# Effizienzsteigerung der Inspektionsprozesse des DOHMH

## Einleitung

Das Department of Health and Mental Hygiene (DOHMH) in New York City steht vor der Herausforderung, über 27.000 Restaurants mit begrenzten personellen Ressourcen effektiv zu überwachen. Diese Situation erschwert es, alle Betriebe gleichermaßen zu kontrollieren, wodurch potenzielle Verstöße in besonders risikoreichen Bereichen unerkannt bleiben können. Der vorliegende Bericht stellt eine datenbasierte Lösung vor, die darauf abzielt, die Effizienz, Relevanz und Qualität der Inspektionsprozesse zu steigern.

## Use Case

Die Überwachung von Restaurants in einer Millionenstadt wie New York ist eine logistische und personelle Herausforderung. Der Use Case zielt darauf ab, mithilfe bestehender Daten und moderner Analysetools die Arbeit des DOHMH zu optimieren. Die Grundlage für die Analysen bildeten Datenquellen wie Yelp-Bewertungen und die Inspektionsberichte des DOHMH.

Ein zentraler Bestandteil des Projekts war die Erstellung geografischer Karten, um Hotspots für Gesundheitsverstöße zu identifizieren. Neben den geografischen Hotspots wurden Restaurantkategorien untersucht, um Muster in den Gesundheitsverstößen zu erkennen. Zusätzlich wurde die Beziehung zwischen Gesundheitscores und Yelp-Bewertungen analysiert.

Insgesamt demonstriert der Use Case, wie bestehende Datenquellen und einfache analytische Ansätze kombiniert werden können, um gezielte Maßnahmen für Inspektionsprozesse zu entwickeln. Der Fokus lag darauf, die Effizienz zu steigern und die verfügbaren Ressourcen des DOHMH optimal einzusetzen.

## Datenaufbereitung

Die Datenaufbereitung spielte eine zentrale Rolle in der Effizienzsteigerung der Inspektionsprozesse des DOHMH. Um valide und aussagekräftige Ergebnisse zu erzielen, wurden mehrere Schritte unternommen, um die Datenqualität zu verbessern und die Analyse zu ermöglichen. Zwei Hauptquellen standen im Mittelpunkt: die Yelp-Datenbank und die Inspektionsberichte des DOHMH. Diese wurden durch eine strukturierte Herangehensweise verknüpft und bereinigt.

### Zusammenführung und Bereinigung der Daten

Die Yelp-Daten lieferten Informationen über Kundenbewertungen und Standortdaten der Restaurants, während die Inspektionsdaten des DOHMH Gesundheitsbewertungen und Inspektionsergebnisse bereitstellten. Beide Quellen wurden mithilfe von MongoDB und Python (Pandas) zusammengeführt. Dabei wurden fehlende Informationen ergänzt, etwa durch die Verknüpfung von Yelp-Einträgen mit den DOHMH-Daten anhand gemeinsamer Parameter wie Standort oder Restaurantname.

Um sicherzustellen, dass die Ergebnisse auf einer soliden Basis beruhen, wurden unvollständige oder fehlerhafte Datensätze entfernt. Dazu zählten Einträge mit fehlenden Geo-Daten, nicht vorhandenen Gesundheitsscores oder inkorrekten Inspektionsinformationen (z. B. mit dem Platzhalterdatum „1/1/1900“). Zusätzlich wurden Duplikate identifiziert und entfernt, um Verzerrungen in der Analyse zu vermeiden.

### Transformation der Daten

Ein weiterer wichtiger Schritt war die Transformation der Daten. Gesundheitsscores der Restaurants wurden in Noten umgerechnet, um eine einheitliche Skala zu schaffen, die leichter interpretierbar ist. Diese Transformation erleichterte die anschließende Analyse und erlaubte eine verständlichere Visualisierung der Ergebnisse. Ebenso wurden die Yelp-Daten angepasst, um sie mit den Inspektionsdaten konsistent zu machen. Hierbei wurden verschiedene Formate harmonisiert, um eine reibungslose Verarbeitung zu gewährleisten.

### Optimierung der Abfragen

Für die effiziente Verarbeitung und Abfrage der kombinierten Daten wurden in MySQL Indizes auf den Primär- und Fremdschlüsseln gesetzt. Dies beschleunigte die Durchführung komplexer Joins und ermöglichte eine schnellere Datenanalyse. Besonders bei großen Datensätzen, wie sie in diesem Projekt verarbeitet wurden, war dieser Schritt entscheidend, um die Performance zu verbessern.

### Ergebnisse der Datenaufbereitung

Die sorgfältige Datenaufbereitung ermöglichte die Erstellung eines konsistenten, bereinigten und analysierbaren Datensatzes, der als Grundlage für alle weiteren Schritte diente. Durch die Bereinigung von Fehlern und die Transformation der Daten konnten verlässliche Hotspot-Analysen durchgeführt und Muster in den Restaurantbewertungen und Gesundheitscores identifiziert werden. Die aufbereiteten Daten legten damit die Basis für eine fundierte Entscheidungsfindung im Rahmen der Effizienzsteigerung der Inspektionsprozesse des DOHMH.

## Analyse

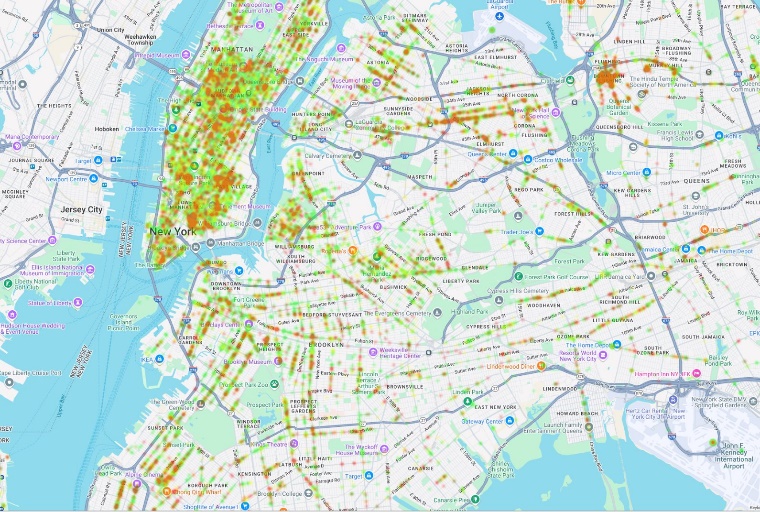
### Geografische Hotspot-Analyse

Die erste Phase der Analyse konzentrierte sich auf die Identifikation von geografischen Hotspots, in denen Gesundheitsverstöße besonders häufig auftreten. Mit Hilfe von Mapping-Tools wie Maptive wurden interaktive Karten erstellt, die es ermöglichten, diese Hotspots visuell darzustellen und zu analysieren.

Drei Regionen traten dabei als besonders problematisch hervor:

1. **Midtown Manhattan (Times Square):** Die hohe Restaurantdichte und der touristische Druck führen zu einem erhöhten Risiko von Gesundheitsverstößen. Diese Region steht besonders unter Inspektionsdruck, da sie eine zentrale Rolle für die Gastronomie und den Tourismus in New York City spielt.
2. **Downtown Manhattan (Chinatown):** Hier erschweren kulturelle und sprachliche Barrieren sowie die vermehrte Nutzung traditioneller Strukturen und Bewertungssysteme eine effiziente Inspektion. Mangelnde Renovierungen in älteren Gebäuden verschärfen die hygienischen Herausforderungen.
3. **Flushing Downtown:** Ähnlich wie in Chinatown führt eine hohe ethnische Vielfalt und die Fokussierung auf traditionelle Küchen zu vergleichbaren Risiken.

Die Hotspot-Analyse lieferte somit klare Indikatoren dafür, wo die begrenzten Inspektionsressourcen priorisiert eingesetzt werden sollten.



### Analyse der Restaurantkategorien

Zusätzlich zur geografischen Untersuchung wurden auch verschiedene Küchenarten analysiert, um herauszufinden, welche Kategorien ein höheres Risiko für Gesundheitsverstöße aufweisen. Dabei zeigten sich deutliche Unterschiede zwischen den Median-Scores verschiedener Restauranttypen.

* **Höheres Risiko:** Küchen wie die chinesische, lateinamerikanische und afrikanische wiesen höhere Median-Scores auf, was auf häufigere Hygienemängel hinweist. Diese Küchen könnten gezielt stärker überwacht werden, um Verstöße frühzeitig zu identifizieren und zu adressieren.
* **Geringeres Risiko:** Amerikanische und vegetarische Küchen erzielten niedrigere Median-Scores, was auf eine bessere Einhaltung der Hygienevorschriften hinweist. Diese Restaurants benötigen weniger intensive Inspektionsmaßnahmen.

Die Ergebnisse zeigen, dass eine themenspezifische Fokussierung auf problematische Küchenarten einen wichtigen Beitrag zur Steigerung der Inspektionseffizienz leisten kann.

Ein Bild, das Text, Screenshot, Diagramm enthält.

Beschreibung automatisch generiert.

### Untersuchung der Korrelation zwischen Gesundheitscores und Yelp-Bewertungen

Die Beziehung zwischen den Gesundheitscores und den Yelp-Bewertungen der Restaurants wurde ebenfalls untersucht. Ziel war es, zu prüfen, ob Yelp-Bewertungen als Indikator für Gesundheitsprobleme genutzt werden können.

Die Analyse ergab, dass Restaurants mit niedrigeren Yelp-Bewertungen (z. B. 1 Stern) tendenziell schlechtere Gesundheitsscores aufwiesen. Allerdings war die Datenbasis für Restaurants mit sehr niedrigen Bewertungen begrenzt, da nur wenige Betriebe dieser Kategorie angehörten. Insgesamt zeigte sich, dass die Korrelation zwischen Bewertungen und Gesundheitsscores nur gering war und andere Faktoren wahrscheinlich eine größere Rolle bei den Verstößen spielen.

Ein Bild, das Text, Screenshot, Farbigkeit, Reihe enthält.

Beschreibung automatisch generiert.

### Erkenntnisse aus der Analyse

Die Analyse lieferte wesentliche Erkenntnisse, die für die Optimierung der Inspektionsprozesse von entscheidender Bedeutung sind. Die geografische und kategorische Risikoeinschätzung ermöglicht es, Inspektionen gezielter und effizienter durchzuführen. Gleichzeitig zeigte die begrenzte Korrelation zwischen Gesundheitscores und Yelp-Bewertungen, dass externe Bewertungsplattformen nur bedingt als alleinige Grundlage für Entscheidungen herangezogen werden können. Die Ergebnisse unterstreichen die Notwendigkeit, datenbasierte Strategien mit umfassenderen Inspektionsprogrammen zu kombinieren, um die öffentliche Gesundheit nachhaltig zu schützen.

## Empfehlung

Eine zentrale Empfehlung ist die gezielte Ressourcenallokation. Inspektionen sollten prioritär in den identifizierten Hotspots wie Midtown Manhattan, Downtown Manhattan (Chinatown) und Flushing Downtown stattfinden. Die hohe Konzentration von Verstößen in diesen Gebieten zeigt, dass eine verstärkte Überwachung dort den größten Effekt auf die Einhaltung der Gesundheitsvorschriften haben kann. Parallel dazu sollten Restaurantkategorien mit einem höheren Risiko, wie chinesische, lateinamerikanische und afrikanische Küchen, stärker in den Fokus der Inspektionen rücken.

Es wird ebenfalls empfohlen, langfristig Schulungsprogramme für Restaurantbesitzer und -mitarbeiter zu entwickeln, insbesondere in den problematischen Regionen und Kategorien. Diese Schulungen könnten dazu beitragen, grundlegende Hygienestandards besser zu vermitteln und nachhaltige Verbesserungen zu erzielen. Eine regelmäßige Kommunikation zwischen dem DOHMH und den betroffenen Betrieben könnte das Bewusstsein für die Bedeutung von Gesundheitsstandards stärken.

Die Korrelation zwischen Yelp-Bewertungen und Gesundheitscores ist zwar schwach ausgeprägt, dennoch könnten Plattformen wie Yelp als ergänzende Informationsquelle dienen. Beispielsweise können negative Trends in Bewertungen ein erster Hinweis auf potenzielle Problemfälle sein, die durch Inspektionen überprüft werden könnten. Es empfiehlt sich, solche externen Daten in Kombination mit den Inspektionsberichten des DOHMH zu verwenden, jedoch nicht als alleinige Entscheidungsgrundlage.

## Fazit

Zusammenfassend zeigt das Projekt, wie datenbasierte Ansätze die Effizienz der Inspektionsprozesse des DOHMH erheblich verbessern können. Durch die gezielte Allokation von Ressourcen in problematische Regionen und Kategorien sowie die Einführung unterstützender Maßnahmen wie Schulungen und Monitoring können die Inspektionen relevanter und wirkungsvoller gestaltet werden.

Die Ergebnisse der Analyse liefern ein starkes Fundament für strategische Entscheidungen und zeigen auf, dass datengetriebene Lösungen nicht nur kurzfristige Effizienzgewinne, sondern auch langfristige Verbesserungen der öffentlichen Gesundheit ermöglichen können. Mit diesen Maßnahmen kann das DOHMH nicht nur die Einhaltung von Gesundheitsstandards erhöhen, sondern auch das Vertrauen der Öffentlichkeit in die Gastronomie in New York City stärken.