

Analisi delle opinioni dei consumatori riguardo
l'implementazione di ulteriori servizi e
miglioramenti alla piattaforma

NETFLIX

ALBANO COSIMO 0022613
DE SIMONE FRANCESCA 0022726
GALEAZZI CHIARA 0022609
PLUTINO ANDREA 0023168

INDICE:

➤ LA STORIA	pagina 3
➤ COS'E' NETFLIX?	pagina 7
➤ NETFLIX IN CIFRE	pagina 8
➤ MERCATO DELLE PIATTAFORME DI STREAMING ONLINE	pagina 9
➤ ANALISI DEI COMPETITOR	pagina 11
➤ OBIETTIVI	pagina 12
➤ TARGET DI RIFERIMENTO	pagina 13
➤ ANALISI DEL QUESTIONARIO	pagina 16
➤ ANALISI DELLE VARIABILI RILEVATE	pagina 17
➤ ANALISI UNIVARIATA	pagina 25
➤ ANALISI BIVARIATA	pagina 38
➤ DIFFERENZIAZIONE DELL' OFFERTA	pagina 49
➤ ANALISI MULTIVARIATA (REGRESSIONE LINEARE)	pagina 53
➤ ANALISI MULTIVARIATA (REGRESSIONE LOGISTICA)	pagina 66
➤ CONCLUSIONE	pagina 83

NETFLIX: LA STORIA 1/4

Netflix è un'azienda che opera nella distribuzione via Internet di film, serie televisive e altri contenuti d'intrattenimento.

Fondata il 29 Agosto del 1997 da Reed Hastings e Marc Randolph a Scotts Valley in California.



NETFLIX: LA STORIA 2/4



La volontà di fondare Netflix nasce da Reed Hastings, ingegnere informatico laureato a Stanford che, a seguito della vendita della sua società Pure Software a Rational Software, decide di lanciarsi in una nuova avventura.

L'idea gli venne quando, dopo la ritardata restituzione del film Apollo 11, fu costretto a pagare una penale di 40\$.

Inizialmente Netflix voleva essere un servizio per il noleggio di DVD, VHS e videogiochi tramite posta: infatti bastava collegarsi al sito internet, scegliere il film preferito tra quelli disponibili e attendere l'arrivo del prodotto.

Tutto questo era possibile pagando un prezzo pari a **6\$**, comprese le spese di spedizione.

NETFLIX: LA STORIA 3/4

L'idea avuta inizialmente stenta però a decollare. Così Hastings decide di cambiare strada e di approdare ad una nuova formula, basata su un abbonamento mensile senza limiti che consentiva di poter scegliere 3 titoli di prodotti offerti, riceverli comunque via posta, restituirli e godere dello stesso servizio anche più volte al mese.

Si introduce quindi una tariffa mensile, viene revocata la multa per il ritardo e il costo di spedizione.

Grazie a questa nuova formula si avvia la piattaforma Netflix.

NETFLIX: LA STORIA 4/4



NETFLIX
STORY



Oggi si sceglie di vedere lo spettacolo che ci piace su qualsiasi schermo possibile e ormai con gli smartphone o i tablet è possibile farlo dappertutto. Tali opportunità sono sublimite grazie ai sottotitoli. L'internet tv è una tv che funziona molto meglio.

Reed Hastings, co-founder e CEO di Netflix.

Nel 2005 Netflix spediva **UN MILIONE** di DVD al giorno; nel 2009 aveva un catalogo di ormai **100.000** titoli solo su DVD e superato i **10 MILIONI di abbonati**.

Dopo la spedizione del miliardesimo DVD, nel febbraio del 2007, si decise per una svolta al servizio Netflix, affiancando al noleggio video una piattaforma per lo streaming video, utilizzando le stesse modalità. Sempre attraverso l'utilizzo di un abbonamento mensile fu dunque possibile guardare online, sul sito di Netflix, tutti i film, serie tv ed altri contenuti offerti dalla piattaforma stessa quando e dove si preferisce senza alcun limite.

COS'È NETFLIX?

Ad oggi Netflix è considerata l'azienda leader nel mercato dello **streaming online**.

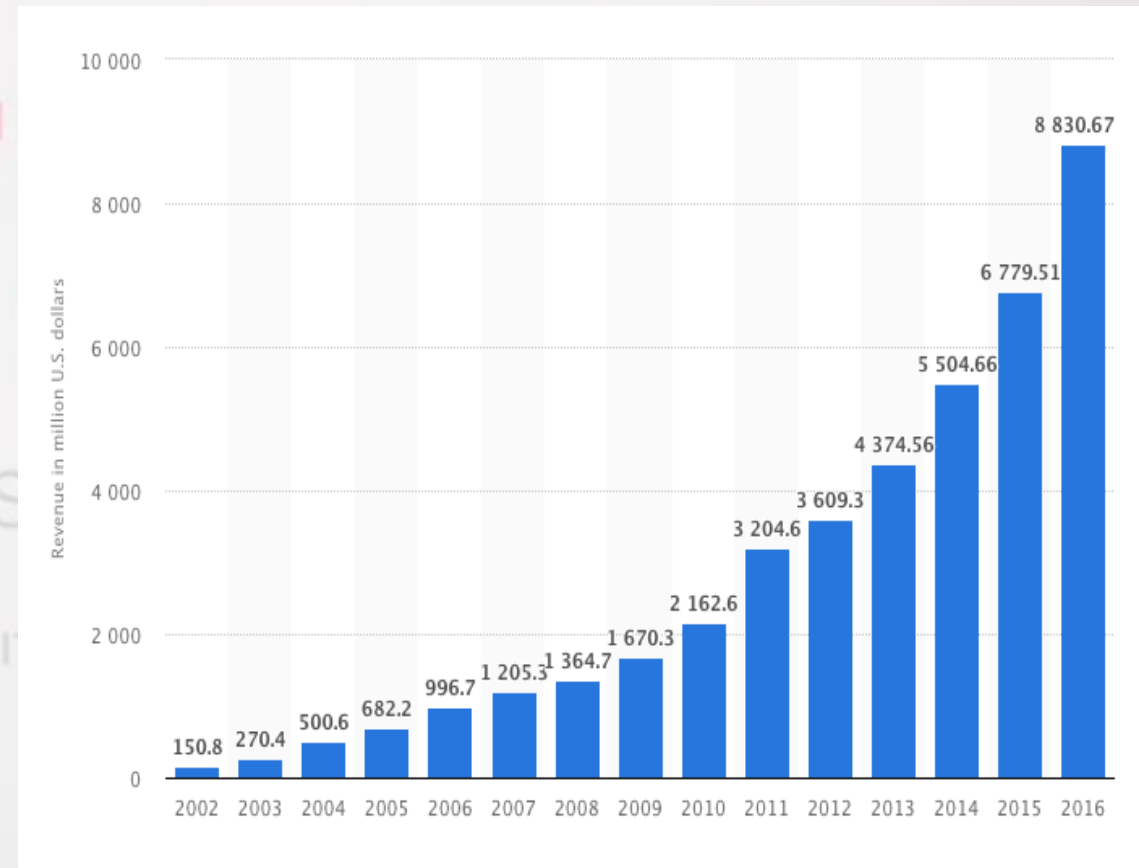
A seguito della registrazione alla piattaforma, il consumatore potrà usufruire di un mese di **prova gratuita**, dopo il quale potrà decidere se disdire l'abbonamento o sottoscriverlo per continuare ad usufruire del servizio online pagando un corrispettivo mensile.

Netflix conta **110 milioni** di utenti a fine settembre 2017.

La piattaforma presenta un'ampia offerta di contenuti: dalle serie TV ai film di ogni genere, dai documentari ai canali per bambini.

Da qualche anno l'offerta proposta dall'azienda contiene **contenuti** (serie TV, film) **originali**, ovvero prodotti da Netflix stesso, che nella maggior parte dei casi si sono rivelati di successo.

Il fatturato di Netflix è cresciuto dal 2002 in modo esponenziale:



The Netflix logo is displayed in white, bold, sans-serif capital letters on a red background at the top of a television screen. The screen shows a grid of movie and TV show thumbnails, including titles like 'Desperado', 'Jarhead', 'Half Baked', 'Requiem for a Dream', and 'Ronin'. The background of the entire image is a dark brick wall. In the foreground, a person's feet wearing blue socks and dark pants are visible, resting on a purple rug. A large white bowl of popcorn is in the lower center.The Netflix logo is positioned on the left side of a white rectangular box. The text 'NETFLIX IN CIFRE' is written in red, bold, sans-serif capital letters to the right of the logo.

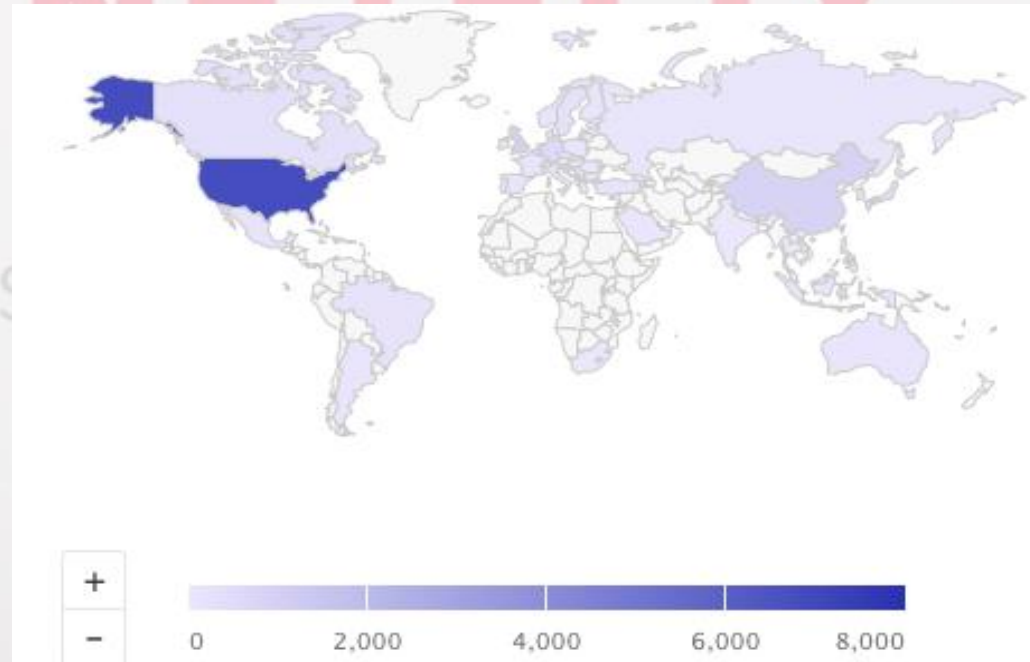
NETFLIX IN CIFRE

- **110 MLN** DI **UTENTI** WORLDWIDE
- **190 PAESI** IN CUI È PRESENTE
- **8,831 MLD** DI **REVENUES** (2016)
- **19 MLN** DI **ATTIVAZIONI** (2016)

MERCATO DELLE PIATTAFORME DI STREAMING ONLINE - MONDO

Il settore delle piattaforme di streaming online è in costante crescita. Nel 2017 ha generato un giro di affari per un totale di **US\$12,5billion**. Il mercato di riferimento prevede un tasso di crescita annuale dell'**8,3%** tra il 2017 e il 2022.

Da una globale prospettiva comparativa, è mostrato che gli Stati Uniti si attestano come mercato leader con un ricavo nel corso del 2017 di circa **US\$6,7billion**.



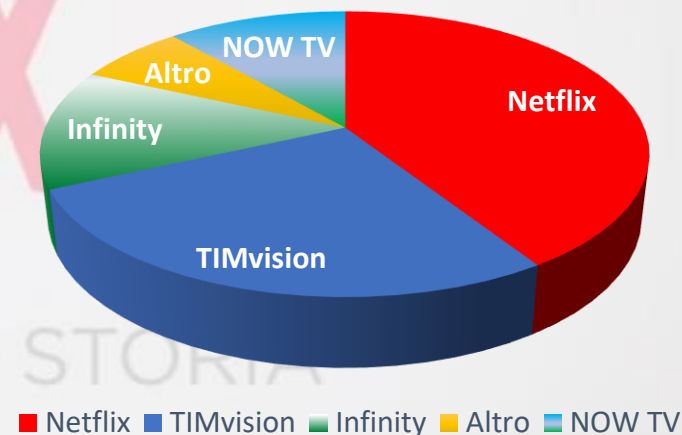
MERCATO DELLE PIATTAFORME DI STREAMING ONLINE - ITALIA

In Italia il fatturato del settore ad oggi risulta essere pari a US\$98million.

È prevista una crescita annuale del settore che si aggira intorno al **18,3%**, portando il fatturato a **US\$227million** nel 2022.

Una survey fatta dall'osservatorio EY e pubblicata nel mese d'ottobre 2017 ha rilevato i seguenti aspetti:

- Considerando i servizi OTT in Italia (Over The Top, ovvero tutti quelli veicolati con streaming online), gratis e pagamento, il numero di utenti a giugno 2017 è cresciuto di **2,9 milioni** rispetto alla rilevazione di settembre 2016.
- Gli utenti interessati ad usufruire dei servizi delle piattaforme di streaming online erano **4,3 milioni** a giugno 2017, con **1,9 milioni di abbonamenti attivi**.
- Si distingue Netflix: **800.000** sottoscrizioni a giugno 2017.



ANALISI DEI COMPETITOR



AMAZON PRIME VIDEO

- Prezzo: 20€ all'anno
- Offerta: contenuti di nicchia
- Economico e molto vantaggioso per utenti Prime
- Catalogo non completo



CHILI

- PayPerView
- Offerta *CINEMA* molto ampia
- Visione di pellicole in anteprima e ultime uscite
- Debole su film datati



INFINITY

- Prezzo: basso (5,99€-7,99€)
- Offerta *CINEMA* molto ampia
- Contenuti Mediaset
- Contenuti per *kids*
- Debole su serie TV e Documentari



NOW TV

- Prezzo: elevato (6,99€-19,99€)
- Offerta: contenuti in esclusiva
- Pacchetto sport acquistabile
- Possibilità di utilizzare 2 dispositivi/ HD con un supplemento



TIM VISION

- Prezzo: più vantaggioso (4,99€ - 10€)
- NO consumo traffico dati per clienti Tim
- Offerta limitata
- Servizi limitati
- Vantaggi per clienti TIM

OBIETTIVI

L'Azienda punta ad analizzare l'opinione dei consumatori cercando di individuare quali sono i punti di forza e debolezza del prodotto ed eventuali vuoti d'offerta/opportunità nell'area di Milano e provincia.

Le finalità dell'analisi sono da ricercarsi nel miglioramento dell'offerta del nostro servizio con eventuali implementazioni finalizzate ad una migliore fruibilità del prodotto per mantenere il vantaggio competitivo nel mercato.

In particolare l'Azienda vuole concentrarsi su:

- Analisi del profilo del consumatore;
- Analisi della willingness-to-pay;
- Analisi della conoscenza e utilizzo dell'HD e ULTRA HD;
- Sviluppo dell'offerta in termini di contenuti: eventi sportivi, concerti in differita e contenuti musicali;
- Sviluppo di BUSINESS ETHICS all'interno dell'azienda;
- Estensione dell'offerta di intrattenimento online nelle strutture turistiche.

TARGET DI RIFERIMENTO 1/3

La popolazione di riferimento presa in esame è rappresentata da un target con età compresa tra i 20 e i 60 anni, residente a Milano e provincia.

Questo al fine di verificare se le nostre proposte possono essere sviluppate nel territorio di riferimento e successivamente esportate su larga scala.

- Il limite massimo d'età fissato per la nostra analisi è di 60 anni in quanto la conoscenza del mondo digital e social, oltre i 60 anni appunto, potrebbe essere limitata.
- Il gap 20-60 è dovuto ad un'analisi mirata ad ottenere informazioni a 360° riguardanti le attitudini, le tendenze, i gusti e le preferenze di una popolazione eterogenea e stratificata.
- Si tratta di un target di genere maschile e femminile proprio perché queste piattaforme offrono contenuti rivolti ad entrambi i sessi.

Di conseguenza abbiamo proceduto con l'analisi dei dati ISTAT riguardanti la popolazione residente a Milano e provincia con età compresa tra i 20 e i 60 anni, confrontandola con il nostro campione di riferimento che comprende 200 persone.

TARGET DI RIFERIMENTO 2/3

POPOLAZIONE DI RIFERIMENTO

Età	Femmine	% Femmine	Maschi	% Maschi	Tot. Popolazione	% Popolazione
20-27	115128	13,06%	123331	13,97%	238459	13,52%
28-35	148705	16,86%	153713	17,42%	302418	17,14%
36-43	196591	22,30%	199251	22,58%	395842	22,44%
44-51	220042	24,96%	218205	24,72%	438247	24,84%
52-60	201288	22,83%	188044	21,31%	389332	22,07%
TOT	881754	100,00%	882544	100,00%	1764298	100,00%

NETFLIX.IT

Fonte ISTAT

TARGET DI RIFERIMENTO 3/3

CAMPIONE DI RIFERIMENTO

Età	Femmine	% Femmine	Maschi	% Maschi	Tot. Popolazione	% Popolazione
20-27	13	13%	14	14%	27	13,5%
28-35	17	17%	17	17%	34	17%
36-43	22	22%	23	23%	45	22,5%
44-51	25	25%	25	25%	50	25%
52-60	23	23%	21	21%	44	22%
TOT	100	100%	100	100%	200	100%

NETFLIX.IT

ANALISI DEL QUESTIONARIO

I NUMERI:

309 QUESTIONARI
SOMMINISTRATI

109 eliminati in quanto non rappresentativi la nostra popolazione di riferimento o svolti senza alcun criterio logico.

200 analizzati

MODALITA' DI
SOMMISTRAZIONE

230 Online Survey

79 Face to Face

→ DOMANDE COMPORTAMENTALI **19**

→ DOMANDE ATTITUDINALI **10**

→ DOMANDE ANAGRAFICHE **5**

ANALISI DELLE VARIABILI RILEVATE

NOME VARIABILE	SPIEGAZIONE VARIABILE	TIPOLOGA
CONOSC_PIATT	Conoscenza piattaforme	Quantitativa
QUALE_PIATT_1	Conosco "Netflix"	Quantitativa
QUALE_PIATT_2	Conosco "Amazon Prime"	Quantitativa
QUALE_PIATT_3	Conosco "Chili"	Quantitativa
QUALE_PIATT_4	Conosco "Mediaset Infinity"	Quantitativa
QUALE_PIATT_5	Conosco "SkyOnline Cinema"	Quantitativa
QUALE_PIATT_6	Conosco "Tim Vision"	Quantitativa
QUALE_PIATT_7	Conosco "Altro"	Quantitativa
QUALE_PIATT_8	Conosco "Nessuna"	Quantitativa

USA_PIATT	Usa le piattaforme	Dicotomica
USA_NETF	Usa Netflix	Quantitativa
CONTENUTI_PIATT_1	Uso le piattaforme per i "Film"	Quantitativa
CONTENUTI_PIATT_2	Uso le piattaforme per "Serie TV"	Quantitativa
CONTENUTI_PIATT_3	Uso le piattaforme per "Documentari"	Quantitativa
CONTENUTI_PIATT_4	Uso le piattaforme per "Contenuti per bambini"	Quantitativa
CONTENUTI_PIATT_5	Uso le piattaforme per "Sport"	Quantitativa
CONTENUTI_PIATT_6	Uso le piattaforme per "Notiziari e informazioni"	Quantitativa
CONTENUTI_PIATT_7	Uso le piattaforme per "Canali musicali"	Quantitativa
LINGUA	Guarda i contenuti in altre lingue	Quantitativa
MOTIVI_LINGUA_1	Lingua diversa per "Motivi di studio "	Quantitativa
MOTIVI_LINGUA_2	Lingua diversa per "Motivi di lavoro "	Quantitativa
MOTIVI_LINGUA_3	Lingua diversa per "Migliorare lingua"	Quantitativa
MOTIVI_LINGUA_4	Lingua diversa in quanto "Bilingue"	Quantitativa
MOTIVI_LINGUA_5	Lingua diversa per "Vedere contenuto in lingua originale"	Quantitativa

HD	Importanza Alta Definizione (HD)	Quantitativa
UHD	Usa l'ultra HD	Quantitativa
ELEM_IMP_PIATT_1	Importanza della "Connessione"	Quantitativa
ELEM_IMP_PIATT_2	Importanza dei "Contenuti"	Quantitativa
ELEM_IMP_PIATT_3	Importanza delle "condizioni meteo"	Quantitativa
ELEM_IMP_PIATT_4	Importanza "visione in ogni momento dei contenuti"	Quantitativa
DISP_PREF_1	Vedo contenuti con la "Tv"	Quantitativa
DISP_PREF_2	Vedo contenuti con il "Computer"	Quantitativa
DISP_PREF_3	Vedo contenuti con il "Tablet"	Quantitativa
DISP_PREF_4	Vedo contenuti con lo "Smartphone"	Quantitativa
WTP	Disponibilità a pagare per un servizio che comprende "Film", "Serie Tv", "Documentari" e "Contenuti per bambini"	Quantitativa
WTP_PLUS	Disponibilità a pagare un servizio con l'aggiunta di "Ultra HD", "Eventi sportivi", "Concerti in diretta" e "Contenuti musicali"	Quantitativa

TEMPO_PIATT_1	Uso la piattaforma il "Mattino"	Quantitativa
TEMPO_PIATT_2	Uso la piattaforma il "Pomeriggio"	Quantitativa
TEMPO_PIATT_3	Uso la piattaforma la "Sera"	Quantitativa
TEMPO_PATT_4	Uso la piattaforma la "Notte"	Quantitativa
QUALIT_PIATT_1	Importanza "Qualità/Prezzo "	Quantitativa
QUALIT_PIATT_2	Importanza "Qualità contenuti/Prezzo "	Quantitativa
QUALIT_PIATT_3	Importanza "Design"	Quantitativa
QUALIT_PIATT_4	Importanza "Facilità d'uso"	Quantitativa
QUALIT_PIATT_5	Importanza "Mese di prova gratuito"	Quantitativa
QUALIT_PIATT_6	Importanza "Contenuti originali"	Quantitativa
QUALIT_PIATT_7	Importanza "Programmi in esclusiva"	Quantitativa
DLOAD_CONTENUTO	Importanza sulla scelta d'acquisto di salvare un contenuto multimediale	Quantitativa
IMPORT_SOCIALE	Importanza sulla scelta d'acquisto l'investimento nel sociale	Quantitativa
HOTEL	Importanza uso piattaforme in strutture ospitanti	Dicotomica

IMPORT_HOTEL	Importanza piattaforme nella scelta della struttura ospitante	Quantitativa
BENEF_SOCIALE	Importanza sociale sostenibilità nella sua vita	Quantitativa
TEMPO_LIBERO	Tempo libero in una giornata	Quantitativa
ATTIVITA_TEMPO_LIB_1	Cosa faccio nel tempo libero "Cinema"	Quantitativa
ATTIVITA_TEMPO_LIB_2	Cosa faccio nel tempo libero "Sport"	Quantitativa
ATTIVITA_TEMPO_LIB_3	Cosa faccio nel tempo libero "Shopping"	Quantitativa
ATTIVITA_TEMPO_LIB_4	Cosa faccio nel tempo libero "Lettura"	Quantitativa
ATTIVITA_TEMPO_LIB_5	Cosa faccio nel tempo libero "Cucina"	Quantitativa
ATTIVITA_TEMPO_LIB_6	Cosa faccio nel tempo libero "Musica"	Quantitativa
TEMPO_LIB_PIATT	Quante ore passo a guardare contenuti multimediali nel tempo libero	Quantitativa
CINEMA	Volte che vado al cinema	Quantitativa
GENERI_MULT_1	Importanza del genere "Sentimentale"	Quantitativa
GENERI_MULT_2	Importanza del genere "Commedia"	Quantitativa

GENERI_MULT_3	Importanza del genere "Drammatico"	Quantitativa
GENERI_MULT_4	Importanza del genere "Thriller"	Quantitativa
GENERI_MULT_5	Importanza del genere "Avventura/Azione"	Quantitativa
GENERI_MULT_6	Importanza del genere "Horror"	Quantitativa
GENERI_MULT_7	Importanza del genere "Crime/Poliziesco"	Quantitativa
GENERI_MULT_8	Importanza del genere "Biografico/Documentari"	Quantitativa
GENERI_MULT_9	Importanza del genere "Fantasy"	Quantitativa
GENERI_MULT_10	Importanza del genere "Family "	Quantitativa
GENERI_MULT_11	Importanza del genere "Cinema Italiano"	Quantitativa
GENERI_MULT_12	Importanza del genere "Comico"	Quantitativa
TECH	Propensione uso strumenti tecnologici	Quantitativa

USO_TECH_1	Essere in grado di "Navigare in Internet"	Quantitativa
USO_TECH_2	Essere in grado di "Scaricare applicazioni"	Quantitativa
USO_TECH_3	Essere in grado di "Condividere contenuti"	Quantitativa
USO_SOCIAL	Quanto usa i social	Quantitativa
CASELLA_POSTA_1	Controlla la posta elettronica "Tutti i giorni"	Quantitativa
CASELLA_POSTA_2	Controlla la posta elettronica "2/3 giorni a settimana"	Quantitativa
CASELLA_POSTA_3	Controlla la posta elettronica "Una volta a settimana"	Quantitativa
CASELLA_POSTA_4	Controlla la posta elettronica "Poche volte al mese"	Quantitativa
CASELLA_POSTA_5	Controlla la posta elettronica "Mai"	Quantitativa

SESSO	Sesso	Qualitative
ETA	Età	Quantitativa
TIT_STUD	Titolo di studio	Qualitative
PROFESS	Professione	Qualitative
N_FAM	Nucleo familiare	Quantitativo

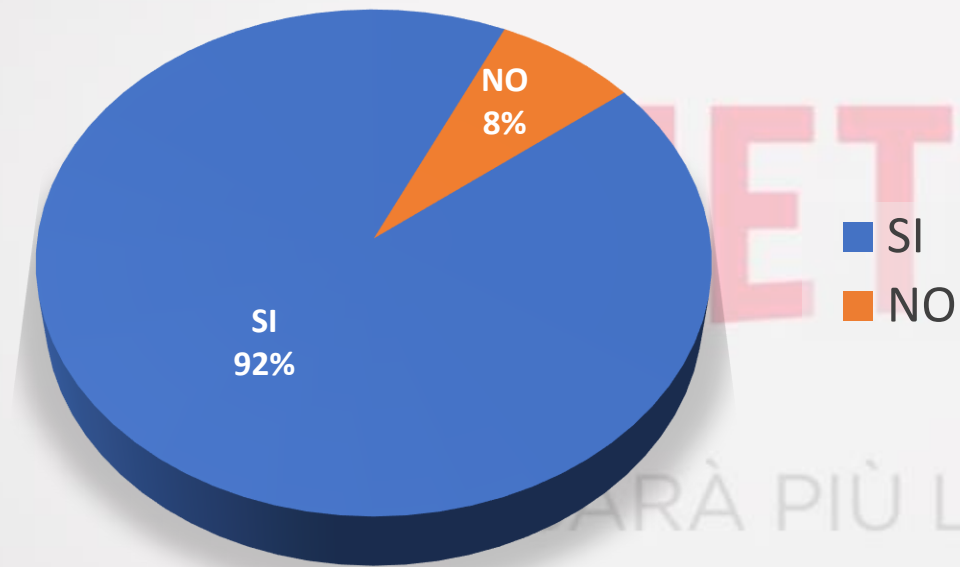
NETFLIX.IT

ANALISI UNIVARIATA

NON SARÀ PIÙ LA SOLITA STORIA

NETFLIX.IT

CONOSCENZA DELLE PIATTAFORME DI INTRATTENIMENTO ONLINE



netf\$CONOSC_PIATT		
	Frequency	Percent
0	15	7.5
1	185	92.5
Total	200	100.0

Dall'analisi della variabile dicotomica si osserva l'elevata conoscenza delle piattaforme di intrattenimento.

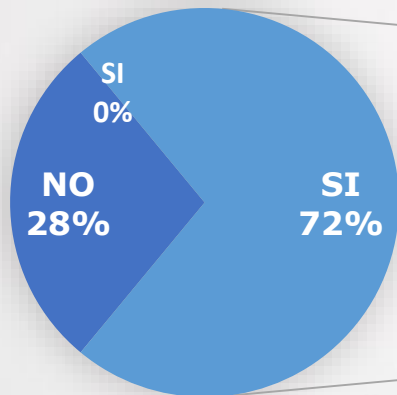
Questa conoscenza, come emerge dal grafico, si conferma molto elevata all'interno del nostro campione di riferimento.

Solo **l'8%** del nostro campione non conosce le Piattaforme di Intrattenimento Online, contro il **92%** di risposte affermative.

Quindi possiamo affermare che la maggioranza del nostro campione conosce effettivamente le piattaforme.

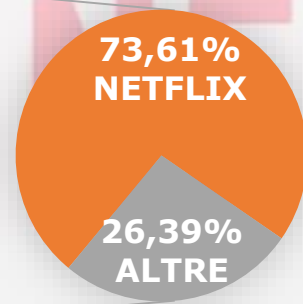
QUANTI USANO: - LE PIATTAFORME DI INTRATTENIMENTO ONLINE - QUANTI USANO NETFLIX

PIATTAFORME
IN GENERALE



■ NO ■ Netflix ■ Altre Piatt. ■ SI

NETFLIX



Dall'analisi di variabili entro categoria, effettuata attraverso l'uso della funzione *frequencyBy*, abbiamo voluto calcolare la percentuale di coloro che godono del servizio Netflix.

È emerso che :

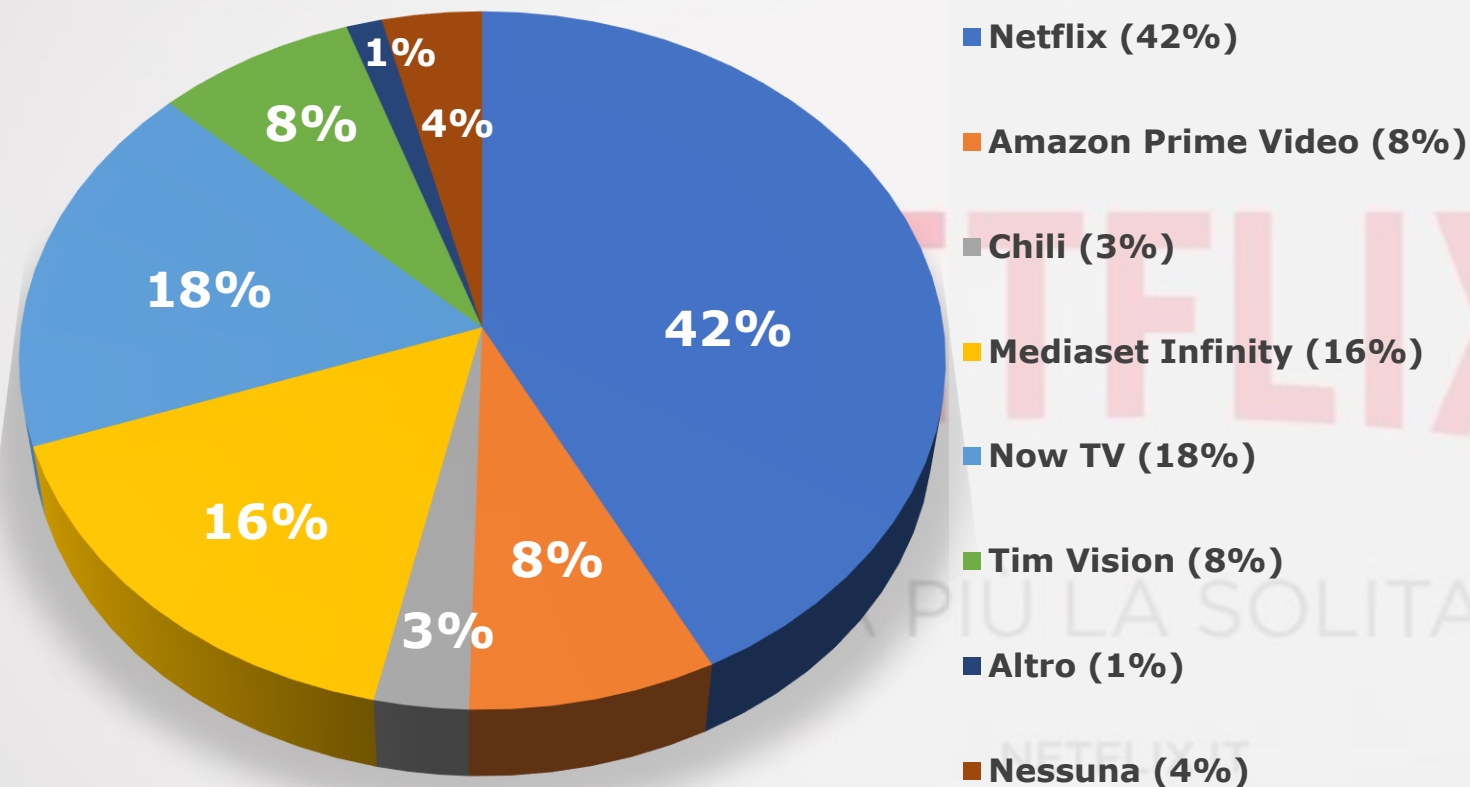
- Il **72%** del nostro campione usa le piattaforme di intrattenimento online.
- Di questo 72%, il **73,61%** usufruisce dei contenuti di Netflix.

\$`1`		
na.exclude(x)	Frequency	Percent
0	38	26.39
1	106	73.61
Total	144	100.00

\$`0`		
na.exclude(x)	Frequency	Percent
0	56	100
Total	56	100

Quindi possiamo affermare che il nostro servizio rappresenta uno dei principali leader di mercato nel settore delle piattaforme di intrattenimento online.

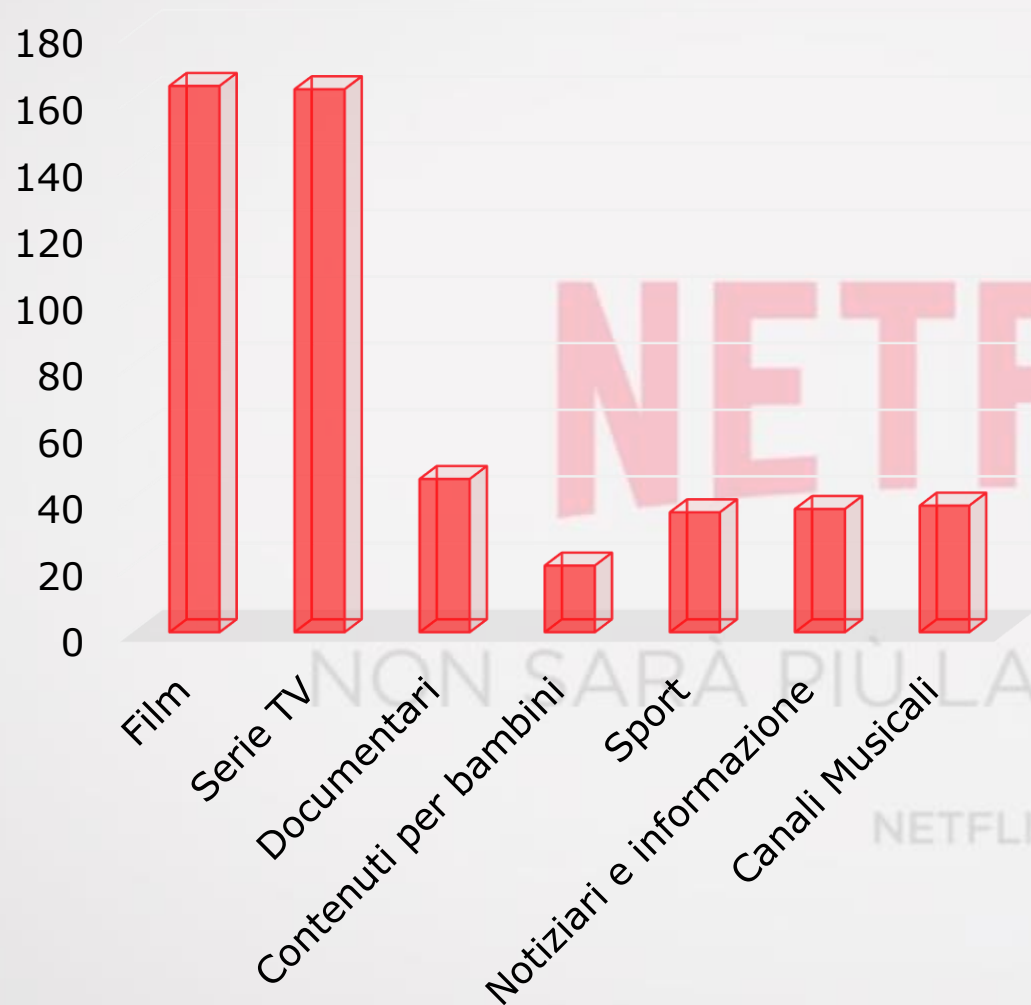
CONOSCENZA DELLE SINGOLE PIATTAFORME



Analisi della brand awareness:

si riconferma ancora una volta il primato di Netflix come leader del mercato nel settore di riferimento: il **42%** del nostro campione conosce la piattaforma Netflix, seguita a distanza da Now TV e Mediaset Infinity, rispettivamente **18%** e **16%** del tasso di conoscenza.

CONTENUTI PREFERITI

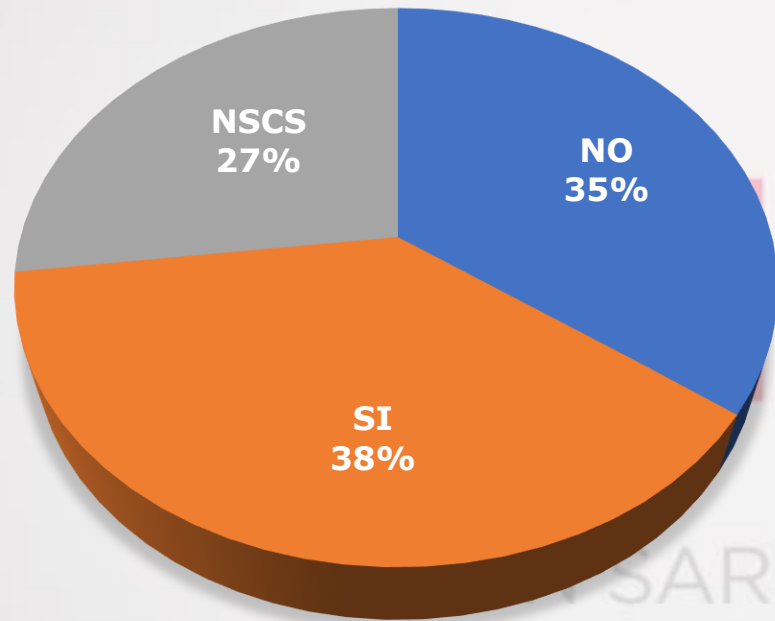


Valutando i contenuti effettivamente presenti e potenziali, possiamo affermare che l'offerta di Netflix è in linea con le richieste dei consumatori: film, serie TV, documentari e contenuti per bambini.

Dal grafico si evidenzia come sport, notiziari e canali musicali siano graditi al pubblico anche se con percentuali inferiori.

Dunque, questi contenuti potrebbero essere un possibile obiettivo da perseguire nel lungo periodo, ovvero aggiungendoli all'offerta della nostra piattaforma.

CONOSCENZA DELL'ULTRA HD



■ SI ■ NO ■ NSCS*

*NSCS = Non so cosa sia

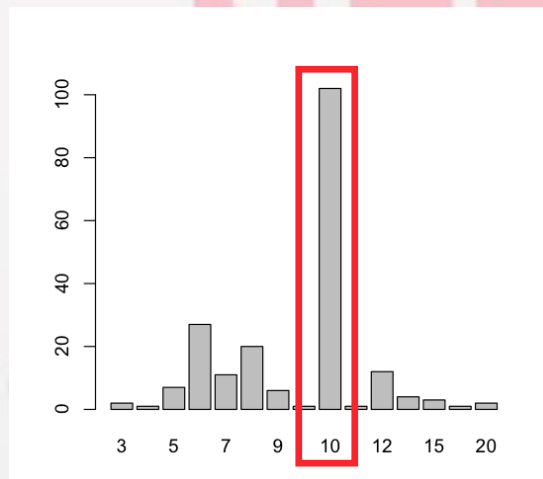
L'analisi proposta evidenzia come la funzione UHD offerta da Netflix non abbia ancora raggiunto un'elevata visibilità nel campione di riferimento, nonostante rappresenti un cambio di orizzonte in termini tecnologici.

Funzione innovativa, dal momento che lascia alle spalle l'HD Ready (720p) e i Full HD (1080p). Il maggior numero di pixel consente infatti di cogliere un dettaglio estremamente superiore.

Risulta quindi evidente un bisogno di sensibilizzazione dell'utente all'importanza dell'UltraHD, valore aggiunto della piattaforma.

WILLINGNESS TO PAY PER IL PACCHETTO BASE

```
> basicStats(netf$WTP)
X..netf.WTP
nobs      200.000000
NAs       0.000000
Minimum   3.000000
Maximum   20.000000
1. Quartile 8.000000
3. Quartile 10.000000
Mean      9.194500
Median    10.000000
Sum       1838.900000
SE Mean   0.176314
LCL Mean  8.846817
UCL Mean  9.542183
Variance  6.217306
Stdev     2.493453
Skewness  0.785459
Kurtosis  3.532271
```



Secondo i dati ricavati con gli outliers è emerso che la distribuzione del pacchetto base secondo la *willingness to pay* è **IPERNORMALE** ($\beta > 3$), infatti l'indice di Kurtosis risulta essere pari a **3.532271**.

L'indice di Skewness pari a **0.785** ($\gamma > 0$) evidenzia la presenza di una distribuzione **QUASI SIMMETRICA**.

Attraverso la *willingness to pay*, vogliamo analizzare quale sia la percezione del pacchetto base in termini di prezzo da parte dei consumatori.

Con riferimento alla **media**, possiamo affermare che è pari a **9.1945 €**, per quanto riguarda invece la **mediana** essa è pari a **10.00€**. Notiamo quindi la presenza di una **ASIMMETRIA NEGATIVA**. La **moda** è pari a **10 €**.

WILLINGNESS TO PAY PER IL PACCHETTO BASE

Valutando il boxplot della variabile WTP, possiamo osservare la presenza di outlier.

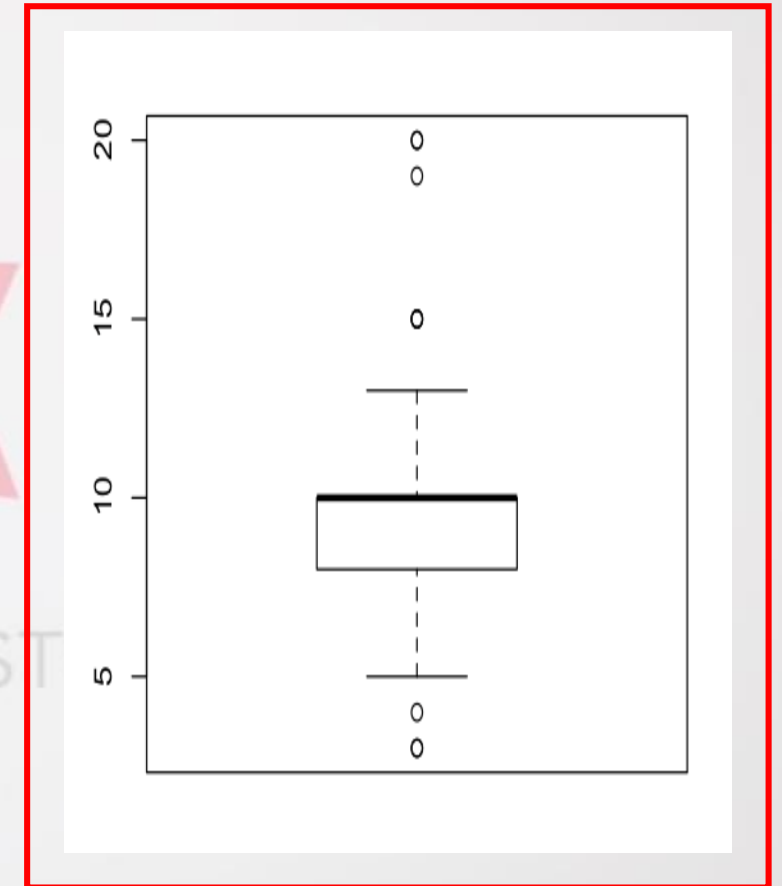
I valori outlier sono individuabili calcolando il minimo e il massimo teorici attraverso queste formule:

$$\text{Max} = [Q3 + 1,5(Q3 - Q1)]$$

$$\text{Min} = [Q1 - 1,5(Q3 - Q1)]$$

Effettuato il calcolo, come possiamo notare anche dal nostro boxplot, il valore massimo teorico è **13** mentre il nostro valore minimo teorico è **5**.

Eliminiamo dal dataset i valori outlier, che risultano essere **9**, e proseguiamo con l'analisi del dataset senza outliers.



```
> quantile(netf$WTP,c(.01,.05,.10,.25,.50,.75,.90,.95,.99))
 1%   5%   10%  25%  50%  75%  90%  95%  99%
3.99 5.95 6.00 8.00 10.00 10.00 12.00 12.05 19.01
```

WILLINGNESS TO PAY PER IL PACCHETTO BASE

NO OUTLIERS

```
> basicStats(na.exclude(netf$WTP))
na.exclude.netf.WTP
nobs      191.000000
NAs       0.000000
Minimum   5.000000
Maximum   13.000000
1. Quartile 8.000000
3. Quartile 10.000000
Mean      9.030890
Median    10.000000
Sum       1724.900000
SE Mean   0.138650
LCL Mean  8.757400
UCL Mean  9.304380
Variance  3.671725
Stdev     1.916175
Skewness  -0.490341
Kurtosis  -0.527328
```

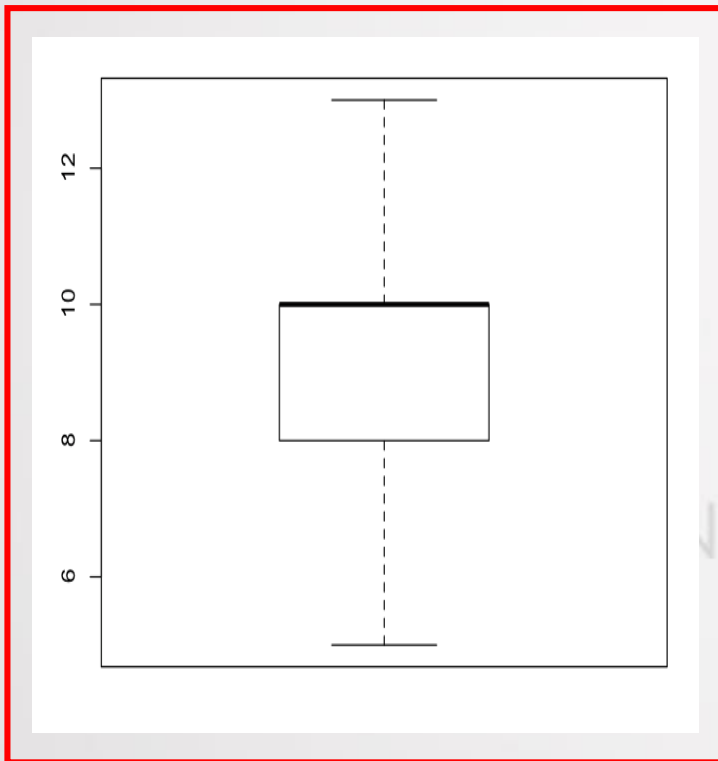
SI OUTLIERS

```
nobs      200.000000
NAs       0.000000
Minimum   3.000000
Maximum   20.000000
1. Quartile 8.000000
3. Quartile 10.000000
Mean      9.194500
Median    10.000000
Sum       1838.900000
SE Mean   0.176314
LCL Mean  8.846817
UCL Mean  9.542183
Variance  6.217306
Stdev     2.493453
Skewness  0.785459
Kurtosis  3.532271
```

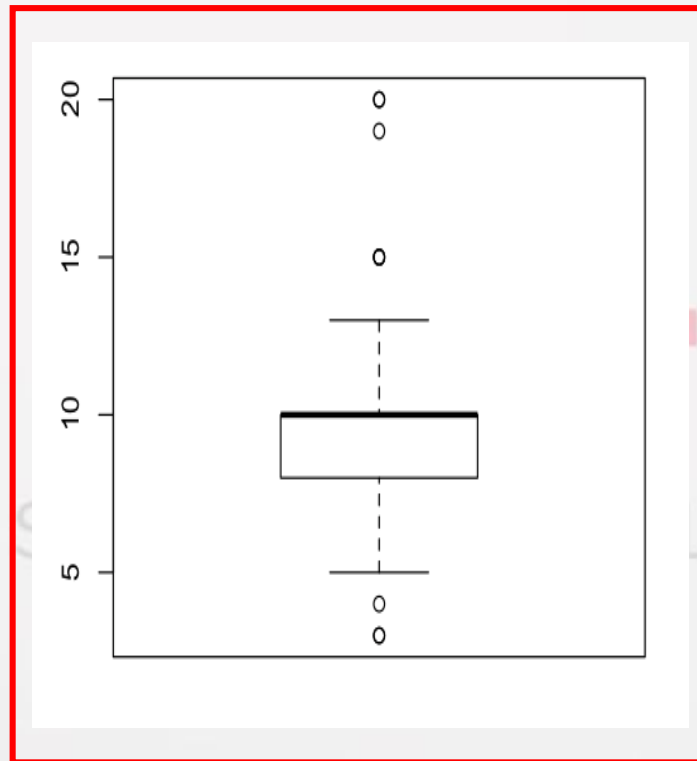
Possiamo notare come nel dataset senza outliers la media differisce di poco da quella del dataset con outliers (da **9.19** a **9.03**). Possiamo dunque affermare che resta comunque una leggera **ASIMMETRIA NEGATIVA (MEDIA < MEDIANA)**. Invece, la *Skewness* diventa leggermente negativa (**-0,49**). La presenza di questo valore suggerisce una distribuzione **QUASI SIMMETRICA**.

WILLINGNESS TO PAY PER IL PACCHETTO BASE

SENZA OUTLIER



CON OUTLIER



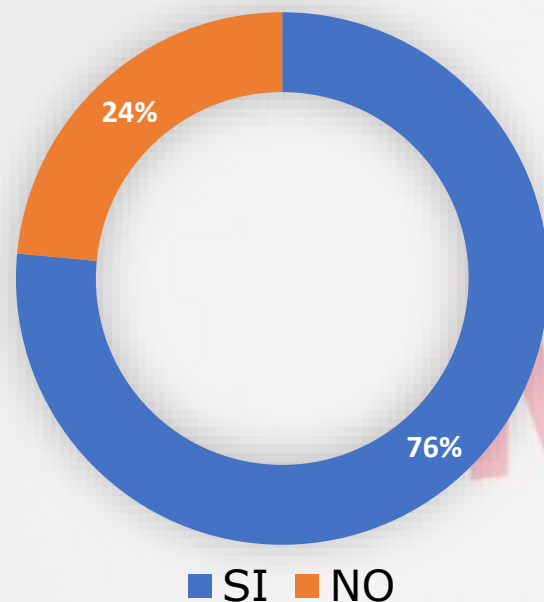
Confrontando i due boxplot, con e senza outlier, possiamo notare come la differenza senza outlier è irrilevante.

La media passa da **9.19** a **9.03** mentre la deviazione standard passa da **2,49** a **1,91**.

Decidiamo, di conseguenza, di rigettare l'ipotesi di eliminare gli outliers dalle seguenti valutazioni, poiché con o senza gli outliers, il risultato non sarebbe significativamente mutato.

NETFLIX NEGLI HOTEL

NETFLIX NELLE STRUTTURE
DI SOGGIORNO



Il risultato ottenuto dall'analisi di frequenza riferita alla variabile `IMPORT_HOTEL` riflette la volontà degli interlocutori considerati di usufruire di Netflix nelle strutture di soggiorno.

È emerso che il **76%** della popolazione di riferimento è a favore di questa implementazione.

Le frequenze cumulate, in relazione alla variabile **IMPORT_HOTEL**, ci suggeriscono che la presenza del servizio Netflix nelle strutture ospitanti, influisce in modo non molto significativo.

Solo il **34,5%** ha affermato che la presenza del servizio sarebbe stata importante nella scelta della struttura ospitante conferendo alla variabile un valore superiore a 7.

Quindi risulta comunque un'opportunità interessante quella di stringere collaborazioni con strutture ospitanti.

```
> cbind(cumsum(table(netf$IMPORT_HOTEL)), cumsum(table(netf$IMPORT_HOTEL)/200))
  [,1] [,2]
1    25 0.125
2    42 0.210
3    54 0.270
4    76 0.380
5   106 0.530
6   131 0.655
7   159 0.795
8   176 0.880
9   185 0.925
10  200 1.000
```


IMPORTANZA DELLA SOCIAL ACCOUNTABILITY PER IL CONSUMATORE

```
> cbind(cumsum(table(netf$IMPORT_SOCIALE)),cumsum(table(netf$IMPORT_SOCIALE)/200))  
  [,1] [,2]  
1     8 0.040  
2    14 0.070  
3    16 0.080  
4    25 0.125  
5    48 0.240  
6    81 0.405  
7   124 0.620  
8   161 0.805  
9   177 0.885  
10  200 1.000
```

Attraverso la funzione *cbind* della variabile *IMPORT_SOCIALE* abbiamo voluto valutare quanto influisse nella scelta di un consumatore la responsabilità sociale di un'azienda.

Il risultato è che il **60%** degli intervistati (con un voto dal 7 in su) attribuisce molta importanza alla responsabilità sociale d'azienda.

L'azienda, quindi, deve aumentare l'attenzione verso il rispetto dei criteri etici, sociali e ambientali promuovendo le iniziative alle quali aderisce, cercando di aumentare così il grado di rispettabilità sociale della stessa.

CONCLUSIONE ANALISI UNIVARIATA

- **ANALISI UNIVARIATA:** Le piattaforme di intrattenimento online sono ben conosciute e usate. In particolare, Netflix si conferma leader di mercato nel settore (più conosciuto e più usato). La sua offerta è in linea con le richieste dei consumatori: Sport, Notiziari e canali musicali risultano essere graditi, quindi potenziali obiettivi da sviluppare e implementare nell'offerta. L'HD ed in particolar modo l'UHD necessitano di maggior sensibilizzazione attraverso campagne di marketing mirate. I nostri interlocutori hanno evidenziato la volontà di poter godere del servizio negli hotel, sebbene non si tratti di un'influenza significativa sulla scelta della struttura. Il sociale, infine, risulta essere molto importante nella vita degli interlocutori, aspetto che deve essere quindi ripreso anche da Netflix.

NETFLIX.IT

ANALISI BIVARIATA

NON SARÀ PIÙ LA SOLITA STORIA

NETFLIX.IT

WILLINGNESS TO PAY - ETA'

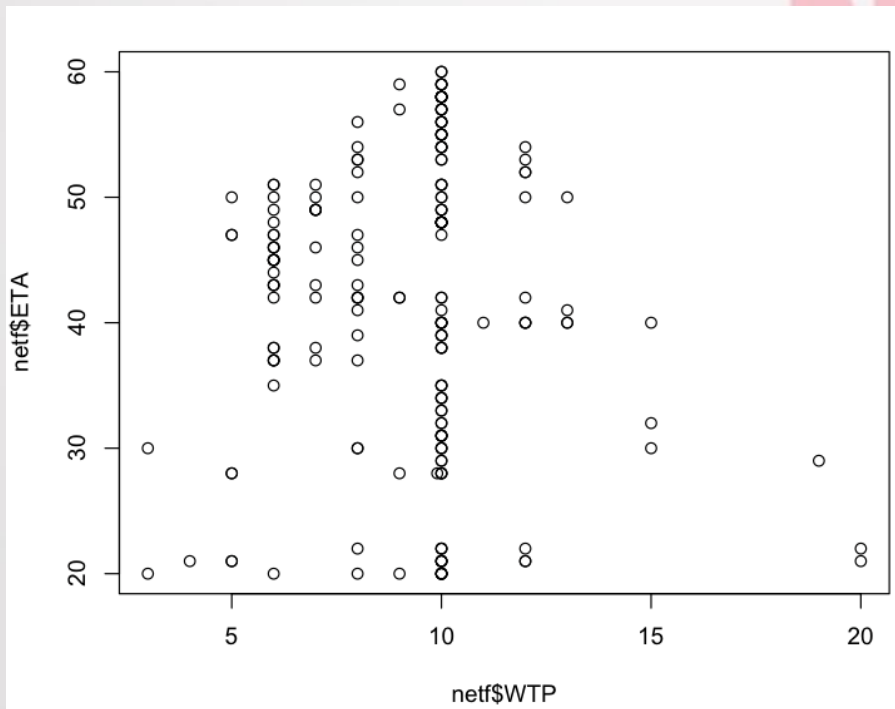
Pearson's product-moment correlation

```
data: netf$WTP and netf$ETA
t = -0.91774, df = 198, p-value = 0.3599
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2019997  0.0743295
sample estimates:
cor
-0.06508269
```

Variabili quantitative → CorTest e plot

H0 ipotesi nulla = INDIPENDENZA LINEARE
H1 ipotesi alternativa = DIPENDENZA

Alfa = 0,05



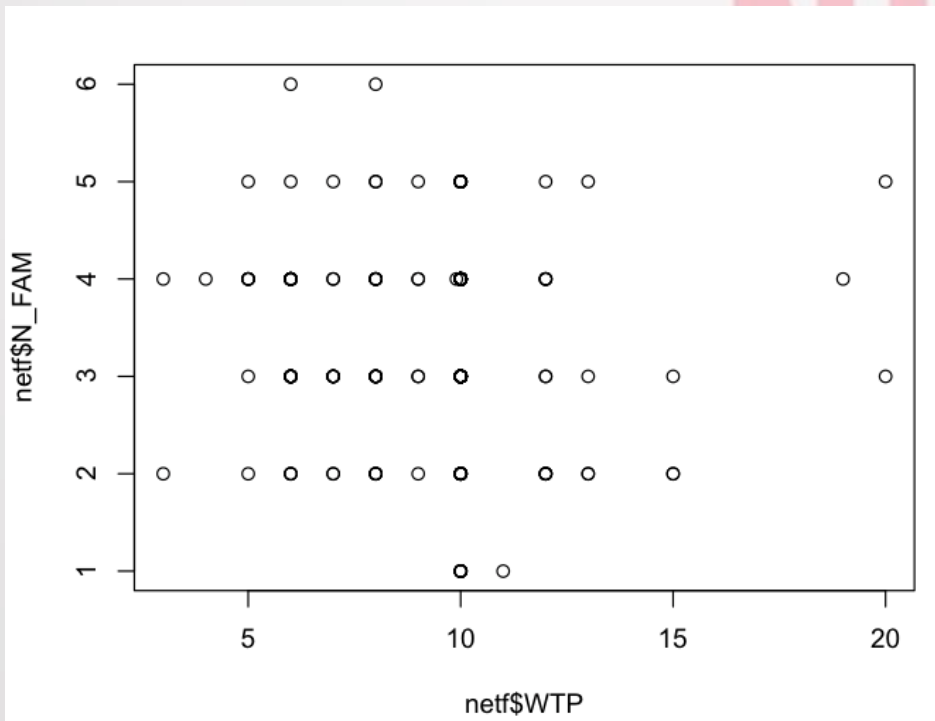
Risultato di questa prima analisi bivariata risulta essere un P-value (= **0.3599**) MAGGIORE di 0,05. Quindi: accettiamo H0, che corrisponde all' **INDIPENDENZA TRA VARIABILI**.

Abbiamo ottenuto inoltre un indice di Pearson MINORE di 0,2. Questo indica che **NON C'È RELAZIONE LINEARE** come si evince anche dal box plot.

*La conclusione di questa analisi è che la propensione alla spesa **non aumenta** al crescere dell'età'.*

WILLINGNESS TO PAY – NUCLEO FAMILIARE

```
Pearson's product-moment correlation  
data: netf$WTP and netf$N_FAM  
t = -1.0124, df = 198, p-value = 0.3126  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.20842889 0.06765149  
sample estimates:  
cor  
-0.0717631
```



Variabili quantitative → CorTest e plot

H0 ipotesi nulla = INDIPENDENZA LINEARE
H1 ipotesi alternativa = DIPENDENZA

Alfa = 0,05

Anche in questa analisi, il P-Value (= **0.3126**) risulta essere **MAGGIORE** di 0,05. Quindi si conferma l'**INDIPENDENZA TRA VARIABILI**.

L'indice di Pearson, **MINORE** di 0,2, indica che **NON C'È RELAZIONE LINEARE** come si evince anche dal box plot.

*In conclusione possiamo affermare che anche in questo caso la propensione alla spesa **NON AUMENTA** all'aumentare del nucleo familiare.*

WILLINGNESS TO PAY – SESSO

Una variabile quantitativa e una qualitativa → aov, model.table, summary ed etaSquared

```
> summary(anova)
      Df Sum Sq Mean Sq F value Pr(>F)
SESSO   1    4.8   4.774   0.767  0.382
Residuals 198 1232.5   6.225
```

```
> anova=aov(WTP~SESSO, data=netf)
> model.tables(anova, type='means')
Tables of means
Grand mean

9.1945

      SESSO
SESSO
      F      M
9.349 9.040
```

```
> etaSquared(anova)
      eta.sq eta.sq.part
SESSO 0.003858617 0.003858617
```

HO ipotesi nulla = INDIPENDENZA IN MEDIA
H1 ipotesi alternativa = DIPENDENZA

Alfa=0,05

Il P-Value è MAGGIORE di 0,05. **Accettiamo** quindi l'ipotesi **H0**. Il medesimo risultato si ottiene osservando l'ETASQUARED (=inferiore a 0,2)

Analizzando la media, si osserva che non esiste un legame con il GENERE: non c'è un sesso al quale indirizzare maggiormente l'offerta, nonostante risultino le donne leggermente in "vantaggio" sugli uomini.

WILLINGNESS TO PAY – PROFESSIONE

```
> anova=aov(WTP~PROFESS, data=netf)
> model.tables(anova, type='means')
Tables of means
Grand mean

9.1945

PROFESS
  ALTRO CASA  DIR IMPIEGA IMPREND  LIB OPERAIO QUADRO  STUD
  11.25  9.5  9.647  8.421  10.63  9.188  9.467  8.231  9.392
rep  4.00 14.0 17.000 38.000  16.00 32.000 15.000 26.000 38.000
```

Una variabile quantitativa e una qualitativa → aov, model.table, summary ed etaSquared

HO ipotesi nulla = INDIPENDENZA IN MEDIA
H1 ipotesi alternativa = DIPENDENZA

Alfa=0,05

```
> summary(anova)
      Df Sum Sq Mean Sq F value Pr(>F)
PROFESS    8  103.9   12.988   2.189  0.03 *
Residuals 191 1133.3    5.934
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il P-Value risulta essere MINORE di 0,05. Vi è quindi **DIPENDENZA IN MEDIA**, ma “eta” è pari a 0.08398 che è MINORE di 0,2 (= indicatore soglia). Questo indica una dipendenza **molto debole** (= categorie molto simili tra di loro).

```
> etaSquared(anova)
      eta.sq eta.sq.part
PROFESS 0.0839827  0.0839827
```

Hanno una WTP maggiore gli IMPRENDITORI e i DIRIGENTI.

Il dato “ALTRO” risulta essere rilevante in questa analisi.

WILLINGNESS TO PAY – DESIGN

Variabili quantitative → Cor.test e plot

```
> cor.test(netf$WTP, netf$QUALIT_PIATT_3)

Pearson's product-moment correlation

data: netf$WTP and netf$QUALIT_PIATT_3
t = 5.2433, df = 198, p-value = 4.032e-07
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2211413 0.4653661
sample estimates:
cor
0.3491693
```

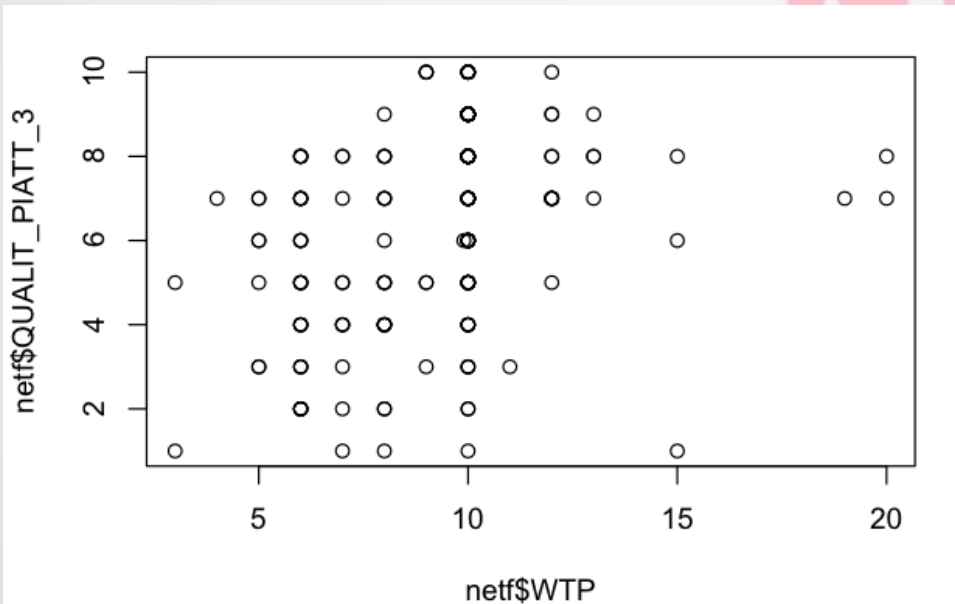
H0 ipotesi nulla = INDIPENDENZA LINEARE
H1 ipotesi alternativa = DIPENDENZA

Alfa = 0,05

Anche in questo caso Il P-value risulta essere MINORE di 0,05. Quindi rifiuto l'ipotesi nulla H0 e accettiamo l'ipotesi alternativa H1. Abbiamo quindi una **DIPENDENZA TRA VARIABILI**.

L'indice di Pearson è pari a **0,35**. Essendo tale indice maggiore di 0,2, possiamo affermare che c'è una leggera **CORRELAZIONE LINEARE POSITIVA**.

Possiamo di conseguenza affermare che la willingness to pay, per una piattaforma di intrattenimento on line, aumenterà all'aumentare della qualità del design della piattaforma stessa. Continuare ad investire sul design innovativo e accattivante può risultare una strategia vincente.



WILLINGNESS TO PAY – FACILITA' USO DELLA PIATTAFORMA

```
> cor.test(netf$WTP, netf$QUALIT_PIATT_4)
```

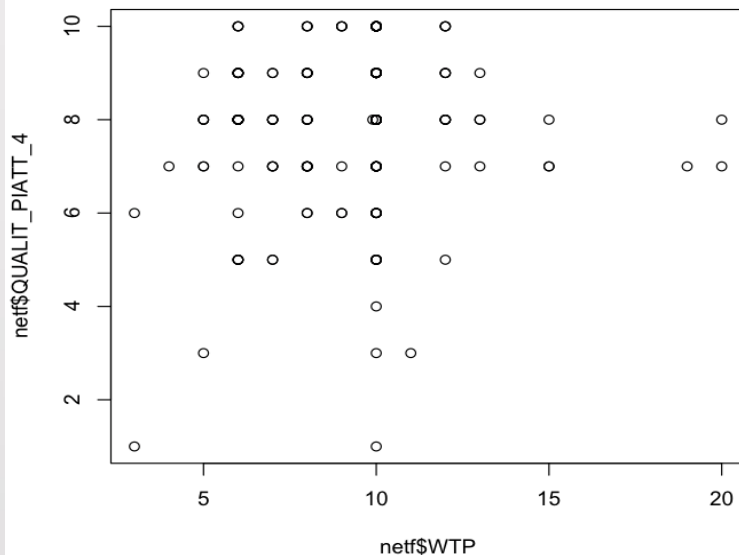
Pearson's product-moment correlation

```
data: netf$WTP and netf$QUALIT_PIATT_4  
t = 1.4845, df = 198, p-value = 0.1393  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
-0.03432082 0.24016481  
sample estimates:  
cor  
0.1049198
```

Variabili quantitative → Cor.test e plot

H0 ipotesi nulla = INDIPENDENZA LINEARE
H1 ipotesi alternativa = DIPENDENZA

Alfa = 0,05



In questo caso, il P-value è **MAGGIORE** di 0,05. **Accettiamo** l'ipotesi H0, e si conferma quindi l'**INDIPENDENZA LINEARE** tra le variabili.

Il **coefficiente di correlazione di paesron** risulta essere invece inferiore a **0,2**, quindi non ci sta correlazione lineare tra le due variabili.

Possiamo quindi affermare, che da tale analisi emerge come la facilità d'uso del servizio, non influenza la propensione all'acquisto dello stesso.

WILLINGNESS TO PAY – MESE PROVA GRATUITO

Variabili quantitative → Cor.test e plot

```
Pearson's product-moment correlation  
data: netf$WTP and netf$QUALIT_PIATT_5  
t = 3.8638, df = 198, p-value = 0.0001513  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.1308504 0.3892266  
sample estimates:  
 cor  
0.2647844
```

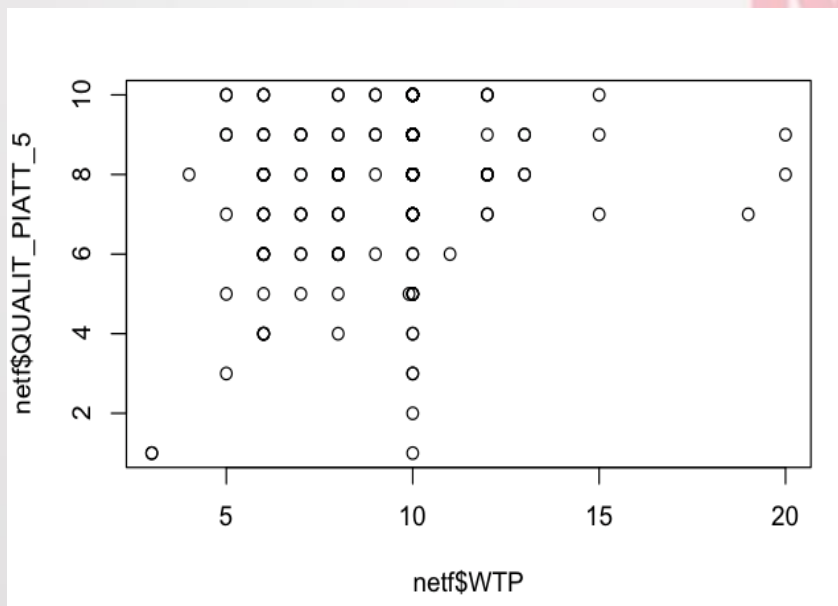
H0 ipotesi nulla = INDIPENDENZA LINEARE
H1 ipotesi alternativa = DIPENDENZA

Alfa = 0,05

In quest'analisi, il P-value ritorna ad essere MINORE di 0,05, quindi rifiuto l'ipotesi H0 e accettiamo l'ipotesi H1. Vi è quindi **DIPENDENZA** tra le variabili.

Il **coefficiente di correlazione** risulta essere in questo caso di poco superiore al valore soglia (0,2), quindi possiamo affermare che tra le due variabili sussiste una relazione lineare molto debole.

Il mese di prova gratuito si conferma così, sulla base dei risultati ricavati da quest'analisi, un elemento rilevante per la propensione all'acquisto.



WILLINGNESS TO PAY – CONTENUTI ORIGINALI

Variabili quantitative → Cor.test e plot

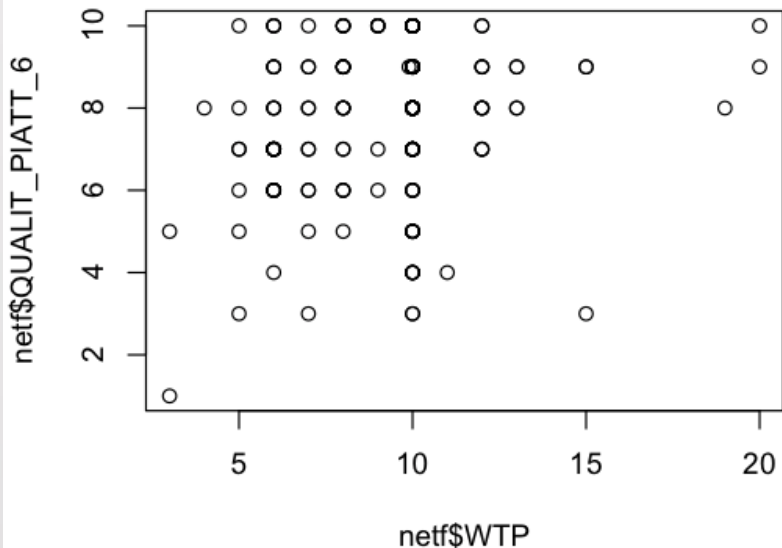
```
> cor.test(netf$WTP, netf$QUALIT_PIAATT_6)

Pearson's product-moment correlation

data: netf$WTP and netf$QUALIT_PIAATT_6
t = 3.1359, df = 198, p-value = 0.001974
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.08123384 0.34582742
sample estimates:
cor
0.2175233
```

H0 ipotesi nulla = INDIPENDENZA LINEARE
H1 ipotesi alternativa = DIPENDENZA

Alfa = 0,05



Il P-value risulta essere MINORE 0,05, abbiamo quindi una **DIPENDENZA IN MEDIA**.

Il coefficiente di correlazione è pari a **0,2175233**, di poco superiore al valore soglia (0,2). Questo valore rileva una **DIPENDENZA MOLTO DEBOLE**.

In tal caso, l'originalità dei contenuti risulta quindi essere un fattore abbastanza rilevante. È quindi importante continuare ad investire su quest'aspetto.

WILLINGNESS TO PAY – DOWNLOAD CONTENUTI PER VISUALIZZAZIONE OFFLINE

Variabili quantitative → Cor.Test e plot

H0 ipotesi nulla = indipendenza lineare
H1 ipotesi alternativa = dipendenza

```
> cor.test(netf$WTP, netf$DLOAD_CONTENTUTO)
```

Pearson's product-moment correlation

data: netf\$WTP and netf\$DLOAD_CONTENTUTO

t = 0.22279, df = 198, p-value = 0.8239

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.1231807 0.1542331

sample estimates:

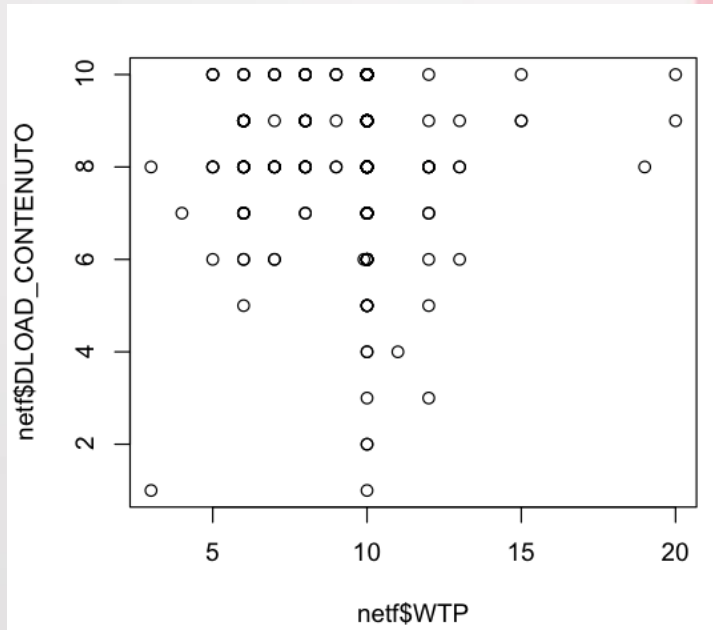
cor
0.01583087

Alfa=0,05

Risultato di questa analisi bivariata è un P-Value
MAGGIORE di 0,05: questo indica una **FORTE
INDIPENDENZA.**

Inoltre, il coefficiente di correlazione risulta essere
MINORE di 0,2.
Ciò comporta la completa **ASSENZA** di
CORRELAZIONE.

*Risultato inaspettato risulta quindi essere il fatto che la
possibilità di effettuare il download dei contenuti per
visualizzarli offline non sia un fattore significativo.
Quindi la willingness to pay non aumenta all' aumentare
della possibilità di poter scaricare i contenuti.*

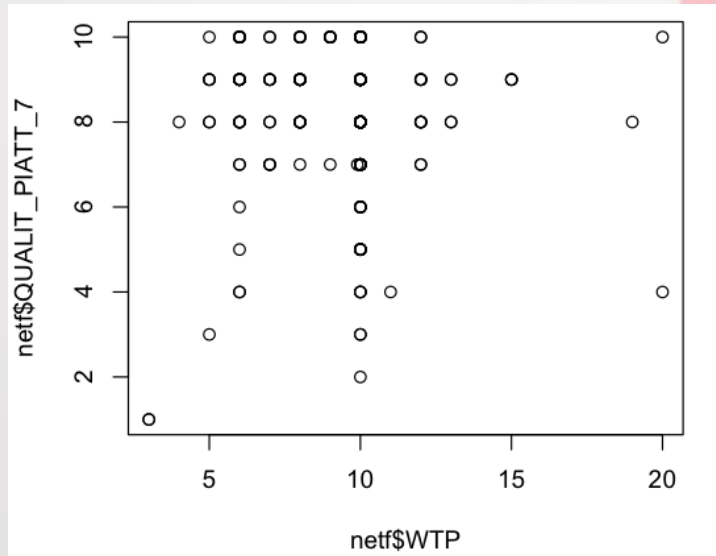


WILLINGNESS TO PAY – CONTENUTI IN ESCLUSIVA

```
> cor.test(netf$WTP, netf$QUALIT_PIATT_7)

Pearson's product-moment correlation

data: netf$WTP and netf$QUALIT_PIATT_7
t = 0.4484, df = 198, p-value = 0.6544
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1073649  0.1698410
sample estimates:
 cor
0.03185054
```



Variabili quantitative → Cor.Test e Plot

H0 ipotesi nulla = indipendenza lineare
H1 ipotesi alternativa = dipendenza

Alfa = 0,05

In questa analisi il P-Value calcolato è **MAGGIORE** di 0,05. C'è quindi una **FORTE INDIPENDENZA**.

Come nel caso precedente, anche in questa analisi il coefficiente di correlazione risulta essere **MINORE** di 0,2. Si evidenzia quindi la **MANCANZA di CORRELAZIONE**.

Anche in questo caso l'esclusività dei contenuti non è rilevante. Quindi possiamo affermare che la propensione all'acquisto non aumenta all'aumentare della presenza sulla piattaforma di contenuti in esclusiva.

DIFFERENZIAZIONE DELL'OFFERTA

NETFLIX

NON SARÀ PIÙ LA SOLITA STORIA

NETFLIX.IT

USO DELLE PIATTAFORME NEGLI HOTEL

```
> CrossTable(netf$USA_PIATT, netf$HOTEL, prop.chisq=FALSE, chisq=TRUE)
```

Cell Contents

		netf\$HOTEL		
		NO	SI	Total
netf\$USA_PIATT	NO	16	40	56
		0.286	0.714	0.280
		0.340	0.261	
		0.080	0.200	
SI	SI	31	113	144
		0.215	0.785	0.720
		0.660	0.739	
		0.155	0.565	
Total		47	153	200
		0.235	0.765	

Statistics for All Table Factors

Pearson's Chi-squared test

Chi^2 = 1.112722 d.f. = 1 p = 0.291

Pearson's Chi-squared test with Yates' continuity correction

Chi^2 = 0.7554085 d.f. = 1 p = 0.385

Quest'analisi ha evidenziato la **mancanza di connessione** tra le variabili, che risultano essere quindi **INDIPENDENTI**.

Quindi il possesso delle piattaforme da parte delle strutture ospitanti non rappresenta un termine di influenza per gli ospiti.

La **mancanza di correlazione** viene testimoniata dalla V di Cramer, che risulta essere MINORE di 0,2.

```
> CramerV(netf$USA_PIATT, netf$HOTEL)
```

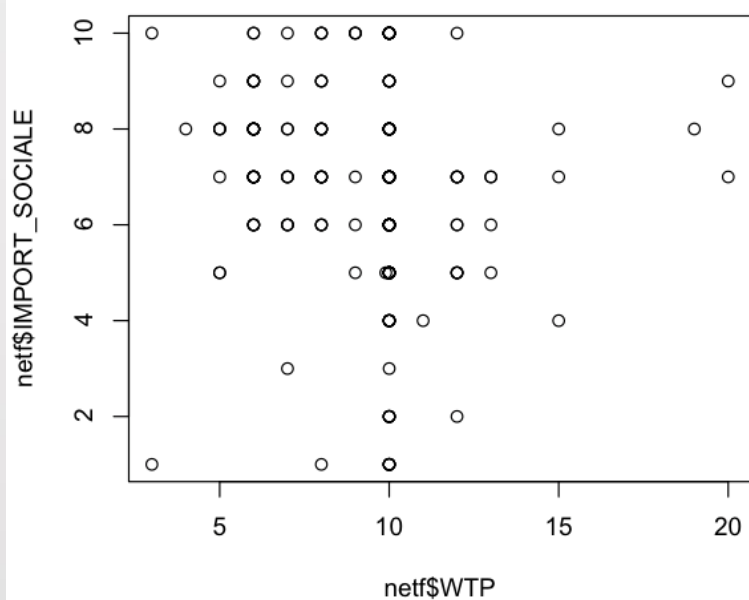
```
[1] 0.07458963
```

WILLINGNESS TO PAY – IMPORTANZA DEL SOCIALE

```
> cor.test(netf$WTP, netf$IMPORT_SOCIALE)

Pearson's product-moment correlation

data: netf$WTP and netf$IMPORT_SOCIALE
t = -2.1085, df = 198, p-value = 0.03624
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.281153101 -0.009650103
sample estimates:
 cor
-0.1481927
```



Il P-Value ottenuto d questa analisi bivariata risulta essere MINORE di 0,5, il quale indica quindi la presenza di **DIPENDENZA**.

Le variabili sono CORRELATE.

Tuttavia, l'indice di Pearson evidenzia come tale correlazione (lineare negativa) sia **MOLTO DEBOLE** (-0.148).

Non risulta essere quindi interessante l'investimento sul sociale da parte della piattaforma.

CONCLUSIONE ANALISI BIVARIATA

- **ANALISI BIVARIATA:** È emerso come la propensione alla spesa non aumenti al crescere dell'età e al numero di componenti del nucleo familiare. La willingness to pay non ha riscontrato alcun legame con il genere. Tuttavia, WTP maggiore è risultata a capo di imprenditori e dirigenti. Il design, che rappresenta un aspetto qualificante della piattaforma, ha evidenziato dipendenza con la WTP: all'aumentare della qualità del design aumenterà la WTP. Dipendenza riscontrata anche tra il mese di prova, elemento rilevante per la propensione all'acquisto. La facilità d'uso non influenza la WTP. La WTP risulta infine essere caratterizzata da dipendenza con i contenuti originali, mentre contenuti in esclusiva e il download per la visualizzazione dei contenuti offline sono caratterizzati da forte indipendenza.

NETFLIX.IT

ANALISI MULTIVARIATA

NON SARÀ PIÙ LA SOLITA STORIA

NETFLIX.IT

REGRESSIONE LINEARE MULTIPLA 1/2

L'obiettivo del nostro lavoro è quello di analizzare come la **variabile dipendente** "**willingness to pay**" venga influenzata dalle esigenze funzionali dei consumatori. Abbiamo quindi proseguito con l'analisi fattoriale in modo tale da poter facilitare l'interpretazione dei nostri output.

Per spiegare la nostra variabile dipendente abbiamo individuato e spiegato nella tabella seguente le variabili indipendenti utilizzate per lo svolgimento di quest'analisi.

VARIABILI	SPIEGAZIONE VARIABILI
HD	Importanza Alta Definizione (HD)
ELEM_IMP_PIATT_1	Importanza della "Connessione"
ELEM_IMP_PIATT_2	Importanza dei "Contenuti"
ELEM_IMP_PIATT_3	Importanza delle "condizioni meteo"
ELEM_IMP_PIATT_4	Importanza "visione in ogni momento dei contenuti"
QUALIT_PIATT_1	Importanza "Qualità/Prezzo "
QUALIT_PIATT_2	Importanza "Qualità contenuti/Prezzo "
QUALIT_PIATT_3	Importanza "Design"
QUALIT_PIATT_4	Importanza "Facilità d'uso"

REGRESSIONE LINEARE MULTIPLA 2/2

VARIABILI	SPIEGAZINE VARIABILI
QUALIT_PIATT_5	Importanza "Mese di prova gratuito"
QUALIT_PIATT_6	Importanza "Contenuti originali"
QUALIT_PIATT_7	Importanza "Programmi in esclusiva"
DLOAD_CONTENTUTO	Importanza sulla scelta d'acquisto di salvare un contenuto multimediale
IMPORT_SOCIALE	Importanza sulla scelta d'acquisto l'investimento nel sociale
IMPORT_HOTEL	Importanza piattaforme nella scelta della struttura ospitante
BENEF_SOCIALE	Importanza sociale sostenibilità nella sua vita
TEMPO_LIBERO	Tempo libero in una giornata
TEMPO_LIB_PIATT	Quante ore passo a guardare contenuti multimediali nel tempo libero

ANALISI FATTORIALE

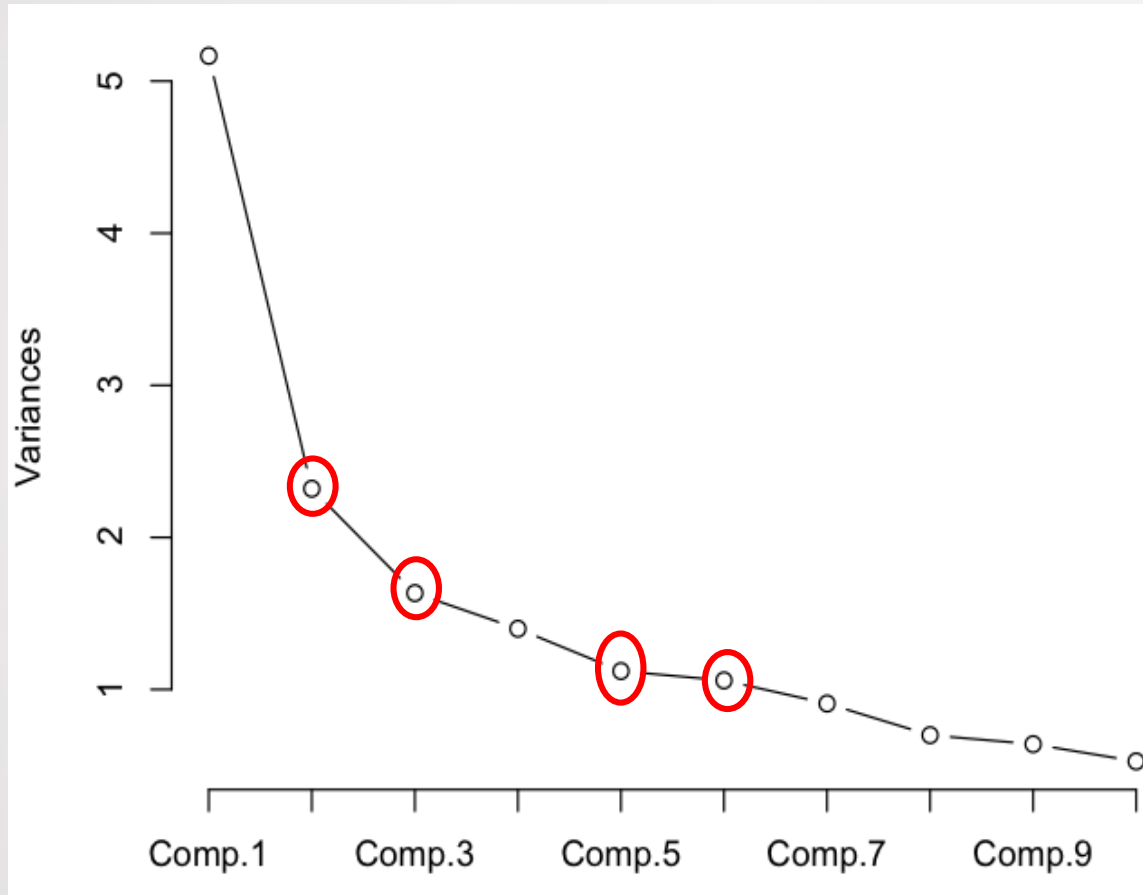
```
> eig.val<-get_eigenvalue(fit)
> eig.val
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	5.1663273	28.701818	28.70182
Dim.2	2.3200698	12.889277	41.59109
Dim.3	1.6336775	9.075986	50.66708
Dim.4	1.3993179	7.773989	58.44107
Dim.5	1.1211397	6.228554	64.66962
Dim.6	1.0596377	5.886876	70.55650
Dim.7	0.9076959	5.042755	75.59925
Dim.8	0.6997443	3.887468	79.48672
Dim.9	0.6405685	3.558714	83.04544
Dim.10	0.5274503	2.930280	85.97572
Dim.11	0.4549698	2.527610	88.50333
Dim.12	0.4248826	2.360459	90.86379
Dim.13	0.3399870	1.888817	92.75260
Dim.14	0.3095896	1.719942	94.47254
Dim.15	0.3011770	1.673205	96.14575
Dim.16	0.2603280	1.446267	97.59202
Dim.17	0.2284981	1.269434	98.86145
Dim.18	0.2049389	1.138550	100.00000

Per capire qual' è il numero di variabili da tenere in considerazione, abbiamo a disposizione tre diversi metodi:

- **METODO EMPIRICO** → Tale metodo (1/3 del totale) ci suggerirebbe di utilizzare **6** variabili.
- **VARIABILITA' SPIEGATA** → Il metodo della percentuale di variabilità spiegata, espressa tra il 60%-70% della variabilità totale, ci suggerisce di tenere in considerazione **5** o **6** variabili.
- **AUTOVALORI** → Il metodo degli autovalori ci suggerisce di tenere in considerazione i valori che hanno autovalore maggiore di 1. E in tal caso bisogna considerare le prime **6** variabili.

ANALISI FATTORIALE – SCREE PLOT



Attraverso il metodo dello *scree plot*, possiamo notare la presenza di diversi gomiti. Il primo all'altezza di **5** componenti (che spiegano il 64%); il secondo all'altezza di **6** componenti (che spiegano il 70%).

Si hanno inoltre altri due gomiti, rispettivamente all'altezza di **2** e **3** componenti. Confrontando questi due valori con i metodi spiegati precedentemente, possiamo affermare che essi non sono rilevanti.

L'analisi prosegue dunque prendendo in considerazione **5-6** componenti.

Verrà aggiunta anche l'analisi a **7** componenti, per valutare se con una variabile in più sia possibile riuscire ad interpretare meglio l'output, o se sia sufficiente l'utilizzo di solo 6 variabili.

ANALISI FATTORIALE

	Z=5	Z=6	Z=7
HD	0.6085202	0.6892331	0.7109140
ELEM_IMP_PIATT_1	0.7679480	0.7694497	0.7751361
ELEM_IMP_PIATT_2	0.7072057	0.7526189	0.7829608
ELEM_IMP_PIATT_3	0.5404230	0.6290209	0.6539139
ELEM_IMP_PIATT_4	0.7496385	0.7635660	0.7949304
QUALIT_PIATT_1	0.8008990	0.8027787	0.8128774
QUALIT_PIATT_2	0.6392324	0.6856831	0.6934808
QUALIT_PIATT_3	0.5437339	0.5451012	0.8039200
QUALIT_PIATT_4	0.6846914	0.6867161	0.7216970
QUALIT_PIATT_5	0.6207770	0.6208530	0.6224445
QUALIT_PIATT_6	0.6534039	0.6629610	0.7335075
QUALIT_PIATT_7	0.6175986	0.6415681	0.8656974
DLOAD CONTENUTO	0.6400287	0.7234828	0.7487616
IMPORT_SOCIALE	0.5265585	0.7014952	0.7725811
IMPORT_HOTEL	0.6202342	0.6218445	0.6422942
BENEF_SOCIALE	0.5095104	0.8313306	0.8620776
TEMPO_LIBERO	0.6753490	0.8174989	0.8219950
TEMPO_LIB_PIATT	0.7347799	0.7549683	0.7886763

Attraverso il confronto delle soluzioni a 5-6-7 fattori possiamo notare che, nelle celle evidenziate in giallo, abbiamo avuto un significativo **AUMENTO DI COMUNALITA'**. La soluzione con la comunalità di regressori più alta, risulta essere la soluzione a **7** fattori. Ma, in virtù dei criteri presi precedentemente in considerazione, possiamo affermare che anche la soluzione a **6** regressori ha un'ottima comunalità, in quanto abbiamo avuto aumenti rilevanti rispetto all'analisi a 5 fattori, che risulta infatti essere meno esplicativa delle altre.

Ai fini interpretativi, la scelta ottimale risulta quindi essere la soluzione a **6** regressori, ed è quella che noi prenderemo in considerazione.

ANALISI FATTORIALE

```
> z6=principal(netf2, nfactors = 6, residuals = FALSE, rotate='varimax')
> print(z6$loadings,sort=TRUE,cutoff=0.3)
```

Loadings:

	RC1	RC3	RC2	RC5	RC4	RC6
QUALIT_PIATT_3	0.633					
QUALIT_PIATT_4	0.769					
QUALIT_PIATT_5	0.743					
QUALIT_PIATT_6	0.763					
QUALIT_PIATT_7	0.740					
ELEM_IMP_PIATT_1	0.340	0.590		0.378	0.343	
ELEM_IMP_PIATT_2		0.836				
ELEM_IMP_PIATT_3		0.714				
ELEM_IMP_PIATT_4		0.788				
HD			0.769			
DLOAD_CONTENUTO			0.797			
IMPORT_HOTEL			0.668			0.369
QUALIT_PIATT_1				0.855		
QUALIT_PIATT_2	0.429			0.605		
TEMPO_LIBERO			-0.351		0.810	
TEMPO_LIB_PIATT			0.304		0.791	
IMPORT_SOCIALE			0.458			0.634
BENEF_SOCIALE						0.860

- **SERVIZI ADDIZIONALI:** Possiamo notare come la classe Servizi Addizionali è maggiormente influenzata dalle variabili *QUALIT_PIATT_4, QUALIT_PIATT_5, QUALIT_PIATT_6, QUALIT_PIATT_7*.
- **MOTIVI UTILIZZO:** Possiamo notare come la variabile *ELEM_IMP_PIATT_2* influenzi maggiormente la classe Motivi di Utilizzo.
- **TECNOLOGIE:** Notiamo come la variabile *HD* e *DLOAD_CONTENUTO* influenzi maggiormente la classe Tecnologie.
- **CONVENIENZA:** Notiamo come la classe Convenienza sia maggiormente influenzata dalla variabile *QUALIT_PIATT_1*.
- **TEMPO LIBERO:** In questo caso le variabili *TEMPO_LIBERO* e *TEMPO_LIB_PIATT* influenzano allo stesso modo la classe Tempo Libero.
- **IMPORTANZA SOCIALE:** Notiamo che la classe Importanza sociale è maggiormente influenzata dalla variabile *BENEF_SOCIALE*.

REGRESSIONE LINEARE MULTIPLA

Dall' analisi fattoriale sono emersi **6** fattori (componenti principali) che verranno utilizzati per la **regressione lineare**.

RC1	SERVIZI_ADDIZIONALI
RC2	TECNOLOGIE
RC3	MOTIVI_UTILIZZO
RC4	TEMP_LIB_RC
RC5	CONVENIENZA
RC6	IMPORTANZA_SOCIALE

A questi fattori vengono aggiunti altre variabili che si ritiene possano non apportare problemi di **multicollinearità**.

TECH	ETA	N_FAM
------	-----	-------

Equazione di regressione lineare multipla

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$$

Y = variabile dipendente (WTP)

X = variabili indipendenti (SERVIZI_ADDIZIONALI, IMPORTANZA_SOCIALE, MOTIVI_UTILIZZO, TEMPO_LIB_RC, CONVENIENZA, TECH, ETA, N_FAM)

VALUTAZIONE DEL MODELLO

Call:

```
lm(formula = WTP ~ SERVIZI_ADDIZIONALI + IMPORTANZA_SOCIALE +  
MOTIVI_UTILIZZO + TEMPO_LIB_RC + CONVENIENZA + TECNOLOGIE +  
TECH + ETA + N_FAM, data = nuovodataset)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.1880	-1.0556	-0.0770	0.8637	10.9599

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.466649	1.322643	6.401	1.18e-09	***
SERVIZI_ADDIZIONALI	0.665634	0.164668	4.042	7.68e-05	***
IMPORTANZA_SOCIALE	-0.405097	0.155658	-2.602	0.00999	**
MOTIVI_UTILIZZO	-0.464660	0.158642	-2.929	0.00382	**
TEMPO_LIB_RC	0.888821	0.170869	5.202	5.09e-07	***
CONVENIENZA	-0.137447	0.165876	-0.829	0.40836	
TECNOLOGIE	-0.348809	0.187328	-1.862	0.06414	.
TECH	0.135647	0.119567	1.134	0.25802	
ETA	0.007415	0.014445	0.513	0.60832	
N_FAM	-0.202147	0.157246	-1.286	0.20016	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.13 on 190 degrees of freedom

Multiple R-squared: 0.3034, Adjusted R-squared: 0.2704

F-statistic: 9.195 on 9 and 190 DF, p-value: 1.58e-11

R-quadro=0.3034 → Il modello spiega il **30%** della variabilità della disponibilità a pagare. Dal momento che il modello deve avere valore compreso da 0 e 1, possiamo affermare che il modello ha una discreta capacità esplicativa.

R-quadro Adj =0.2704 → Il **27%** della variabilità della willingness to pay può essere spiegato dal modello proposto, considerando il numero dei regressori e l'ampiezza campionaria.

P-value= 1.58e-11 → Il P-value associato al test F è inferiore al nostro livello di significatività 0,05.

Rifiutiamo quindi l'ipotesi H0 (in cui tutti i coefficienti sono pari a zero), e accetto l'ipotesi alternativa H1 (almeno uno dei coefficienti non è nullo), ed affermiamo che il modello ha capacità esplicativa poiché almeno una variabile all'interno dello stesso è esplicativa.

Pr(>|t|)= Tramite il metodo **stepwise**, elimineremo tutte quelle variabili che non hanno capacità esplicativa.

VALUTAZIONE DEL MODELLO 1/2 – STEPWISE

```
Call:
lm(formula = WTP ~ SERVIZI_ADDIZIONALI + IMPORTANZA_SOCIALE +
    MOTIVI_UTILIZZO + TEMPO_LIB_RC + TECNOLOGIE, data = nuovodataset)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-7.0963 -1.0571 -0.0341  0.8526 10.5837
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.1945     0.1504  61.127 < 2e-16 ***
SERVIZI_ADDIZIONALI  0.6775     0.1508   4.493 1.20e-05 ***
IMPORTANZA_SOCIALE -0.4475     0.1508  -2.968  0.00338 **
MOTIVI_UTILIZZO   -0.4325     0.1508  -2.868  0.00459 **
TEMPO_LIB_RC      0.9476     0.1508   6.284 2.13e-09 ***
TECNOLOGIE       -0.2484     0.1508  -1.647  0.10108
---
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.1945	0.1504	61.127	< 2e-16 ***
SERVIZI_ADDIZIONALI	0.6775	0.1508	4.493	1.20e-05 ***
IMPORTANZA_SOCIALE	-0.4475	0.1508	-2.968	0.00338 **
MOTIVI_UTILIZZO	-0.4325	0.1508	-2.868	0.00459 **
TEMPO_LIB_RC	0.9476	0.1508	6.284	2.13e-09 ***
TECNOLOGIE	-0.2484	0.1508	-1.647	0.10108

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.127 on 194 degrees of freedom
Multiple R-squared:  0.2905, Adjusted R-squared:  0.2722
F-statistic: 15.88 on 5 and 194 DF, p-value: 4.168e-13
```

Notiamo come, anche al seguito dell'applicazione del modello stepwise, **R-quadro**, **R-quadro Adj** e il **P-value(<0,05)** rimangono come valori rilevanti.

Attraverso la stepwise, ci concentreremo sui regressori che hanno un'elevata capacità esplicativa (**TEST T < 0,05**), che risultano essere **SERVIZI_ADDIZIONALI**, **IMPORTANZA_SOCIALE**, **MOTIVI_UTILIZZO** e **TEMPO_LIB_RC**.

Poiché al seguito metodo stepwise ci rimane ancora la variabile **TECNOLOGIE**, che non risulta essere rilevante a fini esplicativi. Stimeremo quindi un nuovo modello di regressione lineare con le sole variabili significative.

VALUTAZIONE DEL MODELLO 2/2

```
Call:
lm(formula = WTP ~ SERVIZI_ADDIZIONALI + IMPORTANZA_SOCIALE +
    MOTIVI_UTILIZZO + TEMPO_LIB_RC, data = nuovodataset)

Residuals:
    Min       1Q   Median       3Q      Max
-7.0140 -1.1012 -0.0156  0.9729 10.3898

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.1945     0.1511  60.860 < 2e-16 ***
SERVIZI_ADDIZIONALI  0.6775     0.1515   4.473 1.31e-05 ***
IMPORTANZA_SOCIALE -0.4475     0.1515  -2.955 0.00352 **
MOTIVI_UTILIZZO   -0.4325     0.1515  -2.856 0.00476 **
TEMPO_LIB_RC      0.9476     0.1515   6.256 2.45e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.137 on 195 degrees of freedom
Multiple R-squared:  0.2805. Adjusted R-squared:  0.2658
F-statistic: 19.01 on 4 and 195 DF, p-value: 3.241e-13
```

```
> v
      Variables VIF
1  SERVIZI_ADDIZIONALI 1
2  IMPORTANZA_SOCIALE 1
3   MOTIVI_UTILIZZO 1
4   TEMPO_LIB_RC 1
```

Rivalutiamo la bontà del modello tramite il test **R-quadro (28%)** ed **R-quadro Adj (26,5%)** che rimangono su un livello accettabile di esplicabilità.

Tramite il test **P-value**, che rimane inferiore a 0.05, possiamo dire che almeno una variabile nel modello è esplicativa.

Verifichiamo la presenza di multicollinearità anche se tra le variabili significative prese in considerazione, nessuna è stata aggiunta successivamente alla fattoriale.

In seguito all'analisi fatta possiamo affermare l'assenza di multicollinearità tra le variabili esplicative, essendo i valori dei VIF pari a 1.

VALUTAZIONE DELL'IMPATTO DEI REGRESSORI E INTERPRETAZIONE 1/2

```
> lm.beta(REGRESSIONE)
SERVIZI_ADDIZIONALI  IMPORTANZA_SOCIALE  MOTIVI_UTILIZZO  TEMPO_LIB_RC
0.2717206            -0.1794660        -0.1734536       0.3800190
```

Vogliamo ora valutare l'**impatto dei regressori** ed effettuare la relativa **interpretazione** tramite la funzione `lm.beta`.

Osserveremo nell'analisi un'influenza positiva e una negativa di due regressori sulla nostra variabile dipendente.

- **TEMPO_LIB_RC** = All'aumentare del tempo libero di un'unità osserviamo che la WTP aumenta di **0.38 punti**. Possiamo quindi affermare che, per esempio, focalizzarsi maggiormente su campagne pubblicitarie mirate alla sensibilizzazione di studenti universitari o soggetti il cui tempo libero viene occupato in gran parte dallo streaming online.
- **SERVIZI_ADDIZIONALI** = All'aumentare dei servizi offerti di un'unità aumenta la WTP di **0.27 punti**, dato molto interessante. Quest'ultimo suggerisce infatti di enfatizzare maggiormente il **valore aggiunto** di Netflix. Si mettono dunque in evidenza i già esistenti mese di prova gratuito, contenuti originali e contenuti in esclusiva.

VALUTAZIONE DELL'IMPATTO DEI REGRESSORI E INTERPRETAZIONE 2/2

- **MOTIVI_UTILIZZO** = All'aumentare dei motivi di utilizzo di un'unità, la WTP diminuisce di **0.17 punti**. I motivi esogeni di fatto non sono controllabili da Netflix: condizioni metereologiche, qualità della connessione Internet, qualità dei contenuti televisivi. Si osserva infatti che, come anche rilevato dall'analisi dei competitors, le piattaforme televisive iniziano a muoversi all'interno di questo mercato dello streaming online. Essendoci più offerta e quindi concorrenza, i consumatori cercano spendere di meno. Più scarsa è la qualità dei programmi tv, maggiore sarà la probabilità che i consumatori si orientino verso le piattaforme di intrattenimento online, il cui mercato, essendo in netta ascesa, vedrà potenzialmente l'inserimento di nuovi competitors che faranno aumentare ancora la concorrenza.
- **IMPORTANZA_SOCIALE** = Un risultato inaspettato è stato il seguente: all'aumentare dell'importanza del sociale per i consumatori di un'unità, diminuisce la WTP di **0.17 punti**. La WTP non viene quindi influenzata positivamente da questo fattore e non risulterebbe utile intraprendere una campagna di sensibilizzazione verso un argomento socialmente utile. L'idea sarebbe stata quella di devolvere in beneficenza 0,50€ ogni mese per ogni abbonamento sottoscritto. Nonostante il risultato inaspettato, rifacendoci a quanto ottenuto dall'analisi univariata, la componente etica risulta essere molto importante nella vita della nostra popolazione in esame: da qui la volontà di perseguire aspetti quali business ethics, codice etico, promozione di tematiche sociali...

REGRESSIONE LOGISTICA

La regressione logistica che abbiamo svolto individua la probabilità della seguente variabile dicotomica: scelta di un hotel per un soggiorno sulla base del possesso da parte della struttura di un abbonamento alle piattaforme di intrattenimento online e quanto quest'ultimo incida sulla scelta.

È stata usata la funzione **glm** (mylogit). La variabile DIPENDENTE da noi utilizzata è HOTEL, ovvero se gli ospiti vogliono usufruire delle piattaforme online durante i soggiorni. La nostra analisi vuole individuare appunto come le variabili INDIPENDENTI da noi scelte influenzino la variabile DICOTOMICA.

NETFLIX.IT

REGRESSIONE LOGISTICA

Le percentuali dei nostri intervistati che ritengono (1) e non ritengono (0) importante la presenza di un abbonamento alle piattaforme di intrattenimento online nella struttura ospitante sono le seguenti:

netf\$HOTEL		
	Frequency	Percent
0	47	23.5
1	153	76.5
Total	200	100.0

REGRESSIONE LOGISTICA

Di seguito le
variabili che
possono
influenzare la
variabile
dicotomica:



```
Call:
glm(formula = HOTEL ~ ELEM_IMP_PIATT_2 + ELEM_IMP_PIATT_3 + ELEM_IMP_PIATT_4 +
  DISP_PREF_1 + DISP_PREF_2 + DISP_PREF_3 + QUALIT_PIATT_3 +
  QUALIT_PIATT_4 + QUALIT_PIATT_5 + QUALIT_PIATT_6 + QUALIT_PIATT_7 +
  IMPORT_SOCIALE + BENEF_SOCIALE + TECH + USO_SOCIAL + SESSO +
  ETA + N_FAM + ALTO_REDD + MEDIO_REDD, family = "binomial",
  data = netf)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.61209  0.03592  0.36073  0.61853  2.11339
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.283234  2.105921  -2.984 0.002849 **
ELEM_IMP_PIATT_2 -0.026252  0.149424  -0.176 0.860539
ELEM_IMP_PIATT_3 -0.122543  0.110282  -1.111 0.266490
ELEM_IMP_PIATT_4  0.118340  0.164391   0.720 0.471606
DISP_PREF_1     1.259981  0.567832   2.219 0.026491 *
DISP_PREF_2     0.903391  0.661369   1.366 0.171957
DISP_PREF_3     0.862502  0.541346   1.593 0.111103
QUALIT_PIATT_3  -0.627081  0.189653  -3.306 0.000945 ***
QUALIT_PIATT_4   0.464328  0.219861   2.112 0.034693 *
QUALIT_PIATT_5  -0.149010  0.160626  -0.928 0.353572
QUALIT_PIATT_6  -0.513096  0.219283  -2.340 0.019290 *
QUALIT_PIATT_7   0.513189  0.200520   2.559 0.010488 *
IMPORT_SOCIALE  0.300368  0.115720   2.596 0.009441 **
BENEF_SOCIALE   0.056566  0.108134   0.523 0.600896
TECH             0.375289  0.168158   2.232 0.025630 *
USO_SOCIAL       0.323191  0.150936   2.141 0.032254 *
SESSOM           0.008371  0.460523   0.018 0.985497
ETA              -0.003557  0.025488  -0.140 0.888997
N_FAM            -0.039760  0.236278  -0.168 0.866364
ALTO_REDD        -0.028935  0.702346  -0.041 0.967139
MEDIO_REDD       1.105858  0.775389   1.426 0.153811
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 218.10  on 199  degrees of freedom
Residual deviance: 150.96  on 179  degrees of freedom
AIC: 192.96
```

REGRESSIONE LOGISTICA

```
> summary(a)

Call:
glm(formula = HOTEL ~ DISP_PREF_1 + DISP_PREF_2 + DISP_PREF_3 +
     QUALIT_PIATT_3 + QUALIT_PIATT_4 + QUALIT_PIATT_6 + QUALIT_PIATT_7 +
     IMPORT_SOCIALE + TECH + USO_SOCIAL + MEDIO_REDD, family = "binomial",
     data = netf)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.67002  0.03949  0.35938  0.62701  1.90454

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.6626     1.7316  -3.848 0.000119 ***
DISP_PREF_1   1.4098     0.5260   2.680 0.007362 **
DISP_PREF_2   0.8681     0.6099   1.423 0.154622
DISP_PREF_3   0.8462     0.5095   1.661 0.096714 .
QUALIT_PIATT_3 -0.6845     0.1800  -3.802 0.000143 ***
QUALIT_PIATT_4  0.3842     0.2035   1.888 0.058962 .
QUALIT_PIATT_6 -0.5261     0.2103  -2.502 0.012355 *
QUALIT_PIATT_7  0.5100     0.1914   2.665 0.007701 **
IMPORT_SOCIALE  0.2933     0.1052   2.789 0.005286 **
TECH           0.3618     0.1611   2.245 0.024740 *
USO_SOCIAL     0.3878     0.1406   2.758 0.005812 **
MEDIO_REDD    1.1819     0.5574   2.120 0.033989 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 218.10  on 199  degrees of freedom
Residual deviance: 153.74  on 188  degrees of freedom
AIC: 177.74

Number of Fisher Scoring iterations: 6
```

Attraverso il modello Stepwise individuiamo, tra i regressori presi in considerazione, quali sono effettivamente quelli interpretativi del modello, e quindi valutare la loro capacità esplicativa. Dall'analisi possiamo notare come la procedura Stepwise non ha selezionato solo i regressori interpretativi con P-value inferiore a 0,05.

REGRESSIONE LOGISTICA

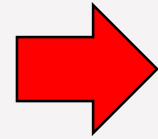
```
> CalculateConcordance(mylogit)
```

```
$Concordance  
[1] 0.854
```

```
$Discordance  
[1] 0.146
```

```
$Tied  
[1] 0
```

```
$Pairs  
[1] 7191
```



Dalla nostra analisi è emerso che la percentuale di Concordance risulta essere pari all' **85%**, maggiore del 70% che rappresenta il livello soglia. *Possiamo quindi affermare che il modello ha una buona capacità di stimare la probabilità che il fenomeno possa verificarsi.*

```
> waldtest(a)
```

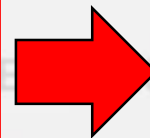
```
Wald test
```

```
Model 1: HOTEL ~ DISP_PREF_1 + DISP_PREF_2 + DISP_PREF_3 + QUALIT_PIATT_3 +  
QUALIT_PIATT_4 + QUALIT_PIATT_6 + QUALIT_PIATT_7 + IMPORT_SOCIALE +  
TECH + USO_SOCIAL + MEDIO_REDD
```

```
Model 2: HOTEL ~ 1
```

	Res.Df	Df	F	Pr(>F)
1	188			
2	199	-11	3.0959	0.0007534 ***

```
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Per analizzare la significatività congiunta dei coefficienti, eseguiamo il **Wald test** (che corrisponde al test F).

Possiamo in questo caso affermare che il P-value del Wald test è inferiore a 0,05, e che quindi il modello ha CAPACITA' ESPLICATIVA.

REGRESSIONE LOGISTICA

In questo caso procediamo, attraverso l'analisi dei V.I.F., ad individuare la presenza o meno di multicollinearità nel modello.

```
> vif(multicoll)
  Variables      VIF
1  DISP_PREF_1  1.057458
2  QUALIT_PIATT_3  1.375115
3  QUALIT_PIATT_6  2.218292
4  QUALIT_PIATT_7  1.949686
5  IMPORT_SOCIALE  1.121764
6           TECH  2.165514
7   USO_SOCIAL  2.037918
8   MEDIO_REDD  1.127236
```

Dall'analisi effettuata notiamo come i valori non sono tutti prossimi all' 1, e si riscontra così la **presenza di multicollinearità**.

Si procede quindi con L'ANALISI FATTORIALE di TUTTE le variabili di partenza, in quanto l'esito della stepwise potrebbe essere influenzato dalla presenza di multicollinearità.

NETFLIX.IT

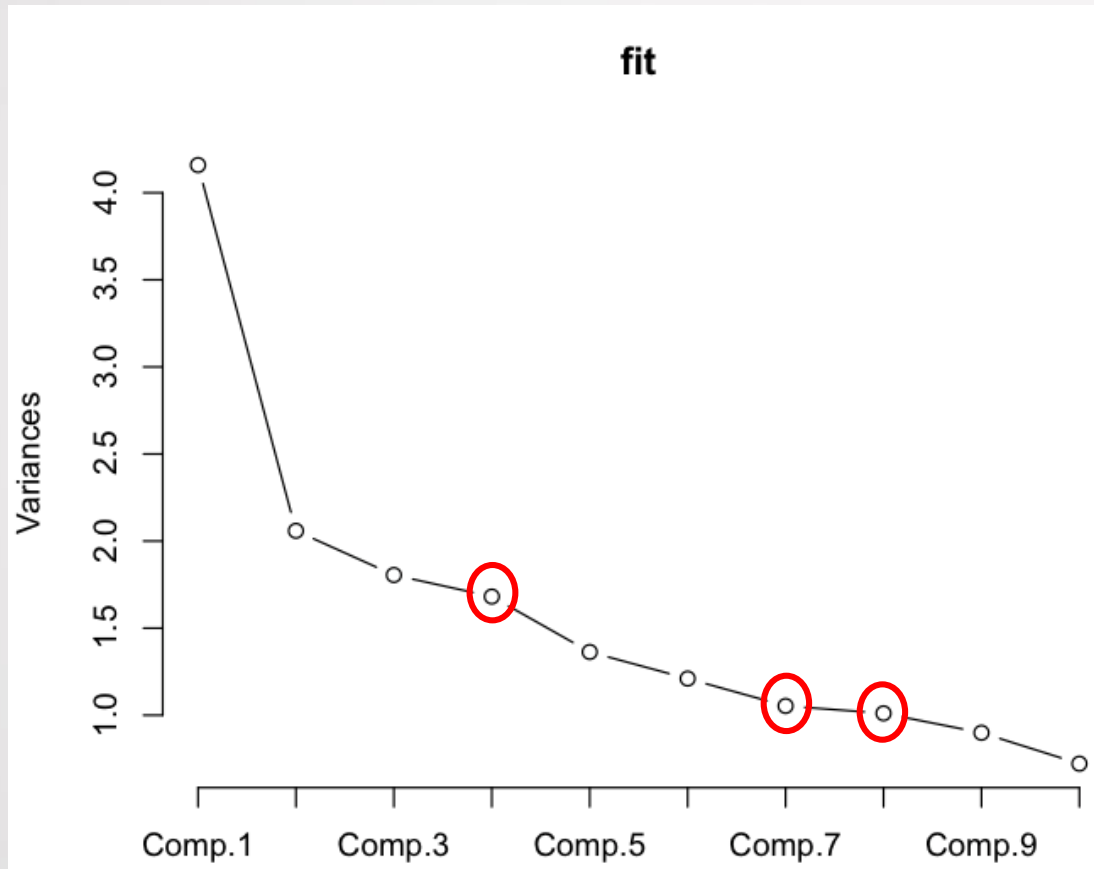
REGRESSIONE LOGISTICA: ANALISI FATTORIALE

Analisi Fattoriale

```
> Eig.val
eigenvalue variance.percent cumulative.variance.percent
Dim.1 4.1590575 21.8897766 21.88978
Dim.2 2.0588429 10.8360153 32.72579
Dim.3 1.8051523 9.5008017 42.22659
Dim.4 1.6814189 8.8495732 51.07617
Dim.5 1.3632040 7.1747577 58.25092
Dim.6 1.2103980 6.3705158 64.62144
Dim.7 1.0532782 5.5435696 70.16501
Dim.8 1.0111086 5.3216242 75.48663
Dim.9 0.8992994 4.7331549 80.21979
Dim.10 0.7217453 3.7986594 84.01845
Dim.11 0.6305464 3.3186651 87.33711
Dim.12 0.4807144 2.5300758 89.86719
Dim.13 0.4060536 2.1371244 92.00431
Dim.14 0.3694546 1.9444981 93.94881
Dim.15 0.2714623 1.4287489 95.37756
Dim.16 0.2618622 1.3782223 96.75578
Dim.17 0.2172683 1.1435174 97.89930
Dim.18 0.2106532 1.1087012 99.00800
Dim.19 0.1884797 0.9919983 100.00000
```

- **METODO EMPIRICO** → Tale metodo (1/3 del totale) ci suggerisce di tenere in considerazione **6 o 7** variabili.
- **VARIABILITA' SPIEGATA** → Il metodo della percentuale di variabilità spiegata, espressa tra il 60%-70% della variabilità totale, ci suggerisce di tenere in considerazione **6 o 7** variabili.
- **AUTOVALORI** → Il metodo degli autovalori ci suggerisce di tenere in considerazione i valori che hanno autovalore maggiore di 1. E in tal caso bisogna considerare le prime **8** variabili.

REGRESSIONE LOGISTICA: ANALISI FATTORIALE



Attraverso il metodo dello *scree plot*, possiamo notare la presenza di diversi gomiti. Il primo all'altezza di **7** componenti (che spiegano il 70%); il secondo all'altezza di **8** componenti (che spiegano il 75%).

Si ha un altro gomito all'altezza di **4** componenti. La variabilità spiegata risulta il 50%, quindi non sufficientemente utile per fini esplicativi.

L'analisi prosegue dunque prendendo in considerazione **7-8** componenti.

Verrà aggiunta anche l'analisi a:

- 6 componenti sulla base dei metodi precedentemente riportati
- 9 componenti per valutare se con una variabile in più sia possibile o meno riuscire ad interpretare meglio l'output.

REGRESSIONE LOGISTICA: ANALISI FATTORIALE

	Z=6	Z=7	Z=8	Z=9
ELEM_IMP_PIATT_2	0.7882680	0.7983319	0.8131306	0.8143621
ELEM_IMP_PIATT_3	0.4358196	0.5758786	0.5836322	0.6922599
ELEM_IMP_PIATT_4	0.7567584	0.7636186	0.8403965	0.8403965
DISP_PREF_1	0.5089073	0.8492167	0.8572513	0.8768977
DISP_PREF_2	0.7671339	0.8127663	0.8168556	0.8651052
DISP_PREF_3	0.7675044	0.7678390	0.8119474	0.8409613
QUALIT_PIATT_3	0.5734065	0.5922665	0.7515132	0.7870000
QUALIT_PIATT_4	0.6535893	0.7178316	0.7458687	0.7813302
QUALIT_PIATT_5	0.6063806	0.6848745	0.6865787	0.6893745
QUALIT_PIATT_6	0.6441735	0.6519352	0.6745154	0.7471516
QUALIT_PIATT_7	0.6476286	0.6559621	0.6837011	0.7902911
IMPORT_SOCIALE	0.5163641	0.7225566	0.7579614	0.7712102
BENEF_SOCIALE	0.4490455	0.4917690	0.4952103	0.6863040
TECH	0.7690953	0.7691077	0.7952957	0.8102381
USO_SOCIAL	0.7782082	0.7947140	0.8090664	0.8152270
ETA	0.6023415	0.6165419	0.8491857	0.8505138
N_FAM	0.5018126	0.5025409	0.6729903	0.8342959
ALTO_REDD	0.8109923	0.8183435	0.8186128	0.8687551
MEDIO_REDD	0.7006439	0.7452572	0.8787471	0.8800764

Attraverso il confronto delle soluzioni a 6-7-8-9 fattori possiamo notare che, nelle celle evidenziate in rosso, abbiamo avuto un significativo **AUMENTO DI COMUNALITA'**.

Ai fini interpretativi la scelta ottimale, che risulta essere maggiormente esplicativa è la soluzione ad **8** regressori, ed è quella che noi prenderemo in considerazione.

REGRESSIONE LOGISTICA: ANALISI FATTORIALE

```
> y<-principal(fattorialelogistica,nfactors=8,residuals=FALSE,rotate="varimax")
> print(y$loadings,sort=T,cutoff=0.3)
```

Loadings:

	RC1	RC5	RC2	RC4	RC3	RC7	RC8	RC6
QUALIT_PIATT_3	0.644		0.419				0.312	
QUALIT_PIATT_4	0.803							
QUALIT_PIATT_5	0.803							
QUALIT_PIATT_6	0.677							
QUALIT_PIATT_7	0.670							
ELEM_IMP_PIATT_2		0.858						
ELEM_IMP_PIATT_3		0.604						
ELEM_IMP_PIATT_4		0.858						
TECH			0.840					
USO_SOCIAL			0.867					
IMPORT_SOCIALE				0.843				
BENEF_SOCIALE				0.672				
ALTO_REDD					0.695		0.472	
MEDIO_REDD					-0.916			
DISP_PREF_1						-0.906		
ETA							0.860	
DISP_PREF_2						0.578		-0.590
DISP_PREF_3								0.882
N_FAM	0.321		-0.322	0.430	0.345		-0.360	
SS loadings	2.910	2.095	1.977	1.632	1.587	1.399	1.384	1.358
Proportion Var	0.153	0.110	0.104	0.086	0.084	0.074	0.073	0.071
Cumulative Var	0.153	0.263	0.367	0.453	0.537	0.611	0.683	0.755

Procediamo con l'interpretazione del modello , che però non risulta essere particolarmente soddisfacente.

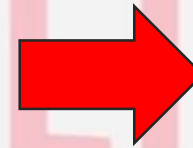
Prendiamo quindi in considerazione l'ipotesi di aggiungere un altro fattore alla nostra analisi. Procederemo quindi con lo svolgimento dell'analisi tenendo in considerazione **9** fattori, in quanto vogliamo prediligere l'interpretabilità del modello.

REGRESSIONE LOGISTICA: ANALISI FATTORIALE

```
> y<-principal(fattorialelogistica,nfactors=9,residuals=FALSE,rotate="varimax")
> print(y$loadings,sort=T,cutoff=0.3)
```

Loadings:

	RC1	RC2	RC5	RC3	RC4	RC9	RC6	RC7	RC8
QUALIT_PIATT_4	0.668					0.363			
QUALIT_PIATT_5	0.754								
QUALIT_PIATT_6	0.769								
QUALIT_PIATT_7	0.796								
QUALIT_PIATT_3	0.482	0.531				0.458			
TECH		0.861							
USO_SOCIAL		0.867							
ELEM_IMP_PIATT_2			0.855						
ELEM_IMP_PIATT_3			0.627			0.307		-0.317	
ELEM_IMP_PIATT_4			0.850						
ALTO_REDD				0.807		0.333			
MEDIO_REDD				-0.876					
IMPORT_SOCIALE					0.761				
BENEF_SOCIALE					0.758				
ETA						0.838			
DISP_PREF_2							-0.700	-0.495	
DISP_PREF_3							0.879		
DISP_PREF_1								0.913	
N_FAM									0.868
SS loadings	2.698	2.076	2.068	1.625	1.468	1.411	1.374	1.340	1.182
Proportion Var	0.142	0.109	0.109	0.086	0.077	0.074	0.072	0.071	0.062
Cumulative Var	0.142	0.251	0.360	0.446	0.523	0.597	0.669	0.740	0.802



RC1	QUALIT_PIATT
RC2	PROPENSIONE_TECH
RC5	MOTIVI UTILIZZO
RC3	REDDITO
RC4	ETICA
RC9	ETA
RC6	PREF_DISP_MOBILI
RC7	PREF_TV
RC8	FAMIGLIA

Poiché l'interpretabilità dei dati risulta essere soddisfacente, proseguiremo con la seguente analisi.

REGRESSIONE LOGISTICA

```
> summary(a2)

Call:
glm(formula = HOTEL ~ RC2 + RC4 + RC6 + RC7 + RC9, family = "binomial",
     data = new_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4529  0.2717  0.5138  0.7178  1.5909

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.4074     0.1989   7.074 1.5e-12 ***
RC2           0.5013     0.1831   2.738 0.00618 **
RC4           0.5877     0.1879   3.128 0.00176 **
RC6           0.3911     0.2038   1.919 0.05495 .
RC7           0.3464     0.1782   1.944 0.05194 .
RC9          -0.4623     0.1936  -2.388 0.01696 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 218.10  on 199  degrees of freedom
Residual deviance: 185.05  on 194  degrees of freedom
AIC: 197.05

Number of Fisher Scoring iterations: 4
```

Considerando il risultato ottenuto da questa nuova analisi, notiamo come la stepwise prende in considerazione anche i valori RC6 (PREF_DISP_MOBILI) ed RC7 (PREF_TV), valori che non risultano essere effettivamente esplicativi per la nostra analisi. Ma considerando che il valore di RC6 ed RC7 sono al limite del valore soglia (0,05), abbiamo deciso comunque di considerare anche questi regressori nella nostra analisi.

Invece considereremo ottimali ed effettivamente esplicativi i regressori **RC2 (PROPENSIONE_TECH), RC4 (ETICA) e RC9 (ETA)**

REGRESSIONE LOGISTICA

```
> CalculateConcordance(mylogit_factor)
$Concordance
[1] 0.7548

$Discordance
[1] 0.2452

$Tied
[1] 0

$Pairs
[1] 7191
```

Verifichiamo la bontà del modello. In questo caso notiamo come la concordanza risulta essere pari al **75%**. Possiamo quindi affermare di avere un buon modello esplicativo.

```
> waldtest(a2)
Wald test

Model 1: HOTEL ~ RC2 + RC4 + RC6 + RC7 + RC9
Model 2: HOTEL ~ 1
  Res.Df Df    F Pr(>F)
1     194   -  -    -   -
2     199  -5 5.1338 0.0001909 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La capacità esplicativa viene confermata dal Wald test che risulta essere pari a **0.0001909**, valore inferiore a 0,05 (P-value).

REGRESSIONE LOGISTICA

```
> vif(multicoll)
Variables VIF
1      RC2    1
2      RC4    1
3      RC9    1
4      RC6    1
5      RC7    1
```

Procediamo con una nuova analisi V.I.F. per individuare se anche in questo caso, con le nuove variabili prese in considerazione, si presenta o meno il fenomeno della multicollinearità.

Dall'analisi effettuata notiamo come tutti i valori risultano in questo caso essere pari ad uno.

Possiamo quindi dire che il problema della multicollinearità è stato risolto.

INTERPRETAZIONE DEI COEFFICIENTI 1/2

```
> lm.beta(a2)
      RC2      RC4      RC6      RC7      RC9
1.1793691 1.3826478 0.9200862 0.8150302 -1.0876751
```

Regressori	Stime Std.
ETICA	1.3826478
PROPENSIONE_TECH	1.1793691
ETA	1.0876751
PREF_DISP_MOBILI	0.9200862
PREF_TV	0.8150302

Per stabilire le variabili più importanti andremo ad ordinare in valore assoluto i parametri standardizzati.

RC4 - RC2 - RC9 - RC6 - RC7

Il regressore ETICA che rappresenta l'importanza della responsabilità sociale è quello che influenza maggiormente il modello.

Seguono Propensione all'uso della tecnologia, Età, Preferenza di uso dei dispositivi mobili (Tablet e Laptop) e della TV.

- All'aumentare dell'importanza della componente etica, aumenta la propensione all'uso delle piattaforme di intrattenimento nelle strutture alberghiere e non. (**ETICA = 1.3826 segno positivo**)
- All'aumentare della propensione alla tecnologia, aumenta la propensione all'uso delle piattaforme di intrattenimento nelle strutture alberghiere e non. (**PROPENSIONE_TECH = 1.1793 segno positivo**)
- All'aumentare dell'età, diminuisce tale propensione (**ETA = 1.0876 segno negativo**)
- All'aumentare della preferenza di uso per dispositivi mobili (Laptop e Tablet), aumenta la propensione all'uso delle piattaforme nelle strutture alberghiere e non. (**PREF_DISP_MOBILI = 0.92 segno positivo**)
- Stesso risultato per l'uso delle TV anche se in via minore. (**PREF_TV = 0.8150 segno positivo**)

INTERPRETAZIONE DEI COEFFICIENTI 2/2

Nell'interpretare le stime Odds-Ratio poniamo soglia 1.

Calcoliamo la differenza tra la stima Odds e la soglia 1 per interpretare tali regressori.

Regressori	Odds Ratio
ETICA	1.7998659
PROPENSIONE_TECH	1.6508759
ETA	0.6298143
PREF_DISP_MOBILI	1.4785981
PREF_TV	1.4140233

```
> exp(a2$coefficient)
(Intercept)      RC2      RC4      RC6      RC7      RC9
 4.0854601  1.6508759  1.7998659  1.4785981  1.4140233  0.6298143
```

- All'aumentare di un'unità dell'importanza della responsabilità sociale (ETICA), aumenta la probabilità che il cliente voglia usufruire di tale servizio all'interno delle strutture alberghiere di **79.98 punti** ($1.7998 - 1 = 79,98$).
- All'aumentare di un'unità della propensione all'uso della tecnologia (PROPENSIONE_TECH), aumenta la probabilità di usufruire dei servizi suddetti di **65 punti** ($1.6508 - 1 = 65$).
- All'aumentare dell'età (ETA), diminuisce la probabilità di usufruire di tali servizi di **37 punti** ($0.6298 - 1 = -0,37$).
- All'aumentare di un'unità della preferenza di uso di dispositivi quali laptop e tablet (PREF_DISP_MOBILI), aumenta la probabilità di usufruire di tale servizio di **47.85 punti** ($1.4785 - 1 = 47,85$).
- All'aumentare di un'unità della preferenza di uso di TV (PREF_TV), aumenta la probabilità di usufruire di tale servizio di **41,40 punti** ($1.4140 - 1 = 41,40$).

CONCLUSIONE ANALISI MULTIVARIATA

- **REGRESSIONE LINEARE:** Abbiamo osservato che all'aumentare del tempo libero e dei servizi offerti che la WTP aumenta. All'aumentare dei motivi di utilizzo si riscontra, invece, una riduzione della WTP. Inaspettatamente abbiamo ottenuto come risultato che All'aumentare dell'importanza del sociale per i consumatori, diminuisce la WTP.
- **REGRESSIONE LOGISTICA:** La maggioranza dei nostri intervistati ha evidenziato la preferenza di strutture di vacanza che abbiano abbonamenti attivi alle piattaforme di intrattenimento online. In particolare, questa preferenza rivolta alle piattaforme nelle strutture ospitanti, è stata riscontrata nei risultati delle analisi svolte, come per esempio: all'aumentare dell'importanza della componente etica aumenta la propensione all'uso delle piattaforme di intrattenimento nelle strutture alberghiere e non; all'aumentare della propensione alla tecnologia aumenta la propensione all'uso delle piattaforme; all'aumentare della propensione all'uso della tecnologia aumenta la probabilità di usufruire delle piattaforme.

CONCLUSIONE

NON SARÀ PIÙ LA SOLITA STORIA

NETFLIX.IT

Possiamo concludere che la conoscenza delle piattaforme è ben sviluppata.

Trattandosi di un servizio di intrattenimento multimediale, il profilo del consumatore risulta essere molto eterogeneo: non abbiamo osservato un genere di riferimento che prevalesse sull'altro.

Il consumatore ha una conoscenza limitata della tecnologia HD (alla quale tuttavia dà molta importanza) e all'UHD, di cui, come detto in precedenza, andrebbe sensibilizzata maggiormente con campagne marketing mirate.

La possibilità di vedere eventi sportivi, concerti in differita e contenuti musicali si sono rivelati importanti per garantire una posizione predominante nel mercato, che risulta essere sempre più competitivo, e per garantire un'offerta a 360°. Questo per rendere ancor più appetibile la piattaforma stessa.

Il mese gratuito prima della sottoscrizione dell'abbonamento e il design della piattaforma risultano essere fattori chiave per l'aumento degli abbonati. Di conseguenza il nostro suggerimento sarebbe quello di investire maggiormente su questi fattori, per esempio aumentando il periodo di prova gratuito e rendendo ancor più avvolgente e accattivante l'estetica della piattaforma.

Si denota anche una forte tensione verso la componente etica, aspetto che ha richiamato la nostra attenzione. Questa componente è collegata con la Willingness-to-pay mediante una correlazione lineare negativa che, seppur in modo labile, ci suggerirebbe di non investire nella Business ethics. Tuttavia, come già detto, l'etica è fortemente presente nell'indole del consumatore.

Il risultato della regressione logistica afferma anch'esso che all'aumentare dell'importanza del sociale per i consumatori, aumenta la probabilità di voler usufruire del servizio di intrattenimento online nelle strutture alberghiere e non.

Per quanto riguarda la possibilità di vedere Netflix nelle strutture turistiche, possiamo affermare che questo servizio non si è rivelato essere una necessità primaria per i consumatori. Ciò nonostante potrebbe risultare un elemento vincente nel lungo periodo: dal momento che i principali destinatari del servizio sono i giovani, la partnership con catene di hotel (Hilton, NH Hotel, Ibis..) potrebbe rivelarsi un valore aggiunto che garantirebbe un aumento del potere di mercato.