

Assignment Summary: Hotel Price Estimation

This document summarizes the approach and results for estimating hotel prices for February 2020, based on the provided historical data (2012-2016).

Approach

The estimation was performed using the Python script [assignment.py](#). The process involved:

- 1. **Data Preparation:** Loading the historical data, handling outliers, and filling missing values.
- 2. **Feature Engineering:** Creating time-based and cyclical features (e.g., year, month, day of the week) to capture trends and seasonality.
- 3. **Feature Selection:** Applied correlation-based filtering to reduce multicollinearity, retaining 7 final features from the original 19 engineered features
- 4. **Modeling:** Training an ensemble of three machine learning models:
 - Linear Regression
 - Random Forest
 - XGBoost
- 5. **Forecasting:** Using the trained ensemble model to predict daily hotel prices for the entire month of February 2020.

Code

The complete code used for this analysis is in the [assignment.py](#) file attached.

Results

The final estimated hotel prices for each day in February 2020 are saved in the [forecast.csv](#) file and also attached as [forecast.png](#).

A more detailed technical report, including model performance metrics and visualizations, is available at [forecasting_report.pdf](#).

	Date	Forecast
0	2020-02-01	163.856
1	2020-02-02	163.104
2	2020-02-03	164.473
3	2020-02-04	167.793
4	2020-02-05	171.577
5	2020-02-06	168.694
6	2020-02-07	168.491
7	2020-02-08	168.239

	Date	Forecast
8	2020-02-09	166.844
9	2020-02-10	166.456
10	2020-02-11	167.959
11	2020-02-12	171.303
12	2020-02-13	175.474
13	2020-02-14	176.374
14	2020-02-15	173.319
15	2020-02-16	171.398
16	2020-02-17	175.126
17	2020-02-18	175.038
18	2020-02-19	174.708
19	2020-02-20	176.49
20	2020-02-21	176.48
21	2020-02-22	175.136
22	2020-02-23	173.288
23	2020-02-24	174.923
24	2020-02-25	173.857
25	2020-02-26	176.407
26	2020-02-27	178.003
27	2020-02-28	178.605
28	2020-02-29	177.053