Analisi dataset
Clustering
Modelli predittivi

CESMA

TelefonoAMICOItalia

Indice

- Contesto e domande di ricerca
- Descrizione delle telefonate
- Ricerca di associazioni tra fattori
- Quali sono le tipologie di chiamate?
- Chi sono gli abituali?
- Campanelli d'allarme: Quali caratteristiche di chiamata e di utenti influiscono sulla probabilità che l'argomento della chiamata sia il suicidio?
  - Modelli

Contesto e domande di ricerca

A partire da una serie di dataset fornitoci dall'associazione Telefono Amico Italia, caratterizzato da **quattro anni di dati (2019-2022)** relativi a circa 198.000 chiamate a supporto delle persone in tema di emergenza emozionale e salute sociale, abbiamo effettuato delle analisi al fine di rispondere alle seguenti domande:

- Potenziale presenza di:
  - Tipologie di chiamate
  - Tipologie degli utenti ricorrenti (con almeno 20 chiamate totali)
  - Fattori legati alla probabilità di osservare una chiamata riguardante il suicidio (analizzando gli effetti semplici e di interazione)
- Comparazione tra modelli parametrici e non parametrici nella capacità esplicativa di chiamate riguardanti il tema del suicidio

Punto di attenzione: Non si ha la possibilità di attribuire le telefonate ad un singolo utente

Descrizione delle telefonate

CHI CHIAMA?

QUANDO CHIAMA?

QUANTO DURA LA CHIAMATA?

PERCHÉ CHIAMA?

Chi chiama?

Sesso

55% Uomini
45% Donne

Età

2% Tra 0 e 18
5% Tra 19 e 25
13% Tra 26 e 35
21% Tra 36 e 45
27% Tra 46 e 55
23% Tra 56 e 65
10% Oltre i 66

Provenienza

61% Rilevato

62% Nord
18% Centro
20% Sud e Isole

Professione

52% Rilevato

32% Pensionato/a
28% Non Occupato/a
22% Lavoratore Dip.
6% Studente/essa
4% In Proprio
4% Precario/a
4% Casalinga/o

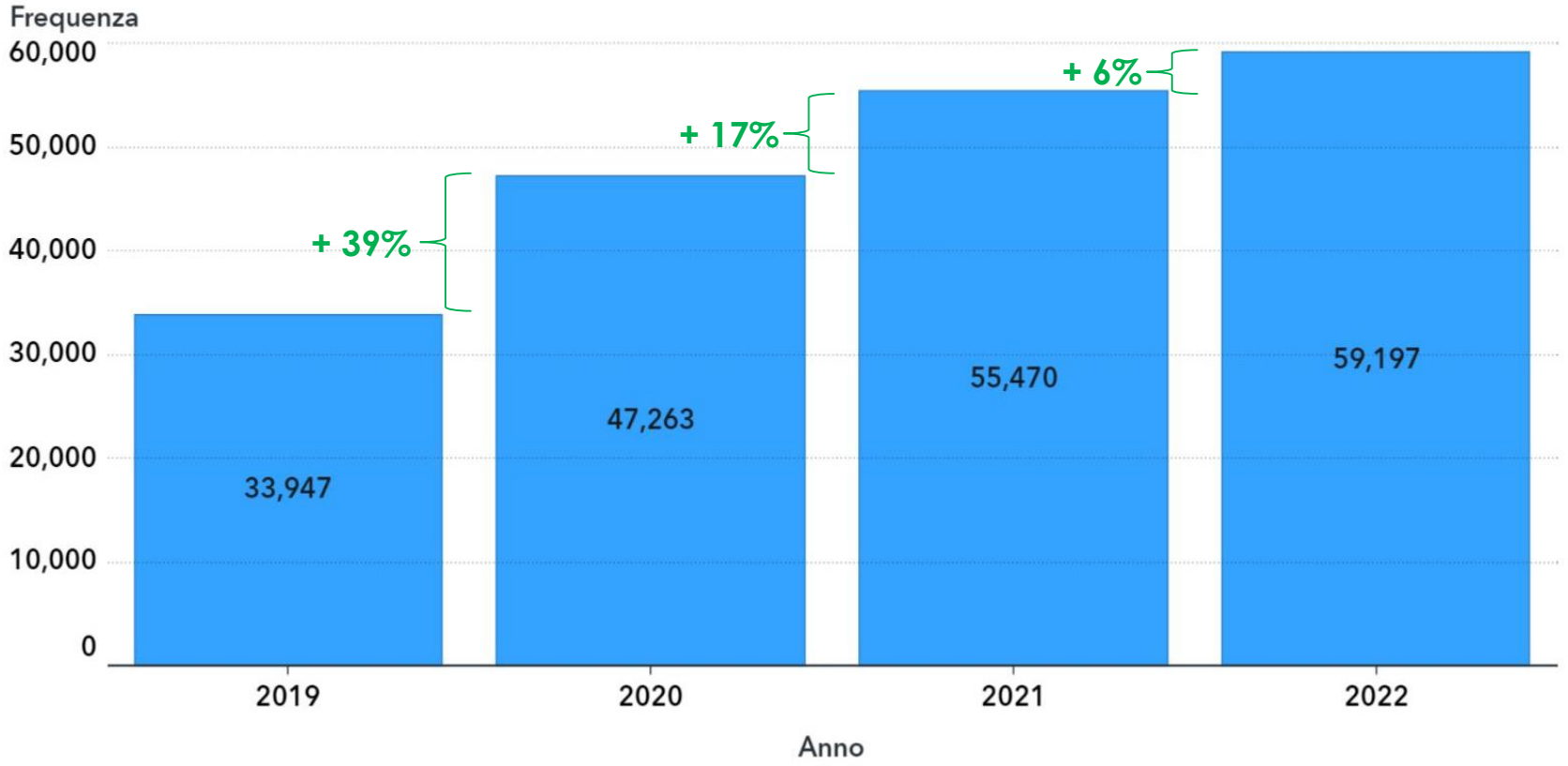
Contesto Relazionale

71% Rilevato

56% Vive Solo
35% Vive in Famiglia o con amici
8% Vive con il Partner
2% Altro

Quando chiama?

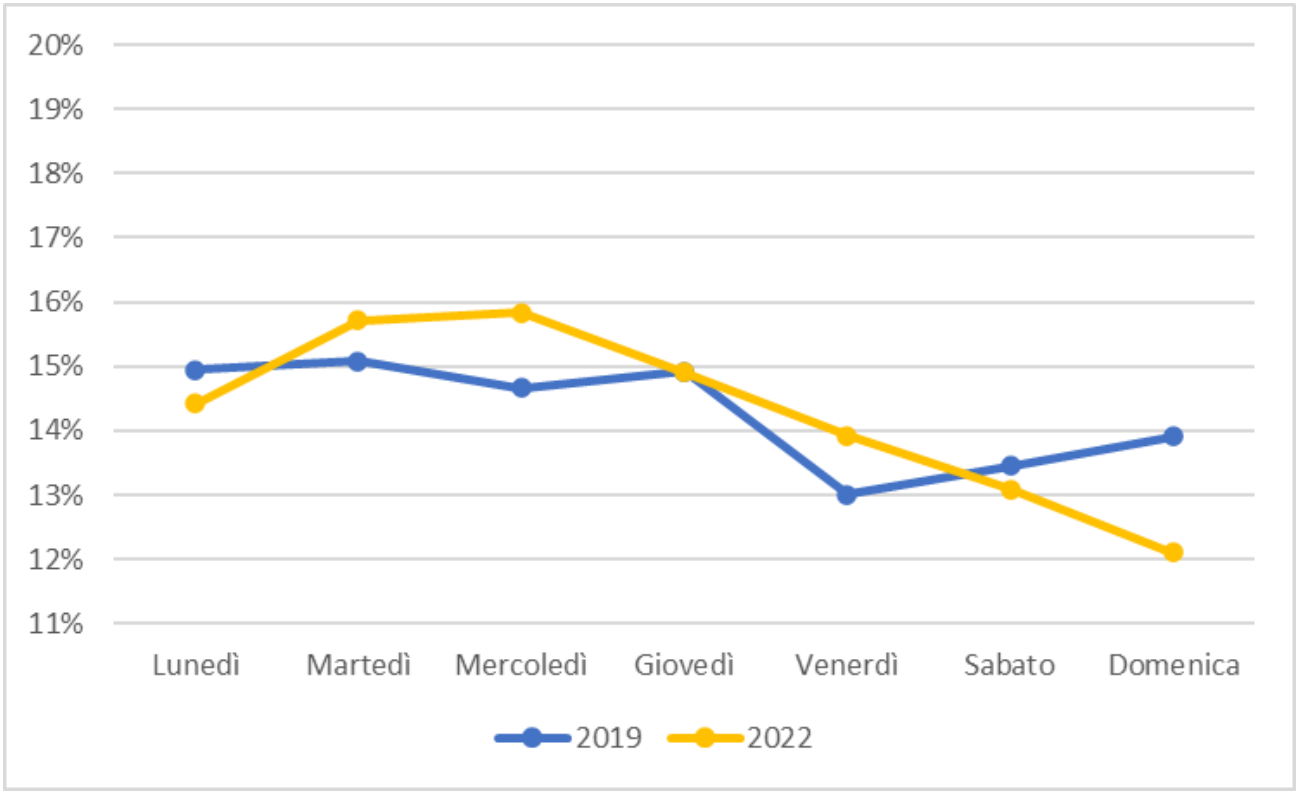
Andamento del numero delle chiamate dal 2019 al 2022



- **Crescita** del numero di chiamate, con un tasso di crescita via via più basso.
- Non è possibile sapere se all'aumento del numero di chiamate corrisponda un **aumento degli utenti o un aumento della frequenza delle chiamate** effettuate dallo stesso utente.

Quando chiama?

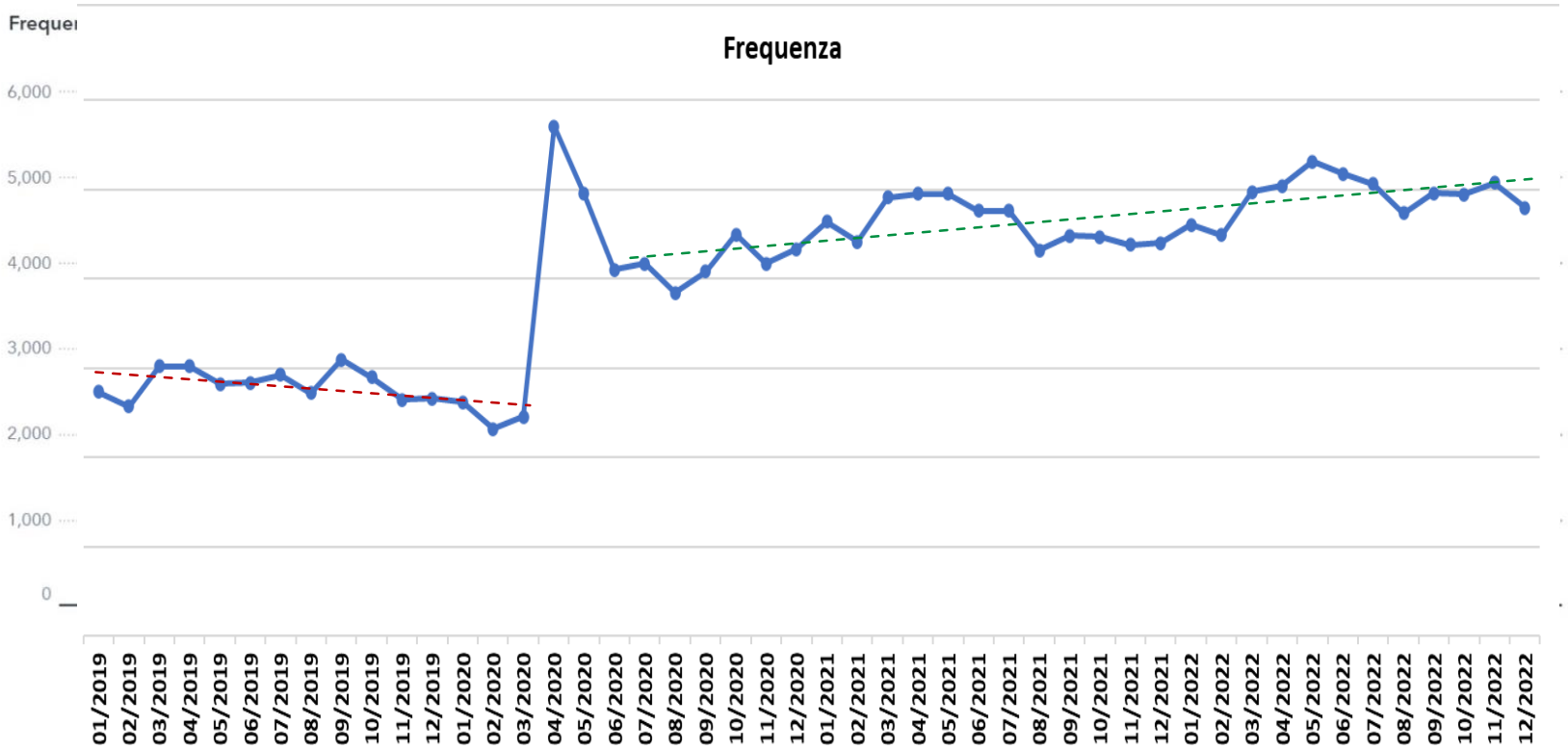
Distribuzione delle chiamate per giorno della settimana (2019 vs 2022)



- Nel **2019** maggiore affluenza di chiamate **da lunedì a giovedì** (come nel 2020).
- Nel **2022** concentrazione delle chiamate nelle giornate di **martedì** e **mercoledì** (come nel 2021)

Quando chiama?

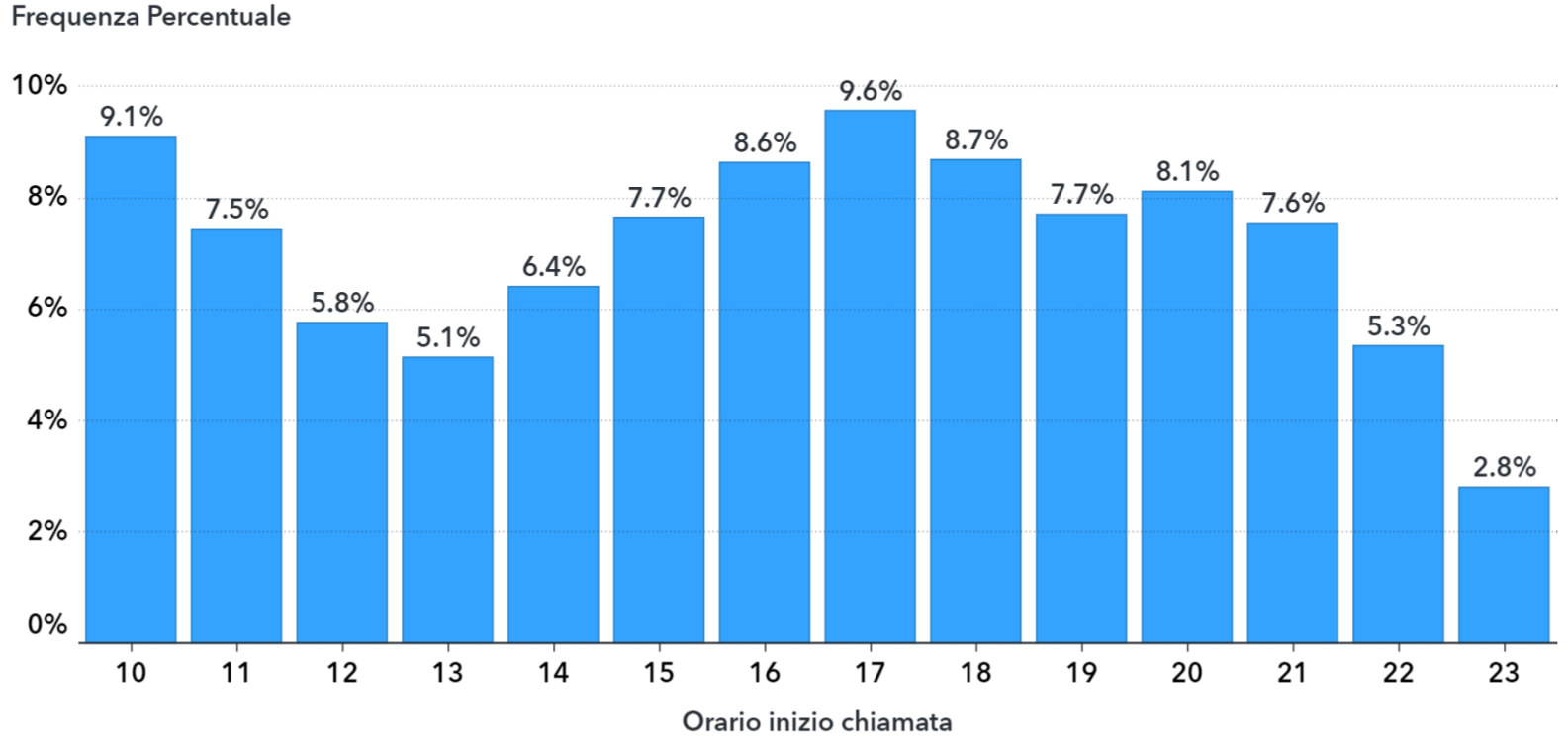
Distribuzione delle chiamate per mese



- Presenza di picco chiamate ad **Aprile-Maggio 2020** (inizio primo lockdown)
- **Da aprile 2020** il volume di chiamate, dopo un assestamento, aumenta costantemente
- Prevalenza di chiamate tra Marzo e Luglio (tutti gli anni)

Quando chiama?

Distribuzione delle telefonate per orario della chiamata



Picchi di chiamate:

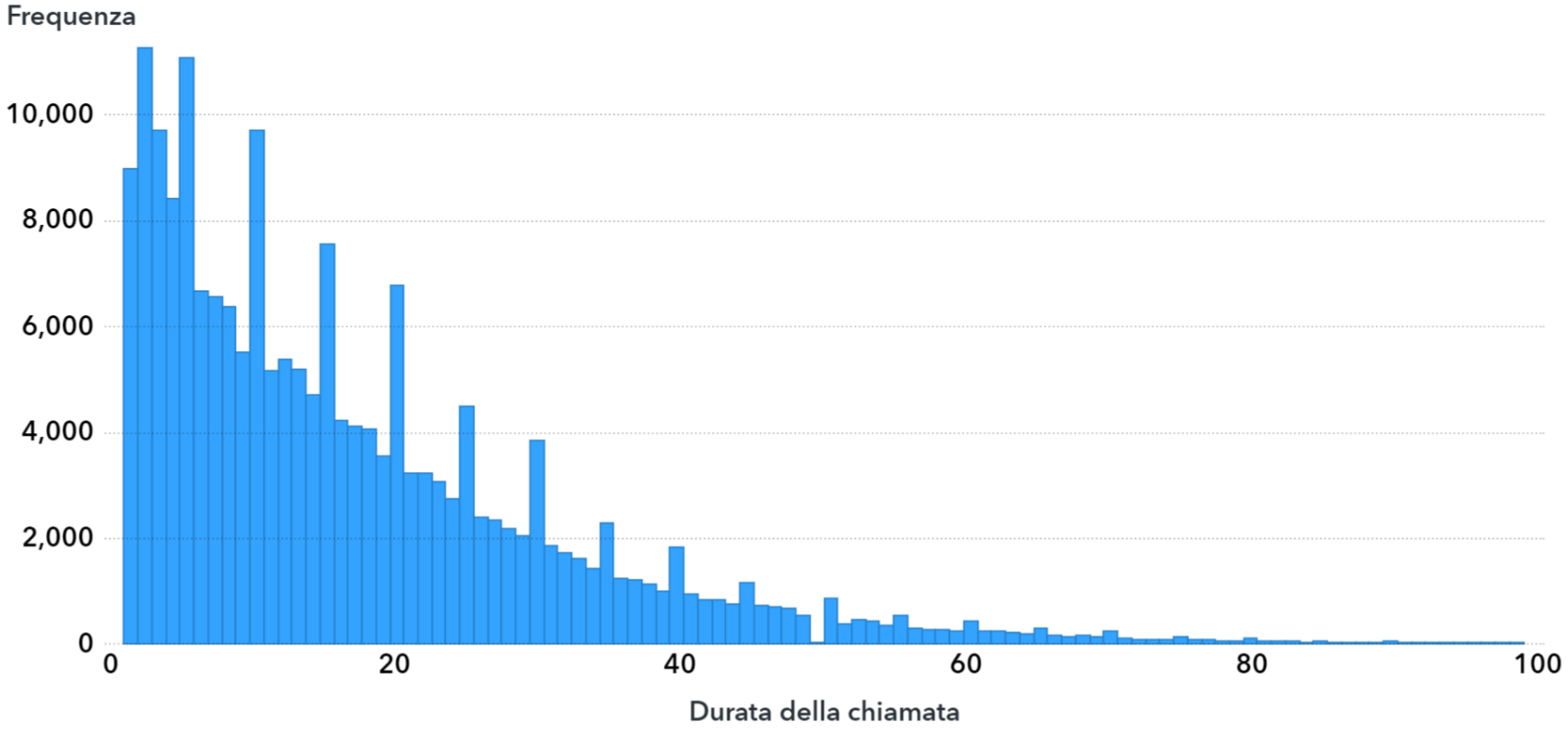
- orario di apertura
- ore 17

Diminuzione frequenza chiamate:

- durante i pasti (ore 13 e ore 19)

Quando dura la chiamata?

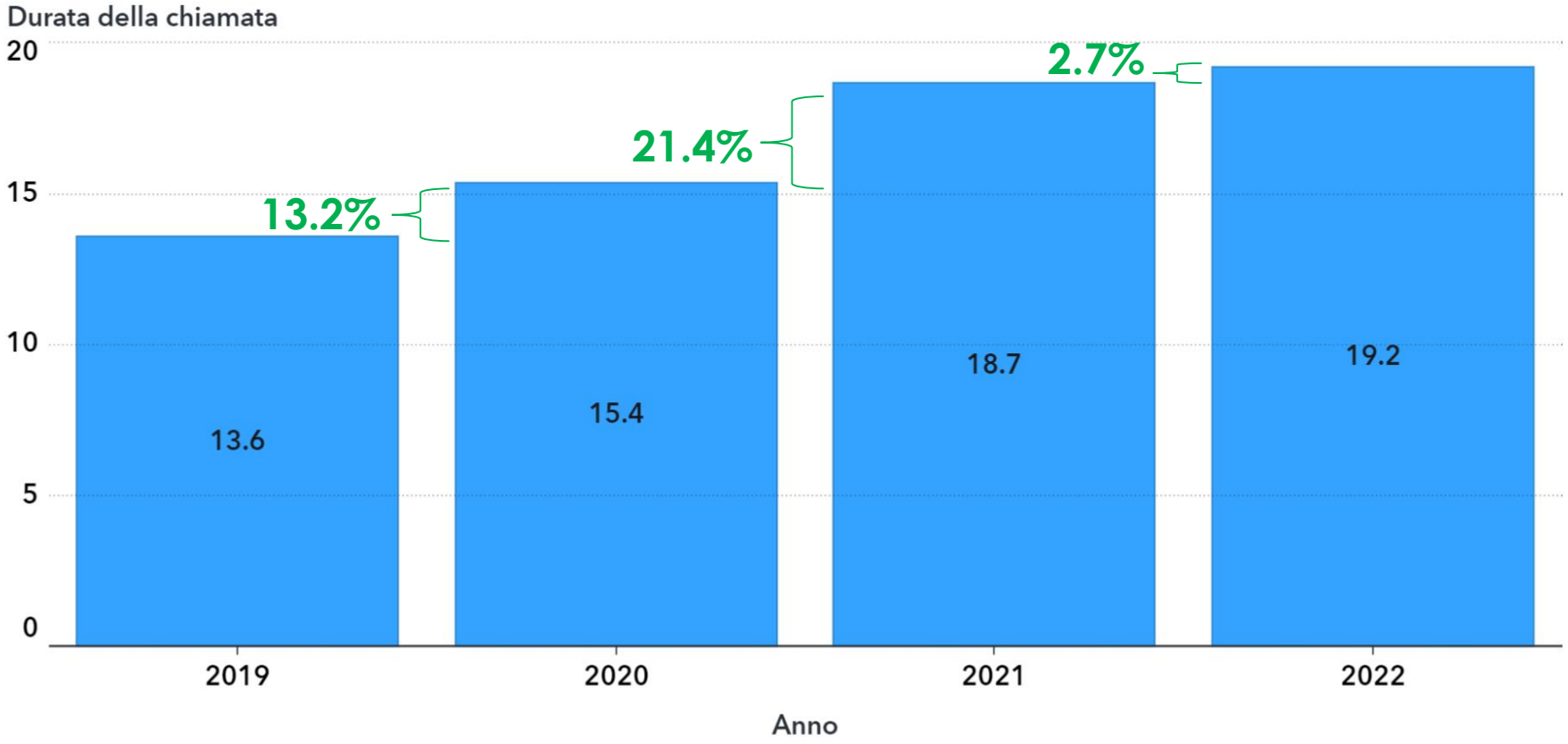
Distribuzione della durata delle chiamate



- Durata chiamata media circa **17 minuti**
- Presenza di **picchi** che rappresentano l'arrotondamento della durata della chiamata nei minuti **"multipli di 5"**

Quando dura la chiamata?

Andamento della durata media delle chiamate dal 2019 al 2022

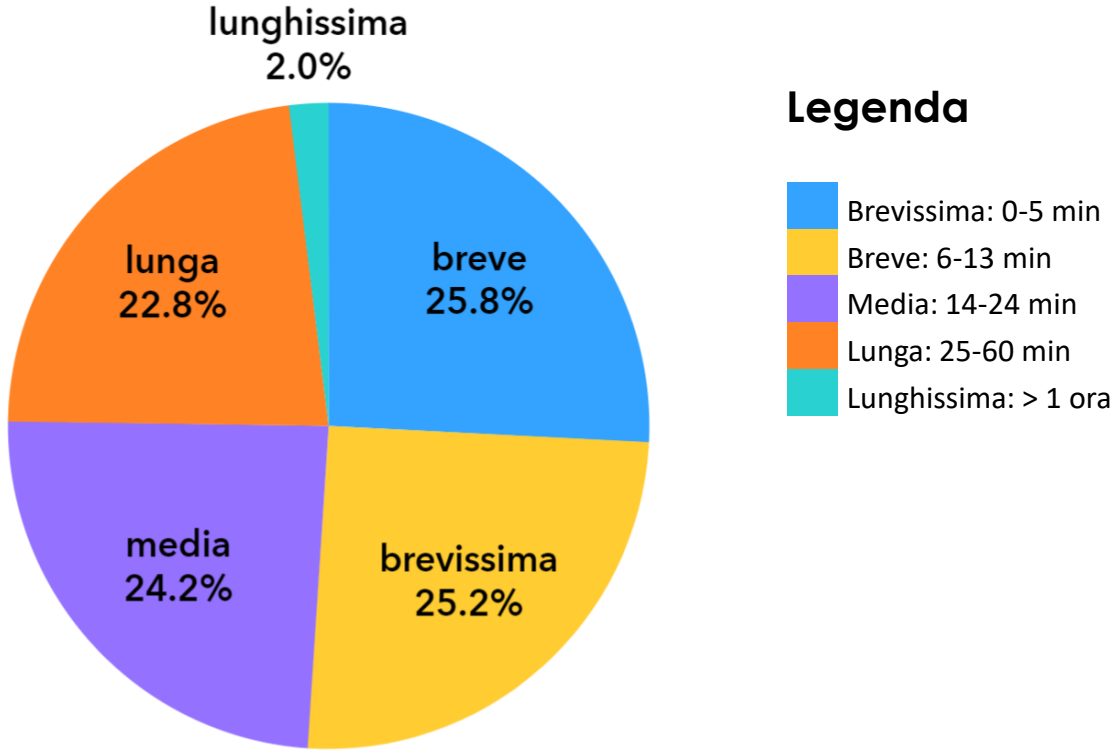


- Con l'avvento del Covid-19 la durata media delle chiamate ha subito un **incremento** considerevole
- Post Covid-19: **stabilizzazione** durata telefonata

Quando dura la chiamata?

- Circa il 25% degli utenti ha effettuato chiamate non superiori a **5 min**
- Per la metà degli utenti la durata è **< 13 minuti**
- Per il 75% degli utenti la durata è **< 25 minuti**
- Il 98% degli utenti ha effettuato chiamate **non superiori a 1 ora**
- Il restante 2% ha effettuato chiamate **superiori a 1 ora**

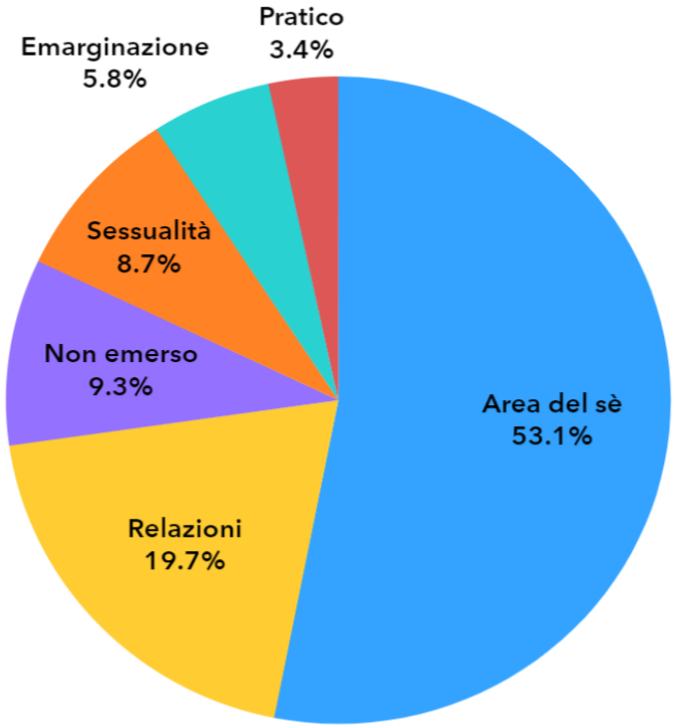
Durata della chiamata per fasce di durata



Perché chiama?

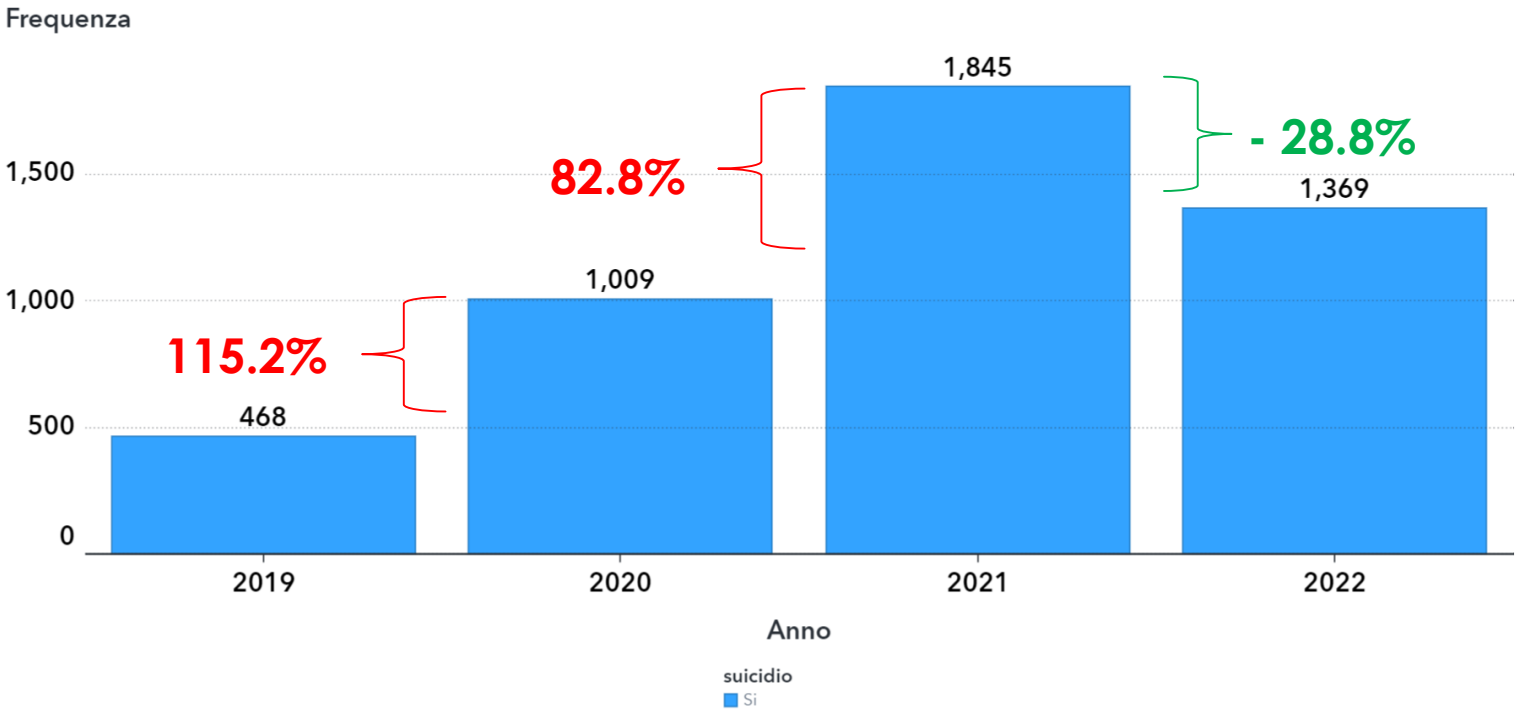
- Composizione percentuale tendenzialmente **stabile** negli anni.
- Variazione sistematica solo per la macro area dei problemi legati alla sfera **Pratica** tra 2019 e 2022 con decremento dal **5%** all'**1.8%**.
  - Si tratta di problemi legati a diversi ambiti: **abitativo**, **economico**, **lavorativo**, **giuridico**, **sanitario**.

Frequenza del problema prevalente oggetto delle chiamate



Perché chiama?

Frequenza di telefonate in cui si tratta il tema del suicidio



- Dal *2019* al *2021* le telefonate legate al tema del suicidio sono quasi **triplicate** (**+294%**)
- **Picco nel 2021**: dopo 1 anno dalla pandemia
- **Calo nel 2022** del **29%**

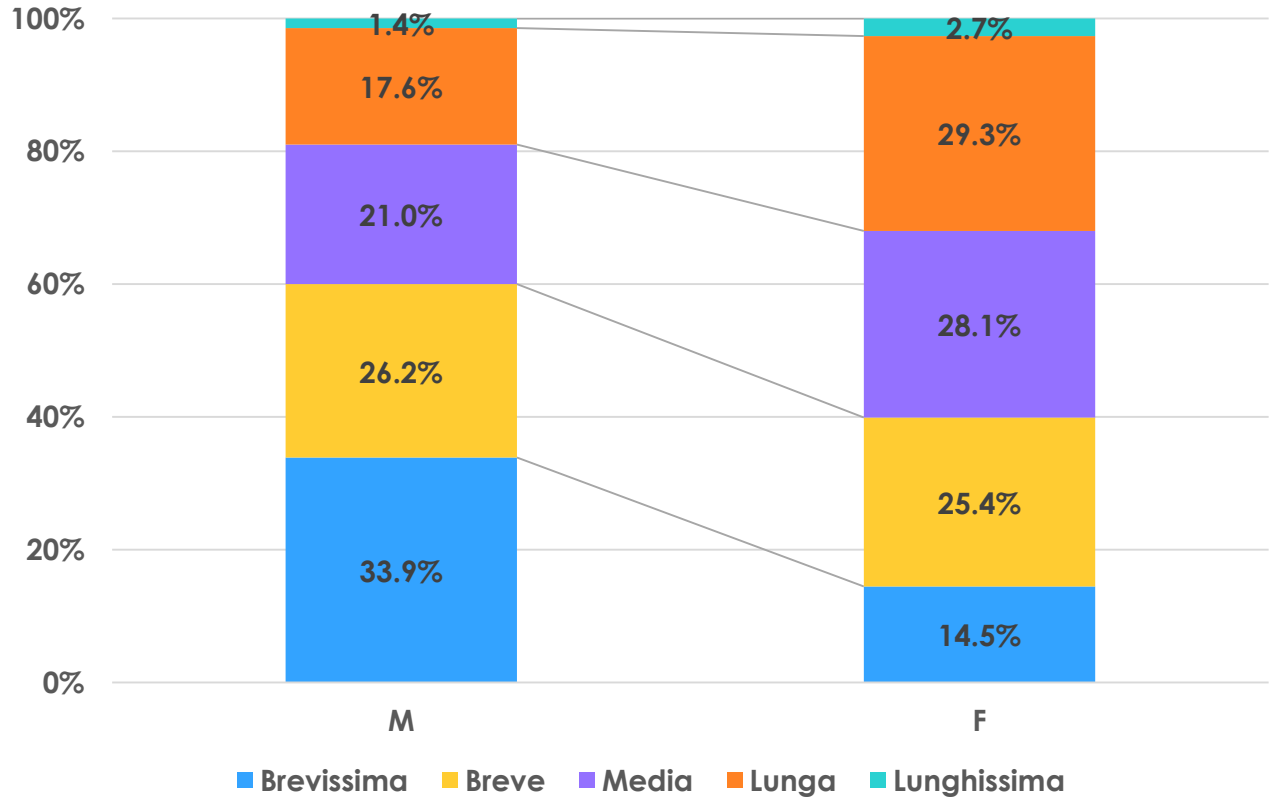
Ricerca di associazioni tra fattori

SELEZIONE DELLE ASSOCIAZIONI PIÙ SIGNIFICATIVE IN BASE AGLI INDICI STATISTICI:

- CHI QUADRO
- V DI CRAMER
- TAU-B DI KENDALL
- GAMMA
- LAMBDA ASIMMETRICA

Ricerca associazione tra fattori

Comparazione della durata delle chiamate in base al sesso

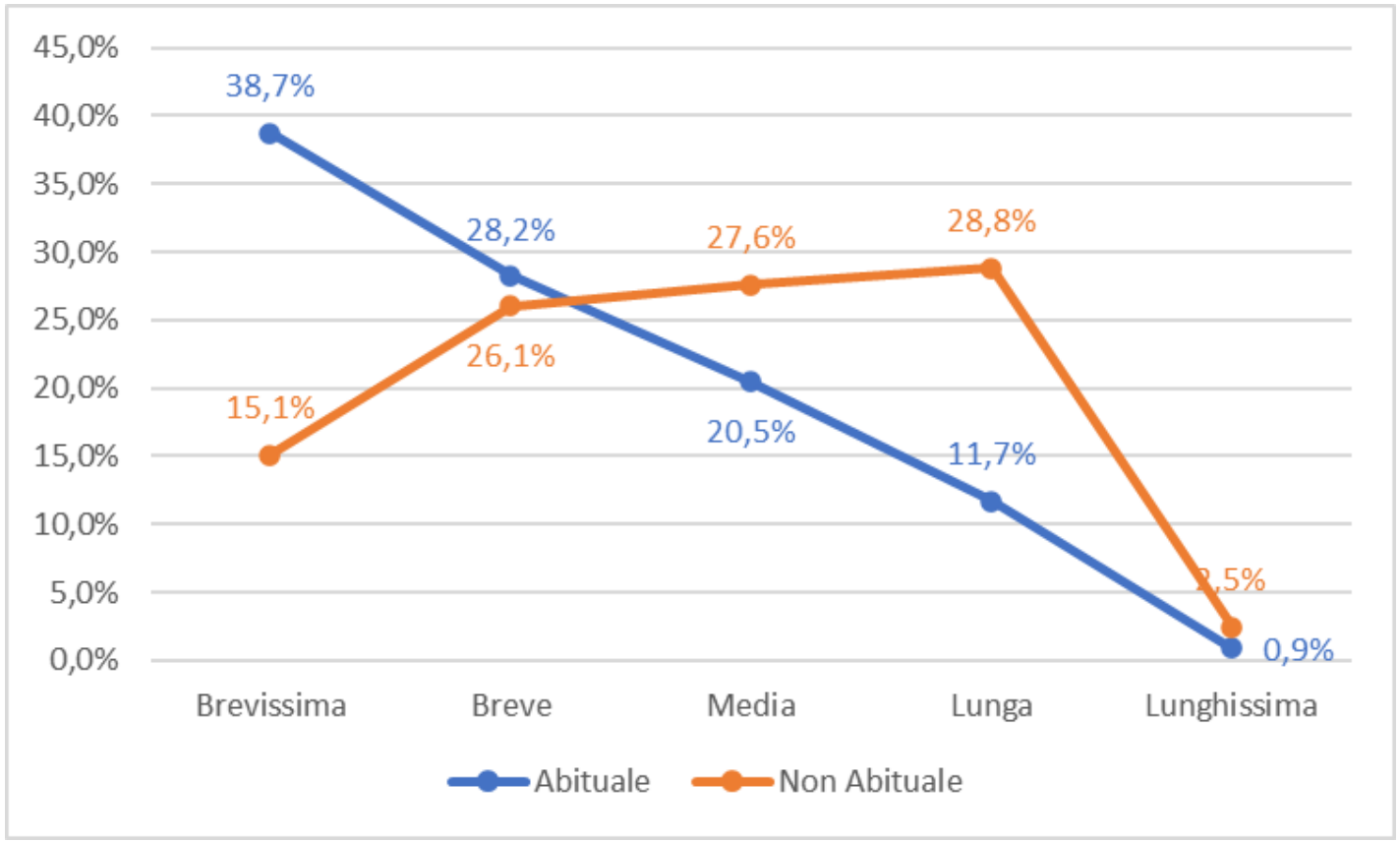


- Il **60% delle donne** effettua chiamate di durata lunghissima, lunga o media (**durata media** chiamata donna **circa 20 minuti**)

- Il ***60% degli uomini*** effettua chiamate di durata brevissima o breve (**durata media** chiamata uomo **circa 14 minuti**)

Ricerca associazione tra fattori

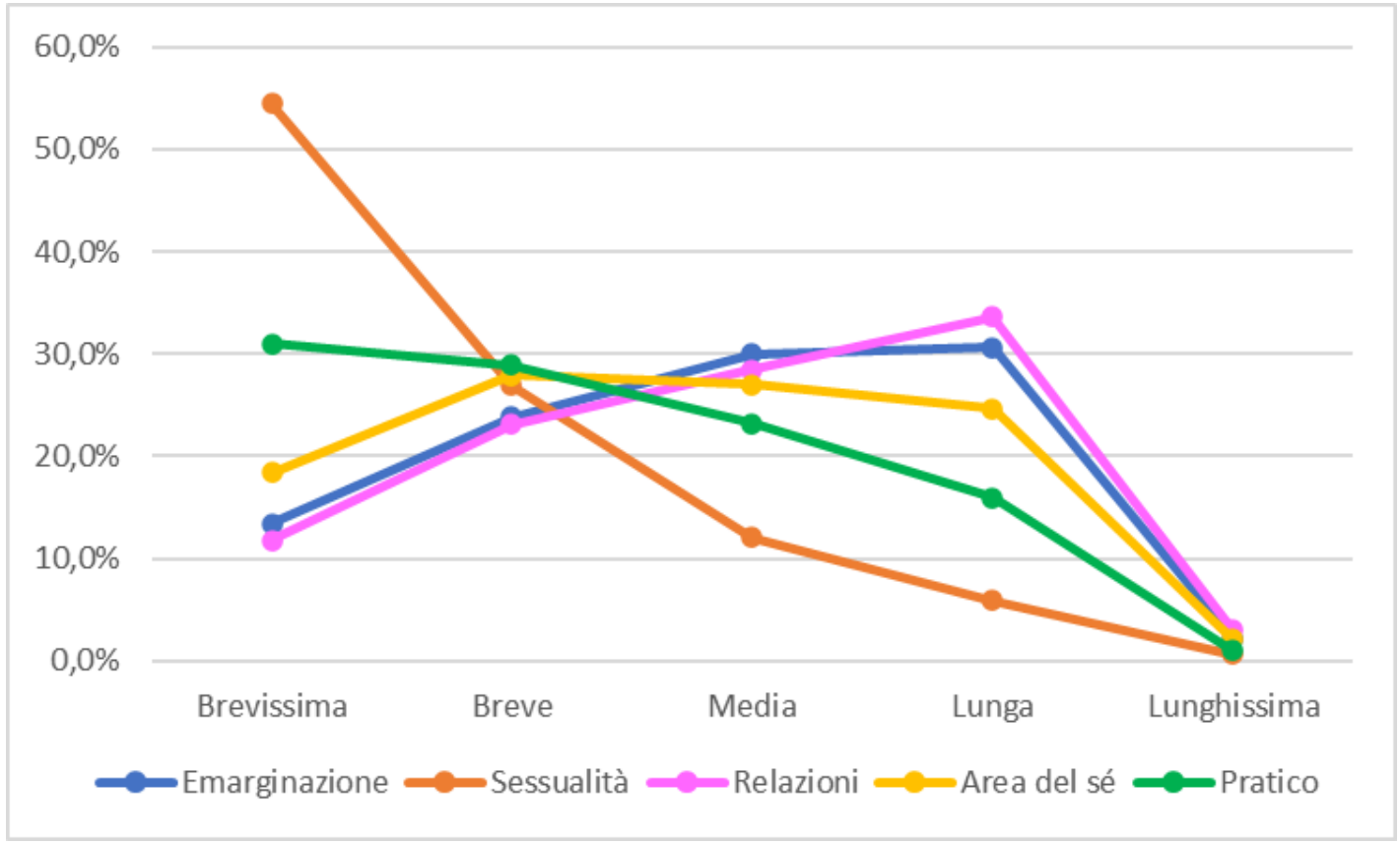
Comparazione della durata chiamata tra utenti abituali e non



- Durata media della telefonata per gli utenti **abituali**: **11 minuti**
- Durata media della telefonata per gli utenti **non abituali**: **19 minuti**
- Per gli **abituali** la maggior parte delle telefonate sono di durata brevissima e breve
- Per i **non abituali** la maggior parte delle telefonate sono di durata media e lunga

Ricerca associazione tra fattori

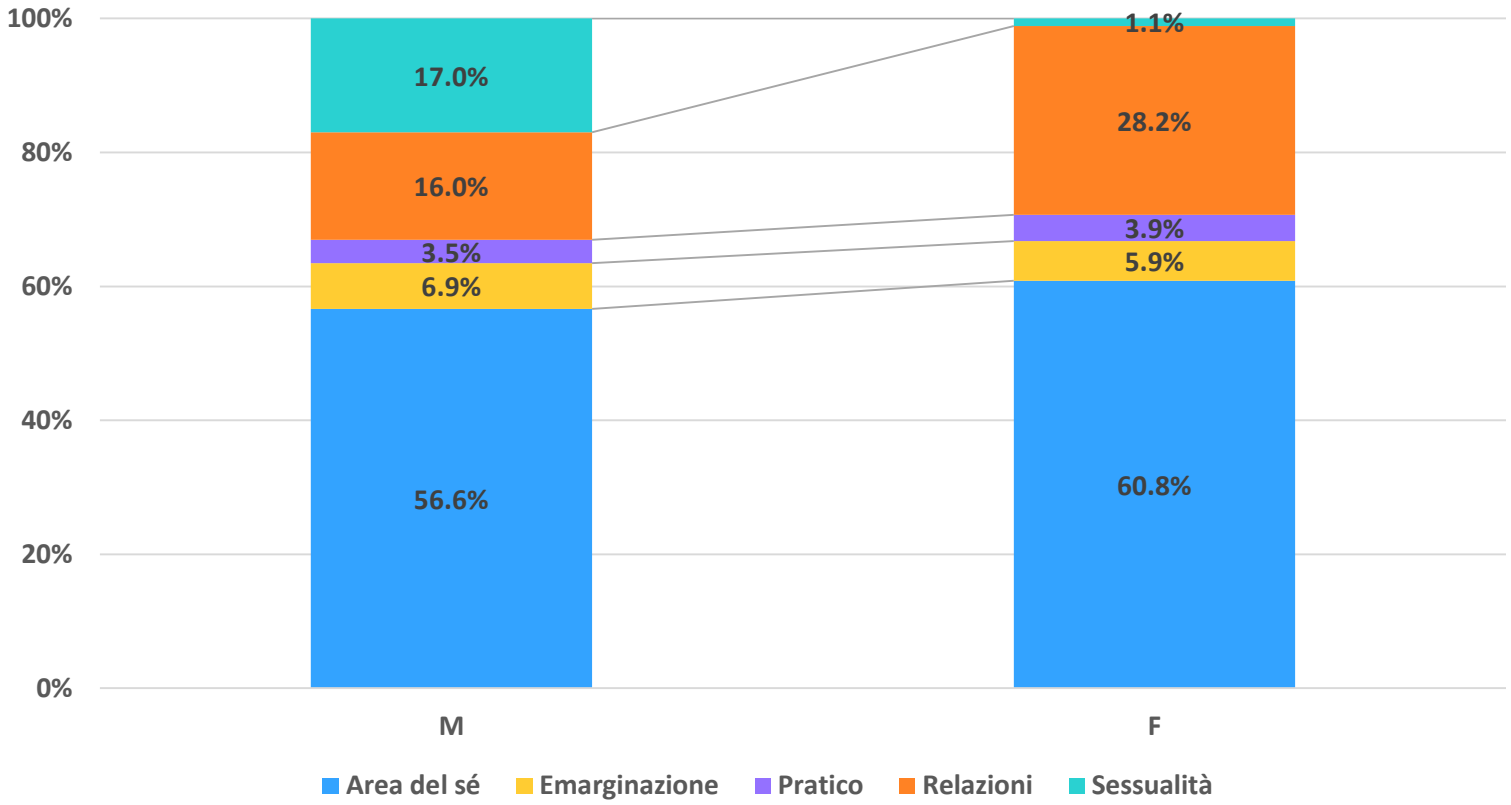
Distribuzione della durata della telefonata per tipo di problema prevalente



- Prevalenza di chiamate **brevissime** nella trattazione di temi legati alla **sessualità**
- Tendenza di chiamate **lunghe** per la trattazione di temi legati a:
  - **Emarginazione**
  - **Relazioni**
- Chiamate di **breve e media** durata nella trattazione di tematiche dell'**area del sé**

Ricerca associazione tra fattori

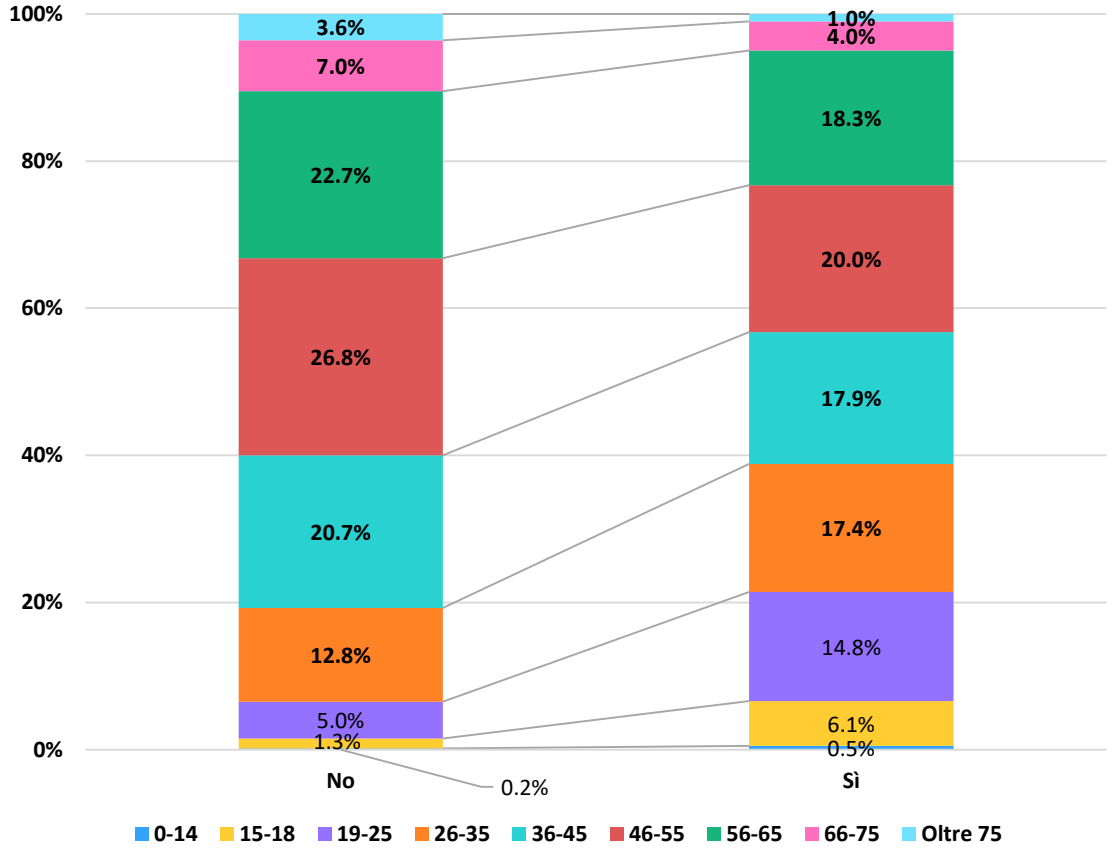
Comparazione del problema prevalente rispetto al sesso



- Problematica caratterizzante gli **uomini** rispetto alle donne:
  - **Sessualità** 17% uomini rispetto a 1% donne
- Problematica caratterizzante le **donne** rispetto agli uomini:
  - **Relazioni** 28% donne rispetto a 16% uomini

Ricerca di associazioni tra fattori

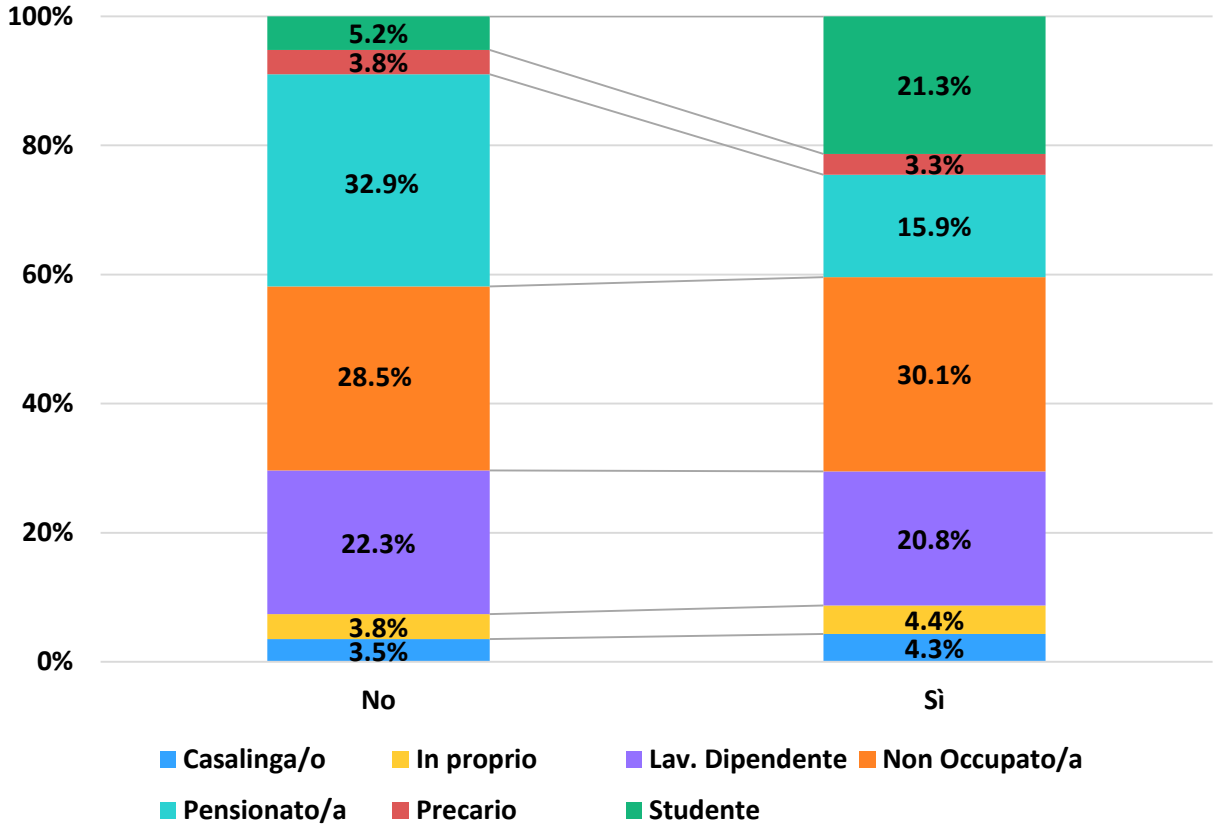
Comparazione delle fasce di età rispetto al tema del suicidio



- Prevalenza fasce di età "**giovani**" nella trattazione di temi di **suicidio** nella telefonata:
  - 15-18: 6% (vs 1.3%)
  - 19-25: circa 15% (vs 5%)
  - 26-35: 17% (vs 13%)
- Non trattazione di temi legati al suicidio per le fasce:
  - 56-65: 23% (vs 18%)
  - 66-75: 7% (vs 4%)
  - Oltre 75: 3.6% (vs 1%)

Ricerca di associazioni tra fattori

Comparazione delle professioni rispetto al tema del suicidio



- Prevalenza nella trattazione di temi legati al **suicidio** da parte di **studenti** (21%)
- I pensionati tendono a trattare di meno i temi legati al suicidio durante le telefonate (33%)

Ricerca di associazioni tra fattori

Distribuzione del problema trattato per tipo di professione

Professione	Area del sé	Emarginazione	Pratico	Relazioni	Sessualità	Totale
Casalinga/o	54,1%	3,1%	2,3%	39,9%	0,5%	100,0%
In proprio	48,2%	6,4%	5,4%	33,3%	6,6%	100,0%
Lav.dipendente	42,5%	8,3%	2,4%	25,7%	21,0%	100,0%
Non occupato/a	72,9%	6,9%	3,6%	14,7%	1,9%	100,0%
Pensionato/a	75,5%	4,6%	5,2%	14,1%	0,7%	100,0%
Precario	62,1%	13,9%	2,5%	19,0%	2,5%	100,0%
Studente	54,7%	5,3%	1,5%	35,6%	2,9%	100,0%

- Problemi prevalenti trattati: **Area del sé e Relazioni** (in particolare da parte di **non occupati** e **pensionati**)
- Argomento trattato per il 21% dalla professione **lavoratore dipendente: Sessualità**
- Argomento trattato per il 14% dalla professione **"precario": Emarginazione**

Quali sono le tipologie di chiamate?

Definizione e **raggruppamento in tipologie di chiamate** "tipo" sulla base delle seguenti **variabili**:

- Orario della chiamata
- Durata della chiamata
- Giorno della settimana
- Mese

Variabili aggiuntive e determinanti nella definizione dei "gruppi" di chiamata:

- Sesso
- Età
- Provenienza
- Professione
- Contesto relazionale
- Problema prevalente
- Tematica del suicidio durante la telefonata

Chiamate

(si riportano solo le caratteristiche significative di ogni cluster)

	Orario della chiamata	Durata della chiamata	Giorno della settimana	Mese
Telefonate diurne di fine settimana (39.6%)	Mattinata	Media (19 min)	Weekend	Maggio - Luglio
Lunghe telefonate serali (14.1%)	Tardo pomeriggio	Lunga (46 min)	-	Gennaio - Maggio
Brevi telefonate serali di fine settimana (46.3%)	Tardo pomeriggio	Breve (13 min)	Giovedì e venerdì	Settembre - Dicembre

Il **primo** e il **terzo** cluster si distinguono solo sulla base delle caratteristiche specifiche delle telefonate:

entrambi si caratterizzano per I seguenti **problemi prevalenti**:

- **Sessualità**
- **Compagnia**
- **Malattia psichica**

	Sesso	Età	Provenienza	Professione	Contesto relazionale	Problema prevalente	Tematica suicidio durante la chiamata
Lunghe telefonate serali	F	Mediamente più giovani (19-35)	Non rilevato basso	Non rilevato basso	Non rilevato basso Tendenzialmente "Vive Solo"	Non rilevato basso Esistenziali, Relazioni familiari, Relazioni sentimentali Assenza "Sessualità"	Alto (doppio rispetto alla media)

Chi sono gli abituali?

Campione di 83 utenti con più di 20 chiamate dal 2019 al 2022

Definizione e **raggruppamento in tipologie di utenti abituali** sulla base delle seguenti **variabili**:

- Orario della chiamata
- Durata della chiamata
- Problema prevalente
- Giorno della settimana
- Mese

Variabili aggiuntive e determinanti nella definizione dei "gruppi" di utenti abituali:

- Sesso
- Età
- Provenienza
- Professione
- Contesto relazionale
- Tematica del suicidio durante la telefonata

Abituali

(si riportano solo le caratteristiche significative di ogni cluster)

	Orario della chiamata	Durata della chiamata	Giorno della settimana	Mese	Problema prevalente	Sesso	Età	Provenienza	Professione	Contesto relazionale	Tematica suicidio durante la chiamata
Anziani soli (28.9%)	Ore 16	Breve	Mercoledì	Maggio Giugno Dicembre	Bisogno di compagnia Malattia psichica Malattia fisica	M	Anziani (66+)	Sud e Isole	Pensionato Non occupato	Residenza Sanitaria Assistenziale (R.S.A) Vive solo	-
Donne esaurite (30.1%)	Ore 19	Lunga	Sabato Domenica	Settembre Dicembre	Relazioni familiari Relazioni amicali Disagio psicologico	F	46-65	Nord Est	Casalinga In proprio	Vive solo	-
Uomini con problemi di coppia (19.3%)	Ore 16	Breve	Venerdì	Agosto Ottobre Novembre	Sessualità Relazioni di coppia	M	36-45	Non rilevato	Non rilevato	Vive con Partner	Molto basso
Immigrati / Emarginati (14.5%)	Ore 19	Lunga	Lunedì Martedì Giovedì	Estate: Luglio - Settembre	Emarginazione Esistenziali Prospettive e cambiamento Disagio psicologico	-	Giovani (26-35)	Sud e Isole	Non occupato	Badanti Vive con famiglia/amici	Basso
Giovani Problematici (7.2%)	Ore 19	Breve	Sabato	Inizio Anno: Gennaio - Aprile	Sessualità Relazioni sentimentali Relazioni di coppia	M	Principalmente Molto giovani (19-25) Ma anche Giovani (26-35)	Centro	Lavoratore dipendente In proprio	Non rilevato Vive solo	Alto

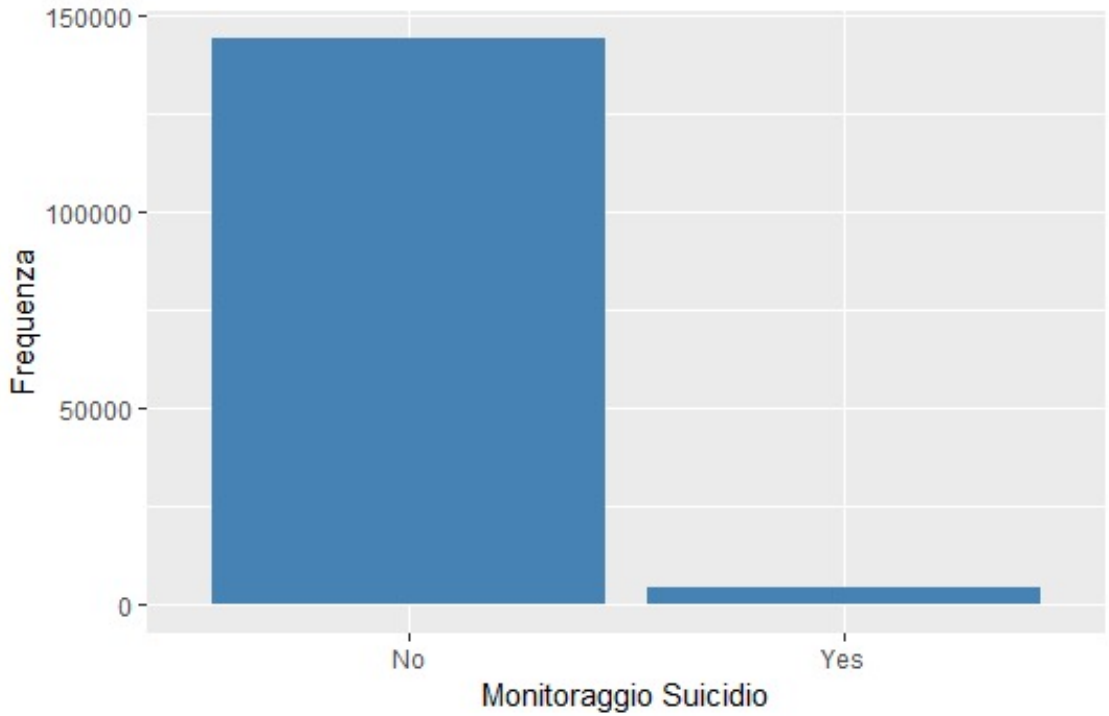
Modelli

La costruzione di **modelli descrittivi** ci consente di determinare attraverso le caratteristiche della telefonata (durata, orario chiamata) e di chi telefona (sesso, professione...) **se la conversazione verterà sul tema suicidio**.

I modelli utilizzati sono:

- Logistico
- Albero di classificazione

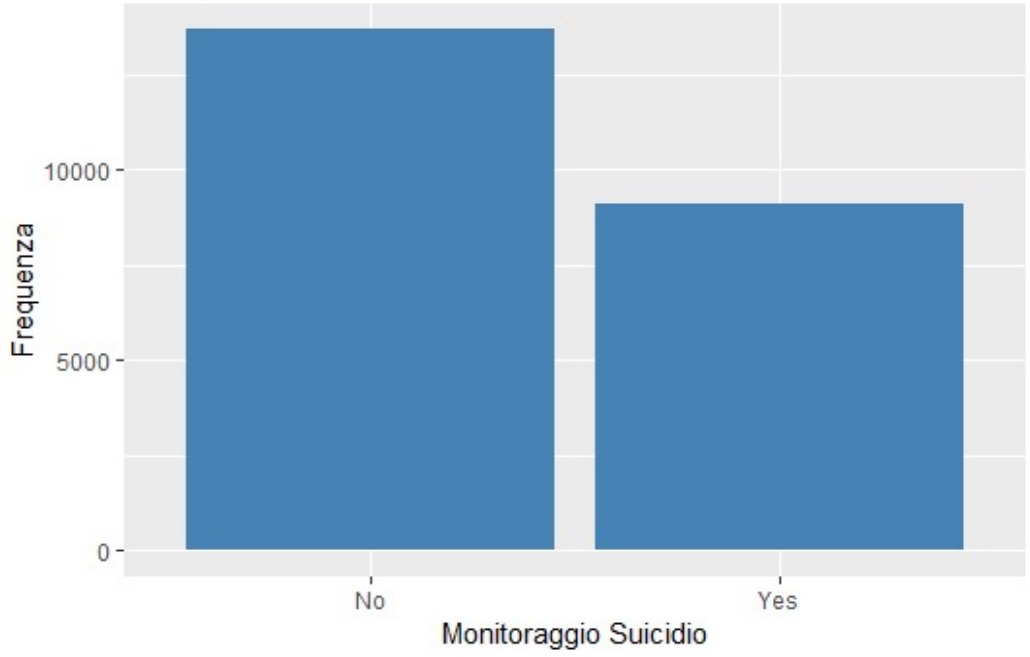
Modelli



- La prima analisi svolta consiste nella verifica della frequenza delle classi della variabile target.
- È possibile notare un **forte sbilanciamento tra la classe 1 e la classe 0 (chi parla e chi non parla di suicidio):**
  - **3.1% parla di suicidio**
  - **96.9% NON parla di suicidio**

Modelli

- Abbiamo applicato al dataset originario una tecnica di ***oversampling (SMOTE)*** per incrementare i casi di telefonate con argomento suicidio e l'***undersampling*** per ridurre le telefonate senza argomento suicidio. In questo modo i modelli sono applicati a un campione sufficientemente bilanciato e potranno generare risultati soddisfacenti.
- Il dataset finale risulta composto: per il **60%** da coloro che durante la chiamata **non parlano di suicidio (No)**, per il **40%** invece, da chi **ne parla (Yes)**.
- Il dataset così ottenuto è stato poi suddiviso in un **Training** e in un **Test** sample rispettivamente pari al **80%** e **20%** del dataset.



Modello Logistico

Per stimare il modello Logistico, sono state dapprima utilizzate tutte le variabili presenti nel dataset. Sono state quindi selezionate solamente quelle **variabili statisticamente significative**:

- Età
- Contesto relazionale
- Provenienza
- Professione

Di queste variabili sono state poi analizzate le **modalità significative**.

Modello Logistico

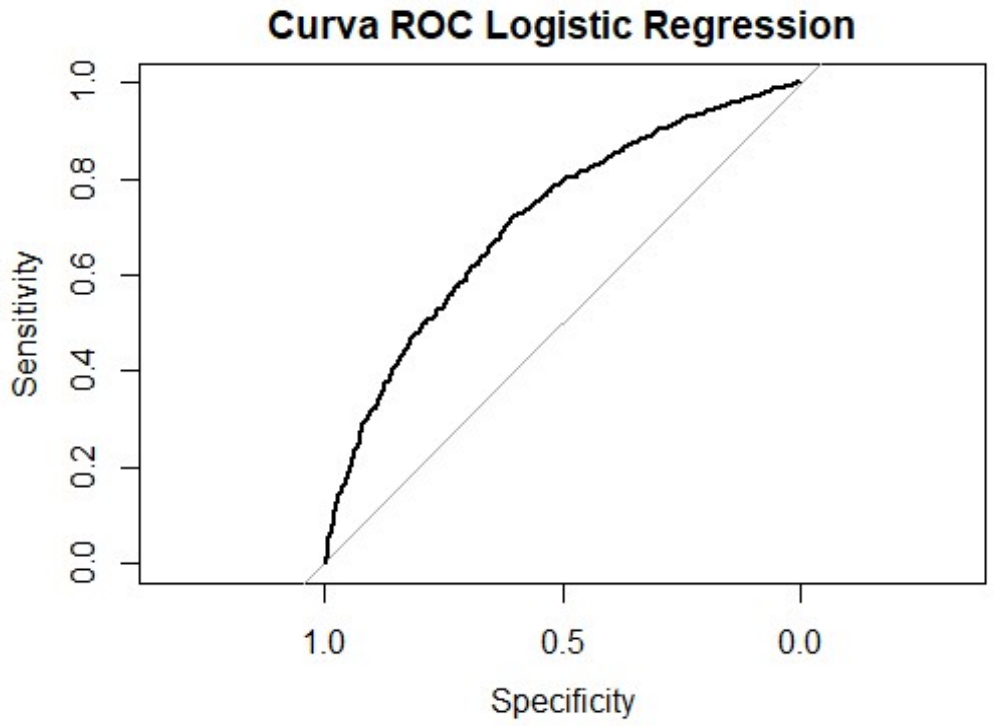
Dall'analisi degli **Odds Ratio** è possibile osservare:

- Rispetto all'**età** si può affermare che, superati i 25 anni di età, la probabilità che il tema suicidio emerga tende a diminuire con l'aumentare dell'età dell'utente.
- Per la **Provenienza**, non si riscontrano differenze significative tra le diverse aree territoriali.
- Riguardo la **Professione**, se la telefonata proviene da uno studente allora aumenta la probabilità che si parli dell'argomento suicidio.
- Per il **contesto relazionale**, la modalità per cui risulta più probabile rilevare il tema del suicidio è la classe Altro (comunità/badante)

Modello Logistico

Dalla ***curva ROC*** (nell'immagine a destra) si evince che il modello stimato ha una capacità di classificare correttamente gli eventi (Yes/No) superiore rispetto ad un modello *at random*, che non utilizza alcune delle variabili esplicative considerate nell'analisi.

Per valutare la capacità di generalizzare i risultati è possibile osservare l'***AUC*** (area sotto la curva ROC). Il valore dell'area è pari a ***0,71*** che indica una buona capacità complessiva di classificare correttamente gli eventi.



Modello Logistico

Metriche utilizzate

Metrica	Valore
Precisione	0.5720207
Richiamo	0.6046002
Specificità	0.7258922
AUC	0.7104413
F1 Score	0.5878594

Per un cut-off=0,5

- La **Precisione** indica tra tutte le telefonate classificate dal modello come quelle che hanno l'argomento suicidio quante sono quelle che effettivamente lo trattano. Qui il 57% delle telefonate classificate con argomento suicidio presentano davvero tale argomento.
- La ***Recall*** (richiamo) indica la quota di telefonate con argomento suicidio che sono state ben classificate dal modello. In questo caso il modello riesce a individuare correttamente il 60% delle telefonate con argomento suicidio.

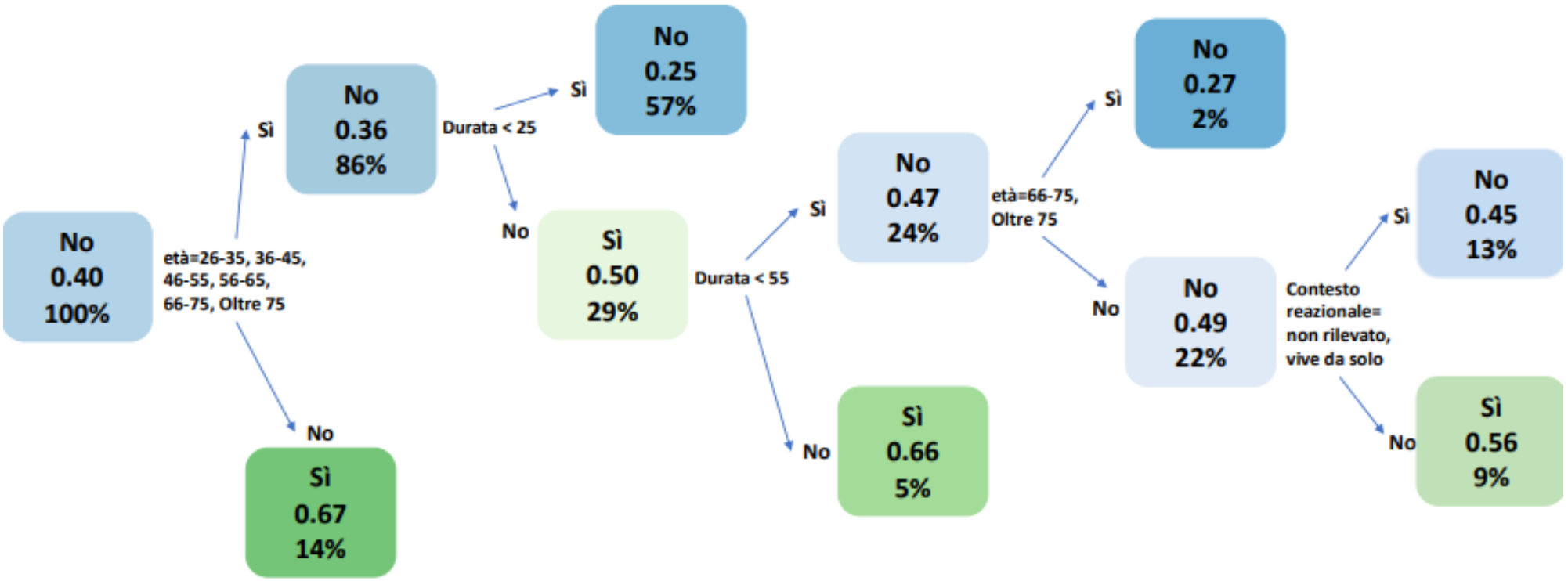
Albero di classificazione

L'albero di classificazione è un modello parametrico, tipico dell'approccio machine learning, efficace nel mettere in luce le eventuali interazioni tra le variabili esplicative del modello

L'output del modello è un insieme di regole basate sui combinazioni di valori delle variabili esplicative che suddividono l'insieme delle telefonate in un determinato numero di sottogruppi (a ciascuno di questi gruppi corrisponde uno specifico insieme di regole)

Alle telefonate di ogni sottogruppo è associata una sola delle due classi (telefonata contenete argomento suicidio vs telefonata senza argomento suicidio)

Albero di classificazione



LEGENDA

SI
0.67
14%

SI/NO = Tema suicidio / non tema suicidio

0.67 = % di osservazioni in cui si riscontra il suicidio nel nodo

14% = % di osservazioni totali nel nodo

Albero di classificazione

L'albero di classificazione è stato utilizzato come **modello descrittivo**. Sono stati determinati sei nodi principali.

- Di seguito i nodi che evidenziano i casi in cui **emerge il tema suicidio**.
  - Il primo (dall'alto verso il basso) comprende persone con **età ≤25 anni**. Questo nodo è formato dal 14% delle osservazioni totali e al suo interno nel 67% delle osservazioni si riscontra il tema suicidio.
  - Il secondo comprende persone con **età >25 anni** con chiamate dalla **durata >55 minuti**. Questo nodo è formato dal 5% delle osservazioni totali e al suo interno nel 66% delle osservazioni si riscontra il tema suicidio.
  - Il terzo comprende persone con **età tra i 26 e i 65 anni** e con una **durata della telefonata tra i 25 e 55 minuti**. In aggiunta il soggetto vive con famiglia, amici, partner o altro (badante, comunità). Questo nodo è formato dal 9% delle osservazioni totali e al suo interno nel 56% delle osservazioni si riscontra il tema suicidio.

Albero di classificazione

- Di seguito i nodi che evidenziano i casi in cui **il tema suicidio emerge in modo meno frequente**.
  - Il quarto comprende persone con **età tra i 26 e i 65 anni** con chiamate dalla **durata tra i 25 e 55 minuti**. In aggiunta il soggetto vive da solo o questa informazione non è stata rilevata. Questo nodo è formato dal 13% delle osservazioni totali e al suo interno nel 55% delle osservazioni non si riscontra il tema suicidio.
  - Il quinto comprende persone con **età >65 anni** con chiamate dalla **durata tra i 25 e 55 minuti**. Questo nodo è formato dal 2% delle osservazioni totali e al suo interno nel 73% delle osservazioni non si riscontra il tema suicidio.
  - Il sesto comprende persone con **età >25 anni** e che hanno una **durata della telefonata >25 minuti**. Questo nodo è formato dal 57% delle osservazioni totali e al suo interno nel 72% delle osservazioni non si riscontra il tema suicidio.

Analisi sui volontari

Di seguito viene presentata un'analisi esplorativa riguardante i volontari che gestiscono la chiamata, al fine di **individuare eventuali bias compilativi** riferiti al volontario che gestisce la chiamata.

Potrebbe verificarsi che i diversi volontari mostrino una maggiore predisposizione e sensibilità nel rilevare il tema del suicidio nella chiamata, nonchè nell'individuare il problema predominante della chiamata e i dati anagrafici correlati.

L'analisi è condotta solo sui volontari che rispondono a chiamate **di utenti non abituali**, per evitare un effetto distorsione causato dalle caratteristiche uniche dell'utente abituale.

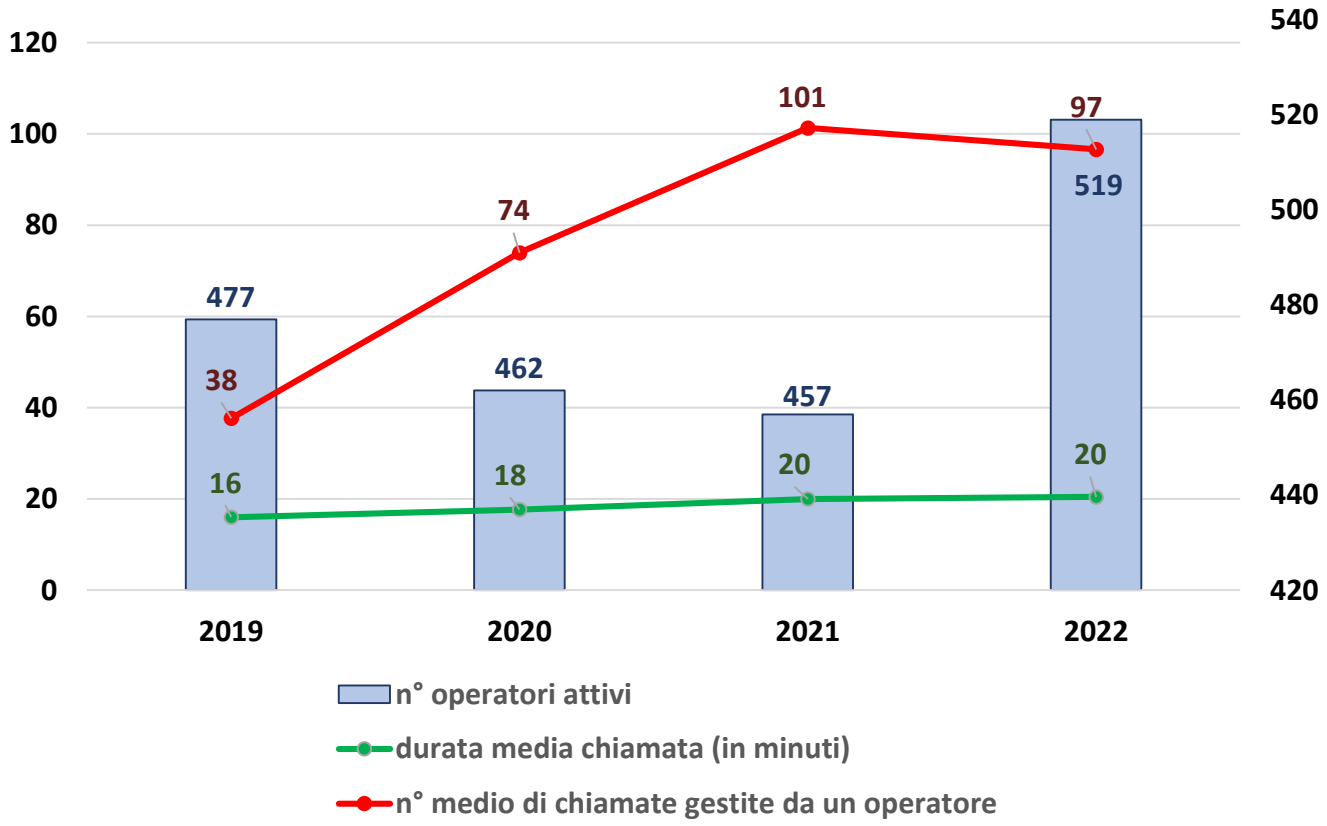
Inoltre è bene sottolineare come una maggior predisposizione a gestire chiamate in cui emerge il tema del suicidio possa essere condizionata anche dal **tipo di utente** che chiama il servizio.

Variabile	% Non Rilevato o Non Emerso
Provenienza	48%
Professione	60%
Contesto Relazionale	35%
Problema Prevalente	8%
Altre Segnalazioni	49%

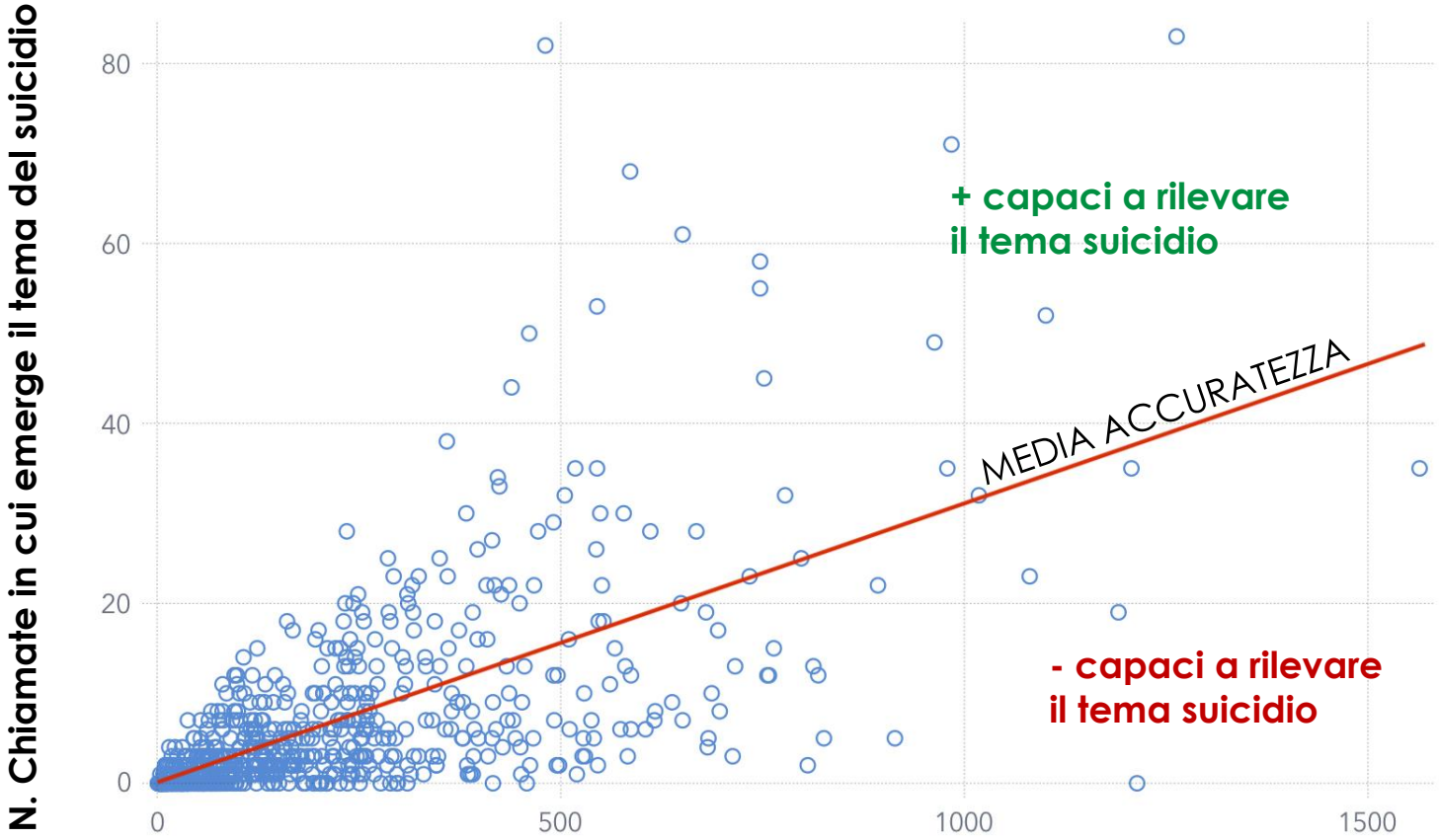
Quanti sono e quante chiamate ricevono?

Notiamo come nei primi 3 anni, sebbene il numero di chiamate gestite sia aumentato, il numero dei volontari attivi è diminuito. Tuttavia nel 2022 c'è stato un netto incremento dei volontari, che ha compensato l'aumento delle chiamate. La durata media delle chiamate, invece, è passata dai 16 minuti del 2019 ai 20 del 2022.

Si ricorda che il grafico riguarda esclusivamente chiamate di **utenti non abituali**.



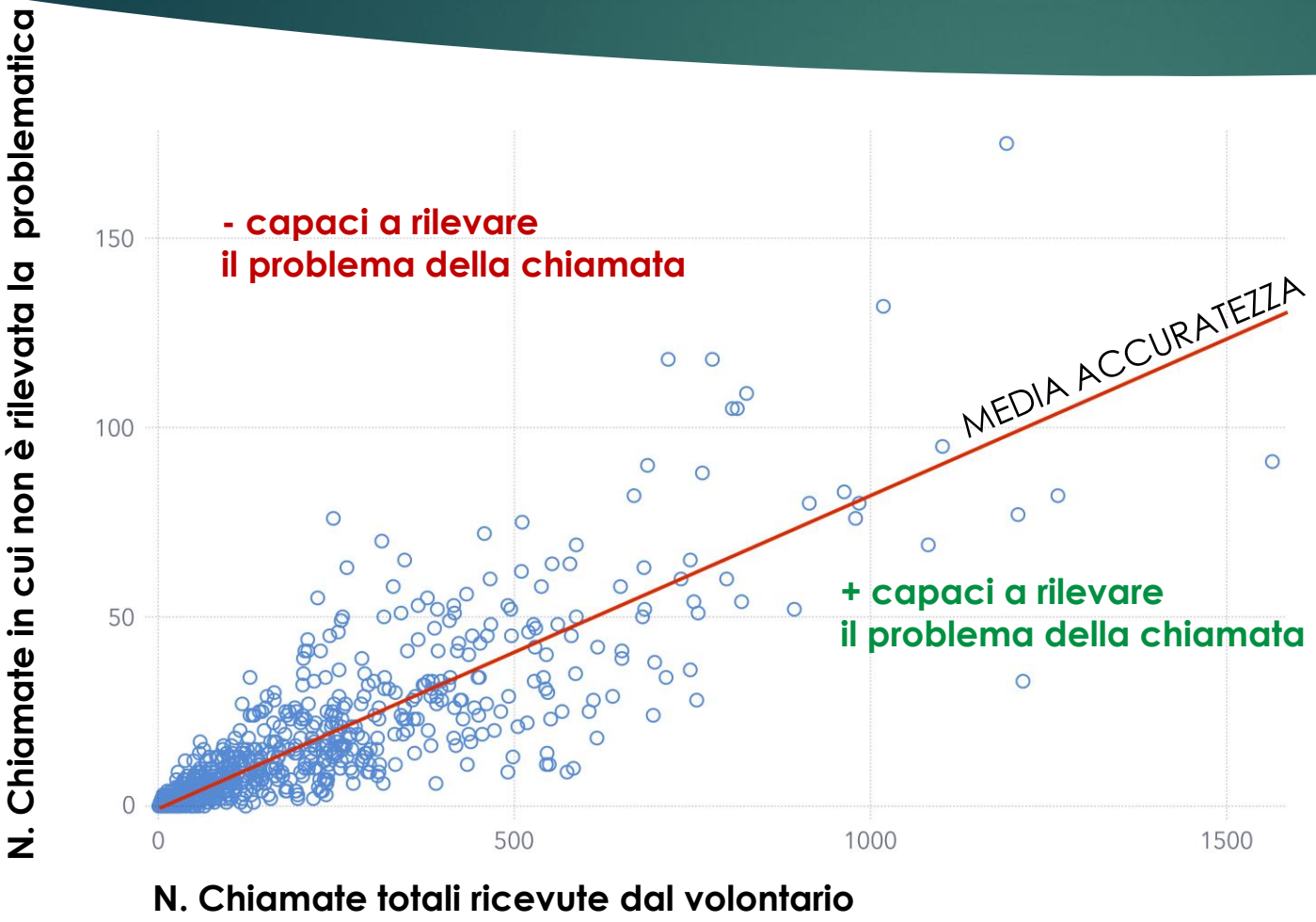
Rilevazione tema del suicidio



Ogni **punto** nel grafico corrisponde ad un **volontario**. Sulle **ascisse** sono presenti le **chiamate totali** gestite da ciascun volontario. Sulle **ordinate** sono riportate il numero di chiamate in cui è emerso il tema del suicidio gestite da ogni singolo volontario. La **retta** rappresenta come varia il numero medio di telefonate in cui appare il suicidio al variare del numero totale di chiamate per volontario.

Al di sopra della retta troveremo quindi tutti quei volontari che hanno una **maggiore propensione e sensibilità nel far emergere dalla chiamata il tema del suicidio**.

Capacità nel rilevare il problema prevalente



In questo caso, il grafico mostra la capacità del volontario a **rilevare il problema prevalente** della chiamata.

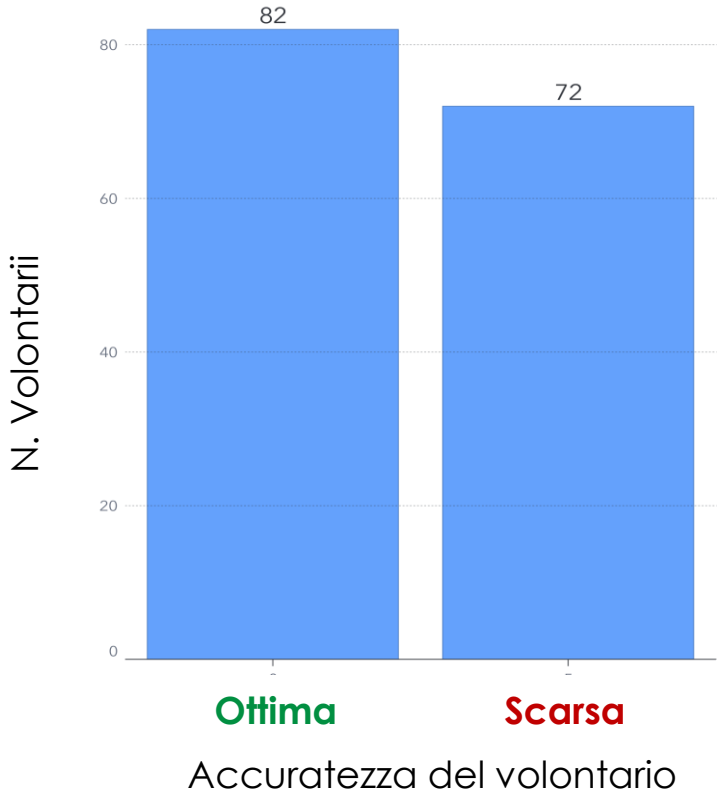
Anche in questo caso avremo volontari che **mostrano una maggiore attenzione** nell'individuazione del problema della chiamata, i quali sono collocati **al di sotto** della retta.

Gli stessi trend si registrano per altre variabili, come la **provenienza**, la **professione**, il **contesto relazionale** e le **segnalazioni**.

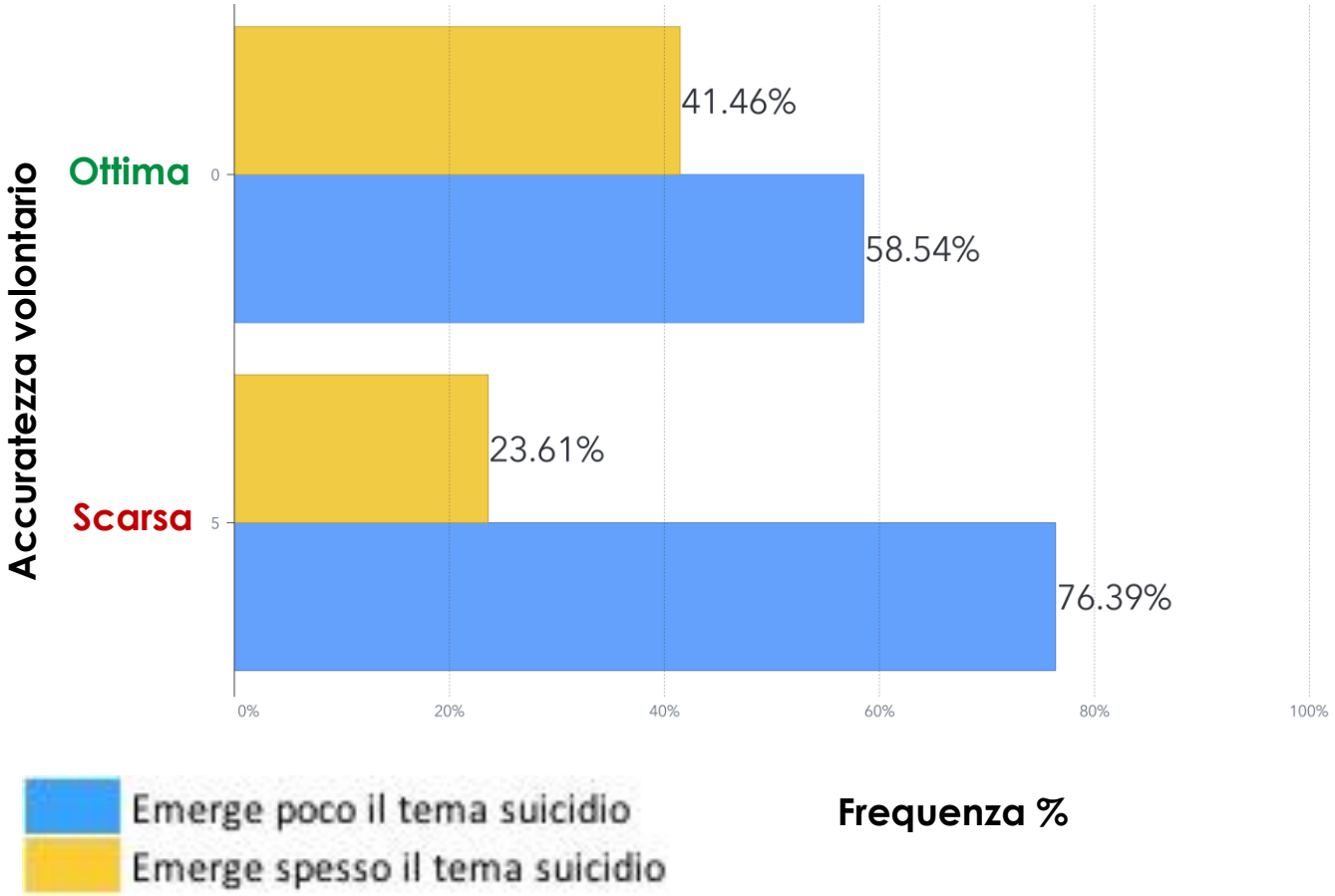
Accuratezza dei volontari

Sono state definite **due diverse classi**, che identificano il grado di accuratezza dei volontari: nella classe "**ottima**" si collocano tutti I volontari che si posizionano sempre al di sotto la retta vista nel grafico precedente, in tutte e 5 le variabili anagrafiche dove è presente la **modalità "Non rilevato**", ovvero **Problema prevalente, Provenienza, Professione, Contesto relazionale** e **Segnalazioni**. Al contrario, nella classe "**scarsa**" sono collocati tutti i volontari che si posizionano per tutte e 5 le variabili sempre al di sopra della retta.

Ottima: Volontari che per tutte e 5 le variabili sono sempre al di sotto della retta: ossia hanno un'accuratezza superiore alla media.
Scarsa: Volontari che per tutte le 5 variabili sono sempre al di sopra della retta, ossia hanno un'accuratezza inferiore alla media.



Rilevazione del tema suicidio: vi è un possibile BIAS?



Ai volontari che presentano una **maggiore propensione a non rilevare sia il problema della chiamata sia le variabili anagrafiche**, **corrisponde un più basso tasso di rilevazione del tema del suicidio?**

Sì, poiché emerge che il gruppo di volontari che mostrano una minore accuratezza nel rilevare i dettagli della chiamata è minore la percentuale di telefonate in cui emerge il tema del suicidio rispetto ai volontari con un'ottima accuratezza.

Fabiana Bracaglia
Nicolò Apolloni
Raffaello Cesetti
Chiara Apuzzo
Luca Barbaro
Nicla De Camillis
Giulio Baldini
Fabrizio Franchitti
Alessandro Piccirillo
Luca Spagnuolo
Giuseppe Giugliano