCONTRINI PER FESTIVITÀ	8
TABELLA 9: STATISTICHE VARIABILE SCONTRINO_MEDIO PER FESTIVITÀ	9
TABELLA 10: VARIABILE SCONTRINO_MEDIO RISPETTO A RISTORANTE E PRE/POST COVID	10
TABELLA 11: STATISTICHE DESCRITTIVE VARIABILE INFLAZIONE PER ANNO	10
TABELLA 12: VARIABILE SCONTRINO_MEDIO PER ANNO	10

Riferimenti

<https://www.kaggle.com/datasets/arthurio/italian-vaccination>

https://github.com/imcatta/restrizioni_regionali_covid

https://ec.europa.eu/eurostat/databrowser/view/PRC_HICP_MANR_custom_6312991/default/table?lang=en

Appunti del corso di Streaming Data Management and Time Series Analysis (LM in Data Science)

<https://towardsdatascience.com/time-series-analysis-with-auto-arma-in-r-2b220b20e8ab>

<https://otexts.com/fppit/stationarity.html>

<https://r-graph-gallery.com/index.html>

<https://pandas.pydata.org/docs/index.html>

<https://python-graph-gallery.com/>