Analisi stagionalità delle vendite del negozio di strumenti musicali "Toni&Suoni"

Università degli studi di Udine Facoltà di IBML Corso Fondamenti di Scienza dei Dati e Laboratorio [MA0682]

Studente [MAT159263]
Fabio Massimo Lattanzio
Anno accademico 2022- 2023

INTRODUZIONE

- In questo progetto verrà sviluppata l'analisi dei dati di vendita, di un piccolo negozio di strumenti musicali, situato all'interno di una scuola di musica. Il negozio nasce alla fine del 2018 e, oltre alla vendita di strumenti musicali e di tutti i relativi accessori, vende libri musicali (metodi, partiture, canzonieri, quaderni di musica, etc.) e fornisce servizi di liuteria (cambio corde, pulizia dello strumento, accordatura, assemblaggio, sostituzione di meccaniche, etc.).
- Tutti i dati a disposizione, sia relativi al magazzino sia relativi alle vendite del negozio, sono raccolti in un software, utilizzato per la gestione del negozio
- Il mio progetto si focalizzerà sull'andamento delle vendite negli anni di osservazione 2019, 2020, 2021, 2022 fino a maggio 2023, con focus su:
- Andamento delle vendite per anno con particolare attenzione a periodi economici contingenti (covid e chiusure forzate)
- Stagionalità delle vendite per periodo quali:
 - 1. apertura anno scolastico;
 - 2. festività (Natale e Pasqua);
 - 3. Periodo estivo e chiusura dell'attività scolastica
- Andamento delle vendite distinta per CATEGORIA (serie: Accessori, Audio, Libri, Liuteria, Strumenti musicali)

ANALISI DEL DATASET

I dataset estrapolati dal programma di gestione del magazzino MaestroGold sono 2:

- 1. Magazzino (Movimenti da gennaio 2019 a maggio 2023)
- 2. Fatturato (Vendite da gennaio 2019 a maggio 2023)

Ho importato i due file .xls in R e li ho salvati nella cartella E:\Data Science\Progetto\Data

I risultati sono salvati nella cartella E:\Data Science\Progetto\Source

Utilizzo il comando str(Dataset) che restituisce informazioni sulla struttura dei dataset.

str(Magazzino)

```
tibble
                  [2,035
                                                                                  tbl df/tbl/data.frame)
                                                                 (S3:
                                               "AROMA00AG15A"
                                                                               "0278" "TCL019271"
   Codice
                                                                "EWAVE00GC10"
                 : chr [1:2035] "AMPLIFICATORE AROMA PER ACUSTICA AG15 15 WATT" "AMPLIFICATORE CHIT.
 $ Descrizione
ELETTRICA EWAVEOOGC10" "BLACKSTAR MINI AMPLI 3W CON ALIMENTATORE" "GUITAR EXAM INITIAL 2020 2023 TRINITY
COLLEGE
                                                  LONDON"
 $ Categoria
                                         "AMPLIFICATORI" "AMPLIFICATORI"
      Ultimo
                                          [1:2035]
                                                      81.15
                                                                          73.77
                 Prezzo
                                            [1:2035]
                                                                        12
     Ouantità
                                                                        130.3
      Totale
                                       [1:2035]
                                                              114.8
                                                                                  65.4
                  Importo:
                               num
                                                    162.3
                                                    [1:2035]
                                                                              86.3
     Utile
                                             num
   Ult. Carico
                        : chr [1:2035] "20/07/2020"
                                                          "20/07/2020"
                                                                        "22/08/2018"
$ ult. scarico : chr [1:2035] "11/12/2020" "17/10/2020" "25/10/2022" "15/09/2022" ...
```

Il dataset Magazzino contiene il numero di articoli in giacenza con indicazione della data ultimo carico, scarico, ultimo prezzo di vendita e relativo utile. Le dimensioni del dataset Magazzino sono: 2035 righe (osservazioni) e 9 colonne (variabili). Non ci sono valori mancanti.

ANALISI DEL DATASET

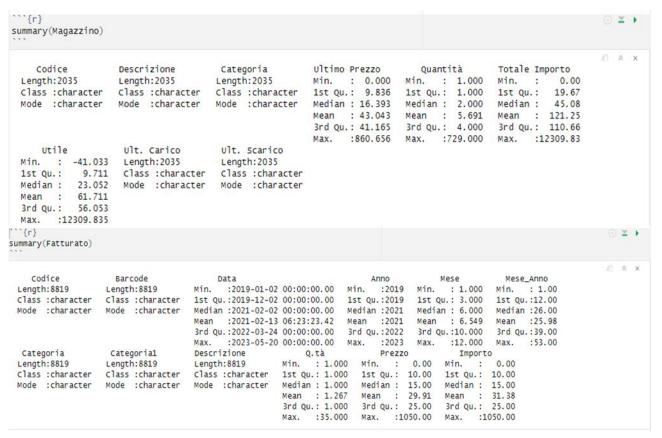
str(Fatturato)

```
tibble
                   [8,819
                                                                                    tbl_df/tbl/data.frame)
      Codice
                                               [1:8819]
                  : chr [1:8819] "8000000109576"
   Barcode
                                                    "800000112620" "9002761038798"
                   : POSIXct[1:8819], format:
   Data
                                                                            "2019-01-02" "2019-01-02"
                                                   [1:8819]
      Anno
                                              [1:8819]
     Mese
                                     [1:8819]
      Mese_Anno
                                     [1:8819]
                                                                 "ACCESSORI"
      Categoria
                             chr
                                                  "ACCESSORI"
                                                                                 "AUDIO"
                                                                                            "LIBRI"
                                                "ACCESSORI"
      Categoria1
                           chr
                                   [1:8819]
                                                                "ACCESSORI"
                                                                                "AUDIO"
                                                                                            "LIBRI"
                                                                      "CUFFIA"
                                                                                 "METODO
                                   "BACCHETTE
                                                     "BACCHETTE
                                                                                          PIANOFORTE"
    Descrizione:
     Q.tà
                                       num
                                              [1:8819]
              : num [1:8819] 15 2 55 22 12 98 13 3 110 5 ...
       Prezzo
                                                       [1:8819]
                                                                                 55
                                                                                         22
                                                                                                12
 $ Importo
```

Il dataset Fatturato contiene pezzi e importi degli articoli venduti giornalmente distinti per categoria.

Le dimensioni del dataset sono: 8819 righe (osservazioni) e 12 colonne (variabili). Non ci sono valori mancanti.

Utilizzo la funzione summary(Dataset) per vedere la struttura dei miei 2 dataset.



La mia analisi si limita a verificare l'andamento nel tempo delle vendite pertanto analizzerò il dataset Fatturato.

Le dimensioni del dataset sono: 8819 righe (osservazioni) e 12 colonne (variabili). Non ci sono valori mancanti.

Conduco le analisi descrittive delle variabili contenute nel dataset: le variabili Q.tà, Prezzo e Importo sono variabili quantitative, potrò quindi calcolare gli indici di posizione.

```
> summary(Fatturato$Q.tà)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  1.000  1.000  1.000  1.267  1.000  35.000
```

La variabile Q.tà assume come valore minimo 1 e come valore massimo 35; la mediana è 1 e la media è 1,267. Sembrerebbe che per la maggior parte delle vendite è richiesto 1 solo articolo, con qualche eccezione.

```
> summary(Fatturato$Prezzo)
   Min. 1st Qu. Median Mean 3rd Qu. Max.
   0.00 10.00 15.00 29.91 25.00 1050.00
```

La variabile Prezzo (espressa in Euro) assume come valore minimo 0 e come valore massimo 1050,00€; la media è 29,91€ e la mediana è 15,00€.

```
> summary(Fatturato$Importo)
  Min. 1st Qu. Median Mean 3rd Qu. Max.
  0.00 10.00 15.00 31.38 25.00 1050.00
```

La variabile Importo (espressa in Euro) assume come valore minimo 0 e come valore massimo 1050,00€; la media è 31,38€ e la mediana è 15,00€. La differenza tra la variabile Prezzo e la variabile Importo è dovuta al fatto che la prima descrive il prezzo di ogni singolo articolo, la seconda moltiplica il numero degli articoli venduti per il loro prezzo di vendita.

Sia la variabile Prezzo, che la variabile Importo, assumo come minimo il valore 0.

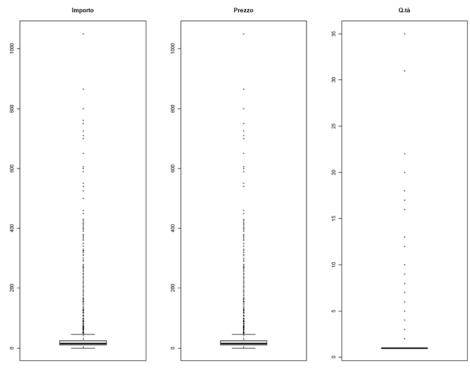
Il fenomeno è dovuto a come sono riportati i dati nel dataset Fatturato che presenta il "mese 16" corrispondente ad aprile 2020, con le variabili Prezzo e Fatturato pari a 0. Si tratta del mese di chiusura totale del negozio per le restrizioni dovute al Covid. Nelle successive analisi sulla stagionalità delle vendite verificherò se la chiusura del negozio ha influito negativamente sul fatturato totale.

Osservo che media e mediana non coincidono nei vari summary quindi realizzo dei boxplot par(mfrow=c(1,3))

boxplot(Fatturato\$Importo, main="Importo")

boxplot(Fatturato\$Prezzo, main="Prezzo")

boxplot(Fatturato\$Q.tà, main="Q.tà")



I Boxplot evidenziano l'asimmetria della distribuzione delle variabili.

Studio la distribuzione della variabili quantitative Q.tà e Fatturato:

par(mfrow=c(1,3))

hist(Fatturato\$Q.tà)

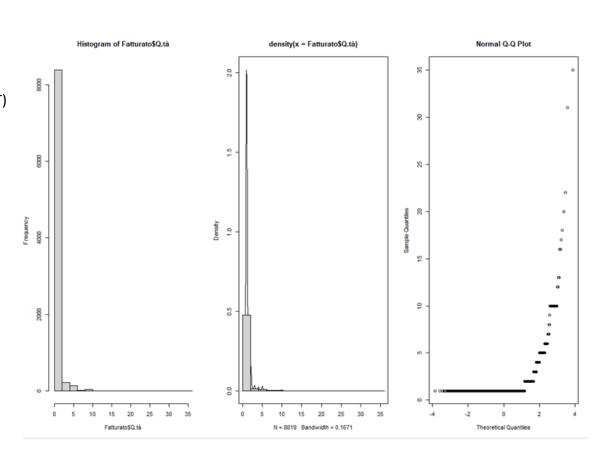
plot(density(Fatturato\$Q.tà))

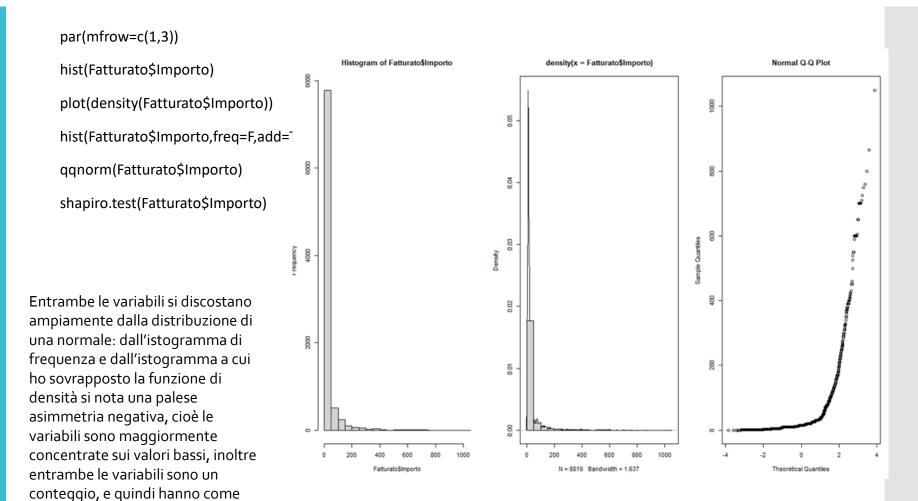
hist(Fatturato\$Q.tà,freq=F,add=T)

qqnorm(Fatturato\$Q.tà)

shapiro.test(Fatturato\$Q.tà)

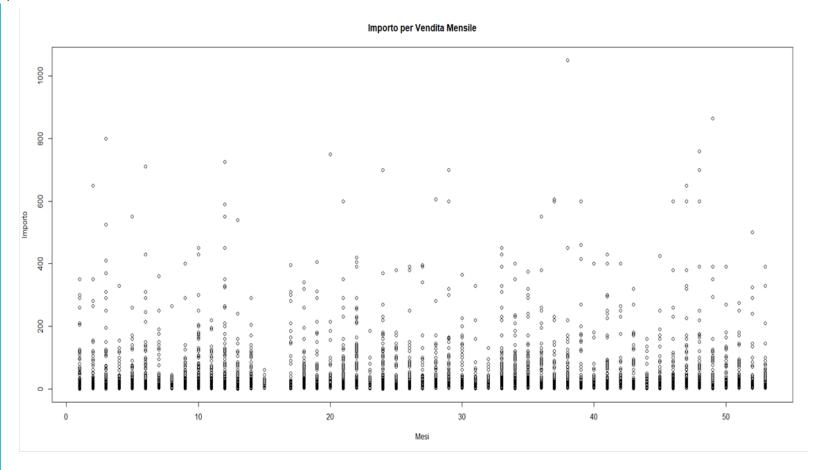
I Boxplot evidenziano l'asimmetria della distribuzione delle variabili.





supporto gli interi non negativi. Dal QQ-plot noto di nuovo che sia la variabile Q.tà che la variabile Importo non assumono le determinazioni di una normale. Conduco le analisi descrittive delle variabili contenute nel dataset Fatturato: plot(Fatturato\$Mese_Anno, Fatturato\$Importo)

ANALISI DEL DATASET Analisi descrittiva bivariata



Approfondisco quindi l'analisi delle vendite andando a guardare la variabile Categoria che suddivide gli articoli del negozio in 5 serie: Accessori, Audio, Libri, Liuteria e Strumenti.

Le informazioni degli indici sulla variabile Mese_Anno non sono significative. Analizzando un periodo di 53 mesi (corrispondenti a 4 anni e 5 mesi) si noterà infatti che media e mediana si posizionano al 27 mese.

```
> summary(Accessori)
                                                                            > summary(Libri)
                                    > summary(Audio)
Q.tà
               Fatturato
                                                                             Q.tà
                                                                                           Fatturato
                                      Q.tà
                                                  Fatturato
                                                                                              Min.
Min.
                 Min.
                                                     Min.
                                                                             Min.
                                                                                   : 0.00
                                                                                                    : 0.0
                                      Min.
                                            : 0.0
1st Qu.: 85.0
                 1st Qu.:1169
                                                                             1st Qu.: 7.00
                                                                                             1st Qu.:118.9
                                     1st Qu.:24.0
                                                     1st Qu.: 609.0
                                                                             Median :15.00
Median :121.0
                  Median :1748
                                                                                              Median :224.5
                                      Median:32.0
                                                     Median : 812.0
        :125.4
                 Mean
                         :1624
                                                            : 931.8
                                                                                   :17.96
                                                                                              Mean
                                                                                                     :254.6
                                            :34.4
                                                                             Mean
                                      Mean
                                                     Mean
3rd Ou.:157.0
                 3rd Qu.:2005
                                                                             3rd Ou.:24.50
                                     3rd Qu.:42.0
                                                                                              3rd Ou.:365.2
                                                     3rd Ou.:1306.0
        :290.0
                 Max.
                         :3240
                                                                                    :66.00
                                                                                                     :813.0
                                                                                              Max.
                                      Max.
                                             :95.0
                                                     Max.
                                                             :2834.0
> summary(Liuteria)
                                      > summary(Strumenti)
0.tà
               Fatturato
                                              Fatturato
        : 0.00
                 Min.
Min.
                        : 0.0
                                       Min.: 0.00 Min.: 0
1st Qu.:13.00
                 1st Qu.: 216.0
                                       1st Qu.: 7.00 1st Qu.:1100
Median :19.00
                 Median : 400.0
                                       Median: 13.00 Median: 1776
        :19.15
                 Mean
                        : 438.1
                                       Mean :14.58 Mean :1983
3rd Qu.:25.00
                 3rd Ou.: 583.0
                                       3rd Qu.:18.00 3rd Qu.:2669
        :40.00
                 Max.
                         :1038.0
                                       Max. :53.00 Max. :6565
```

Più interessanti sono gli indici di Q.tà e di Fatturato definiti su tutte le serie, nel periodo di osservazione considerato.

Per ogni serie, per mese, si vede il valore minimo e massimo (Q.ta e Fatturato), la media, la mediana e i quartili.

Sono informazioni già interessanti, ad esempio osservando la serie Accessori si evidenzia che il valore massimo di Fatturato per mese è stato di 3240€ e il numero massimo di Accessori venduti in un mese è stato di 290 articoli.

La mediana ci dice che i clienti del negozio nel mese, acquistano per la maggior parte 121 Accessori per un Importo pari a 1748€.

Analogamente possiamo discutere tutte le altre serie.

Osservando la serie Strumenti si evidenzia che il valore massimo di Fatturato per mese è stato di 6565€ e il numero massimo di strumenti venduti in un mese è stato di 53 articoli.

La mediana ci dice che i clienti del negozio nel mese, acquistano per la maggior parte 13 Strumenti per un Importo pari a 1776€. Le mediane del Fatturato/mese della serie Accessori e della serie Strumenti sono molto vicine. È evidente che gli Strumenti costano circa 9 volte più degli accessori, ma la tendenza mediana del Fatturato per mese è la stessa.

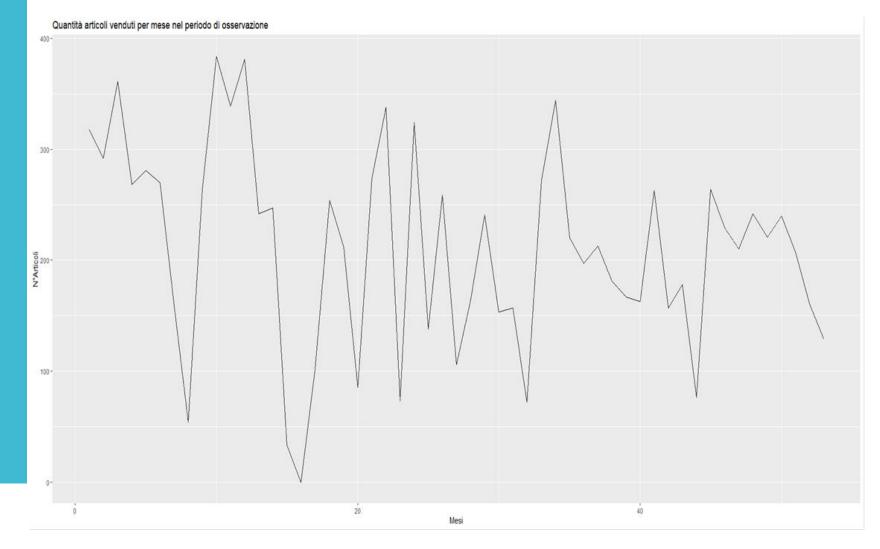
SERIE STORICHE

Una serie storica è una sequenza di osservazioni ordinate rispetto al tempo. Nel mio dataset Fatturato ho il fatturato giornaliero – mensile, la quantità degli articoli venduti giornaliera e mensile distinti per 5 serie di articoli.

Lo scopo dell'analisi delle serie storiche consiste nello studio dell'evoluzione passata del fenomeno rispetto al tempo; la previsione viene ottenuta ipotizzando che tali regolarità di comportamento si ripetano nel futuro. Per prima cosa decido di visualizzare i dati mediante una rappresentazione grafica. Lo scopo è quello di individuare eventuali regolarità di comportamento che sono utili nel suggerire l'approccio modellistico. Il grafico più semplice è il cosiddetto line plot che consiste nella rappresentazione dei dati rispetto al tempo.

1. Serie Q.tà

Analizzo la variabile Q.tà
ggplot(data = TotSum, aes(x = Mese_Anno, y = Q.tà)) +
geom_line() + ggtitle("Quantità articoli venduti per mese nel periodo di osservazione") +
xlab("Mesi") + ylab("N°Articoli")



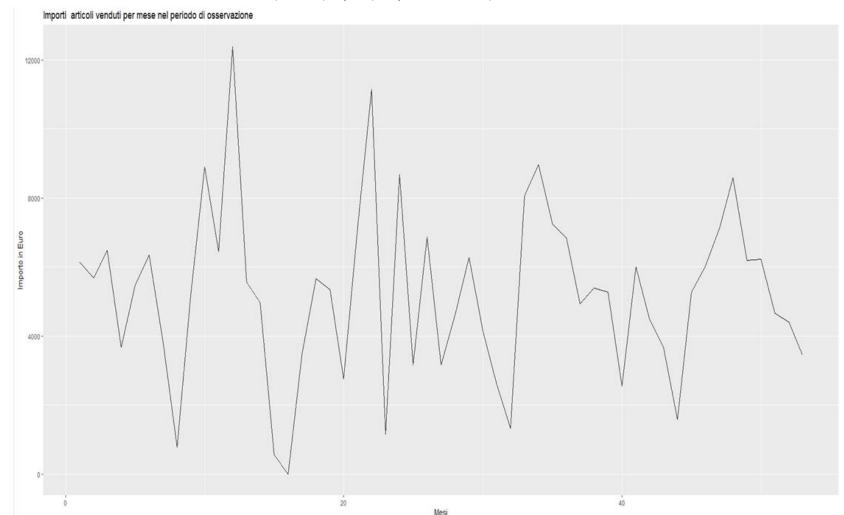
SERIE STORICHE

In entrambi i grafici si evidenzia una andamento non lineare; essendo i dati a cadenza mensile, in essi è presente il fenomeno denominato stagionalità. Si può notare che i picchi elevati si registrano sempre nei mesi caldi (novembre e dicembre) e vicini alle feste natalizie.

Si nota inoltre che entrambe le variabili hanno un andamento nel tempo molto simile e pur essendo grandezze di misura diverse i due grafici hanno degli andamenti comuni.

2. Serie Importo

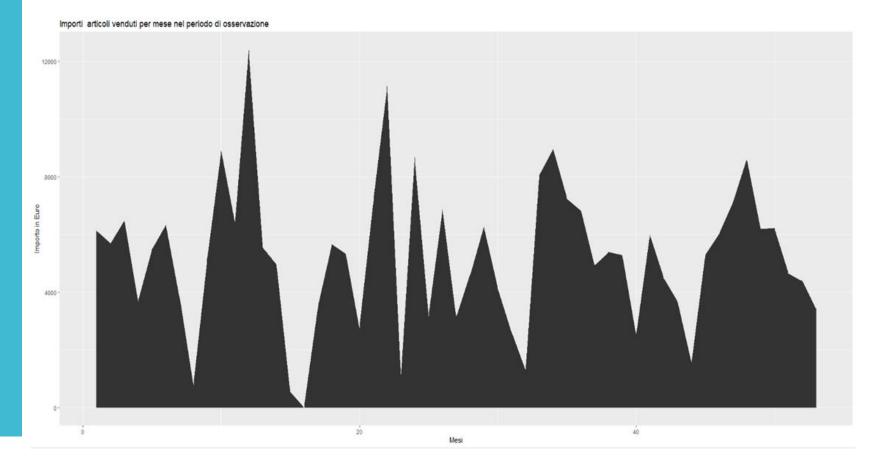
Analizzo la variabile Importo
ggplot(data = TotSum, aes(x = Mese_Anno, y = Fatturato)) +
geom_line() + ggtitle("Importi articoli venduti per mese nel periodo di osservazione") +
xlab("Mesi") + ylab("Importo in Euro")



SERIE STORICHE

Una rappresentazione alternativa è quella data dal grafico sottostante che evidenzia l'area e rende più chiara l'eventuale stagionalità

```
ggplot(data = TotSum, aes(x = Mese_Anno, y = Fatturato)) +
geom_area() + ggtitle("Importi articoli venduti per mese nel periodo di osservazione") +
xlab("Mesi") + ylab("Importo in Euro")
```



SERIE STORICHE Andamenti tipici delle serie storiche

Dai grafici delle serie storiche visti deduciamo che queste possono avere diversi pattern.

- 1. Pattern orizzontale. In questo caso la serie oscilla intorno ad un valore costante (media della serie). Tale serie è detta stazionaria in media. All'inizio della mia analisi abbiamo detto che sia la variabile Q.tà che la variabile Fatturato si posizionano molto più vicine al 1°q. della curva, pertanto questo non è il nostro caso.
- 2. Pattern stagionale. Questo esiste quando la serie è influenzata da fattori stagionali (es. mensile, semestrale, trimestrale, ecc.). Le serie influenzate dalla stagionalità sono dette anche serie periodiche poiché il ciclo stagionale si ripete in un periodo fisso. Nei dati di tipo annuale la stagionalità non è presente. Dai grafici delle serie storiche vediamo che le variabili Q.tà e Fatturato presentano delle punte principalmente nei mesi 12, 24, 36 e 48. È evidente la stagionalità.
- 3. Pattern ciclico. Questo tipo di andamento è presente quando la serie presenta aumenti e diminuzioni che non sono di periodo fisso. Questa è la principale differenza fra le fluttuazioni cicliche e quelle stagionali. Inoltre, l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità. Il pattern ciclico è determinato dalle espansioni e contrazioni dell'economia dovuti a fenomeni congiunturali. Nei grafici delle serie storiche notiamo alcune oscillazioni cicliche che andremo ad analizzare. (vedi la serie Liuteria)
- 4. Trend o tendenza di fondo. È caratterizzato da un andamento crescente o decrescente di lungo periodo. Vedremo di seguito analizzando le diverse serie della variabile Categoria i diversi tipi di pattern.

Vado alla ricerca di una qualche relazione tra le variabili presenti nel dataset studiandole congiuntamente.

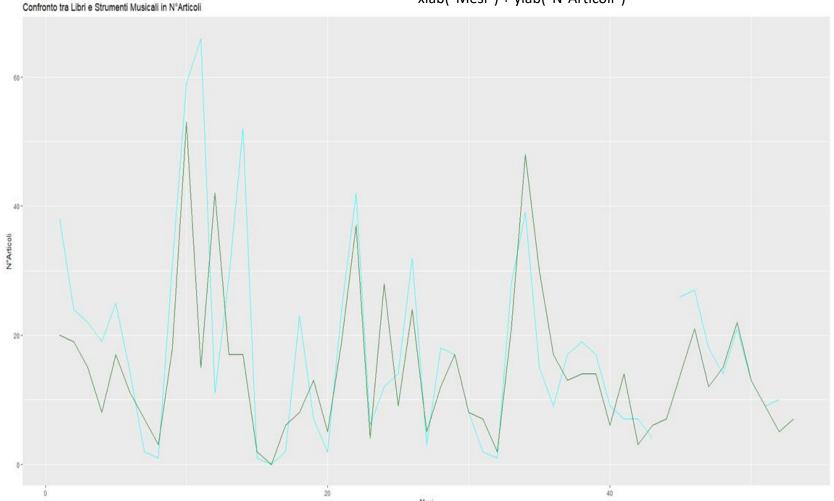
Per far ciò riporto sullo stesso grafico gli andamenti temporali di 2 variabili.

Nel 1° caso analizzo l'andamento Q.tà / mese della serie Libri confrontandola con la serie Strumenti.

Si nota che le due serie hanno un andamento molto simile, quasi sovrapponibile. In questo caso, è possibile ipotizzare che a fronte di ogni strumento acquistato, viene acquistato ugualmente un libro.

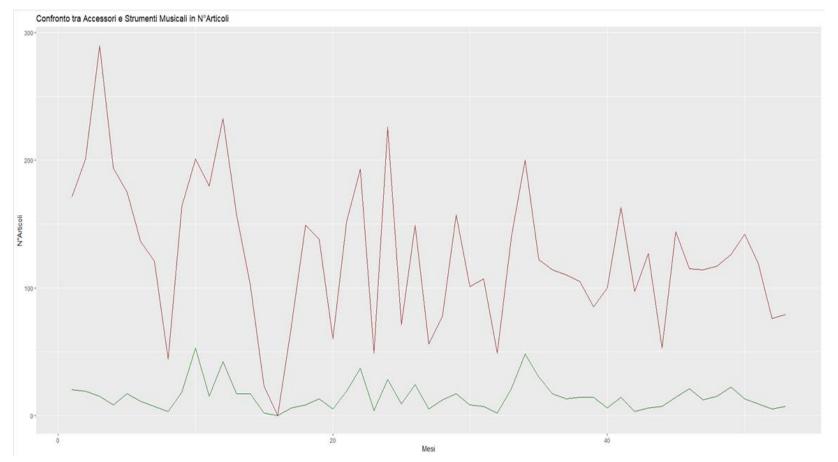
1. Relazione tra serie Libri e Strumenti

ggplot(CattPezz, aes(x=Mese_Anno)) +
 geom_line(aes(y = LIBRI), color = "cyan") +
 geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +
 ggtitle("Confronto tra Libri e Strumenti Musicali in N°Articoli") +
 xlab("Mesi") + ylab("N°Articoli")



2. Relazione tra serie Accessori e Strumenti Musicali

```
ggplot(CattPezz, aes(x=Mese_Anno)) +
geom_line(aes(y = ACCESSORI), color = "darkred") +
geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +
ggtitle("Confronto tra Accessori e Strumenti Musicali in N°Articoli") +
xlab("Mesi") + ylab("N°Articoli")
```

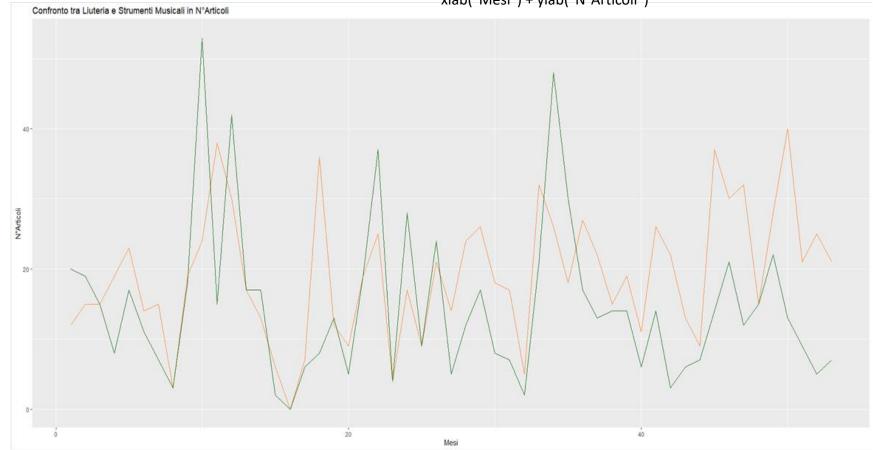


Questo caso è piuttosto particolare. La serie Liuteria ha un pattern stagionale, ma anche un pattern ciclico. Questo tipo di andamento è presente quando la serie presenta aumenti e diminuzioni che non sono di periodo fisso. Inoltre, l'ampiezza delle oscillazioni cicliche

Inoltre, l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità. Nelle serie economiche il pattern ciclico è determinato dalle espansioni e contrazioni dell'economia dovuti a fenomeni congiunturali.

3. Relazione tra serie Liuteria e Strumenti

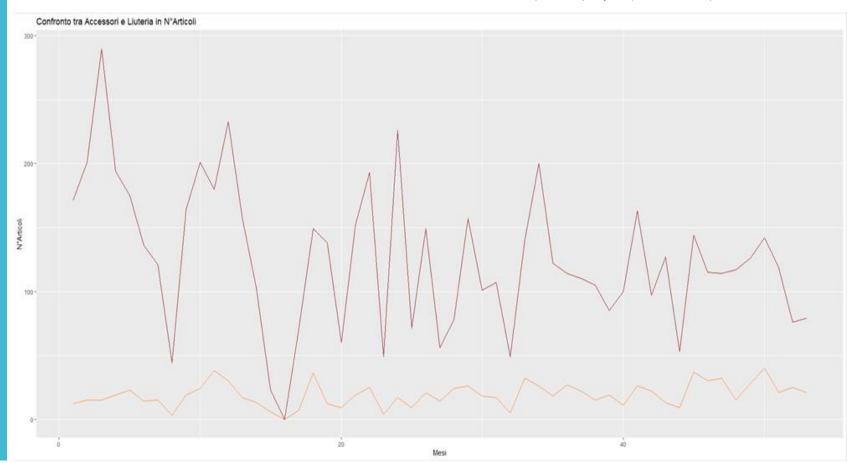
ggplot(CattPezz, aes(x=Mese_Anno)) +
 geom_line(aes(y = LIUTERIA), color = "chocolate1") +
 geom_line(aes(y = STRUMENTI_MUSICALI), color = "darkgreen") +
 ggtitle("Confronto tra Liuteria e Strumenti Musicali in N°Articoli") +
 xlab("Mesi") + ylab("N°Articoli")



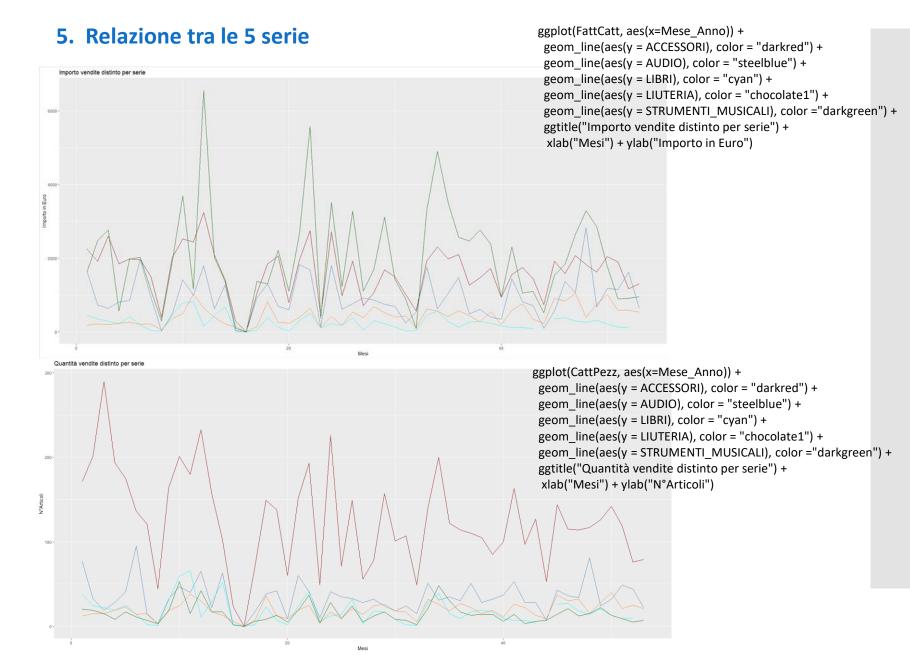
4. Relazione tra serie Accessori e Liuteria

ggplot(CattPezz, aes(x=Mese_Anno)) +
geom_line(aes(y = ACCESSORI), color = "darkred") +
geom_line(aes(y = LIUTERIA), color = "chocolate1") +
ggtitle("Confronto tra Accessori e Liuteria in N°Articoli") +
xlab("Mesi") + ylab("N°Articoli")

RELAZIONI TRA SERIE



Guardando sia il numero Q.tà che gli Importi per ogni sottogruppo della variabile Categoria si nota che il sottogruppo accessori è quello in assoluto più venduto, ma la caratteristica della stagionalità si rileva sempre in tutti i sottogruppi.



CONCLUSIONE

La mia analisi ha evidenziato che serie storiche delle variabili Q.tà e Fatturato presentano una combinazione di diversi pattern delle vendite.

Abbiamo visto che i grafici delle serie storiche delle variabili Q.tà e Fatturato presentano delle punte nei mesi 12, 24, 36 e 48 (dicembre - Festività Natalizie), delle punte leggermente più basse nei mesi 5, 17, 29, 41, 53 (maggio – Comunioni e Cresime) e delle cadute nei mesi 8, 20, 32, 44 (chiusura parziale del mese di agosto). E' evidente la stagionalità. Anche le 5 serie Audio, Accessori, Libri, Liuteria e Strumenti sono influenzate dal pattern stagionale.

La serie Liuteria è un esempio di pattern ciclico. Si vede che l'ampiezza delle oscillazioni cicliche è generalmente più grande di quella dovuta alla stagionalità.

La tendenza di tutte le serie è di tipo crescente, in particolar modo per la serie LIUTERIA, la serie LIBRI e la serie STRUMENTI MUSICALI.

Nessuna serie infine ha un pattern di fondo di tipo orizzontale.