0.1 Question 1: Human Context and Ethics

In this part of the project, we will explore the human context of our housing dataset. **You should watch Lecture 15 before attempting this question.**

0.1.1 Question 1a

Consider the following question: *"How much is a house worth?"*

Who might be interested in an answer to this question? Be sure to list at least three different parties (people or organizations) and state whether each one has an interest in seeing a low or high housing price.

Your response should be approximately 3 to 6 sentences.

1. Homeowners have a dual interest in housing prices. They initially want to see low housing prices when purchasing a house to secure a good deal and minimize initial costs. However, once they own the property, they want its value to increase over time. Since many homeowners view their homes as investments, a higher selling price allows them to maximize profit when they decide to sell. These people may also consider renting out their properties, so a higher property value can lead to higher rental income and long-term financial gain.

2. Real estate agents have a direct financial interest in housing prices since their commissions are based on a percentage of the sale price. Since housing markets can get competitve, they have to find a happy balance that attracts the buyers and sellers. These agents aim for higher prices because it increases their earnings per transaction. However, they must balance this desire with the need for market stability. If prices rise too high, they may struggle to close deals, but if prices drop significantly, their earnings per sale decreases.

3. Renters would want to see low housing price, as rent costs are often tied to property values. In markets where housing prices are high, landlords tend to increase rent to reflect their investment, making housing less affordable for renters. For those in competitive rental markets, high housing prices can make finding affordable rental options increasingly difficult. As a result, renters benefit from lower home prices since they can potentially see more affordable rent costs in the long run.

0.1.2 Question 1b

Which of the following scenarios strikes you as unfair, and why? You can choose more than one. There is no single right answer, but you must explain your reasoning. Would you consider some of these scenarios more (or less) fair than others? Why?

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.

Your response for each chosen scenario should be approximately 2 to 3 sentences.

1. Option A leads to homeowners being taxed above the true market value of their property, which can feel unfair because it violates the expectation that taxes should reflect a property's actual worth. Homeowners may struggle to pay bills or cut essential expenses due to the inflated tax assessment. The inconsistency becomes more evident when similar properties are taxed at lower rates, eroding trust in the tax system and raising concerns about bias.

2. Option B results in homeowners being taxed less than their property's actual worth, benefiting them individually but harming the community. The county collects less revenue, potentially reducing funds for essential services like schools and roads. This can lead to higher taxes for other homeowners or a decline in public services, creating resentment and perceptions of unfairness.

3. Option C places a heavier tax burden on people with lower incomes or less valuable properties, which exacerbates economic inequality. In Cook County, communities of color often own less valuable property, so these tax policies disproportionately impact them. This worsens racial and economic disparities, making it harder for marginalized groups to build wealth and achieve upward mobility.

4. Option D taxes higher-income individuals or those with more valuable properties at a higher rate, promoting fairness by aligning taxes with financial ability. However, some might argue that all property owners should be taxed equally, regardless of wealth, and that high tax rates on expensive properties create an imbalance. This raises questions about the definition of fairness in tax systems, whether it is based on equal treatment or ability to pay.

0.1.3 Question 1d

What were the central problems with the earlier property tax system in Cook County as reported by the Chicago Tribune? What were the primary causes of these problems?

Your response should be approximately 2 to 4 sentences.

Note: Along with reading the paragraph above, you will need to watch Lecture 15 to answer this question.

The Cook County Assessor's Office consistently undervalued more expensive properties and overvalued less expesnive properties, creating a regressive property tax system. This disproportionately impacted less affluent and non-white homeowners, placing an unfair burden on them. The primary cause of this inequity was that wealthier homeowners could afford tax lawyers to contest their property assessments and reduce their tax liabilites. In comparison, homeowners with fewer resources lacked the means to challenge unfair assessments, forcing them to pay higher taxes.

0.1.4 Question 1e

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

Your response should be approximately 3 to 4 sentences.

The Cook County property tax system disproportionately affected non-white homeowners due to historical disparities in property ownership and wealth. Communities of color (ie. Black, immigrants, Latino, etc) faced barriers like redlining, resulting in the ownership of lower-valued properties. When tax assessments in these areas were inflated, non-white homeowners were forced to pay higher taxes than the actual value of their homes, creating financial strain and limiting their ability to build wealth. This flawed system hindered upward mobility and deepened inequality in marginalized communities.

0.2 Question 4a

We can assess a model's performance and quality of fit with a plot of the residuals $(y - \hat{y})$ versus the observed outcomes (y).

In the cell below, use `plt.scatter` (documentation) to plot the **model 2** residuals of `Log Sale Price` versus the original `Log Sale Price` values. For this part, you only need to plot the residuals and outcomes for the **validation data**.

- You should also **ensure that the dot size and opacity in the scatter plot are set appropriately** to reduce the impact of overplotting as much as possible. However, with such a large dataset, it is difficult to avoid overplotting entirely.

```
In [24]: residuals_sale_price = Y_valid_m2 - Y_predicted_m2

         plt.scatter(x=Y_valid_m2, y=residuals_sale_price, s=3, alpha=.3)

         plt.title("Model 2's Residuals vs Original Log Sale Price on Validation Data", fontsize=18);
         plt.xlabel('Original Log Sale Price: $y$', fontsize=18);
         plt.ylabel('Residual Model: $y -\hat{y}$', fontsize=18);
```

Model 2's Residuals vs Original Log Sale Price on Validation Data
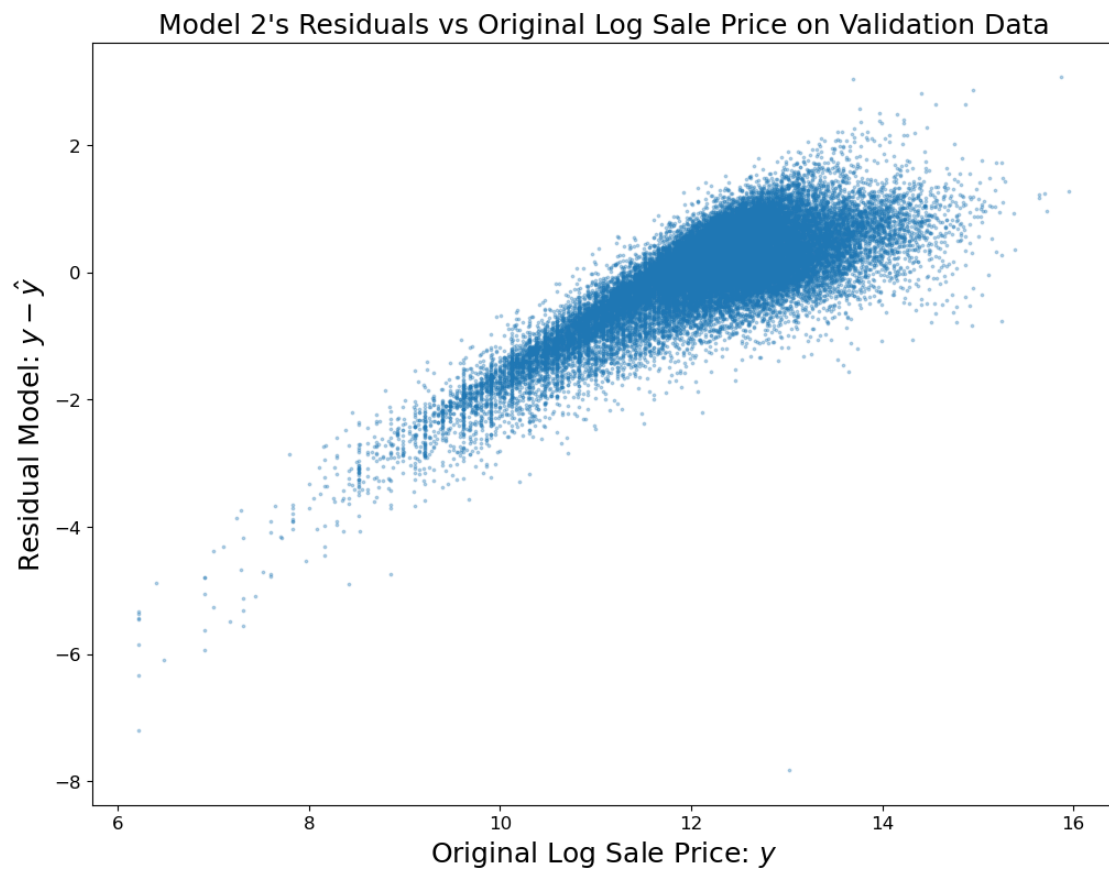
0.2.1 Question 6c

Using the functions above, we can generate visualizations of how the RMSE and proportion of overestimated houses vary for different intervals:

```
In [46]: # RMSE plot
         plt.figure(figsize = (8,5))
         plt.subplot(1, 2, 1)
         rmses = []
         for i in np.arange(8, 14, 0.5):
             rmses.append(rmse_interval(preds_df, i, i + 0.5))
         plt.bar(x = np.arange(8.25, 14.25, 0.5), height = rmses, edgecolor = 'black', width = 0.5)
         plt.title('RMSE For Different Intervals\n of Log Sale Price', fontsize = 10)
         plt.xlabel('Log Sale Price')
         plt.yticks(fontsize = 10)
         plt.xticks(fontsize = 10)
         plt.ylabel('RMSE')

         # Overestimation plot
         plt.subplot(1, 2, 2)
         props = []
         for i in np.arange(8, 14, 0.5):
             props.append(prop_overest_interval(preds_df, i, i + 0.5) * 100)
         plt.bar(x = np.arange(8.25, 14.25, 0.5), height = props, edgecolor = 'black', width = 0.5)
         plt.title('Percentage of House Values Overestimated \n for different intervals of Log Sale Pric
         plt.xlabel('Log Sale Price')
         plt.yticks(fontsize = 10)
         plt.xticks(fontsize = 10)
         plt.ylabel('Percentage of House Values\n that were Overestimated (%)')

         plt.tight_layout()
         plt.show()
```

RMSE For Different Intervals of Log Sale Price

Percentage of House Values Overestimated for different intervals of Log Sale Price
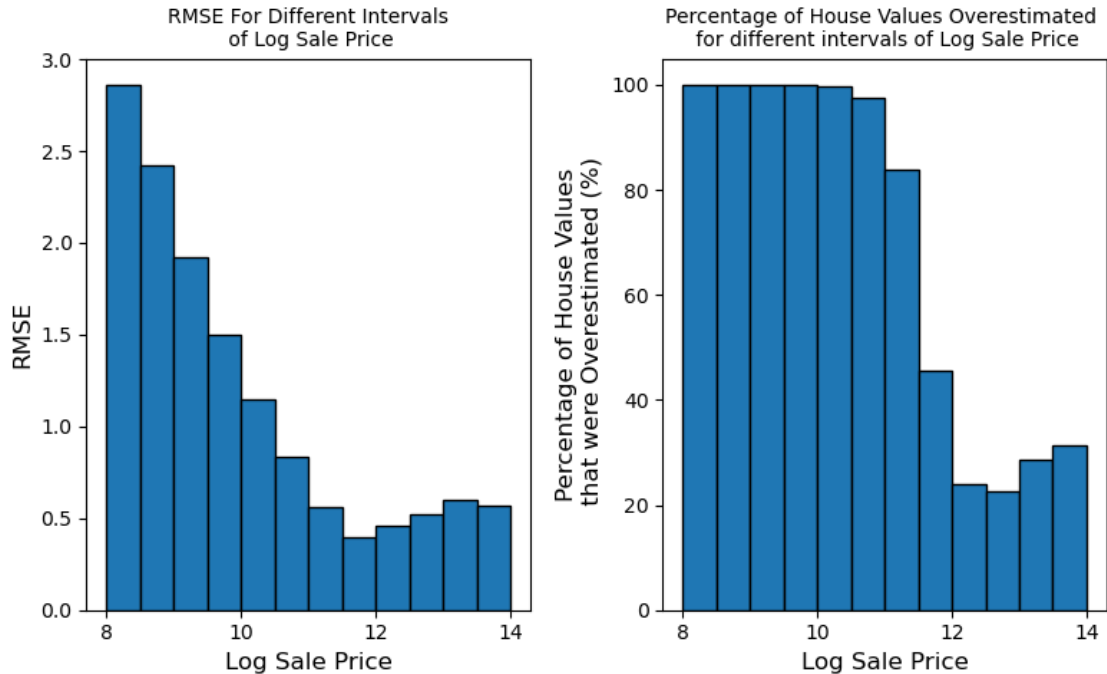
Which of the two plots above would be more useful in ascertaining whether the assessments tended to result in progressive or regressive taxation? Provide a brief explanation to support your choice of plot.

Then, explain whether your chosen plot aligns more closely aligns with scenario C or scenario D from `q1b`:

```
C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive
D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive
```

Your response should be approximately X to Y sentences.

The plot on the right would be more useful in ascertaining whether the assessments tend to result in progressive or regressive taxation. Regressive taxation occurs when lower-priced homes are overvalued relative to higher-priced homes, leading to a disproportionate tax burden on less expensive properties. The right plot shows exactly that: lower-priced homes are overestimated far more frequently than higher-prices homes. This assessment process aligns more closely with scenario C, which describes an assessment process that overvalues inexpensive properties and undervalues expensive ones.

12

0.3 Question 7: Evaluating the Model in Context

_____

0.4 Question 7a

When evaluating your model, we used RMSE. In the context of estimating the value of houses, what does the residual mean for an individual homeowner? How does a positive or negative residual affect them in terms of property taxes? Discuss the cases where the residual is positive and negative separately.

Your response should be approximate 2 to 4 sentences.

The residual represents the difference between the actual market sale price of a home and its predicted value based on the model used by the Cook Count Assessor Office. A positive residual indicates that the property is undervalued, meaning homeowners are paying lower property taxes than they should relative to the true value of their property. In comparison, a negative residual suggests that the property is overvalued, causing homeowners to pay higher property taxes than warranted based on the actual sale price.

0.5 Question 7b

Reflecting back on your exploration in Questions 6 and 7a, in your own words, what makes a model's predictions of property values for tax assessment purposes "fair"?

This question is open-ended and part of your answer may depend on your specific model; we are looking for thoughtfulness and engagement with the material, not correctness.

Hint: Some guiding questions to reflect on as you answer the question above: What is the relationship between RMSE, accuracy, and fairness as you have defined it? Is a model with a low RMSE necessarily accurate? Is a model with a low RMSE necessarily "fair"? Is there any difference between your answers to the previous two questions? And if so, why?

Your response should be approximate 1 to 2 paragraphs. Feel free to answer the questions in the hint to structure your answer.

A model's predictions of property values for tax assessment purposes can be considered "fair" if they are accurate across different groups. This ensures that all homeowners are taxed based on the true value of their property, regardless of location or other biased factors. Fairness means that the model does not favor specific neighborhoods or racial groups, and that errors are evenly distributed without disproportionately affecting certain communities. A model with a lower RMSE indicates higher accuracy, but it doesn't guarantee fairness because a model could still be biased. For example, a model might undervalue properties in lower-income areas and overestimate values in wealthier neighborhoods, leading to unfair tax disparities. The CCAO case study demonstrates that limitations exist in assessing fairness. While minimizing RMSE is important for accuracy, fairness requires ensuring that errors are fairly distributed accross communities. A model must not only reduce overall error but also ensure it doesn't unfairly impact specific groups.