

Über den Datensatz

Der Datensatz „[10000 Restaurant Reviews](#)“ von Joe Beach Capital auf Kaggle enthält Bewertungen von Restaurants. Diese Bewertungen umfassen Texte von Kunden, die ihre Erfahrungen mit Restaurants geteilt haben, sowie Bewertungen in Form von Sternen. Der Datensatz wurde gesammelt, um Einblicke in die Kundenmeinungen und deren Sentiment gegenüber verschiedenen Restaurants zu gewinnen.

Problemstellung

In dieser Analyse möchten wir verstehen, wer häufig Restaurants bewertet und welche Faktoren dieses Verhalten beeinflussen:

- Wir werden die Beziehung zwischen der Anzahl der Bewertungen, die ein Benutzer abgegeben hat, der Anzahl der Bewertungen, die ein Restaurant hat, und der Bewertung, die der Benutzer abgegeben hat, untersuchen.
- Wann gibt ein Benutzer, der normalerweise keine Bewertungen abgibt, eine Bewertung ab?
- Sind diese Bewertungen eher positiv oder negativ, lang oder kurz?
- Gibt der Benutzer eine Bewertung ab, wenn seine Meinung ähnlich oder unterschiedlich zu den anderen Bewertungen ist?
- Gibt der Benutzer eine Bewertung ab, wenn das Restaurant bereits viele oder wenige Bewertungen hat?
- Wie ehrlich sind die Benutzer, die sehr viele Bewertungen abgeben?
 - Kann man dies feststellen, indem man ihre Bewertungen mit den durchschnittlichen
- Bewertungen des Restaurants vergleicht?
- Geben sie immer positive oder negative Bewertungen ab und wie unterscheiden sich ihre Meinungen von denen anderer Benutzer?
- Wir werden auch die Zeit der Bewertung nutzen, um die Art der Mahlzeit (Frühstück, Mittagessen, Abendessen) vorherzusagen.
- Wie korrelieren die Zeiten des Jahres mit der Anzahl und Qualität der Bewertungen?

Wert

Die Erkenntnisse aus dieser Analyse könnten verschiedene wertvolle Anwendungen haben:

- Kundenzufriedenheit steigern: Restaurants können spezifische Aspekte ihres Services oder Angebots verbessern, um die Kundenzufriedenheit zu erhöhen.
- Marketingstrategien optimieren: Durch das Verständnis, welche Wörter in positiven Bewertungen häufig vorkommen, können gezielte Marketingkampagnen entwickelt werden.
- Betriebszeiten anpassen: Wenn bestimmte Tage oder Zeiten identifiziert werden, an denen mehr oder bessere Bewertungen abgegeben werden, können Restaurants ihre Betriebszeiten oder Personalplanung entsprechend anpassen.
- Vertrauenswürdigkeit der Bewertungen: Das Verständnis, wie ehrlich häufige Bewerter sind, kann helfen, die Zuverlässigkeit der Bewertungen zu bewerten und eventuelle Verzerrungen zu erkennen.

Diese Maßnahmen könnten zu einer Steigerung der Kundenzufriedenheit, effizienteren Betriebsabläufen und letztlich zu höheren Umsätzen führen.

Ursprung der Daten und Zugriff

Der Datensatz stammt von der Plattform Kaggle und wurde von Joe Beach Capital erstellt. Der genaue Erhebungszeitraum und die Methode der Datensammlung sind nicht spezifiziert, was einen potenziellen Einfluss auf die Datenqualität und die Analyse haben könnte. Der Zugriff auf die Daten erfolgt über Kaggle, wobei ein Download der CSV-Datei notwendig ist.

Datenformat und Struktur

Der Datensatz liegt im CSV-Format vor und enthält Spalten wie „Review_Text“, „Rating“, „Date“.

Datenqualität

Die Datenqualität kann durch das Vorhandensein von Spalten mit mehreren Werten und unvollständigen Bewertungen beeinträchtigt sein. Geplante Maßnahmen zur Verbesserung der Datenqualität umfassen das separieren von Spalten und das Entfernen von Bewertungen mit fehlenden oder unplausiblen Werten.

Ethik

Es ist wichtig, sicherzustellen, dass alle Analysen im Einklang mit den Datenschutzrichtlinien stehen. Bewertungen enthalten möglicherweise persönliche Meinungen und Informationen, die vertraulich behandelt werden sollten. Es ist entscheidend, dass die Daten anonymisiert werden und dass keine personenbezogenen Daten öffentlich gemacht werden.

Abdelrahman Elsharkawi