**DATA SUMMARY:**

The three provided data sets were used in the study to determine which Victoria, Australia, suburb would be the greatest place to invest in real estate:

• The data in the Apartment\_prices.csv shows the median price of houses in various

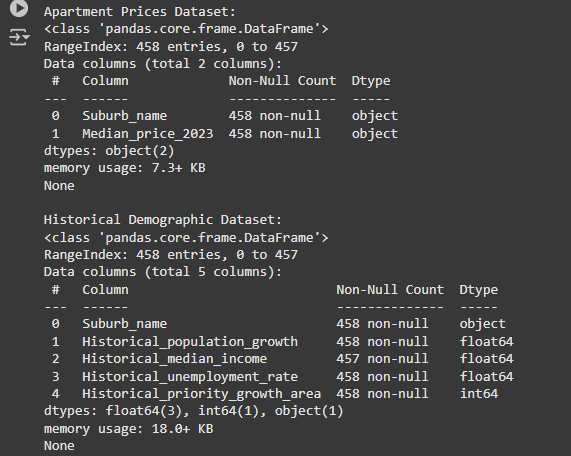
suburbs in 2023.

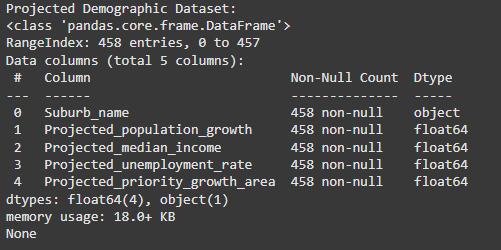
• Historical\_demographics.csv contains data from the previous year's priority growth

areas, median income, unemployment rate, and population growth rate.

• Data on the unemployment rate, population growth rate, median income, and priority

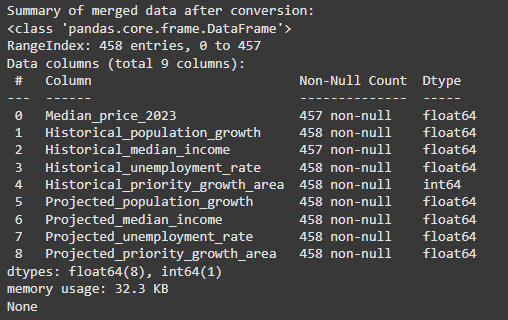
growth area for the upcoming year are provided by projected\_demographics.csv.





The three data sets were first combined on "**Suburb\_name**," after which the data was cleaned and processed.

• **Historical\_median\_income** had one missing value, and Median\_price\_2023 had one incorrect value, which was changed to the column mean.

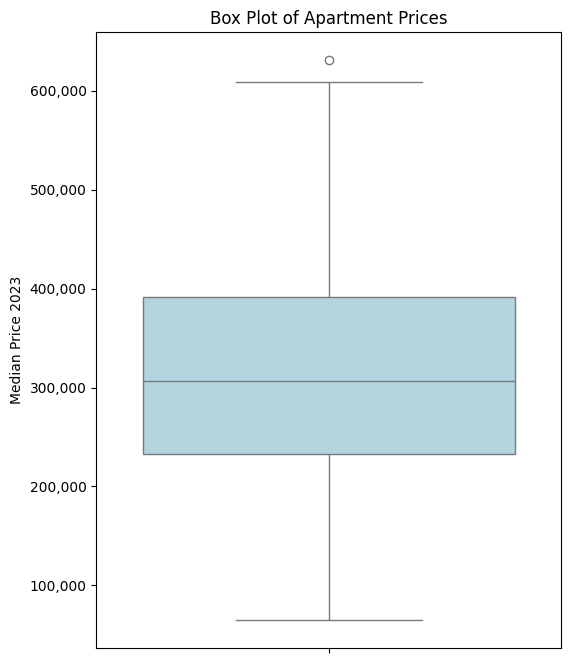


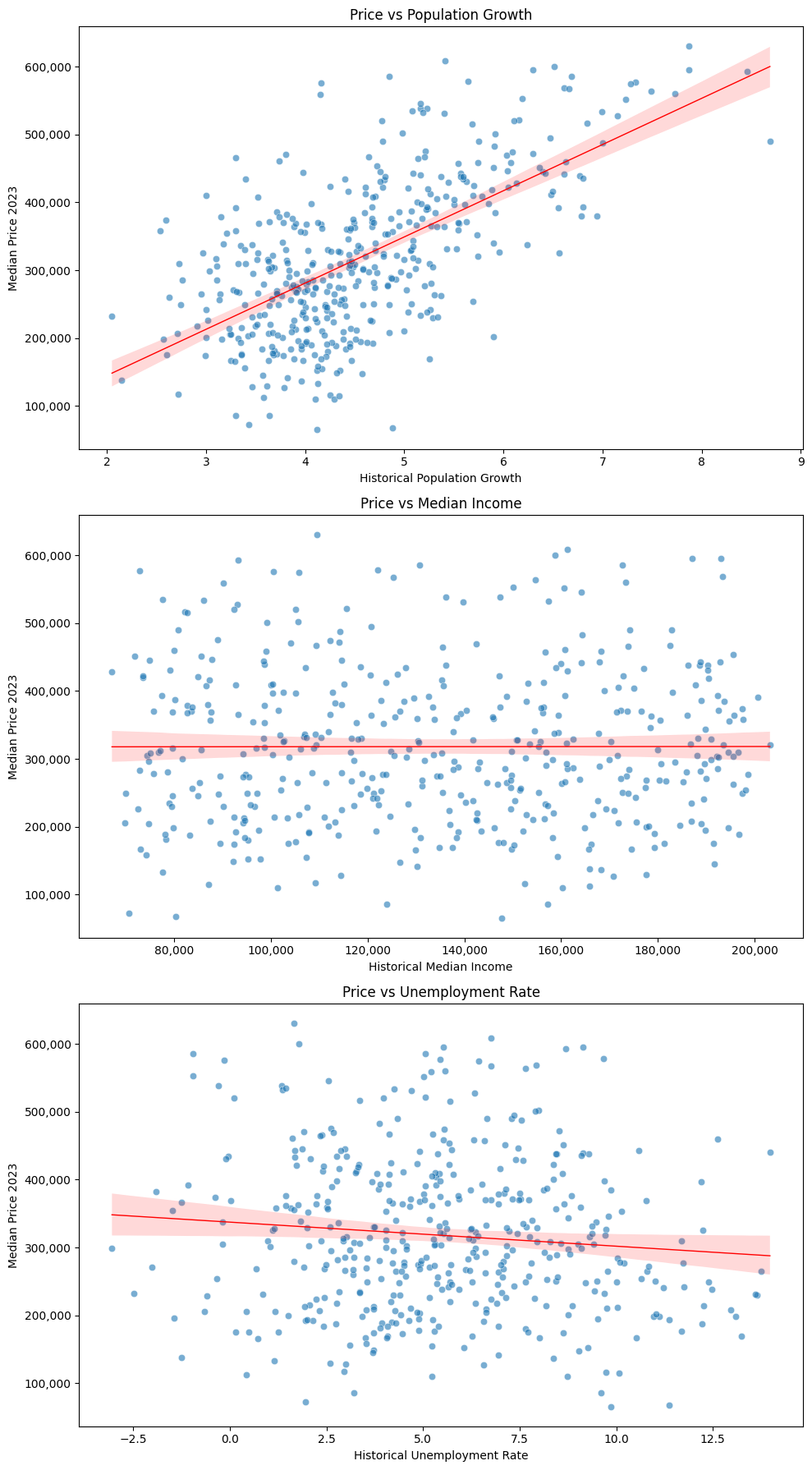
• Outliers were identified using boxplots and handled by excluding the outliers using IQR

method.

After cleaning the data, it is found that the majority of the suburbs have median prices between 200k and 400k, furthermore, the median price of the apartments has no correlation between Historical median income and very slight negative correlation (-0.1) with Historical unemployment rate.

****

****

****

**MODEL ESTIMATION**:

Using the cleaned data, the correlation between the variables was analysed and it was noted that **‘Median\_price\_2023’** had the highest positive correlation of 0.63 between **‘Historical\_population\_growth’** and **‘Projected\_population\_growth’** among other variables.

A linear regression model was selected for its simplicity and interpretability. The independent variables selected were:

• Historical population growth

• Historical unemployment rate

• Historical priority growth area

Historical median income has been excluded from the model as it has no correlation with median prices.

These variables were selected because these are some of the key factors which influence the ROI of an apartment.

The formula for the regression model is :

𝑀𝑒𝑑𝑖𝑎𝑛 𝑝𝑟𝑖𝑐𝑒 = β0 + β1 ∗ Historical\_population\_growth + β2

∗ Historical\_unemployment\_rate + β3 ∗ Historical\_priority\_growth\_area

Where:

β0 = -67585 (Intercept)

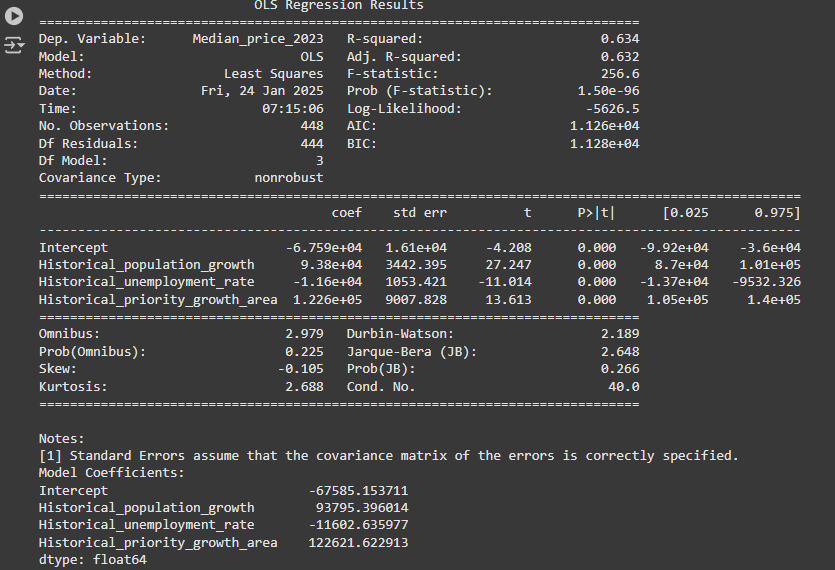
β1 = 93795

β2 = -11063

β3 = 122622

**MODEL INTERPRETATION:**

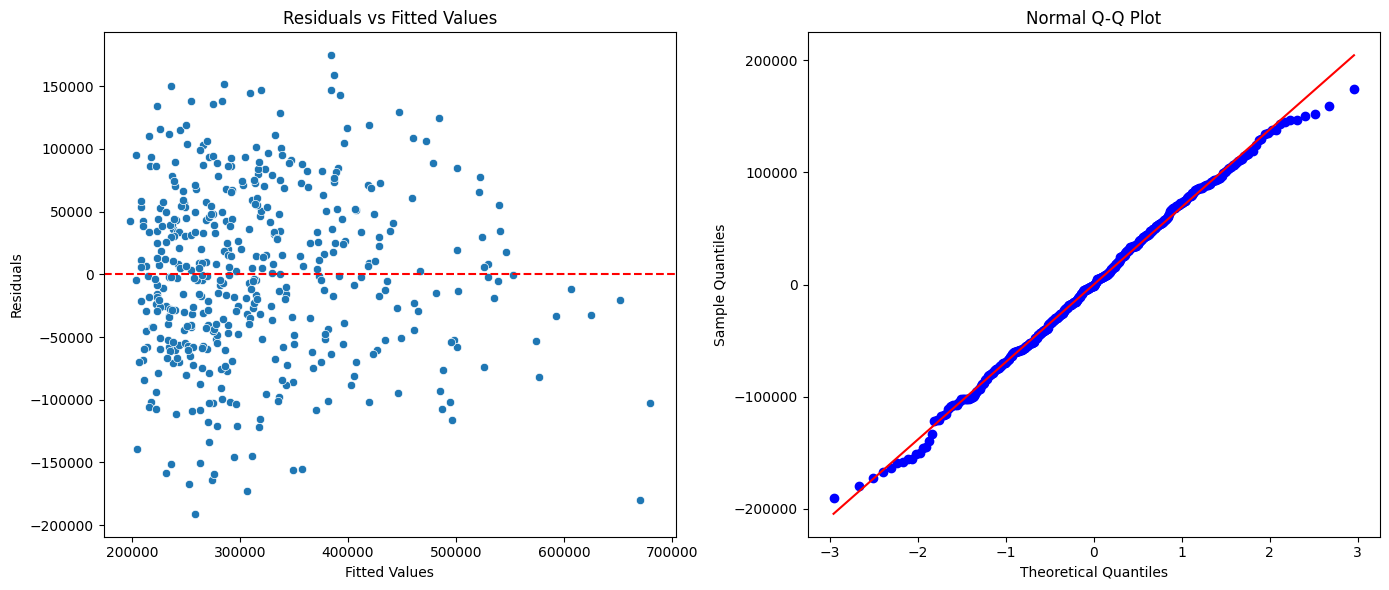
Based on the model summary, it is evident that **Historical\_population\_growth**, **Historical\_unemployment\_rate** and **Historical\_priority\_growth\_area** are significant independent variables.





The suburbs with higher historical rates of population growth and median income are expected to have higher median apartment prices, according to the positive coefficients for both variables. On the other hand, the negative correlation for unemployment rate suggests that lower apartment prices are related to higher unemployment rates.

The multiple R- squared value 0.6342 suggests that this model can interpret 63.42% changes in the median prices based on the used independent variables.



• The lack of a clear pattern suggests that the linear model adequately explains the

relationship between the independent variables and the dependent variable. The

residuals are evenly scattered along the horizontal axis suggesting that the variance of

the residuals is consistent across all levels of fitted values.

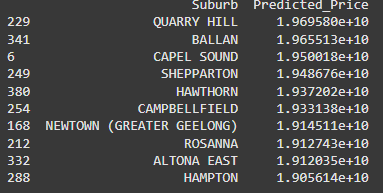
• The Q—Q plot depicts that majority of the residual follow the 45 degree line

suggesting a near normal distribution, validating the assumption required for linear

regression.

**RECOMMENDATIONS**:

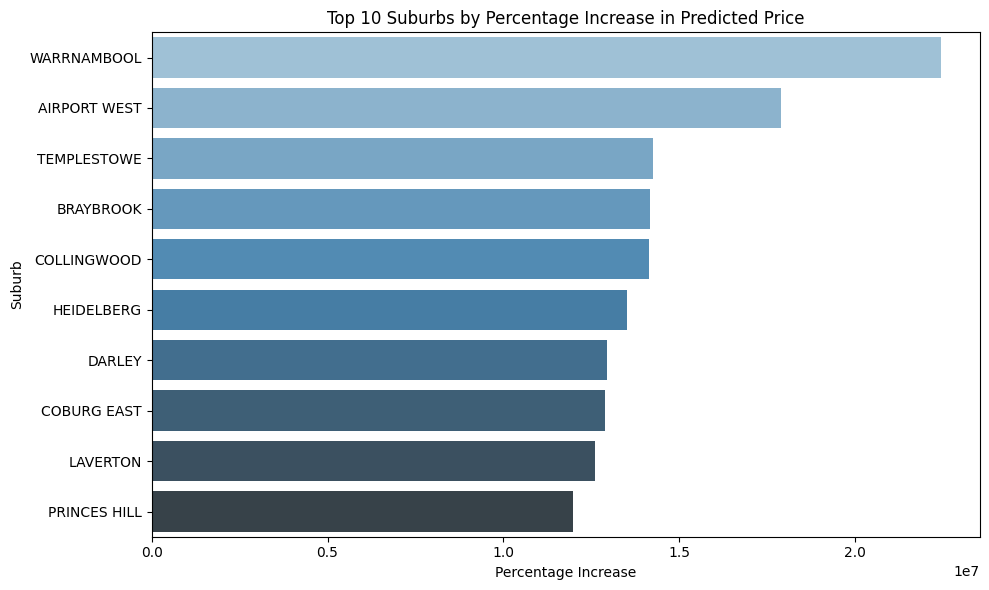
This model is used to find the predicted median price of the apartments in the next year, which are:



Then the percentage increase in the median prices is calculate using the formula



As an investor, the company should invest in the suburb which had the highest percentage increase in the median price



Based on these observations, it is evident **WARRNAMBOOL** is the suburb which had the highest percentage increase in the median price of 297.66%. If this trend continues, investing in Montrose will give the highest ROI for the company.