Hotel Cancellation Prediction Modeling and EDA

The Repository contains a project from my data science portfolio that focuses on predicting cancellation rates for hotels. Which is completed in a Jupyter Notebook.

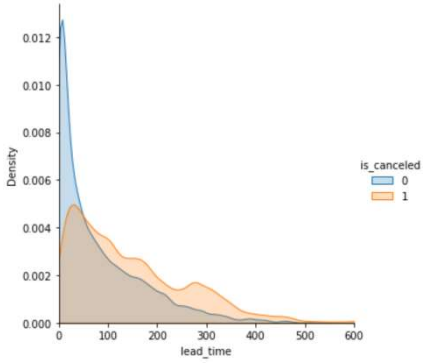
Instructions for Jupyter Notebooks and Content

- If you are unfamiliar with Jupyter Notebook, please follow the link for installation and getting started: [Project Jupyter | Home]
- Next download notebook [Hotel Booking Jupyter Notebook File] and run file in Jupyter Notebook
- Download data from GitHub [hotel_bookings.csv]

Project Overview

The project contains exploratory data analysis that touches on different topics within the hotel industry such as an average daily rate, arrival month, hotel bookings, and more. Seaborn is used in the creation of visualizations such as this one that focuses on cancelations by lead time. That shows viewers the lead time guests give when cancelling which peaks at around 50 days.

```
# I will create a density plot that shows the lead time cancelations
(sns.FacetGrid(Hotel_Bookings, hue = 'is_canceled', height = 5, xlim = (0,600))
    .map(sns.kdeplot, 'lead_time', shade = True)
    .add_legend());
```




```
# I will now print my Classification Repor for KNN
print(knn_class)
```

	precision	recall	f1-score	support
0	0.96	0.98	0.97	26070
1	0.96	0.94	0.95	15487
accuracy			0.96	41557
macro avg	0.96	0.96	0.96	41557
weighted avg	0.96	0.96	0.96	41557

Data modeling is also utilized in this project to predict the rate of cancelation. The K-Nearest Neighbor model showed an accuracy of 96%. Which indicates a great model to utilize when working on predictions for hotel cancelations.

Goals for Project

The goal for this project is to allow users to better understand Python and the tools that can be used to gain a deeper understanding of one's data and how it can be applied to other data sets. If you found this project useful and have any questions regarding any of the projects, please feel free to contact me at [TheodoreKobyHercsky@gmail.com]