

Рейтинги гостиниц и других объектов размещения в разных регионах России: как они формируются и что на них влияет

Игорь Пережогин

Предпосылки к исследованию:

- желание понять, насколько рейтинг соответствует реальному качеству гостиницы
- отсутствие (не нашел) систематизированных данных (отзывов и рейтингов объектов размещения) на русском языке
- желание понять, какие факторы влияют на рейтинг в разных регионах

Задачи, которые возникли в связи с этим:

- получение данных и их обработка
- статистика рейтингов и отзывов по объектам в разных регионах, сравнение
- анализ текста отзывов (TF-IDF) для выявления слов, влияющих на рейтинг; сравнение для разных регионов
- тематическое моделирование отзывов
- поиск аномалий среди отзывов – можно ли увидеть что-то необычное (например, обнаружить сфабрикованные отзывы)?

- **получение данных и их обработка**
- статистика рейтингов и отзывов по объектам в разных регионах, сравнение
- анализ текста отзывов (TF-IDF) для выявления слов, влияющих на рейтинг; сравнение для разных регионов
- тематическое моделирование отзывов
- поиск аномалий среди отзывов – можно ли увидеть что-то необычное (например, обнаружить сфабрикованные отзывы)?

- получение данных и их обработка
- **статистика рейтингов и отзывов по объектам в разных регионах, сравнение**
- анализ текста отзывов (TF-IDF) для выявления слов, влияющих на рейтинг; сравнение для разных регионов
- тематическое моделирование отзывов
- поиск аномалий среди отзывов – можно ли увидеть что-то необычное (например, обнаружить сфабрикованные отзывы)?

- получение данных и их обработка
- статистика рейтингов и отзывов по объектам в разных регионах, сравнение
- **анализ текста отзывов (TF-IDF) для выявления слов, влияющих на рейтинг; сравнение для разных регионов**
- тематическое моделирование отзывов
- поиск аномалий среди отзывов – можно ли увидеть что-то необычное (например, обнаружить сфабрикованные отзывы)?

- получение данных и их обработка
- статистика рейтингов и отзывов по объектам в разных регионах, сравнение
- анализ текста отзывов (TF-IDF) для выявления слов, влияющих на рейтинг; сравнение для разных регионов
- **тематическое моделирование отзывов**
- **поиск аномалий среди отзывов – можно ли увидеть что-то необычное (например, обнаружить сфабрикованные отзывы)?**

Выводы:

- посмотрели, как формируются рейтинги отелей; интересно видеть статистику по разным городам (рейтинги и количество отзывов), что на нее влияет (время между посещением и написанием отзыва, например).
- сравнили лексику, влияющую на отзывы в разных городах. Можно оценить сильные и слабые стороны гостиниц в этих местах.
- поиск аномалий и сфабрикованных отзывов требует более серьезной работы

Проблемы и задачи на будущее:

- Более аккуратная настройка TF-IDF
- Более подробный анализ распределений оценок отзывов

Спасибо за внимание!