

Analisi Sognare insurance S.r.l.

Sommario

<i>Analisi Sognare insurance S.r.l.</i>	1
Preambolo	3
Cosa troviamo nel dataset?	3
Analisi	3
Uso della compagnia assicurativa	3
Spesa assicurativa associata al rischio fumatore e obesità	4
Rischio dei figli a fumo e malnutrizione	4
Conclusioni	6

Preambolo

L'azienda Sognare Insurance S.r.l. è leader per quanto riguarda le assicurazioni. In questo report ci si propone di analizzare i clienti per trovare eventuali criticità o pattern ricorrenti. Abbiamo analizzato il csv "insurance.csv", scaricabile al link:

<https://www.kaggle.com/datasets/ahmettezcantekin/beginner-datasets?resource=download>

Cosa troviamo nel dataset?

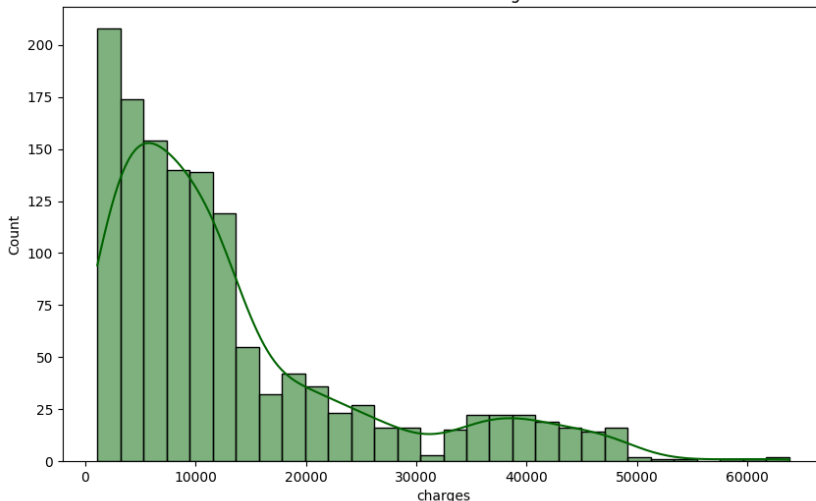
- Troviamo delle informazioni generali della persona che ha sottoscritto l'assicurazione, come età, sesso, figli a carico, peso.
 - La spesa annuale dell'assicurazione
 - Se fumatore oppure no
 - Localizzazione geografica dell'utente
-

Analisi

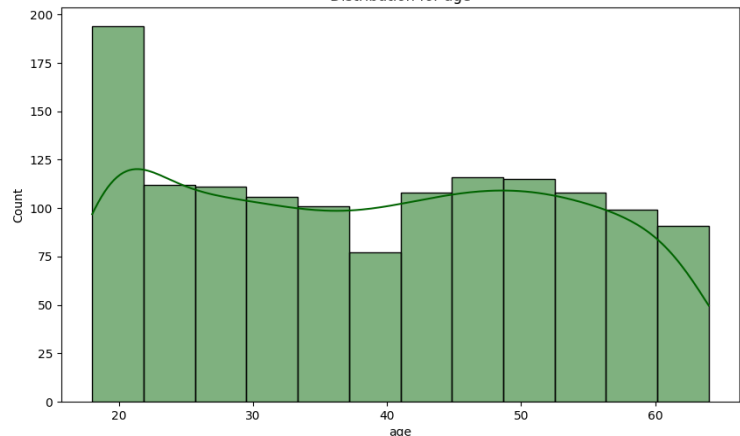
Uso della compagnia assicurativa

Sognare Insurance opera su quattro regioni: northeast, northwest, southeast e southwest, con una fascia d'età tra i 18 e 64 anni. Si noti come la

Distribution for charges



Distribution for age

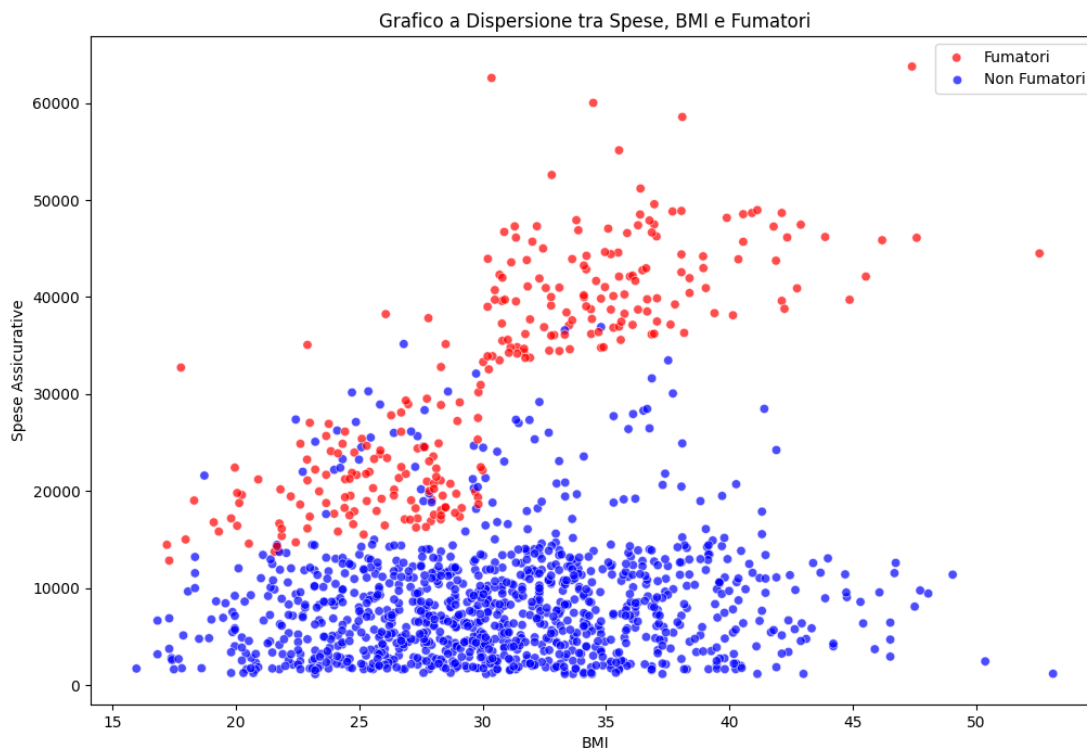
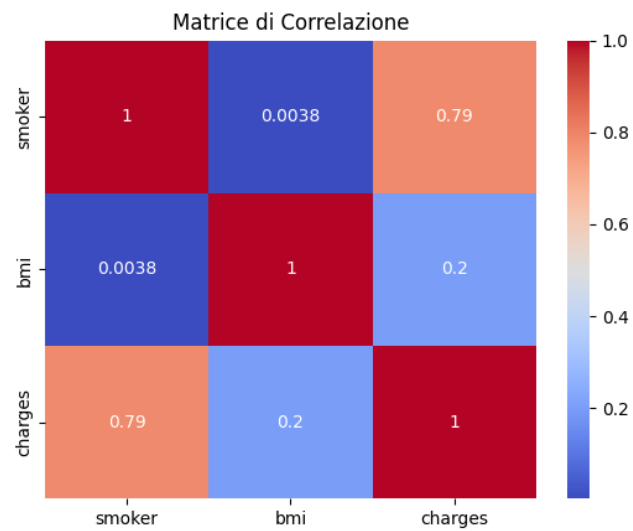


distribuzione per età è abbastanza distribuita, eccezion fatta per la fascia dei 20 anni in positivo e dei 40 in negativo. Il luogo di provenienza invece non presenta picchi, tutte le regioni risultano competitive allo stesso modo.

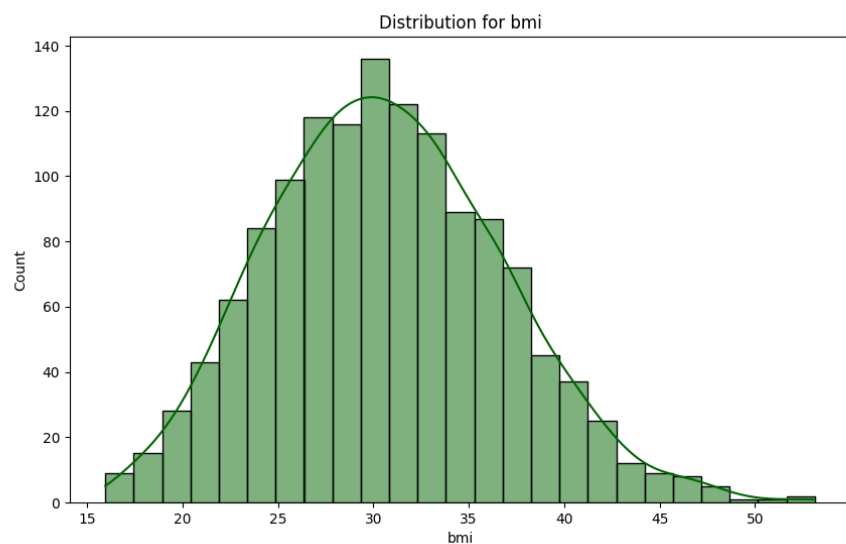
Per quello che riguarda le rate attualmente in essere, la maggior parte si concentra prima dei 10000\$ anno o di poco superiore.

Spesa assicurativa associata al rischio fumatore e obesità

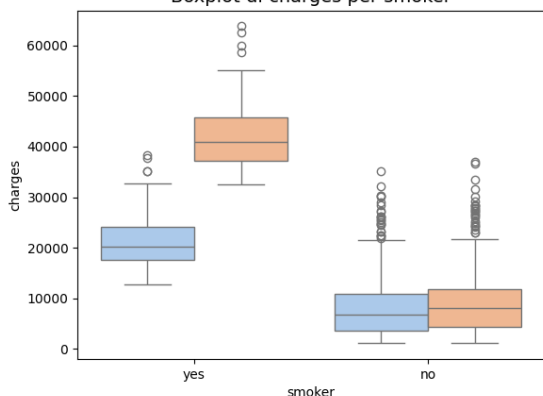
Soprattutto con la spinta delle ricerche scientifiche degli ultimi anni, si è appreso della pericolosità del fumo e della malnutrizione, in questo caso in riferimento alla sovr-alimentazione. Questi devono essere



necessariamente presi in considerazione come elementi di rischio, aumentando le rate. Abbiamo evidenziato come questo processo sia già stato iniziato dall'azienda, come si vede dalla heatmap qui riportata. Abbiamo una forte correlazione tra le rate e i fumatori, al contrario di quanto abbiamo evidenziato con il bmi. Questo possiamo verificarlo anche nella mappa di dispersione qui di seguito, dove i puntini rossi, i fumatori, tendono ad essere sopra ai non fumatori. Il BMI, invece, nell'asse delle ordinate, non presenta particolari trend, nonostante, considerato un valore di BMI massimo per il normopeso pari a 25, la maggior parte dei clienti sia oltre tale soglia.



Boxplot di charges per smoker



Un unico trend visibile è quello dei fumatori e bmi alto, dove risulta, dopo un valore superiore a 30, che le rate sono sensibilmente più alte. Infatti si nota come i punti rossi hanno una crescita importante dopo il valore di 30. Lo vediamo anche nel boxplot a lato, gli smoker hanno un balzo importante dopo il 30, al contrario di quanto accade nei non fumatori.

Rischio dei figli a fumo e malnutrizione

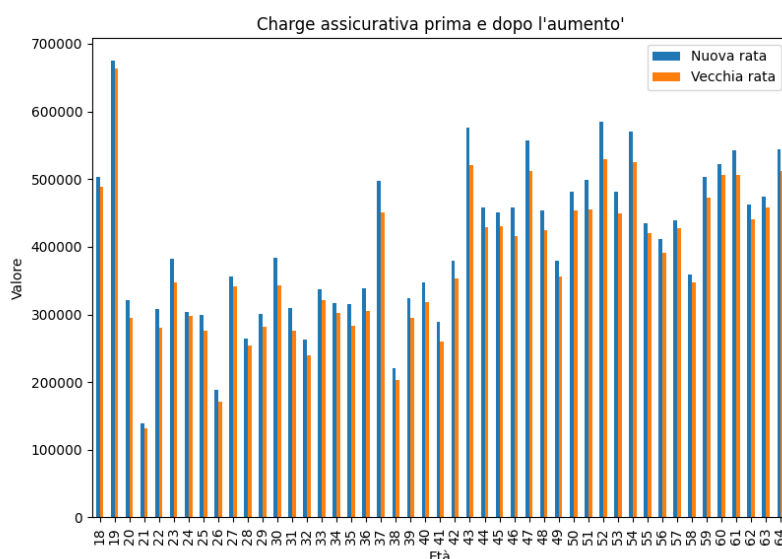
Notiamo che i dataset forniti ci danno informazioni sulle abitudini dei clienti potenzialmente nocive, come una scorretta alimentazione e utilizzo di sigarette. Questa sezione considera il rischio che i figli di questi clienti possano “copiare” questi comportamenti, conseguentemente rischiando di avere problemi vari di salute.

Infatti è più probabile che se un genitore ha una alimentazione scorretta, la porti al figlio, che soprattutto nell'età delicata della crescita, potrà portare alla formazione di malattie. Inoltre, per quel che invece riguarda i fumatori, è stato più volte dimostrato come anche il fumo passivo possa influire negativamente nella salute del ragazzo/a.

La modalità operativa suggerita è quella di aumentare la rata per tutte quelle persone che rispettano i parametri di cui sopra, per esempio del 5% per ogni categoria. Se per esempio un utente con due figli è sia sovrappeso, sia fumatore, si aggiungerà una tassa complessiva del 10%, che poi sarà moltiplicata per il numero dei figli, in questo caso 2. La formula nell'esempio:

$$\text{Aumento} = [(0.05 * \text{charge} * \text{fumatore}) + (0.05 * \text{charge} * \text{risk_bmi})] * n$$

Dove n è il numero dei figli. Quanto proposto inciderebbe come mostrato nel grafico con andamento per età. Si nota come più o meno per tutte le fasce sia abbastanza stabile, che ci indica come questa problematica sia distribuita equamente su tutte le età e che non è necessario fare valutazioni ulteriori.



Conclusioni

Il gruppo di analisi, in conclusione, consiglia:

- Considerare maggiormente l'aspetto del sovrappeso, come è stato considerato il rischio del fumo, soprattutto per la distribuzione che abbiamo evidenziato, dove la maggior parte delle persone è considerata nella categoria "Sovrappeso" o "Obeso", come è stato fatto per le persone fumatrici.
- Considerare il rischio che i figli, visto che l'assicurazione copre anche loro, possano seguire le abitudini dei genitori.

Lavoro di Federico Zanoni