**Creating a report summarizing your findings involves documenting the key steps and insights from your data analysis and modeling process. Below is a structured template you can use to prepare a comprehensive report:**

**---**

**## \*\*House Price Analysis and Prediction Report\*\***

**### \*\*1. Introduction\*\***

**- \*\*Objective\*\*: The objective of this project is to analyze a dataset of house prices, identify key factors influencing these prices, handle outliers, and develop a predictive model to estimate future house prices.**

**### \*\*2. Data Exploration Results\*\***

**- \*\*Dataset Overview\*\*:**

**- The dataset contains \*\*168,446\*\* entries with \*\*20\*\* features including property details such as `property\_type`, `location\_id`, `city`, `province\_name`, `latitude`, `longitude`, `baths`, `bedrooms`, `area`, `purpose`, `Area Size`, etc.**

**- Key features impacting price are `location\_id`, `property\_type`, `Area Size`, `baths`, and `bedrooms`.**

**- \*\*Key Insights from Visualizations\*\*:**

**- \*\*Price Distribution\*\*: The house prices are right-skewed, with a few properties having exceptionally high prices.**

**- \*\*Feature Relationships\*\*: Positive correlations were observed between `Area Size`, `baths`, `bedrooms`, and price, indicating that larger houses with more amenities tend to be priced higher.**

**- \*\*Location Influence\*\*: Specific locations within Islamabad, such as G-10 and E-11, show higher price trends.**

**### \*\*3. Feature Engineering Techniques Used\*\***

**- \*\*Total Rooms\*\*: A new feature `total\_rooms` was created by summing the number of `baths` and `bedrooms`.**

**- \*\*Price per Area Unit\*\*: `price\_per\_area\_unit` was calculated to understand the value per unit area.**

**- \*\*Categorical Encoding\*\*: Categorical features like `property\_type`, `city`, `province\_name`, `area`, `purpose`, and `Area Type` were one-hot encoded to convert them into a numerical format suitable for modeling.**

**### \*\*4. Outlier Analysis\*\***

**- \*\*Identification\*\*:**

**- Outliers were identified in features such as `price`, `baths`, `bedrooms`, and `Area Size`.**

**- Extreme outliers included properties with over 403 baths or a price of 2 billion, which were capped at the 99th percentile.**

**- \*\*Explanation\*\*:**

**- These outliers were likely due to data entry errors or properties with exceptionally luxurious amenities, impacting the model's accuracy.**

**### \*\*5. Model Selection and Evaluation\*\***

**- \*\*Model Chosen\*\*: RandomForestRegressor was selected for its robustness in handling large datasets with numerous features.**

**- \*\*Evaluation Metrics\*\*:**

**- \*\*R-squared\*\*: The model achieved an R-squared value of \*\*0.87\*\*, indicating that it explains 87% of the variance in house prices.**

**- \*\*Mean Squared Error (MSE)\*\*: The MSE was calculated to measure the average squared difference between observed and predicted prices.**

**- \*\*Cross-Validation\*\*: 5-fold cross-validation was performed to ensure the model's reliability across different subsets of the data.**

**### \*\*6. Future Price Prediction Examples\*\***

**- Using the trained model, the price for a 3-bedroom, 3-bathroom flat in Islamabad with an area size of 5 Marla was predicted to be \*\*PKR 1,855,700\*\*.**

**- These predictions help in understanding market trends and making informed decisions for buyers and investors.**

**### \*\*7. Recommendations\*\***

**- \*\*Further Analysis\*\*:**

**- Explore additional features such as proximity to landmarks or public amenities which might further explain variations in price.**

**- Analyze trends over time by incorporating temporal data (e.g., property prices over different years or seasons).**

**- \*\*Data Collection\*\*:**

**- Collect more granular data on property conditions and renovations, as these can significantly impact property values.**

**- Ensure consistent and accurate data entry to reduce the prevalence of outliers.**

**### \*\*8. Deliverables\*\***

**- \*\*Script\*\*: The complete, well-documented script used for data cleaning, exploration, feature engineering, modeling, and prediction is provided.**

**- \*\*Report\*\*: This report summarizes the findings and provides actionable recommendations for further analysis and data collection.**

**---**

**### \*\*Next Steps:\*\***

**1. \*\*Save the Report\*\*: You can document this report in a text file or a PDF document.**

**2. \*\*Compile and Share\*\*: Package the report along with the code script and any necessary data files, and share them as your project deliverables.**

**If you need any further assistance with formatting or generating the final document, let me know!**