11/10/2021

airbnb prices

This project aims to develop a model to predict Airbnb prices in Saudi Arabia.

First, we scraped Airbnb room details from their website. We extracted the room links for 106 cities in Saudi Arabia, we collected 4964 links, then we scraped those links and saved them in csv file.

After that, we started cleaning the data with Pandas. We dropped insignificant rows like country, links, and others. We filled missing data and corrected some data with Arabic names.

The minimum viable product at this stage is the scraped and cleaned dataset. One row will look as the following:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **rating** | **rating count** | **city** | **superhost** | **price (SAR)** | **house type** | **page link** | **guest** | **…** |
| 0.0 | 0 | Buraydah | False | 1100 | Earth House | [the link is here](https://www.airbnb.com/rooms/44085675?previous_page_section_name=1000&translate_ugc=false&federated_search_id=5f2d5336-4a6f-4f81-902a-b7933081484f) | 10 | … |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **bedrooms** | **baths** | **TV** | **wifi** | **kitchen** | **elevator** | **breakfast** | **AC** | **…** |
| 1 | 3 | 1 | 1 | 1 | 0 | 0 | 1 | … |