

Uma Análise Quantitativa dos Meios de Hospedagem do Recife

Juliana Carvalho¹, Ângelo de Sant'Ana², Antônio Carlos³

*Centro de Informática - Universidade Federal de Pernambuco (UFPE)
Caixa Postal 7.851 - 50.732-970 - Recife - PE - Brasil
jco2@cin.ufpe.br, assd@cin.ufpe.br, acpr@cin.ufpe.br*

Abstract

O presente trabalho objetiva analisar a qualidade dos hotéis de acordo com as avaliações dos usuários do site Trivago na Grande Recife - PE. Para alcançar esse objetivo foi realizada uma análise qualitativa e exploratória dos dados. Também foi feita uma pesquisa bibliográfica de caráter exploratório sobre “turismo e hospedagens”, “qualidade e satisfação dos clientes” e “internet e os meios de hospedagens”. O artigo divide-se em quatro seções: introdução, metodologia, análise dos dados e considerações finais. Pretende-se analisar diferentes pontos de um tema a fim de contribuir para a evolução da discussão do tema inovação em meios de hospedagem e fornecer novas bases para a produção de futuras pesquisas acadêmicas.

Keywords: Análise quantitativa, Turismo, Qualidade, Hospedagem, Algoritmos, Ciência dos dados

1. Introdução

Este projeto tem como objetivo colocar em prática os conhecimentos na área de Ciência dos Dados. Escolhemos uma base de dados para que após aplicadas as técnicas de Extração, Pré-processamento e Visualização pudéssemos
5 analisar e extrair informação dos dados. Também utilizamos de algoritmos de

¹Graduanda do curso Bacharelado em Sistemas de Informação da UFPE.

²Graduando do curso Bacharelado em Ciência da Computação da UFPE.

³Graduando do curso Bacharelado em Ciência da Computação da UFPE.

aprendizagem para que aumentar a qualidade da nossa análise. O trabalho está dividido em 5 seções. são apresentados o contexto que desencadeou a escolha do tema de pesquisa, objetivos e estrutura do trabalho aqui apresentado. Na seção 2 deste documento será apresentado quais foram os métodos utilizados para chegar nos resultados desta análise. A seção 3 mostra os resultados da análise dos dados. Na seção 4, são descritas as principais contribuições.

2. Metodologia

Nesta seção serão explicados os métodos utilizados para o desenvolvimento deste trabalho.

2.1. Extração de dados

Neste projeto, para a realização da coleta dos dados, foi necessária a utilização técnicas de data-scraping. Nosso objetivo era recuperar as avaliações e informações de todos os hotéis da região do Grande Recife. O site de hospedagem escolhido foi o Trivago(<https://www.trivago.com.br/>), que coleta a avaliação de diferentes plataformas de reserva de hospedagens, tira uma média, para assim ter uma nota final. Antes de realizar a extração, todas as páginas foram baixadas para evitar problemas com mudanças no código fonte. Foram utilizadas duas bibliotecas de *python*, *BeautifulSoup* e *Pandas*. Com a primeira, extraímos as tags html que continham as categorias de avaliações e informações dos hotéis, e com a segunda agregamos tudo em uma estrutura de *DataFrame*. O arquivo de saída desta etapa serviu de entrada para a fase seguinte, de pré-processamento.

2.2. Pré-processamento

Na etapa de pré-processamento, foram utilizadas algumas bibliotecas *Python*, sendo *Pandas* a principal. Com o auxílio dessas bibliotecas foi possível fazer manipulações e tratamento do nosso conjunto de dados.

O pré-processamento constituiu-se em:

- 35 (1) Substituir os valores não informados por NaN;
- (2) Preencher os valores NaN com a média da coluna;
- (3) Identificar as categorias irrelevantes utilizando correlação
- (4) Remover tais colunas;
- (5) Salvar dados tratados em um novo arquivo CSV nomeado "saidaPreProces-
- 40 samento";

2.3. Técnica de Aprendizagem

Utilizamos regressão linear como técnica de aprendizagem, através da biblioteca *Scipy*. Calculamos a média das notas de cada hotel e plotamos cinco gráficos com as cinco categorias avaliadas pelos clientes: localização, quartos, facilidades do hotel, café da manhã e comida em geral. A análise visual nos sugeriu que tanto o café da manhã quanto a comida não tem relação com a nota obtida pelo hotel, enquanto a localização pareceu influenciar um pouco. No entanto, o gráfico da qualidade dos quartos e principalmente das facilidades do hotel, mostram claramente um aumento na gradativo na nota do hotel a media que a avaliação nessas categorias aumenta. Utilizando o *Scipy*, aplicamos a regressão linear e calculamos as retas que modelam a variação de cada um das categorias com relação a média geral de cada hotel. O que vimos foi uma reta vertical tanto na categoria "café da manhã" quanto na categoria "comida", o que já era esperado, e retas bem comportadas para as categorias "quartos" e "facilidades". Como pode ser visualizado a seguir:

55

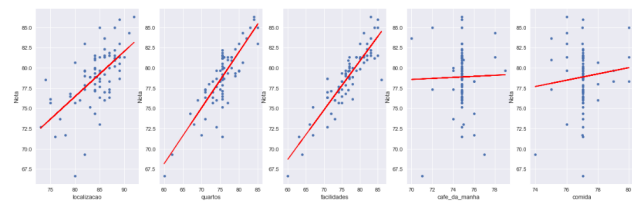


Figura 1: Regressão Linear

2.4. Visualização

Para a visualização foi utilizada a biblioteca *matplotlib*. Criamos 2 gráficos simples para mostrar a diferença entre as médias gerais dos hotéis e as notas em cada categoria.

- 60 (1) Gráfico 1 - 10 hotéis com melhores avaliações da nota geral;

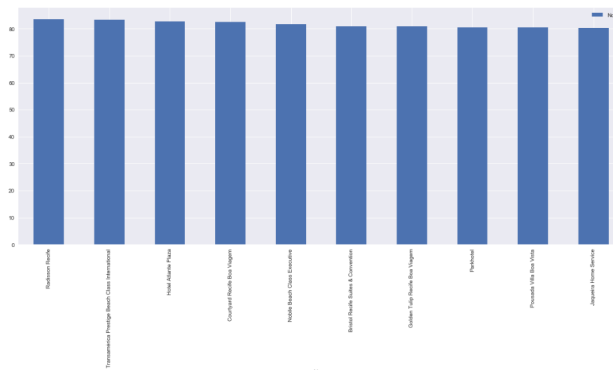


Figura 2: Gráfico 2 - 10 hotéis bem mais avaliados pela nota geral

- (2) Gráfico 2 - 10 hotéis com melhores notas referente a *localização*, *quartos* e *facilidades*.

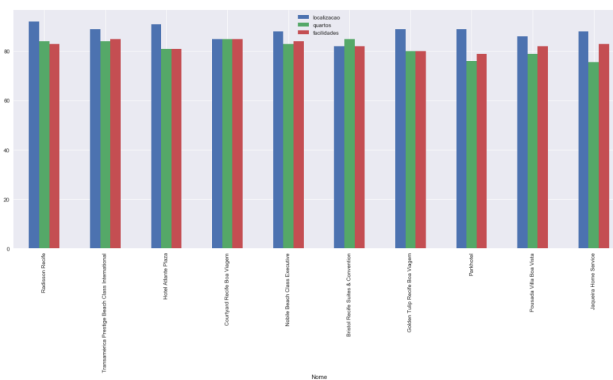


Figura 3: Gráfico 2 - 10 hotéis com melhores notas referente a *localização*, *quartos* e *facilidades*

3. Análise dos Resultados

Embora tenham faltado algumas etapas no processo de avaliação dos atributos relevantes e no processo de aprendizado, percebemos uma contribuição maior da qualidade dos quartos e das facilidades do hotel no grau de satisfação dos clientes.

4. Considerações Finais

O presente trabalho e a disciplina cursada nos mostraram como o estudo dos dados é importante para a geração do conhecimento. Aliado aos constantes debates a respeito da ética na manipulação e comercialização dos dados, este curso e projeto nos mostraram como a partir de uma simples página na web podemos agregar conhecimento e prever os mais diversos cenários do nosso dia-a-dia

Referências

[1]

[1] Trivago, Disponível em: <http://www.trivago.com.br>, acessado em Junho, 2017 (2017).