Kaggle Competition 1
Team Name: Tabs not Spaces
Team: Cassidy Carpenter & Adam Salyers

1. Our Approach
    a. We knew we were given a csv file that contained all the necessary data for us to train our model on. Analyzing the data we though that a linear regression model would be the most straightforward way to test/train our data. First, we had to clean the data. This included removing unnecessary information that would not change the estimation of the 'price' like the unnamed column which contained unique id's and the id column. This cleaning of data also included changing the date column from an object to an int by splitting each row of data at timestamp T. Now that our data was clean we separated it out into our X values and our y values. The y values being the price column and the X values being all the other columns from the clean data. We would use the X values to estimate what value of y should be based on the rest of the X values. We simply put this into a LinearRegression() function given to us through the use of sklearn, and fitted the linear regression to our X and y data. This was our model. We then ran our test values of X from the other csv given to us and wound up with about a 70% accuracy of our model. Not too bad, not great.
    b. We wanted to improve upon the accuracy of our linear regression model so I (Cassidy Carpenter) implemented a K Fold cross validation. I thought this would help randomize the testing and training split into k-number of folds so that I would increase the accuracy. However, I found that it actually lowered the accuracy the more folds I created. With n_folds = 10 making the accuracy decrease to 69%. Thus we abandoned this path.
    c. Next (Adam Salyers) tried to create a decision tree of bins. This would go through all the data and select a bin, then continue to selec bins within the bins and go down further to select a value finally. We worked on this for a while but decided to abandon it when we had no way to match the X values to the y values after the second iteration.
    d. Finally (Adam Salyers and Cassidy Carpenter), we made an SVM classifier that gave us about the same 70-71% accuracy as the linear regression. Since the SVM took about 1-2 minutes to run we decided to just go with the linear regression for the same amount of accuracy for a lower runtime. Then we created the csv file and turned in our results!
2. Results
    a. Linear Regression
        i. 70.4% accurate
    b. Decision Tree
        i. Took to long and the model couldn't predict well. Abandoned before we got accuracy
    c. SVM
        i. 71.2% accurate
3. Conclusion
    a. The linear regression model is definitely not the most accurate of the models we could have used for this project. Perhaps with more time we could have explored more options for the model/ tried different cross validation methods had we had more time for this project! Overall, 70% accuracy is a passing grade so not ba