

Data Scientist Assignment

Overview

Thank you for your interest in the **Data Scientist** role at Outdoorsy!

This assignment provides you with a sample of some of the business challenges we are solving today at Outdoorsy. The goal of this assignment is to better understand how you approach business problems and should not take longer than 2-3 hours to complete.

- Deliverables should include your source code as well as any visualizations or responses for sharing your work and thought process. Example formats can include notebook, slides, docs, etc. **Please also check in your code to GitHub and share the link.**
- Feel free to use any programming language in answering the questions.
- Please include any assumptions you are making when answering the prompts.
- Once completed, please attach any links/deliverables and email back to kellie.jue@outdoorsy.co.

Data Overview

outdoorsy_rental_data.csv: a catalog of rentals and characteristics about each listing

Column Name	Description
rental_id	The ID of the rental
listing_state	The state the rental is listed in
vehicle_type	The type of vehicle being listed
is_motorized	If the vehicle is able to be driven
vehicle_make	The make of the vehicle being listed
vehicle_model	The model of the vehicle being listed
vehicle_year	The year of the vehicle being listed
vehicle_length	The length of the vehicle in feet
cancellation_policy	The cancellation policy set by the owner for their listing
sleeps	The number of adults that can sleep in the rental

is_festival_friendly	If the vehicle can be taken to festivals
is_tailgate_friendly	If the vehicle can be taken to tailgates
is_pet_friendly	If the vehicle is pet friendly
has_toilet	If the vehicle has an indoor toilet
has_generator	If the vehicle has a generator
has_inside_shower	If the vehicle has an inside shower
has_refrigerator	If the vehicle has a refrigerator
has_wifi	If the vehicle has wifi available

outdoorsy_booking_history.csv: a transaction log of rentals booked where each row is a rental that was booked at the specified price during a given month.

Column Name	Description
rental_id	The ID of the rental
booked_price_per_day	The price per day the rental was booked at
month_booked	The month the rental was booked

Assignment

Using the provided sets of data, please answer the following questions in each category:

Analysis & Visualization

- What is the average daily price per month for class c rentals?
- For motorized vs non-motorized vehicles, plot the average daily value of a rental by year model
 - What percent of listings fall outside of the average within 1 standard deviation?
- How does the average value of a listing change over time? Are there any patterns you can detect? How does location affect the average rate? Seasonality?
- The data provided is only a subset of factors available that may factor into the price for a booking - what is the estimated variance in your statistical model?

Modeling

1. Which features in the dataset have the most impact on the average daily value of a rental?
 - a. If you see any missing value, use the approach that you think is the best fit for this use case and explain why.
2. Assume we want to categorize each listing into “luxurious”, “premium”, and “basic”. How would you go about building a model to classify the listings into each of these categories?
 - a. What would you propose are the main characteristics of a “luxurious” rental?
 - b. How would you do feature engineering?
 - c. How would you evaluate the model?
3. Assume we want to reach out to all the owners of “basic” listings to promote how they can upgrade their rental to get a higher return for their daily booking price.
 - a. Outline a testing strategy to see how effective is your model
 - b. How would you estimate the impact of this model?
 - c. How can you further optimize your model to be more accurate?

Please feel free to email if you have any questions about this assignment!