0.1 Question 0

0.1.1 Question 0a

"How much is a house worth?" Who might be interested in an answer to this question? Please list at least three different parties (people or organizations) and state whether each one has an interest in seeing the value be high or low.

Ordinary people, county, real estate agencies. I believe ordinary people either buyer or seller are intrested to see the value of a house. Counties are also interested in looking at house prices in order to predict property taxes and usually they don't care if the price is high or low. Real estate agencies are also the one interested in seeing the value of a house and predict new values.

0.1.2 Question 0b

Which of the following scenarios strike you as unfair and why? You can choose more than one. There is no single right answer but you must explain your reasoning.

A. A homeowner whose home is assessed at a higher price than it would sell for.
B. A homeowner whose home is assessed at a lower price than it would sell for.
C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive properties.
D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive properties.

c. Because it will affect low income families and minorities as they would have to pay more property taxes compared to the majority.

0.1.3 Question 0d

What were the central problems with the earlier property tax system in Cook County as reported by the Chicago Tribune ? And what were the primary causes of these problems? (Note: in addition to reading the paragraph above you will need to watch the lecture to answer this question)

The central problems with the earlier property tax system in Cook County as reported by the Chicago Tribune was that for years the county's property tax system created an unequal burden on residents to homeowners who are well-off while punishing those living in minorities. The primary cuases were producing discrimination against the people of color and working class, specifically, the minorities. For instance, in the lecture we learned that several property owners were affected through this unfair tax burden like Braxton-Williams whose property valued $147550, which she believes at that time his/her property over-valued.

0.1.4 Question 0e

In addition to being regressive, why did the property tax system in Cook County place a disproportionate tax burden on non-white property owners?

To impose racial inequality as real estate is one of the modern key motor of racial inequality in the United States.

0.2   Question 2

Without running any calculation or code, complete the following statement by filling in the blank with one of the comparators below:

$$\geq$$
$$\leq$$
$$=$$

Suppose we quantify the loss on our linear models using MSE (Mean Squared Error). Consider the training loss of the 1st model and the training loss of the 2nd model. We are guaranteed that:

Training Loss of the 1st Model_____Training Loss of the 2nd Model

$$\geq$$

0.3   Question 6

Let's compare the actual parameters (θ_0 and θ_1) from both of our models. As a quick reminder,

for the 1st model,
$$\text{Log Sale Price} = \theta_0 + \theta_1 \cdot (\text{Bedrooms})$$

for the 2nd model,
$$\text{Log Sale Price} = \theta_0 + \theta_1 \cdot (\text{Bedrooms}) + \theta_2 \cdot (\text{Log Building Square Feet})$$

Run the following cell and compare the values of θ_1 from both models. Why does θ_1 change from positive to negative when we introduce an additional feature in our 2nd model?

When we add another feature in our model normally the variance decreases, however, the error of the model incrases. So, in this example θ_1 changes to negative because error incrases.

```
In [ ]: # Parameters from 1st model
        theta0_m1 = linear_model_m1.intercept_
        theta1_m1 = linear_model_m1.coef_[0]

        # Parameters from 2nd model
        theta0_m2 = linear_model_m2.intercept_
        theta1_m2, theta2_m2 = linear_model_m2.coef_

        print("1st Model\n 0: {}\n 1: {}".format(theta0_m1, theta1_m1))
        print("2nd Model\n 0: {}\n 1: {}\n 2: {}".format(theta0_m2, theta1_m2, theta2_m2))
```

0.4 Question 7

0.4.1 Question 7a

Another way of understanding the performance (and appropriateness) of a model is through a plot of the model the residuals versus the observations.

In the cell below, use `plt.scatter` to plot the residuals from predicting `Log Sale Price` using **only the 2nd model** against the original `Log Sale Price` for the **test data**. You should also ensure that the dot size and opacity in the scatter plot are set appropriately to reduce the impact of overplotting.

```
In [ ]: import plotly.express as px
        import plotly.graph_objs as go
        plt.scatter(x=y_predicted_m2, y=y_test_m2, s=3, alpha=0.75)
        plt.xlabel('Predicted Log Sale Price - Y_hat')
        plt.ylabel('Original Log Sale Price - Y')
        plt.title('Log Sale Price vs Sale Price Using Model_2');
```

0.5 Question 9

In building your model in question 8, what different models have you tried? What worked and what did not? Brief discuss your modeling process.

Note: We are looking for a single correct answer. Explain what you did in question 8 and you will get point.

First, I used some features without transforming them which made the error larger. Then I transformed them based on their bulges direction. I have noticed the error getting smaller and smaller. However, I needed to do one hot encoding to make the error more smaller.

0.6 Question 10

When evaluating your model, we used root mean squared error. In the context of estimating the value of houses, what does error mean for an individual homeowner? How does it affect them in terms of property taxes?

For an individual homeowner the error means a difference betweeen the actual value of the house and the predicted value. Therefore, the homeowner would have to pay either less or high property taxes. On the other hand, the error is a bias against homeowners in Cook County.

In the case of the Cook County Assessor's Office, Chief Data Officer Rob Ross states that fair property tax rates are contingent on whether property values are assessed accurately - that they're valued at what they're worth, relative to properties with similar characteristics. This implies that having a more accurate model results in fairer assessments. The goal of the property assessment process for the CCAO, then, is to be as accurate as possible.

When the use of algorithms and statistical modeling has real-world consequences, we often refer to the idea of fairness as a measurement of how socially responsible our work is. But fairness is incredibly multifaceted: Is a fair model one that minimizes loss - one that generates accurate results? Is it one that utilizes "unbiased" data? Or is fairness a broader goal that takes historical contexts into account?

These approaches to fairness are not mutually exclusive. If we look beyond error functions and technical measures of accuracy, we'd not only consider *individual* cases of fairness, but also what fairness - and justice - means to marginalized communities on a broader scale. We'd ask: What does it mean when homes in predominantly Black and Hispanic communities in Cook County are consistently overvalued, resulting in proportionally higher property taxes? When the white neighborhoods in Cook County are consistently undervalued, resulting in proportionally lower property taxes?

Having "accurate" predictions doesn't necessarily address larger historical trends and inequities, and fairness in property assessments in taxes works beyond the CCAO's valuation model. Disassociating accurate predictions from a fair system is vital to approaching justice at multiple levels. Take Evanston, IL - a suburb in Cook County - as an example of housing equity beyond just improving a property valuation model: Their City Council members recently approved reparations for African American residents.

0.7 Question 11

In your own words, describe how you would define fairness in property assessments and taxes.

In my opinion, fairness in property assements and taxes means building good models to predict housing values as accurate as possible. This means property evaluation should be dependent of factors related to the property without considering their owner. It seems in Cook County they didn't practice fairness as well as they didn't consider equality which affected their model and ultimatley the people of color, lower income, and minorities.

0.8 Question 12

Take a look at the Residential Automated Valuation Model files under the Models subgroup in the CCAO's GitLab. Without directly looking at any code, do you feel that the documentation sufficiently explains how the residential valuation model works? Which part(s) of the documentation might be difficult for nontechnical audiences to understand?

The documentation looks ok in terms of how the residential model workds. However, I realized the flow chart is complex and difficult to understand even for technical audiences. Also, I would have suggest to include some explanation to some technical terms (e.g., regularization, over-fitting etc.) in the Model Selection so that it would have become easier for ordinary people/non-technical audiences to understand the terms and ultimatley the model.