**name: Laiba Tauseef , Pirah Agha**

**semester : 6th**

**section: b**

**subject: Data Mining**

**topic: Housing Price Prediction (Islamabad)**

**objective:**

The primary objective of this housing price prediction project is to leverage machine learning algorithms, specifically Decision Trees and Random Forest, to develop an accurate and efficient predictive model. This model aims to analyze and predict housing prices based on relevant features, contributing valuable insights to the real estate sector. The project seeks to enhance prediction accuracy, understand key factors influencing housing prices, and provide a practical tool for stakeholders to make informed decisions in the dynamic real estate market.

**Intro & background:**

The real estate sector plays a pivotal role in global economies, yet traditional pricing models often struggle to capture the complexity of market variables. Machine learning algorithms offer a promising solution This project focuses on leveraging machine learning, specifically Decision Trees and Random Forest algorithms, to enhance the accuracy of housing price forecasts. The goal is to not only develop effective models but also to gain insights that can empower stakeholders in making informed decisions within the dynamic landscape of real estate.

**Data collection:**

Graana.com was chosen as the data source due to its reputation for comprehensive and reliable real estate listings. The platform offers a diverse range of properties, including residential units, which aligns well with the goals of the housing price prediction project.the dataset include variety of feature that can be used as input for predictive model like

Purpose,Area Maarla Size,Area Name , Size Unit,Bedroom , Bathroom,Sub Type, CityName

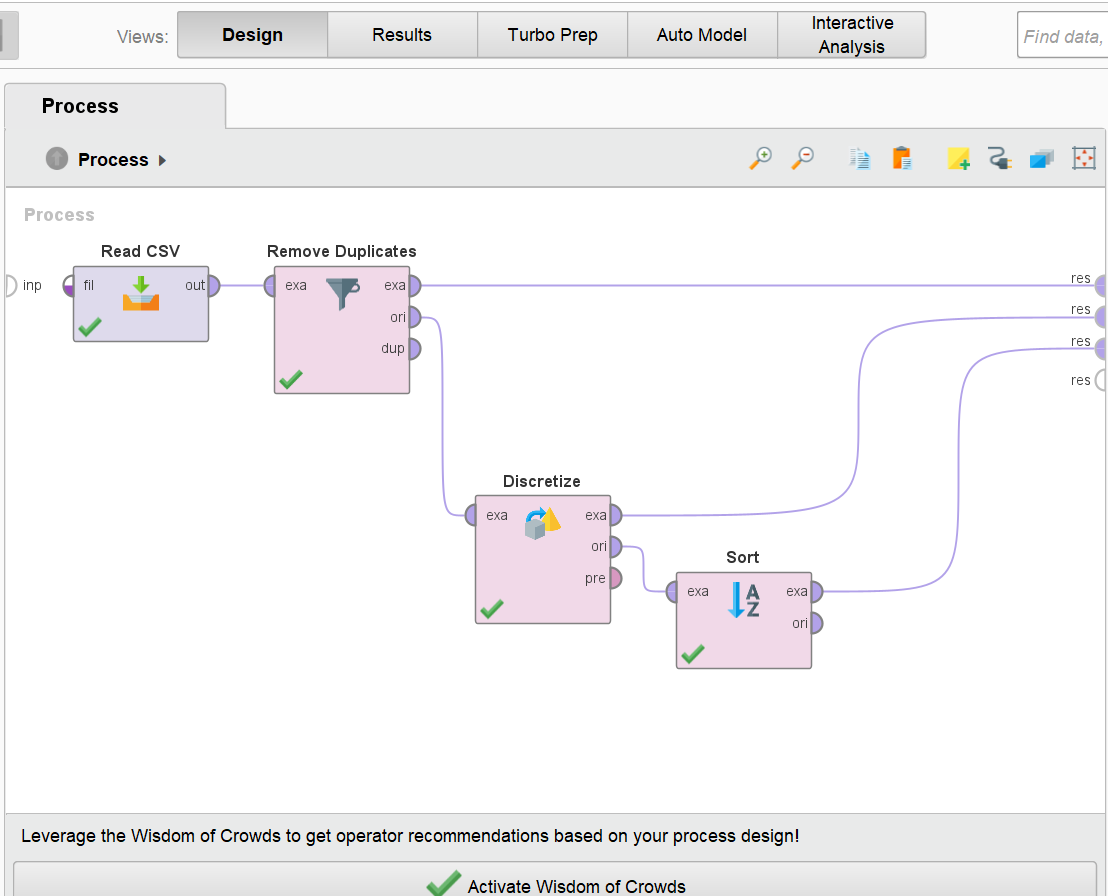
**data preprocessing:**

The initial phase of the project involved comprehensive data preprocessing to ensure the dataset's quality and readiness for predictive modeling. This process was crucial for enhancing the accuracy and reliability of the subsequent predictive models.

**Data Cleaning Techniques:**Various data cleaning techniques were employed to address any anomalies or outliers within the dataset. This step ensured that the input data for the predictive models was reliable and reflective of the real-world housing market.

**Handling Missing Values:** firstly we visualize that our dataset has not missed value.

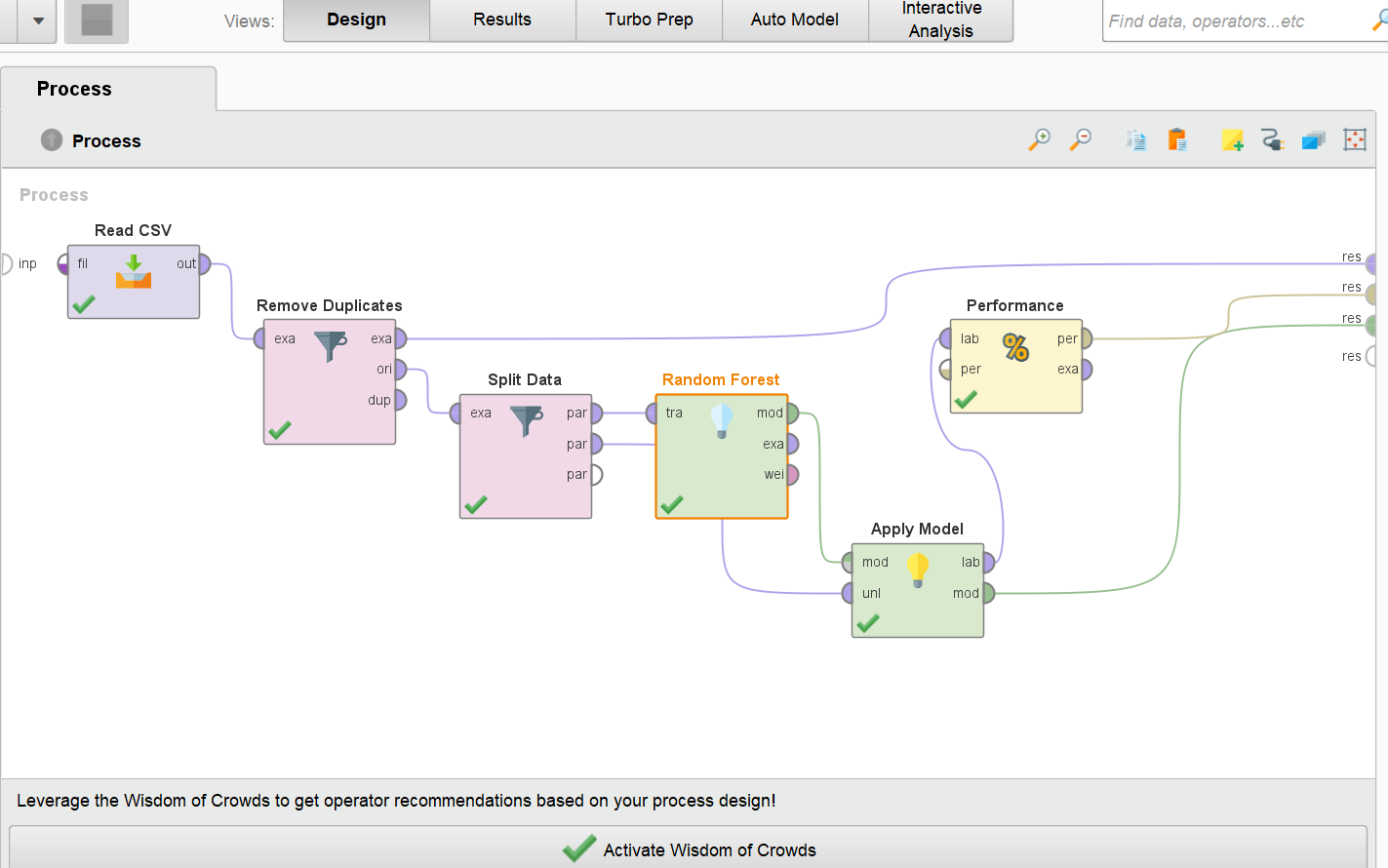
**Redundancy Binning Process:** A strategic redundancy binning process was implemented to identify and eliminate redundant features, promoting a more streamlined and efficient dataset. This step aimed to enhance model performance and interpretability.

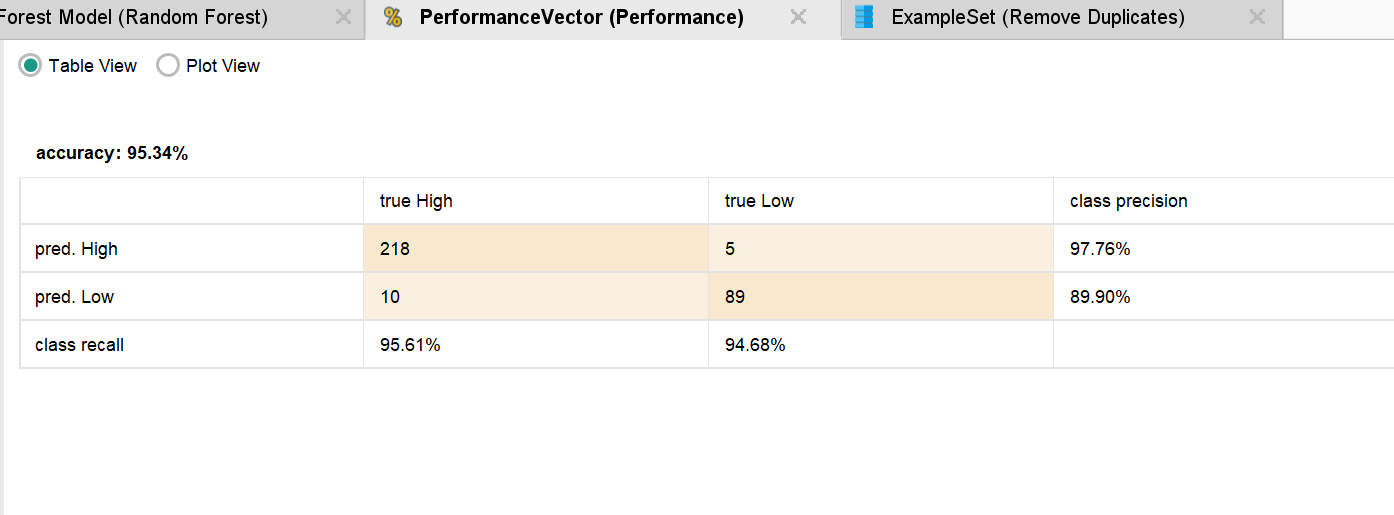


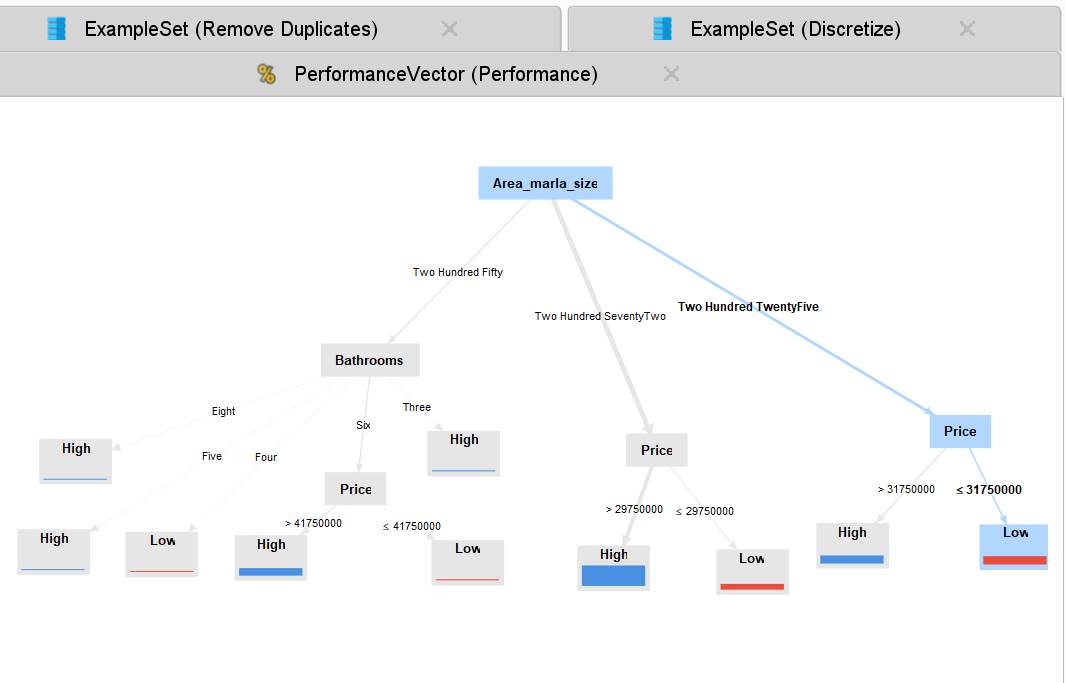
**Model Selection:**

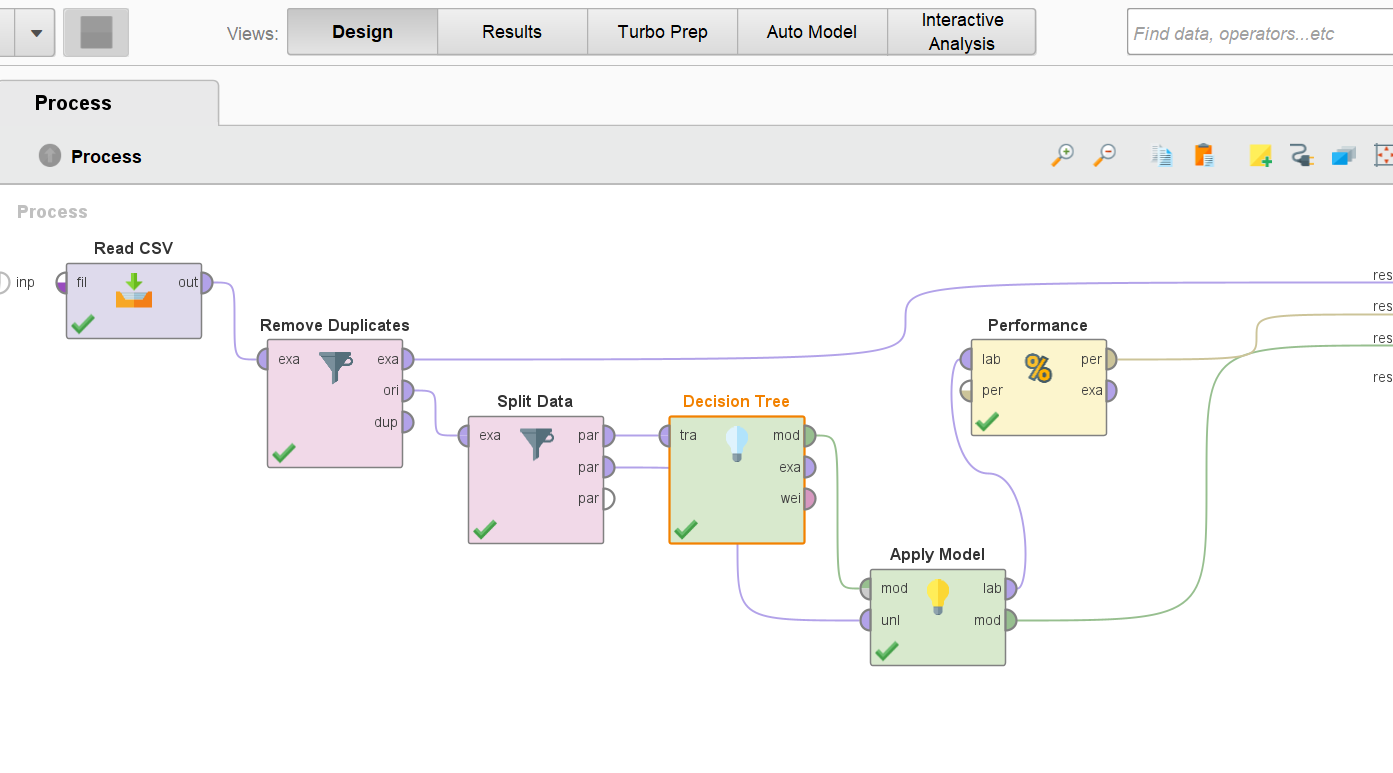
Decision Tree and Random Forest models were selected for housing price prediction due to their ability to handle complex relationships within the data, offering a balance between interpretability and predictive power.Decision tree aree well suitable for this task as they can handle both numerical and categorical value.

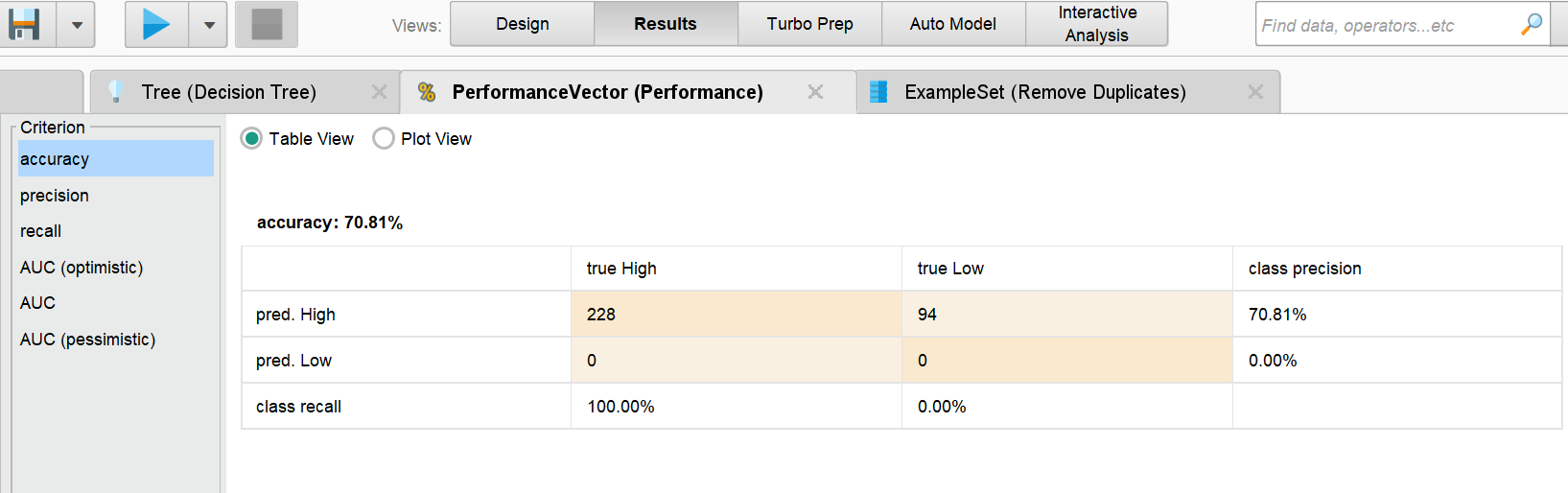
**RESULT**











**Conclusion:**

In conclusion, the application of decision tree and random forest models for housing price prediction has yielded promising results. Through a comprehensive data preprocessing phase that included handling missing values, redundancy binning, and various data cleaning techniques, the models were trained on a robust dataset.The insights gained from this study not only provide a reliable basis for housing price prediction but also highlight the significance of employing ensemble learning techniques like random forests for improved model performance