

1. Data Acquisition

- Explain how the dataset was obtained from *realestate.com.au* (manual HTML saving + BeautifulSoup scraping).
- Mention the raw columns: address, suburb, state, postcode, date_sold, bedroom, bathroom, carspace, size, type, price.

2. Data Preprocessing & Exploratory Data Analysis (EDA)

- **Cleaning:** handling missing values, converting dates, parsing numbers, removing commas.
- **Feature Engineering:** added time-based features (sale_year, sale_month, sale_dow, is_weekend), structural features (rooms_total, size_per_room, carspace_per_rm), frequency encodings (suburb_freq, postcode_freq).
- **Encoding & Standardization:** used OneHotEncoder for categorical features, StandardScaler for numerics in a ColumnTransformer pipeline.
- **Visualizations:**
 - Price distribution (normal and log).
 - Correlation heatmap.
 - Outlier detection with boxplots.
 - Price trend analysis across suburbs.

3. Model Development

- Built three models: Linear Regression, Random Forest, Gradient Boosting.
- Used **5-fold cross-validation** with **MAE, RMSE, and R^2** metrics.
- Show performance table and justify choice of best model (lowest RMSE / best R^2).
- Explain underfitting vs overfitting risk.

4. Feature Importance

- Extracted **feature importances** from tree models.
- Plotted top 20 features.
- Used **SHAP values** to analyze contributions of features like suburb_freq, size, bedroom, sale_year.
- Discussion of interpretability: tree models vs linear regression.


5. Model Deployment

- Built a **Streamlit web app (app.py)**.
- Allows users to enter property details and sale date.
- App computes engineered features and feeds them into the trained pipeline.
- Returns predicted price with simple UI.

```
(base) C:\Users\micha\OneDrive\Desktop>manual backup laptop\deakin stuff\deakintrimester3\SIT720 - Machine Learning\task
8.2>streamlit run app.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.0.43:8501
```



Housing Price Prediction

Enter Property Details

Suburb	Bedrooms
<div>Docklands</div>	<div>4</div>
Postcode	Bathrooms
<div>3008</div>	<div>2</div>
Property Type	Car Spaces
<div>Apartment</div>	<div>2</div>
Size (sqm)	
<div>650</div>	
Sale Date	
Select a sale date	
<div>YYYY/MM/DD</div>	
<div>Predict Price</div>	

Estimated Price: \$1,819,879