Análise exploratória dos imóveis da Cidade de Ribeirão Preto

1- Introdução

Os dados foram extraídos do site vivareal (https://www.vivareal.com.br/), atráves de um processo de web scraping (Mostrar o projeto anterior), gerando uma base de 8395 linhas e 9 colunas.

2 – Análise estatísticas:

Para a exploração da base, foram usadas as seguintes técnicas: Teste de hipóteses, intervalo de confiança, teste da normalidade entre dados, além da análise da média, mediana e moda. Os bairros escolhidos foram: Sumarezinho e Centro, para o tipo de imóvel apartamento.

3- Média, Mediana e Moda:

3.1 – Bairro Centro

Para o centro, foram extraídos 736 imóveis, a média foi de R$ 402076.73, a mediana dos preços é de R$ 360000,00, ela indica que 50 % de dados são menores ou maiores que o valor, a moda apresenta o valor de R$ 450000,00, ou seja, temos vários imóveis com esse valor, o coeficiente de variação registrado apresenta o valor de 55,55%, ou seja uma variabilidade menor entre os dados

3.3 – Bairro Sumarezinho

Para o centro, foram extraídos 118 imóveis, a média foi de R$ 245524,66, a mediana dos preços é de R$ 230000,00, ela indica que 50 % de dados são menores ou maiores que o valor, a moda apresenta o valor de R$ 170000,00, ou seja, temos vários imóveis com esse valor, o coeficiente de variação registrado apresenta o valor de 29,73%, ou seja uma variabilidade menor entre os dados

4- Intervalo de confiança:

O intervalo de confiança é uma medida estatística que fornece uma medida de incerteza para ajudar a avaliar a precisão e a confiabilidade das estimativas, com um valor de probabilidade, geralmente é usada 95 % de probabilidade, ou seja, de cada 100 testes feitos , em 95, apresentaria os resultados dentro da margem de erro.

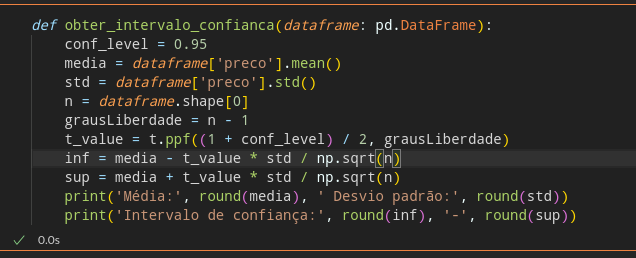


Figura 1: Código com o intervalo de confiança

4.1 – Bairro Centro

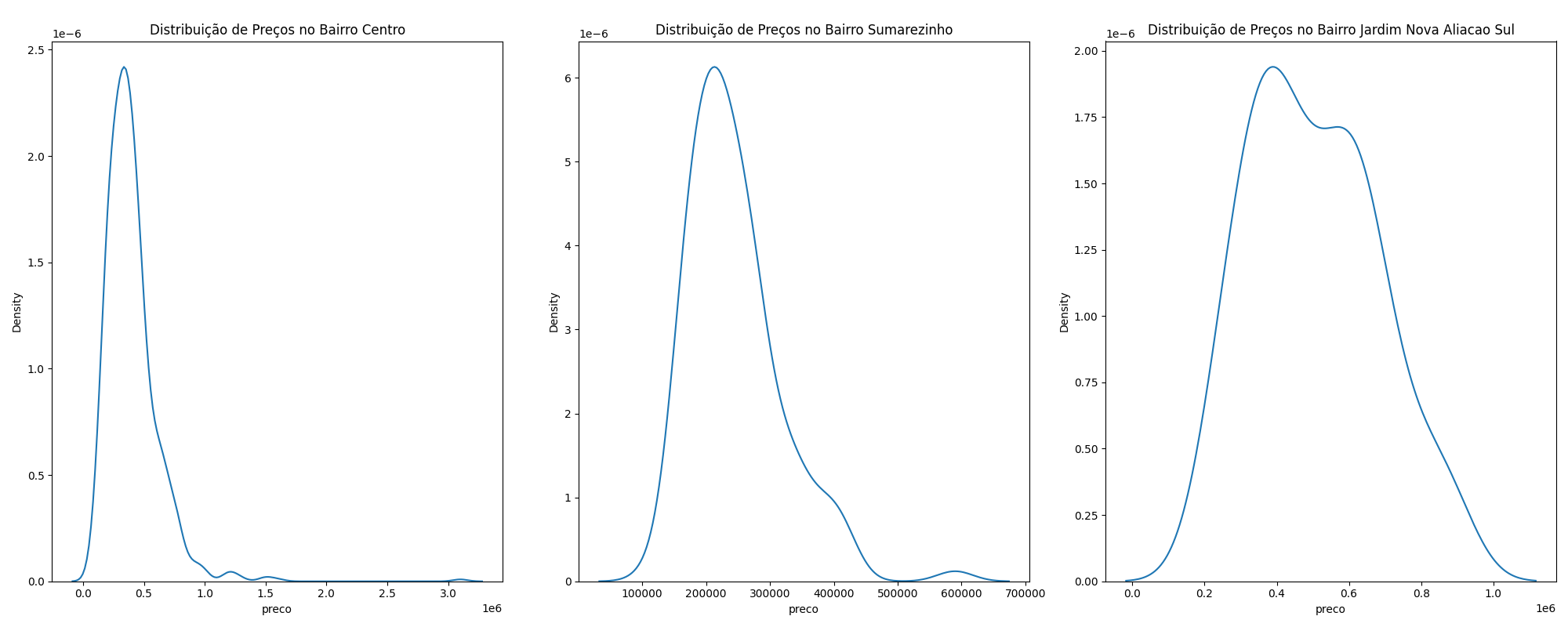
De acordo com o código abaixo, temos uma média de R$ 402077 reais, com o desvio padrão de R$ 223344,00, onde a média verdadeira pode está entre R$ 385915,00 e R$ 418239,00

4.2 – Bairro Sumarezinho

De acordo com o código abaixo, temos uma média de R$ 245525,00 reais, com o desvio padrão de R$ 72993,00 onde a média verdadeira pode está entre R$ 23227,00 e R$ 258832,00

5 – Teste da Normalidade entre os dados

Os testes de distribuição normal são usados para verificar se um conjunto de dados seguem uma distribuição normal. A distribuição normal é usada para descrever a frequência com que os valores aparecem no histograma.



Com base nos histogramas acima, temos uma tendência de normalidade de dados dos preços nos bairros Centro e Sumarezinho e no bairro Jardim nova Aliança Sul, há uma dispersão entre os preços dos apartamentos

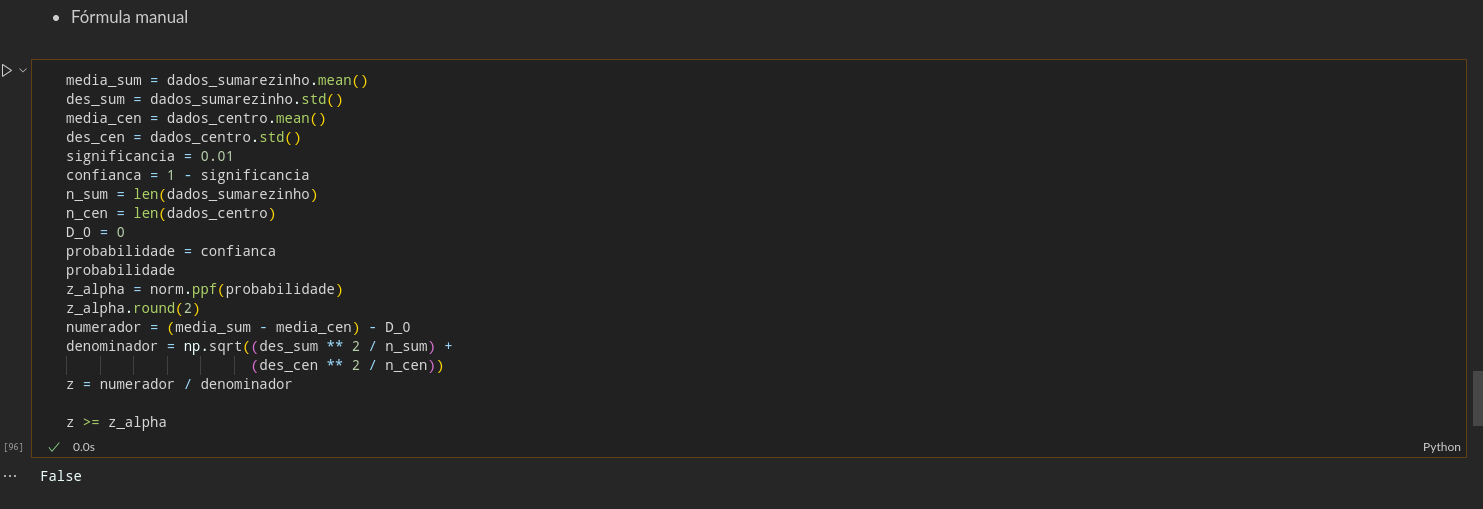
6- Teste de hipótese.

O teste de hipótese é usada para a tomada de decisões sob uma amostra de dados. Ela consiste em formular em duas hipóteses: A hipótese nula H0: É uma afirmação que se assume verdadeira até que seja refutada os dados. A hipótese H1: É a afirmação que se quer testar.

O teste foi proposto da seguinte maneira, os preços dos apartamentos do bairro Sumarezinho, são menores que o bairro centro, as amostras foram selecionadas aleatoriamente e foi utilizado o teste Z.

Teste Z é qualquer teste estatístico no qual a distribuição do teste estatístico sob a hipótese nula pode ser aproximada por uma distribuição normal. É um teste estatístico usado para inferência (afirma a verdade de uma preposição em decorrência de sua ligação com outras já reconhecidas como verdadeiras), capaz de determinar se a diferença entre a média da amostra e da população é grande o suficiente para ser significativa estatisticamente.

* u1: Média dos imóveis do bairro sumarezinho
* u2: Média dos imóveis do bairro Centro
* H0: u1 <= u2
* H1: u1 > u2



Legendas:

Z -> Resultado do teste z

Z alpha -> É a área de rejeição

A sequencia de cálculos foi elaborada da seguinte maneira: Obtemos a a média de preços e os desvios padrão dos bairros Sumarezinho e Centro, logo em seguida calculamos o numerador, a diferença entre as médias e D\_0, que é a diferença entre as duas médias, o denominador, que é a raiz quadra da soma da variância entre as amostras

Com base nos teste de hipóteses (resultado False), aceitamos H0, ou seja, a média de preço dos imóveis do bairro Sumazerinho, são menores que o bairro Centro, com o nível de significancia de 1%.