1. Executive Summary and Implications

Summary of problem:

The problem I am investigating is “How do living area dimensions affect rent prices in Moscow?” This is relevant because being able to accurately predict rent prices in a single city based on basic information about the building and area could give potential renters the ability to compare the information and see if they are getting a fair price. The dataset I am using to answer this was updated in March of 2022, three weeks after the launch of the Russian invasion of Ukraine. I will be using a decision tree regressor to answer this question.

Hypothesis:

I hypothesize that the living area dimensions account for 20% or less of the variation in rent prices. I will test this using the r2 score.

Summary of process:

* Reading in data
* Viewing variable distribution
* Filter out outliers
* Defining variables
* Converting columns to arrays
* Reshaping arrays to remove the index
* Splitting the data into testing and training sets
* Creating the decision tree regressor
* Fitting the model
* Generating predictions
* Calculating r2 score
* Viewing residual plot

Outline of findings:

* There were several outliers for price. To remedy this I removed all rows with a price of 100,000 Rubles or greater.
* Mean price of rent is 50,987 Rubles, which is equivalent to approximately $518.24 U.S dollars. The first quartile mark is at 40,000 Rubles and the third quartile mark is at 60,000 Rubles. This gives us an interquartile range of 20,000 Rubles.
* Correlation of 0.89 for living\_area and kitchen\_area.; correlation of 0.56 for for kitchen\_area and total\_area.
* R2 score is 0.144, meaning about 14% of the variation in rent prices is due to living area alone.
* 14 < 20, so I fail to reject my hypothesis.
* The implication of this analysis is that more than 86% of the variation in rent prices in Moscow at the moment is due to factors other than living area dimensions.

Limitations of technique and tools:

* Very small dataset when compared to the number of properties available for rent at this time.
* Only one type of model was created. Others may be more accurate.

Proposed action:

* Based on my results I recommend running an expanded analysis that looks at all of the available features in relation to rent price.
* I also recommend trying web scraping in an attempt to collect more data.
* Possible future studies include looking at how the closest metro station affects rent and how each property price changes over time.

Benefits of study:

* Established average rent of 50, 987 Rubles with a first quartile of 40,000 Rubles and a third quartile of 60,000 Rubles.
* Established 14% of rent variation is due to living area.
* These numbers serve as reference for those looking for rental properties right now.