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 part.**

0.1.1 Question 1a

"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 housing price to be high or low.**

Homeowners who are looking to sell their properties typically have an interest in seeing housing prices be as high as possible. A higher sale price means a better return on their investment.

Potential homebuyers are generally interested in lower housing prices. Lower prices can make homeownership more affordable and accessible to a broader range of individuals or families.

Local governments and tax authorities may be interested in housing prices, as property taxes are often based on property values. Higher property values can lead to increased tax revenue, which can fund public services and infrastructure.

0.1.2 Question 1b

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. 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.

A. A homeowner whose home is assessed at a higher price than it would sell for: Unfair to the homeowner because they could potentially be paying higher property taxes than they should, given the property's actual market value. It may also be unfair because it places a financial burden on the homeowner.

B. A homeowner whose home is assessed at a lower price than it would sell fo U unfair by other homeowners and local authorities, as it might lead to lower property tax revenue for the community, which in turn could impact public services and infrastructure. However, some may argue that it is fair for homeowners to pay taxes based on the assessed value, even if it's lower than the market value.

C. An assessment process that systematically overvalues inexpensive properties and undervalues expensive propertiUas unfair because it disproportionately affects owners of less expensive properties, who would end up paying higher property taxes compared to the owners of more expensive properties. It can exacerbate socioeconomic disparities and place a heavier financial burden on those who can least afford it.

D. An assessment process that systematically undervalues inexpensive properties and overvalues expensive properUered unfair as it benefits owners of expensive properties while placing a higher tax burden on owners of less expensive properties. It can lead to unequal treatment and may not align with the principle of equitable property taxation.

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? (Note: In addition to reading the paragraph above you will need to watch the lecture to answer this question)

Type your answer here, replacing this text.

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?

Type your answer here, replacing this text.

0.2 Question 4a

One way of understanding a model's performance (and appropriateness) is through a plot of 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 **validation data**. With such a large dataset, it is difficult to avoid overplotting entirely. 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.

In [121]:
```
# Calculate residuals
residuals_m2 = Y_valid_m2 - Y_predicted_m2

# Create the scatter plot
plt.scatter(residuals_m2, Y_valid_m2, s=10, alpha=0.5)

# Set plot labels and title
plt.xlabel("Residuals")
plt.ylabel("Log Sale Price")
plt.title("Residuals vs. Log Sale Price (2nd Model - Validation Data)")
```

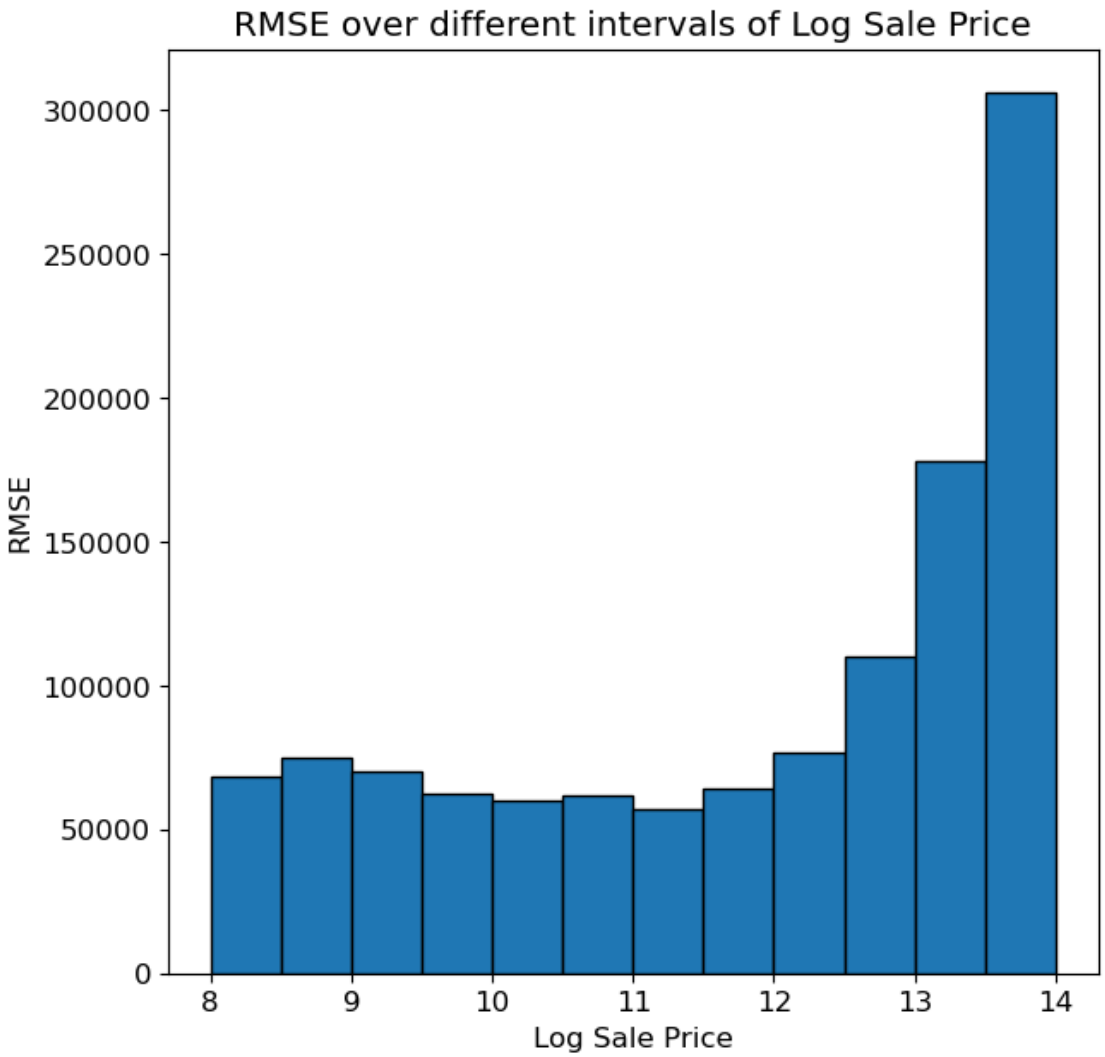
Out[121]:
```
Text(0.5, 1.0, 'Residuals vs. Log Sale Price (2nd Model - Validation Data)')
```


Residuals vs. Log Sale Price (2nd Model - Validation Data)
Log Sale Price
Residuals
16
14
12
10
8
6
−6
−4
−2
0
2

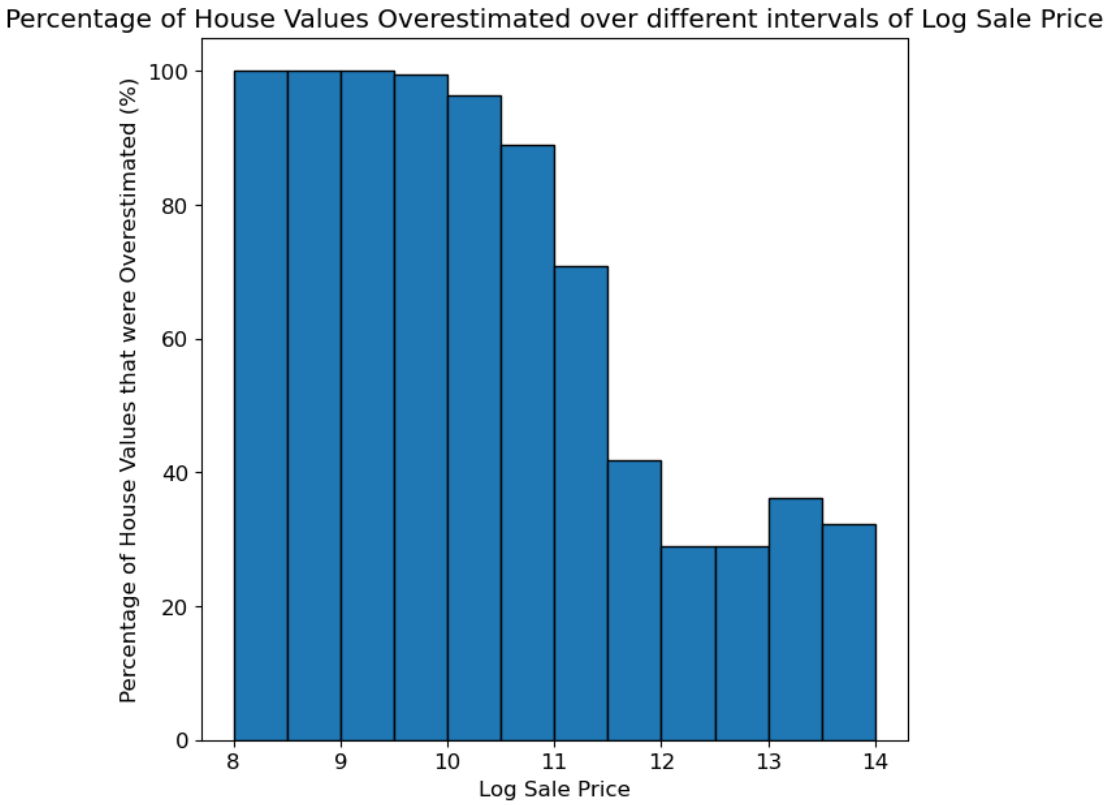
0.2.1 Question 6c

Now that you've defined these functions, let's put them to use and generate some interesting visualizations of how the RMSE and proportion of overestimated houses vary for different intervals.

In [193]:
```
# Run the cell below to generate the plot; no further action is needed
rmses = []
for i in np.arange(8, 14, 0.5):
    rmses.append(rmse_interval(X, Y, i, i + 0.5))
plt.figure(figsize = (7, 7))
plt.bar(x = np.arange(8.25, 14.25, 0.5), height = rmses, edgecolor = 'black', width = 0.5)
plt.title('RMSE over different intervals of Log Sale Price')
plt.xlabel('Log Sale Price')
plt.ylabel('RMSE');
```



In [194]:
```
# Run the cell below to generate the plot; no further action is needed
props = []
for i in np.arange(8, 14, 0.5):
    props.append(prop_overest_interval(X, Y, i, i + 0.5) * 100)
plt.figure(figsize = (7, 7))
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 over different intervals of Log Sale Price')
plt.xlabel('Log Sale Price')
plt.ylabel('Percentage of House Values that were Overestimated (%)');
```



Explicitly referencing **any ONE** of the plots above (using `props` and `rmses`), explain whether the assessments your model predicts more closely align with scenario C or scenario D that we discussed back in `1b`. Which of the two plots 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. For your reference, the scenarios are also shown below:

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

Type your answer here, replacing this text.

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 it affect them in terms of property taxes? Discuss the cases where residual is positive and negative separately.

Type your answer here, replacing this text.

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 upon 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?

Type your answer here, replacing this text.