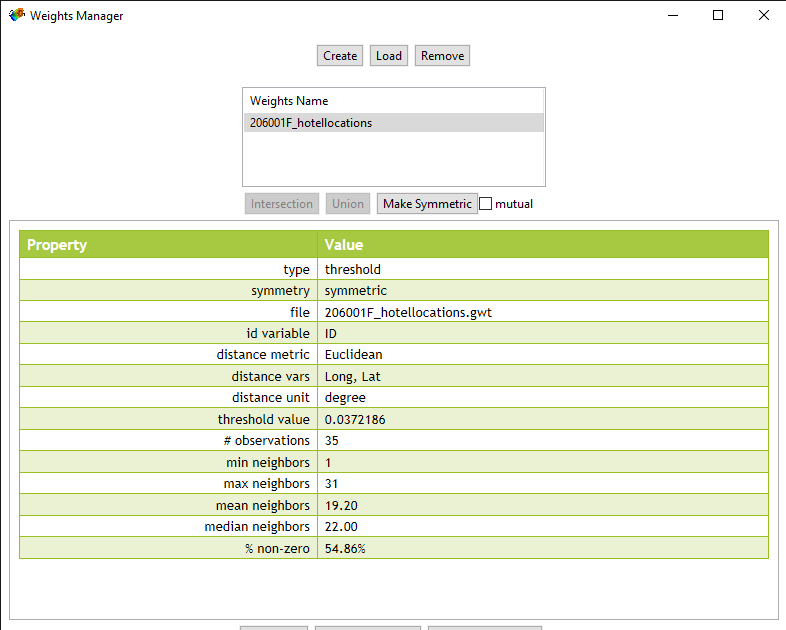
**Exploring Spatial Autocorrelation and Moran’s Index Analysis of Hotel Review Score**

**Name - Aabidh.mm**

**Index No – 206001F**

**Introduction**

This dataset comprises various properties, each uniquely identified by an ID number. These properties encompass both hotels and restaurants, distinguishable by their respective names or titles. The geographic coordinates, expressed as latitude and longitude, pinpoint the precise locations of each property on Earth. Additionally, each property is assigned a review score, likely serving as a measure of customer satisfaction or overall quality. The dataset also includes detailed street addresses, which specify the exact whereabouts of each property, all situated within the city of Colombo, Sri Lanka, and the dataset's overarching country. Moreover, the dataset provides information on expenditure, potentially indicative of the price range or average spending associated with each property. These attributes collectively form a comprehensive dataset that can be utilized for diverse analytical purposes, such as assessing popular properties, exploring geographical patterns, and investigating pricing trends in Colombo's hospitality and restaurant industry.

****

1. Why do you choose Euclidean distance when creating the Distance Metric?

It is permissible to utilize Euclidean distance because our data is projected.

1. What is the critical distance for your point data?

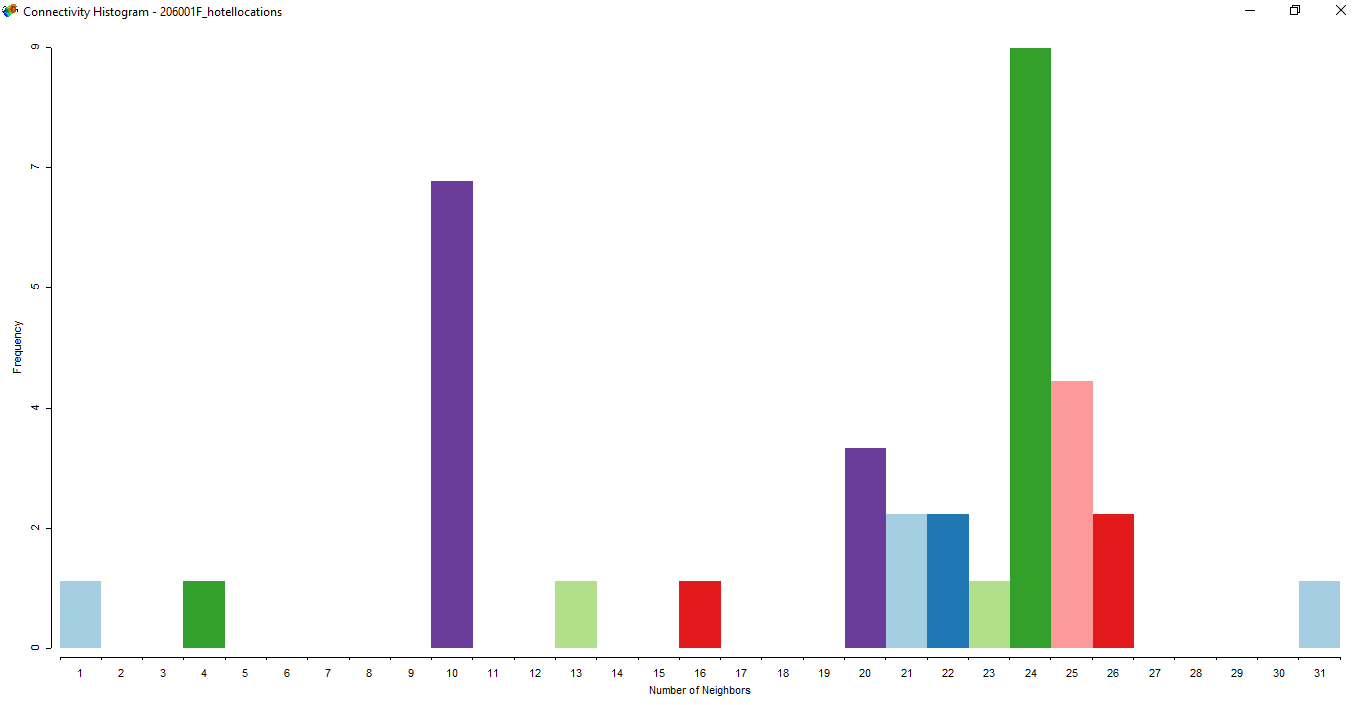
Our point data's critical distance is roughly 0.0372186 degrees.

1. What is the meaning of "critical distance"?

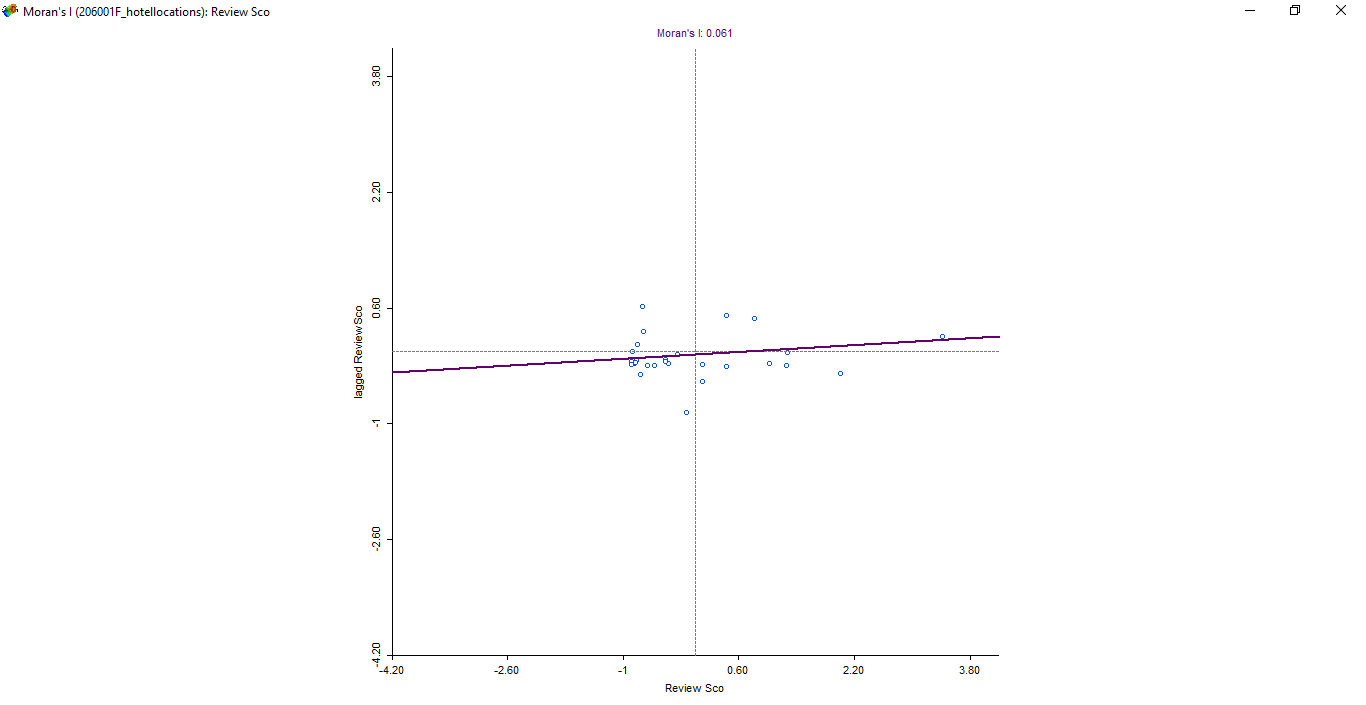
This geographical separation guarantees that each point has at least one neighbor. Critical distance is the distance beyond which willingness to go is significantly influenced by cost, effort, and means.

1. Explain the importance of the "threshold value"?

The only accurate method for recalling the distance cut-off for the weights' distance bands is by this method.



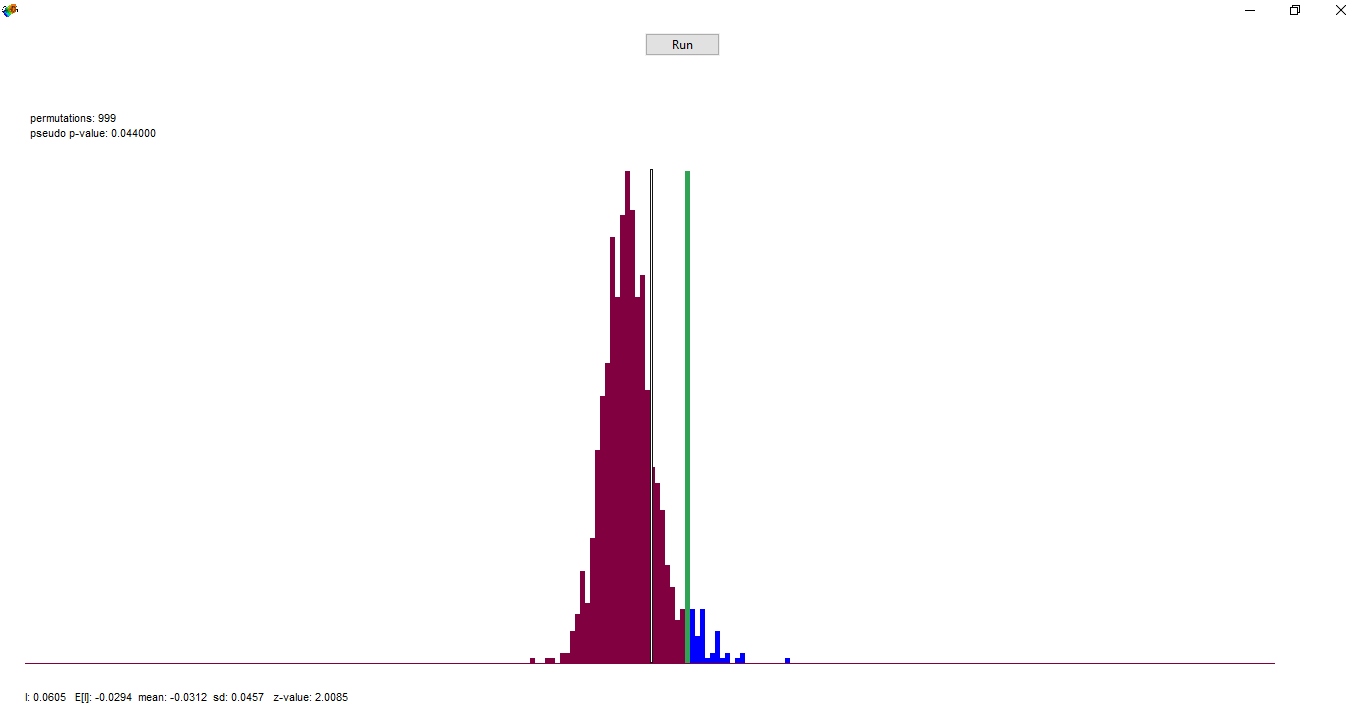
**Create the Moran’s index using the Review Score as the spatial variable.**

****

* What can you tell about the shape of the point cloud?

First, it is determined that Moran's I, a gauge of spatial autocorrelation, is 0.061. This value suggests a weak positive autocorrelation, indicating that although the autocorrelation is not very strong, surrounding data points typically have comparable values. Second, having outliers has a big effect on the dataset. With a review score of 4461, which is three standard deviations above the mean, one data point sticks out. These outliers, or extreme values, have a significant impact on the overall pattern and interpretation of the data. Thirdly, data points are closely clustered at the lower end of the review score range. This shows that the lower range of review scores exhibits less variability since many observations have values that are comparable to or closely clustered. The passage also emphasizes the importance of giving your decision to delete outliers due thought. Outliers should be removed carefully, considering the specific objectives of the study, even though doing so can assist in exposing underlying patterns and relationships in the data.

In conclusion, the dataset displays modest positive autocorrelation, prominent outliers at the high end, and data point clustering at the lower end of the review score distribution. To comprehend any underlying patterns or trends in the data, more investigation or visualization may be necessary.

**Inference**

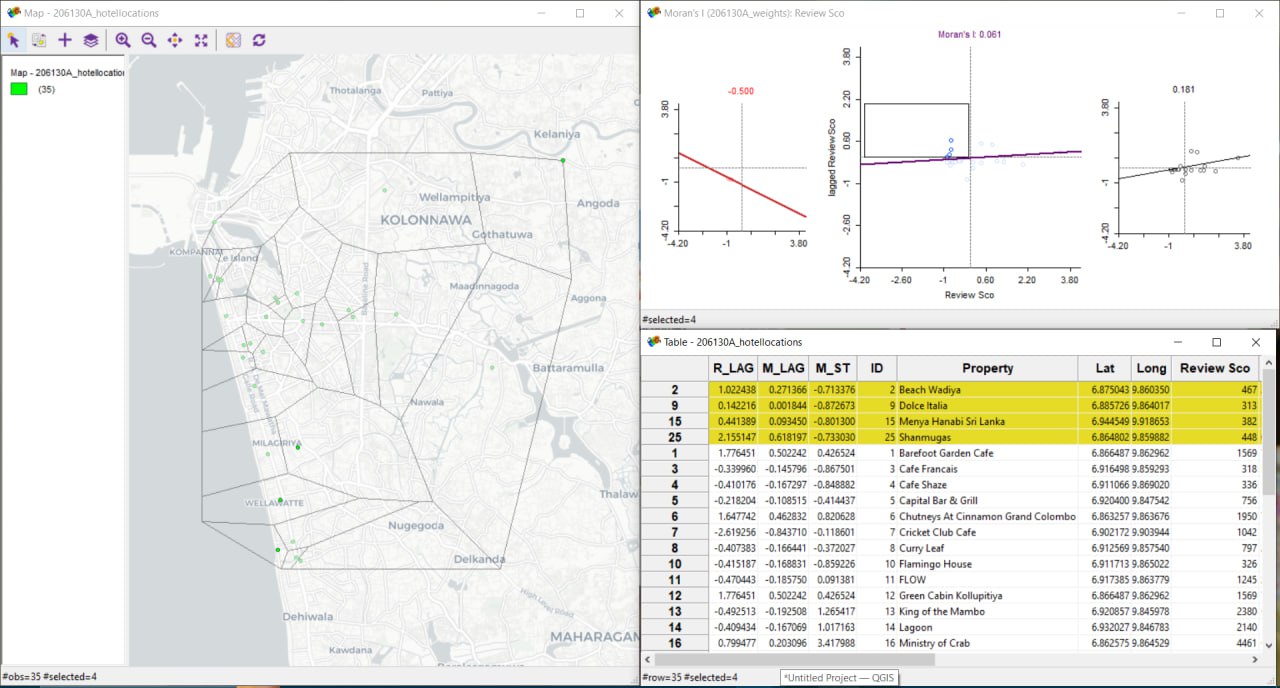
The pseudo p-value in our case is 0. 04400. which is well to the right of the standard distribution, and the green line displays the statistic's value for the actual data, which is set at 0.0605. This shows that the null hypothesis has been rejected. There are numerous summary statistics in the graph. The standard distribution's 999 permutations and corresponding pseudo-p-value are displayed in the top left corner of the graph.

**Identifying locations in the map associated with four types of spatial autocorrelation.**

**A screenshot of a computer

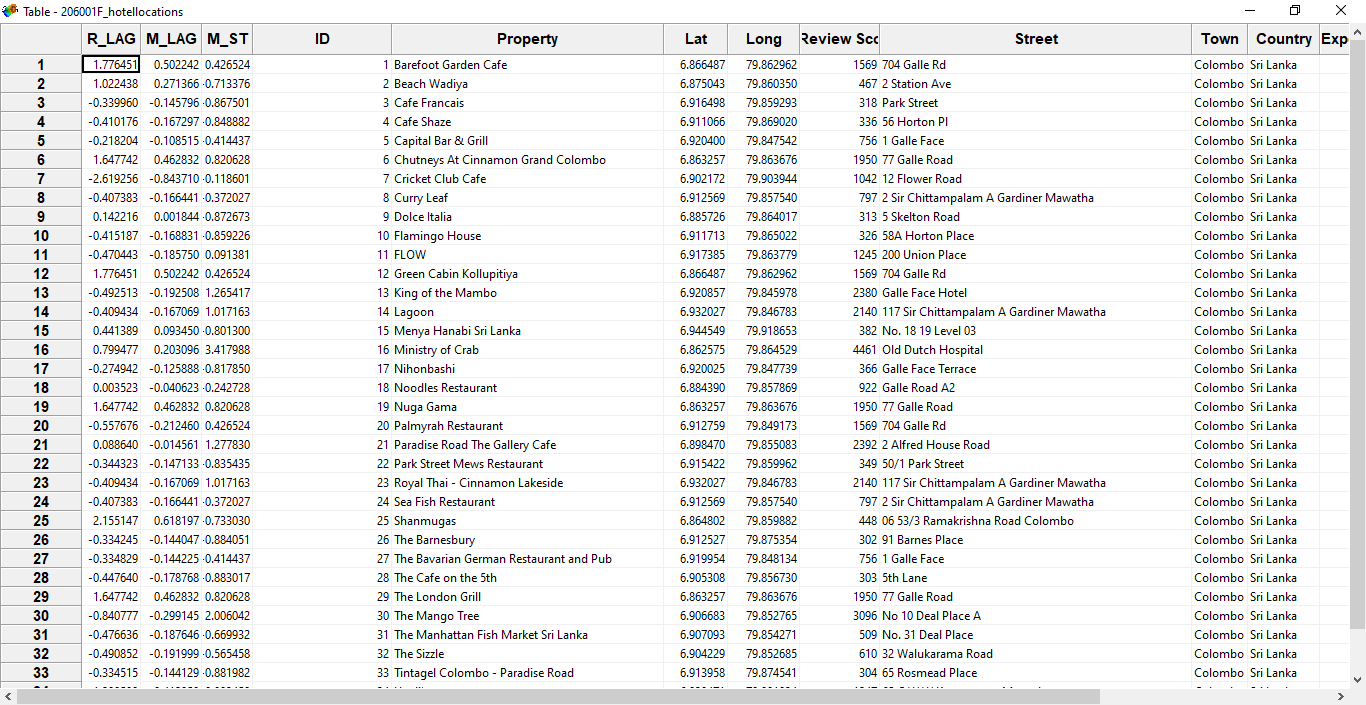
Description automatically generated**

There are six observations in the top-right corner of the Thiessen polygon layer above, indicating a polygon with high ratings and high Lag reviews. In other words, this polygon is surrounded by other polygons that have high values of reviews and have a high value of reviews itself. This explains why the Lag reviews, which measure the average values of these nearby polygons, are also high. Positive spatial autocorrelation is seen in this situation.

A screenshot of a computer

Description automatically generated

A low-review polygon is denoted by a dot in the upper-left corner and is surrounded by high-review polygons. This is a strong indicator of the presence of outliers because the spatial autocorrelation is negative. However, in our dataset, there are few reviews and no outliers, but because of how widely distributed they are geographically, we can still see a negative spatial autocorrelation.

Moran scatter plot attribute table

The Moran scatter plot option added two more columns to the table, bringing the total to five. Here, we can see that the M\_LAG and M\_ST values that we manually determined match the generated MORAN\_STD and MORAN\_LAG values from randomized permutations.