**Introduction:** "Good [morning/afternoon], everyone. Today, I'll be presenting the first section of our report, which focuses on the critical steps of data preprocessing. This phase was essential in transforming raw, unstructured data into a clean and consistent dataset suitable for analysis."

**Slide 1: Data Source** "Our data comes from DataMarket.es, a comprehensive source for real estate listings in Spain. The dataset consists of 1000 instances and 24 variables, with the primary goal of predicting the type of operation for properties—whether they are for sale or rent."

**Slide 2: Data Description** "Before preprocessing, the dataset presented several challenges. Some variables, like website and URL, added little value due to their unique or redundant nature. Others, like price and area, contained missing or inconsistent data. Our goal was to address these issues systematically to ensure reliable results."

**Slide 3: Preprocessing Steps** "Now, let’s talk about the steps we took to clean and prepare the data:

1. **Renaming Variables:** We unified variable names into English for consistency and better visualization.
2. **Handling Missing Data:** For critical variables like price and area, we applied KNN imputation to estimate missing values. Structural missing data, such as rooms for studio apartments, was assigned default values.
3. **Eliminating Redundancy:** For example, we removed duplicate columns like 'floor.1' and variables with nearly 100% missing data.
4. **Adjusting Data Types:** Categorical variables were converted to factors for better compatibility in our analysis."

**Slide 4: Post-Processing Results** "After preprocessing, our dataset was reduced to 21 variables, all of which contribute meaningful information. Importantly, we retained the dataset's original structure of 1000 instances. By resolving inconsistencies and removing noise, we laid a strong foundation for accurate and insightful analysis."

**Conclusion:** "In summary, preprocessing was a critical step that not only cleaned the dataset but also optimized it for advanced analysis. By addressing missing data, unifying variable names, and eliminating redundancy, we’ve ensured that the subsequent steps—like clustering and modeling—will yield reliable insights. Thank you, and I’ll be happy to answer any questions after the presentation!"